NAIST-IS-DD0261005

博士論文

ヒューマノイドロボットのための マルチモーダルインタラクションに関する研究

怡土 順一

2008年8月14日

奈良先端科学技術大学院大学
情報科学研究科 情報システム学専攻

本論文は奈良先端科学技術大学院大学情報科学研究科に
博士 (工学) 授与の要件として提出した博士論文である.

怡土 順一

審査委員：

小笠原 司 教授	（主指導教員）
鹿野 清宏 教授	（指導教員）
高松 淳 准教授	（指導教員）
松本 吉央 客員准教授	（指導教員）

ヒューマノイドロボットのための マルチモーダルインタラクションに関する研究*

怡土 順一

内容梗概

近年，ロボットを我々の生活環境において実際に役立てるための様々な技術が多く研究されている．これらの技術により，実環境において人間と共存可能なヒューマノイドロボットの開発は日進月歩の歩みを続けている．しかし，二足歩行や器用な操作に代表されるようなヒューマノイドロボットの身体的能力に関する研究が盛んに行われ，ある程度の機能を実現している現状に対し，人とヒューマノイドロボットとのコミュニケーションに関する研究はまだ初期段階にあると言わざるを得ない．本研究の目的は，人とヒューマノイドロボットとの対話をより円滑に行うことができるシステムを構築・評価する事である．実現されたシステムは，顔情報計測によるアイコンタクトと，大語彙連続音声認識エンジンによる音声認識を軸としたマルチモーダルなインタラクション機能を特徴とする．

直接対話時に利用される情報は，言語情報と非言語情報の二種類に大別できる．コミュニケーションにおいて前者が重要なのは言うまでもないが，近年，顔向きや視線，ジェスチャなどを含む後者の情報も自然なヒューマンロボットインタラクションを実現される上で重要であることが分かってきた．本研究では，まず，研究プラットフォームとして受付案内ロボット ASKA を構築し，それ用いて発話時における利用者の顔情報の計測，およびその情報の傾向分析を行った．この結果を利用して，アイコンタクトを用いた対話機能を ASKA 上に実装し，その検証実験を行った．また，ジェスチャ認識と音声認識を統合した対話機能を実装することにより，指示語を含む発話の認識を可能にするなど，より自然な対話を可能にした．

次に，対話対象となるヒューマノイドロボットシステムを，自律している否かの視点から考える．ヒューマノイドロボットは，この受付ロボットのような自律シス

*奈良先端科学技術大学院大学 情報科学研究科 情報システム学専攻 博士論文，NAIST-IS-DD0261005, 2008 年 8 月 14 日.

テムとしての利用方法に加えて，遠隔操作システムとして利用する事が可能である．そこで我々は，電話やビデオチャットに代わる新しいメディアとして，ロボット遠隔コミュニケーションシステムを提案した．提案された遠隔コミュニケーションシステムは，受付ロボット ASKA の頭部を利用した顔ロボット上に実装され，他のメディアとの比較実験を通してその有効性の評価が行われた．また，他の表情提示ロボットシステムと異なり，我々の提案したシステムは，計測された表情の解釈というプロセスを通すことなく顔情報をロボットに直接投影する事を特徴としている．このため，基本表上に限定されない顔情報の伝達が可能となった．

さらに，受付案内ロボット ASKA の発展系として，そのインタラクションシステムをヒューマノイドロボット HRP-2 に実装した．主な認識モジュールは前述の ASKA 上で開発されたものを利用しており，これは，我々のシステムの高いポータビリティを示している．HRP-2 も，机を挟んで対面したユーザの視線を，頭部のステレオカメラを利用して検出する．また，ユーザの質問発話や咳やくしゃみなどの非定常雑音を認識する．これらの情報を利用して，ジェスチャを伴う音声合成による応答を行い，あるいは物体の受け渡しなどの指示されたタスクを実行する．さらに，愛・地球博でのデモンストレーション等を通して, 実環境下で実証実験を行った．

これら 3 種類のインタラクションロボットシステムの開発・運用を通じて，アイコンタクトをはじめとする非言語情報と音声情報を人とヒューマノイドロボットとのコミュニケーションに利用する事の有効性を確認した．

キーワード

ヒューマノイド，インタラクション，マルチモーダル，視線, 音声認識

Multi-modal Interaction System for Humanoid Robots *

Junichi Ido

Abstract

In recent years, many researches have focused on various technologies which make robots actually useful in our daily life. These technologies bring us one step closer to the development of a humanoid robot that can be put to practical use in a real world environment requiring coexistence with people. While research about physical ability of humanoid robot such as biped walking and dexterous manipulation have been actively conducted, research on communication with human has not attracted much attention. Goal of this research is to develop and evaluate a humanoid robot system which has more flexible communication ability. The realized system is characterized by a multi-modal interaction ability, which consists mainly of "eye-contact" using a facial information measurement system and "speech recognition" using a large vocabulary continuous speech recognition engine.

The information utilized for face-to-face communication is classified in two major categories, verbal and non-verbal information. Although primary information in communication is the former, the latter information such as facial direction, gaze and gesture is recently emphasized as means of natural human-robot interaction. We first developed a robotic receptionist ASKA as a research platform and observed facial information of users communicating with the humanoid robot. Interaction ability for realizing eye-contact was implemented based on the analysis of human robot communication. In addition to this, integration of gesture recognition and speech recognition enabled the robot to recognize natural utterance containing demonstrative pronouns.

*Doctoral Dissertation, Department of Information Systems, Graduate School of Information Science, Nara Institute of Science and Technology, NAIST-IS-DD0261005, August 14, 2008.

A humanoid robot can be used as a remote controlled system besides an autonomous system such as the robotic receptionist. We proposed a robotic telecommunication system in which a face robot is used in place of conventional media such as a phone and a video chat. The remote communication system was implemented on a robotic head of ASKA, and the effectiveness was confirmed by evaluation experiment. Unlike other robotic systems which express facial expressions, in proposed system, measured facial information is directly projected to the robot without any subjective interpretation.

Based on the interaction system of ASKA, interaction system was implemented on a humanoid robot HRP-2. The fact that most recognition modules are developed directly based on the one of described receptionist robot indicates their portability. HRP-2 sitting opposite to a user across a table can detect gaze direction, head pose and gestures using a stereo camera system attached to the head. In addition, our system can recognize the user's question utterance and non-stationary noise such as coughing and sneezing using microphones. Using efficiently such information, HRP-2 can answer the question by its synthesized voice with gestures and do tasks such as passing objects on the table. We verified the usefulness of our system in the real environment through a demonstration in the Prototype Robot Exhibition at Aichi World EXPO 2005.

The usefulness of multi-modal interaction system for humanoid robot utilizing verbal and non-verbal information was verified through development and operation of these three robot systems.

Keywords:

humanoid robot, multi-modal interaction, gaze, speech recognition

目 次

図 目 次

表 目 次

第1章　序論

1.1. まえがき

人とヒューマノイドロボットとのインタラクションとはなにか．本論に入る前に，この事について簡単に述べることで，本論文の主題を明確にしたいと思う．

人間同士で意思の疎通や理解は通常，“コミュニケーション”という表現で表され，我々は普段無意識にそれを行っている．また，例えば使い慣れない外国語を用いて対話を行う場合や，相手を説得する場合など，強く意識して意図や意見の伝達・理解を行わなければならない場面も多々存在する．つまり，人と人とのコミュニケーションとは，その発信が意識的か否かに関わらないメッセージ集合の伝達であるといえよう．このような事から，車や計算機といった，我々が意識無き道具であると考えるものに対して，“コミュニケーションをとる”という言葉を用いることは希である．逆に言えば，人がコミュニケーションを成り立たせようと試みる場合，相手の意図を汲みとる能力と自己のそれを表出できる機能を対象に期待している，といえる．ここで，対象が必ずしも期待される能力を有している必要はなく，相手となる人間がそう感じるかどうかという主観的な問題がある以上，ロボット，とりわけ人に近しい形状を有するヒューマノイドロボットに対して，コミュニケーションという言葉を用いることは自然であり，これを人と同様に行えるようにする事はロボット研究における到達目標の一つといえる．

では，本題に掲げたインタラクションとは，何を意味するものと捉えるべきだろうか．コミュニケーションが，社会生活を営む人間の間に行われる，言語・文字その他視覚・聴覚に訴える各種のものを媒介とした知覚・感情・思考の伝達と定義されているのに対し，インタラクションは単に，互いに働きかけること，とされている．インタラクションは，物理的作用・応答の意味合いが強く，コミュニケーションを成立させるための要素あるいは手段と考えることができる．つまり，本論文で述べるところの人とヒューマノイドとのインタラクションとは，ヒューマノイドとの

間にコミュニケーションを成立させるための一手段であり，従来の機械に対する入力操作的なものの代替である．

さて，では，ヒューマノイドとのコミュニケーションにおいて適切な手段とは何であろうか．人間のコミュニケーション手段として最も大きな割合を占めるものは，言葉であるが，言葉だけではコミュニケーションは不完全であり，成り立たない場合も存在する．つまり，人間同士のコミュニケーションにおいて，我々は表情や仕草などの様々な情報を統合して意思の疎通を行っている．そこで，言葉以外の情報メディアを認め，それが意識的か非意識的かという事にこだわらずに総合的に活用するマルチモーダルインタラクションが，ヒューマノイドロボットとのコミュニケーションを成り立たせるために重要な手段となるのではないか．このような考察から本研究は出発したといえる．

1.2. 研究の背景と目的

ロボット市場は2003年の約5,000億円から，2010年は約1兆8,000億円，2025年は約6兆2,000億円に拡大．この牽引役となるヒューマン・サポート・ロボットは2010年に普及し始める．経済産業省，新エネルギー・産業技術総合開発機構（NEDO技術開発機構），産業技術総合研究所は，2005年に策定された技術戦略ロードマップ[1]の中で，ロボット分野の導入シナリオをこのように描いた．このために必要な技術項目として，環境構造化・標準化，マニピュレーション，コミュニケーション，移動，パワーマネジメントなどが議論されている（図 1.1）が，特に柔軟なコミュニケーション能力の獲得は難しい問題である．その一因として，人間という不確定な対象を問題にする事の困難さが考えられる．人間が，ロボットに対してどのようなコミュニケーション能力を期待するか，という問題はロボットの機構や役割，時には見た目にも大きく影響を受けてしまう事が，この困難さを生み出している．

従来，ロボットといえば人間の単純労働の代替であった．特化型の機能を持ち効率が重視される産業用ロボットがその代表であり，日本はこの分野において世界一の地位を築いている．しかし，最近ではペット型や二足歩行型などの民生用分野におけるパーソナルロボット[2, 3]が登場しつつあり，また，更に福祉・介護（日本は世界一の長寿国でもある）や災害救助など幅広い分野でのニーズや開発に関心が高まっている．これらのロボットは，その形状や目的などに応じて多岐にわたる分類が可能であるが，このようなロボットが持つ様々な方向性の一つがヒューマノイド

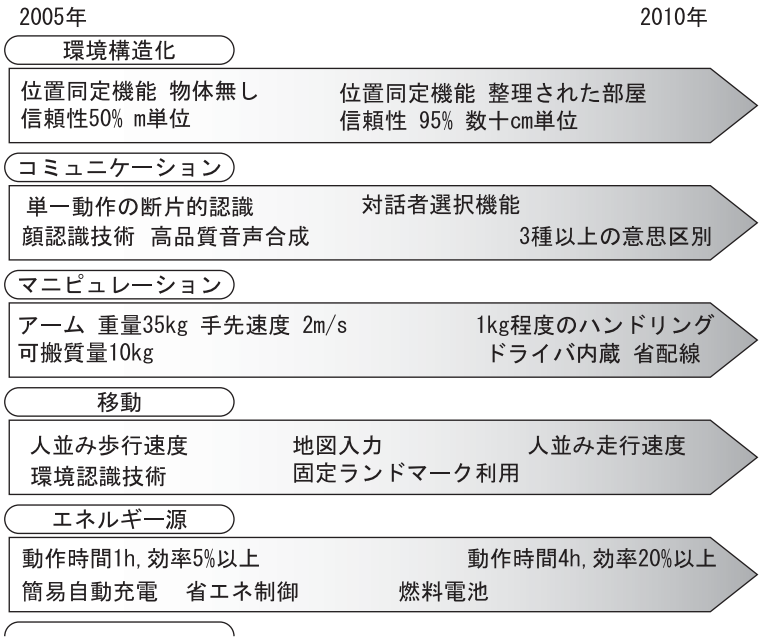


図 1.1 ロボット分野のロードマップにおける技術的課題（一部）

ロボットである．現在，ヒューマノイドロボットの明確な定義というものは存在しないが，その特徴の一つとして，頭部や腕部を備えた比較的人に近い形状を持つという点が挙げられる．これにより，ロボットが人間の生活する環境に適合しやすい，汎用的な作業・支援が期待できるなど，様々な利点が考えられるが，他のロボットと最も大きく異なる点は，ロボットが自身の擬似的な情緒や意図を自然に表出でき，さらに人間がそれを理解することが容易い，というものだろう．これはつまり，言葉だけにとどまらないマルチモーダルなコミュニケーションが可能である事を表しており，将来的に利用者が，この様な存在に対して人と同等あるいは同種のコミュニケーション能力を期待するのは自然な流れである．例えば，代表的な情緒の表出手法であるジェスチャ，あるいはアイコンタクトなどを行うことで自己の意志や情緒を表出するようなヒューマノイドロボットに対して，利用者は，自分側のジェスチャや視線がロボットに認識される事を期待するだろう．

しかし，ヒューマノイドロボットが比較的新しいロボットである事から，移動能力や機械設計などの基礎となるような分野と比較すると，コミュニケーションという分野における研究は数少ない．このようにロボットの機構・外観が人間に近いものに変化することで，人とのコミュニケーション形態も変容しつつある現在におい

も，自然なコミュニケーションと呼べるものは未だ実現されていない．

本論文の最終的な目的は，人とヒューマノイドロボットとのコミュニケーション円滑にするために必要な情報を検討し，その情報を基に実装されたインタラクション機能を実環境下で運用，評価する事である．

1.3. 本論文の構成

本論文は，全 6 章から構成される．以下で，各章の概要について述べる．

第 1 章：序論　本題に掲げたインタラクションの意味について考察し，本研究の背景，目的について述べた．

第 2 章：インタラクションを行うロボット　先ず，人間同士のインタラクションについて考察することによって，人間とロボット間で成立し得るインタラクションについて分類，整理する．さらに，従来より数多く行われてきた人間とロボットとのインタラクションに関する研究について述べ，本研究の位置づけを明確にする．

第 3 章：受付案内ロボット ASKA を用いたインタラクション機能の検討と実装　最初に，人とロボットがインタラクションを行う際の情報計測について述べる．非接触で計測可能な情報について考察し，学内受付で実施したオープンキャンパス時の予備実験を通して，人間がロボットに対してインタラクションを試みた場合の特徴について述べる．受付案内ロボット ASKA のシステムの構築・評価とその運用について述べる．

第 4 章：顔ロボットを用いたインタラクション　顔ロボットを用いたインタラクションについて述べる．前章で実装された顔情報を用いた人とロボットのインタラクションにおいて，ロボットの表情に注目し，これを遠隔対話システムに応用したシステムを構築する．このシステムを用いて，顔計測による表情の再現が対話者に与える影響を，人間同士の直接対話やビデオチャットと比較することで検証し，ロボットを利用した遠隔対話システムの特徴を明らかにする．

第 5 章：ヒューマノイドロボット HRP-2 を用いたインタラクション　3 章のシステムを発展させたシステムについて言及する．ロボットの全身モーションや把持が可能なヒューマノイドロボット HRP-2 を利用し，より多様なインタラクションを可能にしたシステムの詳細について述べる．また，HRP-2 の特徴の一つである二足歩行機能を利用した屋内ナビゲーション手法についても詳細を述べる．

第 6 章：結論　本論文を結ぶ．本研究を通して得られた知見を総括し，今後の展開と可能性について言及する．

第2章　インタラクションを行うロボット

2.1. 人間同士のインタラクション

人と人との対面コミュニケーションを成り立たせる情報には，言語情報と非言語情報に分けることが出来る．言語情報は，対面コミュニケーションにおいては大抵の場合，“音声言語情報” を指す．これは，音素，形態素に分割可能であるという明確な二重文節性を持ち，これにより我々は任意のメッセージをやり取りすることが出来る．これに対して，非言語情報に分類される情報は非常に多様であり，沈黙や衣服の種類，化粧や皮膚の色なども非言語情報とされる．また，非言語情報には明確な文節性を持たないものも多い．

人間は，このような様々な情報を常に送受信しながらコミュニケーションを成立させている．例えば，人物のひととなりを判断する面接という場においては，その対話の内容はもちろん，会話における発音や抑揚，姿勢，身振り，視線を含む態度，服装，髪型など多くの情報に注目する．もし，これをロボットシステムとして実装させようとする場合，センシングを行うだけでも非常に大規模なものになるに違いない．まして，一つ一つの行為における意味を全て判断する事は，ほぼ不可能である．しかし，単一のメディアだけに注目したのでは，人間同士にあるような自然なコミュニケーションは望めない．次節では，人とロボットとのインタラクションにおいて，どのような情報が用いられてきたのか，という観点に基づいて従来研究の概要を述べる．

2.2. 人とロボットとのインタラクション

近年，実環境下での使用および人間との共存を目的とした共存型ロボットの研究が盛んに行われている．それに伴い，人とロボットの間において制約の無い自然な

図 2.1 インタラクションに利用され得る情報の多様性

対話を実現することの重要性が増している．前述したように，人間同士のコミュニケーションにおいて用いられる情報は非常に多彩であるが，もちろんその全てが人とロボットとのコミュニケーションモデルに対して適用できるわけではない．

例えば，ロボットと人間の自然なコミュニケーション手段として音声対話を用いることは非常に有用であり研究も多くなされている．しかし，我々人間が，服装や年齢などから様々な情報を推測してコミュニケーションの取り方を変える事が常であるにも関わらず，それらをコミュニケーションに寄与する情報として利用するようなロボットの研究は，おそらくない．これは，それらの情報が計測し難いという事実もさることながら，ある程度の文節性および時間的変化を持つ情報の方が利用しやすいという理由からだろう．例えば，手話は明らかな二重文節性を持つといえるし，表情や視線，ある種のジェスチャ等も，程度の違いこそあれ文節性を持っているといえる．文節性を持つと言うことは，分類や数値化が比較的容易である事を示しており，これはロボットに実装する上で都合が良い．さらに，これらはいずれも比較的短時間において変化を伴う情報であり，この点からも対人コミュニケーションの場において利用しやすいと言える．

横山ら [4] は，人間同士の対話における発話交替の際の非言語情報の役割に注目し，

表 2.1 インタフェースの種類

分類基準	インタフェース	インタフェースの例
人の対面コミュニケーションに用いられるのと同じまたは類似のメディアをサポート	音声（バーバル）インタフェース	音声入力装置，音声合成装置＋スピーカ
	ノンバーバルインタフェース	身振り入力装置，視線入力装置，位置検出装置，表情 CG
人の対面コミュニケーションに用いられるのとは異なるメディアをサポート	ハプティックインタフェース	キーボード，マウス，ジョイスティック，ペン，レバー，力フィードバック装置
	ノンハプティックインタフェース	呼気検出装置，テキスト表示，（通常の）CG，音声合成装置＋スピーカ，計器

対話時の非言語情報の現れ方を分析し，発話交替における影響の違いを考察した．また，実ロボットに視線制御を適用しロボットインターフェイスにおける非言語情報の有効性を示した．その他に，人と人のコミュニケーションにおける視線に関する研究として，Kendon[5] は被験者の対話を分析することで，視線の機能には対象の視覚情報を取得する機能，会話における発話権の授受を調整する機能，その他の情報を相手に伝達する機能があるという分析結果を示した．深山ら [6] は，擬人化エージェントが視線パラメータ (凝視量，凝視持続時間，非凝視時視線位置) に基づいた視線を出力することで，ユーザが擬人化エージェントに対して持つ印象を操作可能であることを示した．対話時の頭部姿勢を調べた研究として，綿貫ら [7] はモーションキャプチャシステムを用いて頭部，胴体，手の動きを計測し解析を行った．解析の結果，聞き手よりも話し手の方で頭部，胴体，手の大きい動きがあり，特に手にその傾向が強いことを示した．ジェスチャに関しては，Cassell ら [8] は，コミュニケーションにおけるジェスチャを，ジェスチャの役割から偶像的機能，比喩的機能，指示的機能，動作的機能に分類した．

黒川は，インタフェースを表 2.1 で示されるような四種類に分類した．現行の計算機に置いては，ハプティックインタフェースがその大部分を占めるが，その他のインタフェースも数多く研究されており，音声入力や呼気検出，など製品として利用

されているものも少なくない．対して，人の対面インタラクションにおいては当然，バーバルおよびノンバーバルインタフェースにあたるものが利用される．では，ロボットに対してのインタフェースはどのようなものが利用されているのだろうか．もちろん，通常の計算機に対するインタフェースも利用されるが，先に述べたように，バーバル，ノンバーバルなインタフェースも計算機以上に利用されている．以下に，従来のヒューマンロボットインタラクションに関する研究を，この分類に従って述べる．

2.2.1 バーバルインタフェース

バーバル（音声）情報を用いてロボットとのコミュニケーションを試みる研究は数多く [9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21] ，近年では複数言語の認識あるいは発話を行うロボット [22, 23, 24] も存在する．

例えば，Jijo-2[25] はオフィス環境を動き回り，周囲の人間との対話や環境のセンシングにより知識を獲得する自律能動学習ロボットである．主に文脈表現を用いた音声認識 [9] によって，人の言葉を理解，学習し，インタラクションを行う．SIG[18] は，頭部と胸部からなる上半身のみのヒューマノイドロボットであり，人間の耳に当たる部分の内外部に取り付けられた二組のマイクロフォンを用いて，音声認識を行う．HERMES[23] では，語彙クラスを少数のサブクラスに分割し，コンテキストに応じて内部状態を変化させ利用する語彙クラスを制限することで，ロバストな音声認識を実現している．

2.2.2 ノンバーバルインタフェース

先に述べたように，人間はコミュニケーションを成り立たせるために多種多様なノンバーバル情報を活用しているが，従来のロボット研究で用いられるそれはある程度限定的であるといえる．従来の研究において代表的といえるノンバーバル情報は，視覚を利用して得られるジェスチャや顔位置，向きなどの情報である．

23 自由度のアームを持つ移動ロボット ARMAR[26] を用いて，マルチモーダルなインタラクションを実現する研究が行われている [10]．肌色抽出と視差画像生成によるポインティングジェスチャ認識，ニューラルネットワークを用いた顔向きの推定 [27] などを利用してインタラクションを行う．さらに，音声認識の結果とこれらを

総合的に利用することができる．例えば，“このランプのスイッチをオンにして下さい” というような発話に対して，ジェスチャと音声の認識結果が互いを補完し合う事が可能である．AlBERT は [11]，音声認識および輪郭抽出を用いたハンドジェスチャ認識によるインタラクションを行うアシスタントロボットである．指示されたテーブル上のオブジェクトを色情報を用いて識別する．もちろん，ジェスチャの認識だけではなく表出についても研究されている [12]．CERO[17] は，主に運搬用のサービスロボットであるが，その前部に Co-operative Embodied Robot Operator (cero) と呼ばれる人型の小型ロボットが搭載されており，対話の内容に応じて簡単なジェスチャを行う．

顔向きなどの顔情報計測だけではなく，機械的に表情の表出を行うことが出来るロボット [28, 24, 29, 30] も多く，その表現の複雑さは年々増している．Saya[15] は，その顔部の 18 個の制御点を動かすことで，Action Unit（AU）と呼ばれる 44 の基本動作に分類されるもののうち，6 基本表情（驚き，恐怖，嫌悪，怒り，幸福，悲しみ）を表出するのに必要とされる 14 の AU を実現する．また，機械的な頭部を持たず，ディスレプレイ上に表示された CG により表情を表すロボット [31, 32] も存在する．実在する顔と比較すると方向を指示する際の曖昧さや，存在感の欠如が欠点とされるが，複雑かつ微妙な表情をより簡単に再現できる事や実際に顔色を変化させるなどの誇張した表現が可能である事などの利点も存在する．

視覚以外から獲得されるノンバーバル情報として，音源（話者）位置が挙げられる．推定された話者の位置や方向を用いて，ロボットが対話に適切な位置や姿勢をとることが可能になり，これもノンバーバルな情報を有効に活用したインタラクションであるといえる．マイクロフォンアレイを搭載したロボットの多くが音源方向の推定を行っている [9, 13] が，さらに視覚情報を用いて話者の顔を検出し，それらの情報を併せて複数人との対話に利用する研究 [18, 33] なども行われている．また，応答発話の韻律と頭部ジェスチャから，発話対態度（肯定的であるか否定的であるか）を識別するロボット [34] の研究など，人間の発話において，その内容以外の情報を利用する研究も行われている．

触覚情報を利用したコミュニケーションでは，腕部や全身に取り付けられた接触センサにより，抱擁や握手などの行動 [12, 13] を行うものがある．

2.2.3 ハプティックインタフェース

ハプティックインタフェースは，キーボードやマウスなど，通常人間同士の対面コミュニケーションで用いられることのないものと定義されている．しかし，ロボットとのインタラクションにおいては，その安定性から利用される事も多い．タッチスクリーン [19, 16]，Web インタフェース [20, 21]，PDA[17] などを利用したロボットが研究されている．

Minerva[28] は，長期的に公共の実環境下（博物館）で稼働させるために，入力に音声認識やジェスチャ認識などの技術を利用していない．ユーザによる入力は，タッチパネルを用いた直接入力，あるいは Web インタフェースを用いた遠隔地からの入力を受け付け，移動によるナビゲーション等を行う．

Valerie は，液晶ディスプレイの頭部を持つ受付案内ロボット [32] である．入力はキーボードおよび ID カードのスキャンにより行い，合成音声と CG により合成された表情の表出によりインタラクションを行う．脚本を専攻する学生らの手による複数の複雑なシナリオが用意されている事が特徴で，実環境下で長期的な稼働が可能なロボットである．また，ID カードのスキャンを利用して，個人別に対応を変化させる事ができる．

RoboX[22] は，簡単な音声認識に加えて，数種類のボタンによる入力と，感情を表すシンボルやアイコンを表示する LED マトリックスによる感情の表出によりインタラクションを行う．ボタンの入力は，言語の選択やロボットの質問に対する応答などに利用される．

2.2.4 ノンハプティックインタフェース

音声合成やディスプレイによる CG，テキストの表示を行うロボットは，このインタフェースを搭載しているものとして分類できる [22, 28, 16]．しかし，それらは全てロボット側の情報の表示に用いられている．呼気や体温など，人間側から発せられるこれらの情報を計測し，ロボットとのコミュニケーションに利用することは困難である事が原因だと考えられる．

2.3. 本研究のアプローチ

本章では，人間とロボットとの間に自然なコミュニケーションを成立させるためのインタラクションに関する研究を，そのインタラクションに利用されるインタフェース毎に分類しまとめた．

まず，人とロボット間に期待されるインタラクションは，人と計算機等との間に期待されるインタラクションよりもむしろ，人同士におけるそれに近いことは想像に難くない．従来研究を鑑みても，人型に近いロボットではそうでないものと比較して，音声やジェスチャ，アイコンタクトなどの人の対面インタラクションに用いられるインタフェースが利用される割合が高い．そこで基本となるのは，やはり言葉による情報伝達（音声コミュニケーション）である．そこで本研究では，人とヒューマノイドロボットとのインタラクションを実現するにあたり，音声と画像という二種類の情報を用いたアプローチ，いわゆるマルチモーダルインタラクションに注目する．

第3章　受付案内ロボットASKAを用いたインタラクション機能の検討と実装

ASKAは特に，人とロボットのインタラクションにおけるマルチモーダルインタフェースの研究のためのプラットフォームとして開発された．我々が目的とする受付案内というタスク（“ロボティクス講座にはどう行けば良いですか”，“音情報処理学講座の内線番号を教えて下さい”という様な質問に対する音声応答やジェスチャ指示等）をロボットが実現するためには，人間とのコミュニケーションが重要な課題となり，その実現のために，様々な要素技術を利用した機能が実装されている．

3.1. デザインコンセプト

上述した様に，ASKAは，(1）研究プラットフォームの提供，(2）自然なヒューマンロボットインタラクションの実現という二種類のコンセプトを持ってデザインされている．

3.1.1 研究プラットフォーム

従来，研究プラットフォームとして開発されたロボットとして，認知科学の研究を目的としたもの[35]や脳科学の研究に用いられていたもの[36]，さらに，ロボットの身体性や人とのコミュニケーションに関する研究に用いられているもの[12]が存在する．ASKA開発当初において，異なる研究分野における要素技術を実装する共通のプラットフォームを提供するという事が主題であった．このコンセプトが，ソフトウェア基本構成を決定している．

図 3.2 に示すように，モジュール間の通信インタフェースのみを提供し，各機能モジュールが互いの状態のみを通信する事により，研究分野毎の独立したモジュール開発を可能にした．

現在は，リアルタイムの音声認識や画像処理を単一の CPU で同時に処理させる事は困難である理由から，3 台の外付けコンピュータを用意している．しかし，将来的に必要とされるコンピュータの処理能力が変化した場合でも，各モジュールを任意のコンピュータで動作させ，コンピュータ数を容易に増減させる事ができる．各モジュールは独立して動作しているので，それぞれ任意のタイミングで動作の開始/停止を決定することができ，さらにモジュールの数も任意に変更が可能である．この独立性が，全く異なる技術を基礎とした各モジュールの容易な開発とメンテナンスを可能にしている．

3.1.2 ヒューマンロボットインタラクション

対面コミュニケーションにおいて重要な要素は，言語情報と非言語情報とに分類できる．対話の基本となるのは前者であるが，近年，円滑なコミュニケーションを実現する為の手段として，顔向きや視線，ジェスチャなどの非言語情報が重要視されている．本論文では，特に視線情報とジェスチャに注目し，これらと言語情報を組み合わせる事でより自然なインタラクションを実現する事をコンセプトの一つとしている．音声対話部分には，大語彙連続音声認識エンジンである Julius[37] を用いることで，煩雑な文法定義を行うことなしに，自然な発話認識を可能にしている．顔向きや視線，ジェスチャなどの非言語情報を計測するために，眼部に小型ステレオカメラを 2 セット内蔵し，ロボットの外観を損なうことなく対話者と環境の情報を同時に取得する事が可能である．

ここで，機械的に口や目が動作する顔機構が搭載されている事が ASKA の特徴の一つである．機械的故障や生成できる表情の豊かさを考慮すると，頭部にディスプレイを搭載し CG で顔を表現する手法が優れているといえるが，これはヒューマノイドロボットという存在から遠ざかる一因となる．また，頭部に実ロボットを利用する事による利点も存在する．例えば，顔向きによる指示方向が明確になり，これは受付案内指示明瞭化の一助となる．

3.2. 受付案内ロボット ASKA の概要

3.3. システム構成

3.3.1 ハードウェア構成

ASKA のシステムは図 3.1 に示す様に，ロボット本体と 4 台の PC，およびカメラ，マイクロホン，スピーカから構成されている．ASKA 本体の詳細なスペックは，表 3.1 に示す．ロボット本体部分には Tmsuk-04（テムザック社 [38]）をベースとして利用し，そこに Linux を OS とする PC を内蔵している．頭部には通信総合研究所（現 NICT）で開発された Infanoid[39] の頭部を搭載しており，また眼球部には広角と望遠の 2 組のステレオカメラセットが内蔵されている．頭部は本来 8 の自由度を持つが，ステレオカメラのキャリブレーション情報を保つために，眼球の自由度は利用されていない．Tmsuk-04 は本来，遠隔操作型のロボットであるが，本システムにおいてはロボット内部に PC を内蔵することで，胴体部および頭部をコントロール可能な自律型ロボットとしている．

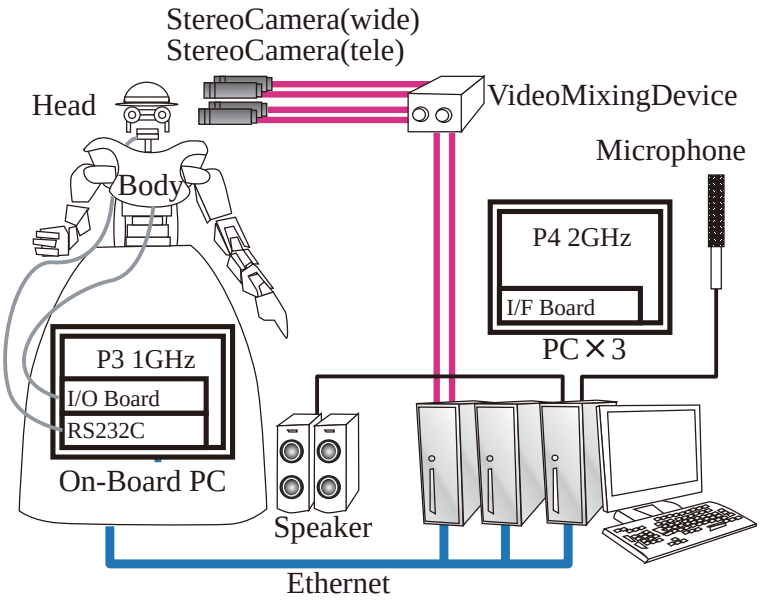


図 3.1 ASKA のハードウェア構成

表 3.1 Tmsuk04 データシート

寸法・重量	
全長	約 750mm
全巾	約 600mm
全高	約 1200mm
総重量	約 100kg
動作自由度	
頭部	2
胸部	1
腕部	$7\times2=14$
手部	$3\times2=6$
走行機能	
駆動輪	左右独立 2 輪駆動
補助輪	前後無軌道補助輪
走行速度	約 3km/h
撮影機能	
カメラ	CCD カメラ
有効画素数	25 万画素
水平画角	114 °
その他	
発話機能	4 音声
安全機能	周辺検知センサ 5ヶ所による走行規制
電源	Ni-Cd バッテリー搭載
通信方式	簡易携帯電話 (PHS) 網
ディスプレイ	液晶カラーディスプレイ
操作装置	指部反力機能付ジョイスティック (3 コントロール ×2) リアルムーブメントアームコントローラ (6 コントロール ×2) リアルムーブメントヘッドコントローラ (2 コントロール) ペダル式ホイールコントローラ (3 コントロール $+\alpha$)

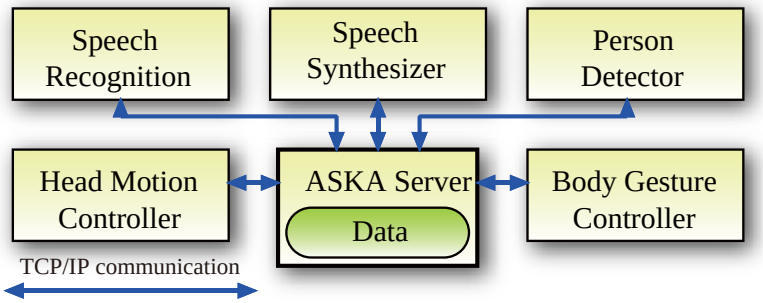


図 3.2 ASKA のソフトウェア構成

3.3.2 ソフトウェア構成

ASKA の基本ソフトウェアは，大きく分けて次の 7 つのモジュールから構成される：

1. 人発見，追跡モジュール
2. ジェスチャ認識モジュール
3. 顔情報計測モジュール
4. 音声認識モジュール
5. 音声合成モジュール
6. 胴体ジェスチャ制御モジュール
7. 頭部ジェスチャ制御モジュール

前述のように，各モジュールはソケット通信によって互いの状態を伝えながら動作している．この状態伝達には，単純な黒板モデル [40] が用いられている．サーバには全モジュールの状態（動作中，待機中，出力値など）が常に蓄積されており，各モジュールは，それを参照して次の状態を決定する．各モジュールに置いて，これらの状態を取得・蓄積するためのインタフェースが定義されているが，モジュールの実装自体は自由である．利用するアルゴリズムやセンサなど変更や言語の別が他のモジュールに影響を与えることはないので，各研究分野の資産を生かして実装を行うことができる．

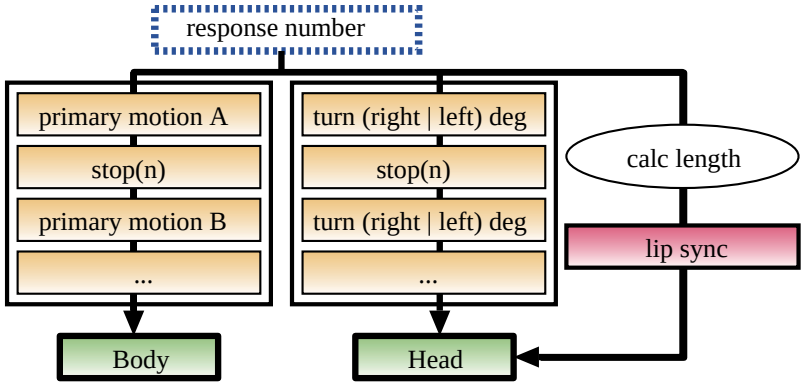


図 3.3 腕部および頭部の応答ジェスチャ処理

3.3.3 インタラクションシナリオ

ASKA は上述の様なシステムにより，訪問者とのインタラクションを行う．その基本的なシナリオは，以下のようになる．

1. ASKA がビジョンにより人の発見，追跡を行う
2. 訪問者が音声とジェスチャにより質問を行う
3. ASKA が質問を認識し，音声とジェスチャにより応答する

ここで，ASKA からの応答では，あらかじめ定められた応答文を発話し，ジェスチャを実行する．応答ジェスチャは，頭部，胴部および両腕の動作から構成される．図 3.3 に示されるように，各ジェスチャは予め規定された基本姿勢のシーケンスとして定義されており，これが，応答文に対応付けされている．基本姿勢自体は，GUI を用いて容易に作成可能である．また，音声合成モジュールにより生成された音声ファイルの長さから口の開閉動作時間を計算し，口の開閉動作を行っている．

3.4. インタラクション機能

3.4.1 音声認識

音声認識モジュールは，李らによって開発された大語彙連続音声認識システム Julius[41] を用いて実装されている．近年，大語彙連続音声認識のパフォーマンスは飛躍的に改善されており，Julius おいては，20,000 語彙の読み上げ音声の場合 94.7%以

100 こんにちは.
303 <is-staff:3>の部屋は，<is-staff:5>にあります.
415 バス停は，そこ玄関をでて左手にあります.

図 3.4 応答文例

上，口語形式の読み上げで専用言語モデルを使用した場合 89.9%の認識率が得られている [42].

言語モデル

受付案内というタスクにおいては，固有名詞（教官名，講座名，施設など）や専門用語など特有の単語がキーワードとして用いられることが多く，これらを的確に認識する必要がある．従って，新聞記事などから学習した既存の N-gram 言語モデルでは，十分な認識性能を得ることができない．そこで，本学の受付案内のタスクのための言語モデルを新たに作成した．N-gram 言語モデルの学習に必要な学習用テキストには，以下のリソースから取得したテキストを結合したものを用いた.

- Web ページ
- 学内メーリングリスト
- 学内教職員データベース

作成した N-gram 言語モデルは，Julius 向けの 2-gram 及び逆向き 3-gram モデルであり，学習に使用した語彙は，学習様テキスト中の出現頻度の上位 2 万語である．言語モデル評価のために後述の応答文と「連続音声認識コンソーシアム 2000 年度版ソフトウェア」[43] に含まれる PTM triphone, 64 混合, 3000 状態の男性用 JNAS モデルを用いて実験を行った結果，約 90%の単語正解精度を確認した.

応答生成

ASKA では，受付案内に必要だと思われる質問文を決定するためのアンケートを事前に行い，その結果から応答文を作成した．図 3.4 に，その応答文例を示す．図 3.4 がその応答文例であり，文頭の 3 桁の番号は応答文インデックスである．このイン

100 p こんにちは
303 k 教授
303 k 研究室
303 k 部屋
303 k 番号

図 3.5 部屋番号に関する質問に対するキーワードリスト

181 小笠原 司 オガサワラ ツカサ A514 A 5 5370 ロボティクス 小笠原ロボティクスコウザ オガサワラケン

図 3.6 is-staff ファイルの例

デックスは，他のモジュールとのメッセージ交換に用いられる．さらに応答文は，1, 3 行目の様な定型文と 2 行目の様な挿入型文に分類される．例えば，この挿入文のスロット<is-staff:n>には，is-staff ファイル（図 3.6）から n 番目のデータが抽出，挿入される．スロットに挿入される情報は，予め作成された学内案内情報のみに限られない．また，web サイトからリアルタイムに “天気” や “時刻” などの情報を得て応答文を生成することで，“明日の奈良の天気はどうですか” 等の質問にも答えることが可能になっている．認識結果から応答文を選択するのに用いられるキーワードリストは，図 3.5 に例を示すように，予め応答文毎にキーワードを定義して作成してある．ここで，行頭の番号は前述の応答文インデックスを表す．キーワードリスト内で，インデックスに続き “k” が指定されている文字列はキーワードを表し，形態素に分割された認識結果とキーワードの一致数がもっとも多い応答文が選択される．音声認識結果の出力には，N-best を用いている．また，“p” に続く文字列はパターンマッチワードを表し，認識結果の一部がこの文字列と一致する場合には，キーワードよりも優先して応答文の選択が行われる．

3.4.2 人の発見と追跡

今回用いた人発見手法の特徴として，視差画像生成の際に SAD(Sum of Absolute Difference) 形式の相関演算と再帰相関演算を利用した高速局所相関演算アルゴリズム [44] を用いることで，ソフトウェアによるリアルタイム処理が可能であることが挙げられる．発見の処理手順を次に示す．

1. 広角カメラ特有のレンズ歪みを補正
2. 広角カメラを用いた平行ステレオ法による視差画像の生成
3. 視差画像から視差の大きさ (距離情報) に基づき関心領域を抽出
4. 関心領域から人の特徴を備えた領域を人領域として選出

広角レンズの歪み補正

人の発見・追跡処理に使用する広角カメラは，広角カメラ特有のレンズ歪みを持っているため，画像処理を行う前に補正を行う必要がある．レンズの補正手法として Tsai[45] のカメラモデルを用いた補正手法が多く用いられるが，この手法は CCD 素子数，CCD 素子間隔等のカメラの内部パラメータを必要とする．ここでは，Luca[46] の提案する手法を用いてカメラのレンズ歪み補正を行う．この手法は，カメラの内部パラメータを必要とせず，Hough 変換によってキャプチャした画像の歪みを評価し，最適なパラメータを推定する手法である．Luca の手法では，カメラを以下の様にモデル化している．

$$\begin{aligned} i &= (i_d + O_x) \times \alpha \times (1 + K_1 \times rd^2) + O_x \\ j &= (j_d + O_y) \times (1 + K_1 \times rd^2) + O_y \end{aligned} \tag{3.1}$$

$$rd^2 = (\alpha(i_d - O_x))^2 + (j_d - O_y)^2 \tag{3.2}$$

ここで，K_1 は歪み係数，$(i, j), (i_d, j_d)$ はそれぞれ歪み補正後の座標，補正前の座標，(O_x, O_y) はカメラの中心座標，α は縦横比を吸収する変数となっている．通常，(O_x, O_y) はキャプチャサイズの中心とし，$\alpha = 1$ とする．このモデルを用いてレンズの歪み補正を行った結果を図 3.7 に示す．

図 3.7 レンズ歪み補正前と補正後

平行ステレオ法による視差画像の生成

既知の位置関係にある複数のカメラで対象を観測し，その対象の三次元位置を再構成するのがいわゆるステレオ視である．平行ステレオ法とは，ステレオ視において各カメラの光軸の方向を平行とし，光学中心の高さ並びに光軸方向の位置をそれぞれそろえた手法である．このような場合には，エピポーラ線はすべて各カメラの水平走査線と一致するため，対応点探索は一律に水平走査となり高速な探索が可能となる．視差は，基準カメラでの画素と別のカメラでの対応点の水平方向のずれで表される．平行ステレオ法を用いた視差画像の生成手順を以下に示す．

視差画像生成：

1. 前処理
   - 移動平均法による 3×3 フィルタマスク平滑化
   - 正規化オペレータ・強度変化検出器である LoG フィルタ [47] による正規化・強度変化検出
2. 局所相関演算法に基づく視差計算
   - SAD 形式の相関演算による高速な演算
   - 再帰相関演算による高速な演算
3. 後処理
   - 一貫性評価法による誤対応点の排除

図 3.41 に，実際にキャプチャした画像 (解像度 312×234) を示す．また， 図 3.42 には平滑化フィルタの出力画像 (解像度 104×78) を示し，図 3.43 には LoG フィルタ

の出力画像を示す．最後に，図 3.44 に生成した視差画像を示す．

生成した視差画像から人領域を選出

視覚情報のみに基づいて人の発見を行う場合，画像中の背景と人をいかにして分離するかが重要な課題となる．本研究では，以下の手順で人領域を選出する．

1. 視差画像を一定の幅をもった視差単位にレイヤ化
2. 各レイヤに対する領域分割処理，及び人領域の抽出
3. 全レイヤを通して最適な人領域を選出

人領域を抽出するために設定した条件を以下に示す．

- 人は背景より前の前景にいるため，視差が大きい
- 人は視差画像中で，ある視差の範囲に分布する連続領域として現れる
- 人らしい特徴を身長 1.2〜2.0[m]，横幅 0.3〜0.7[m]，頭の長さ 25[cm] 程度で寸胴とする

平行ステレオ法では，各画像面座標での位置座標 $(x_l, y), (x_r, y)$ がわかると，3 次元座標 (X, Y, Z) は以下に示す逆透視変換により計算できる．(図 3.8)

$$\begin{aligned} X &= \frac{b((x_l - x_{lc}) + (x_r - x_{rc}))}{2d} \\ Y &= \frac{b(y - y_c)}{d} \\ Z &= \frac{bf}{d} \end{aligned} \tag{3.3}$$

ただし，f はカメラの焦点距離，bはカメラ間の距離 (基線長) であり，$(x_{lc}, y_c), (x_{rc}, y_c)$ は各画像面の画像中心を表す．図 3.9 に実際に視差画像をもとに作成した視差レイヤの一部を示す．

3.4.3 人の追跡

人の発見処理によって人の存在を認識するだけでなく，発見した人を追跡する事によって，後の画像処理が容易になるだけでなく，ロボット近辺にいる人物へロボッ

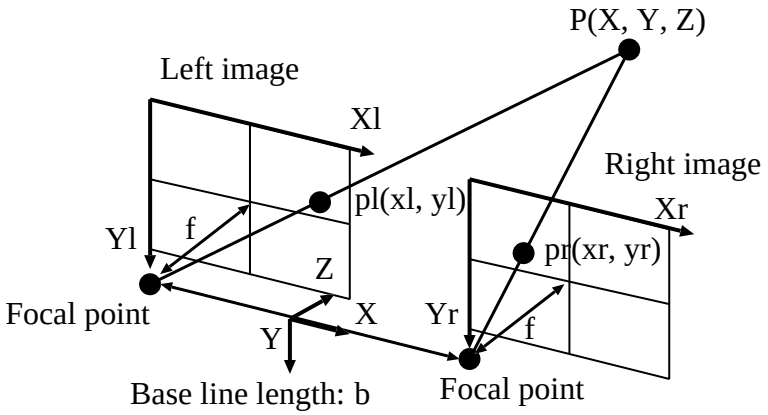


図 3.8 平行ステレオ法のモデル

トが注意を向けている事を知らせることができる．ここでの人追跡は，カメラを制御する頭部姿勢制御からなり，追跡対象となる人領域の重心位置をカメラ画像座標中の中心位置に合わせることで実現する．人の追跡の処理手順を以下に示す.

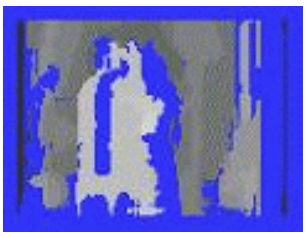
1. 選出した人領域の頭部位置を計測
2. 頭部位置を追跡対象に設定し，頭部位置とカメラ方向の差を計算
3. 頭部位置とカメラ方向の差を元に，カメラ姿勢制御

注視点の設定

追跡する対象となる注視点として，人の発見処理で人領域と認識された領域から頭部領域を特定し，頭部領域の画像座標上での中心座標を注視点と設定する．頭部領域は，人領域をもとに次のように求める．

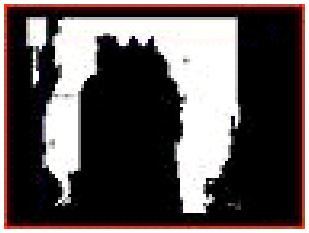
$$face = height \times 1/6$$

$face$ は頭部の高さ（単位：pixel）であり，$height$ は人領域の高さ（単位：pixel）である．頭部の位置は，人領域の上から $face$ 分の領域の幅と設定する．以上のように求めた頭部領域の中心座標を注視点と設定する．

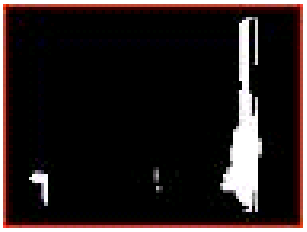


(a) 視差画像

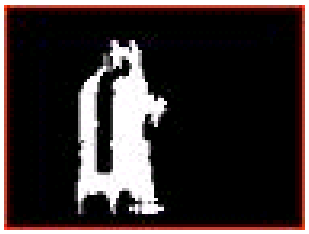
(b) 視差レイヤ ($2 \leq d \leq 5$)



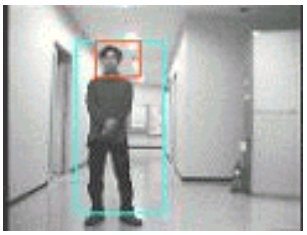
(c) 視差レイヤ ($6 \leq d \leq 9$)



(d) 視差レイヤ ($10 \leq d \leq 13$)



(e) 視差レイヤ ($14 \leq d \leq 17$)



(f) 発見した人領域

図 3.9 視差画像からレイヤ化

制御角の計算

注視点とカメラの関係を考えると，ステレオカメラから見て左側に注視点がある場合は，視差画像上も画像中心よりも左側に注視点が現れる（図 3.10）．また，注視点とカメラの三次元位置は図 3.11 のようになるため，Pan 角は式 (3.4) で求めることができる．

$$\begin{aligned} X &= \frac{b((x_l - x_l c) + (x_r - x_r c))}{2d} \\ Z &= \frac{bf}{d} \\ \theta &= sin^{-1}(\frac{X}{Z}) \end{aligned} \tag{3.4}$$

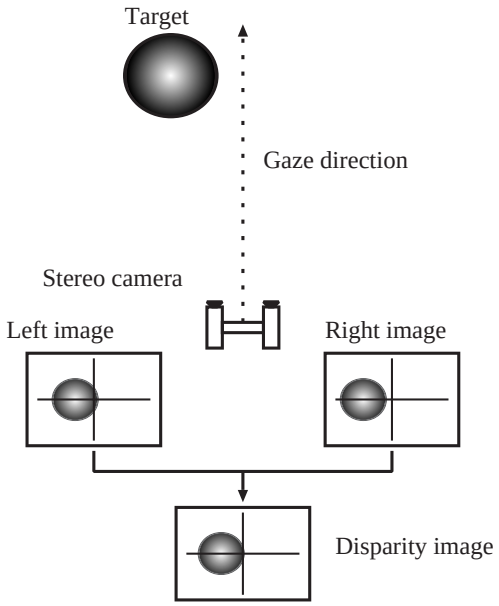


図 3.10 注視点とカメラの関係

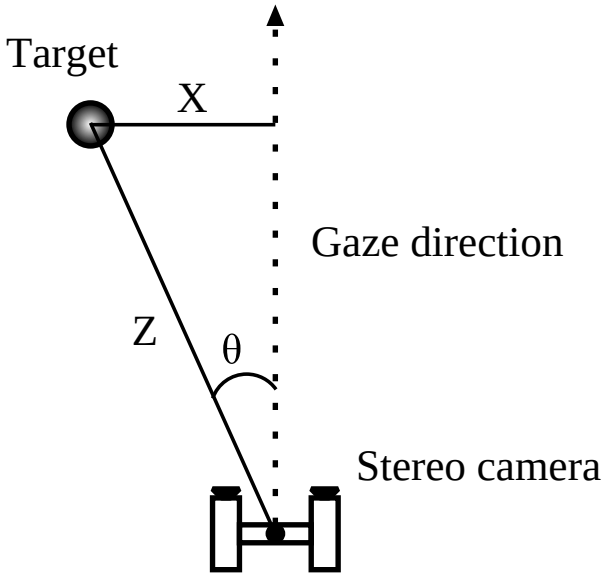


図 3.11 注視点とカメラの位置関係

$$= \quad sin^{-1}(\frac{(x_l - x_l c) + (x_r - x_r c)}{2f})$$

X, Y は注視点の三次元座標，b はベースライン長，f は焦点距離，d は視差，θ は Pan 角となっている．ここから，図 3.12 に示すように，カメラ姿勢の制御を行う．

3.4.4 アイコンタクトと発話区間推定

対話時の非言語情報計測実験

非言語情報を用いて円滑な対話システムを構築するためには，人とロボットが対話をするときの非言語情報の現れ方を分析する必要がある．非言語情報の現れ方，特に発話前後の非言語情報に注目して計測を行う．

予備実験　人とロボットの受付対話の計測実験の予備実験として，2002 年 11 月 2 日に本学で行われたオープンキャンパス（一般来場者を対象としたデモ展示）での対話デモ中の被験者の対話状況をビデオ撮影した後に分析した．この予備実験は，正式な計測実験では，被験者が実験であることを意識して自然な対話が計測できない可能性があるために一般対象者を被験者として行われた．また，顔情報計測のため

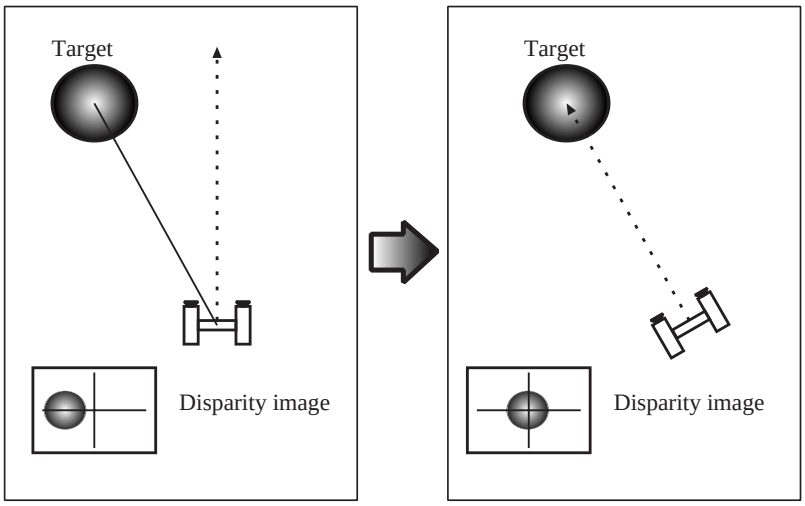


図 3.12 カメラ姿勢制御による追跡

の初期化作業（特徴領域テンプレートの主導登録）を全利用者に対して行うことは困難であるため，今回はビデオ撮影された動画の人手での分析を行った．実験の目的は，人が受付案内ロボットと対話するときに現れる非言語情報の特徴を分析することである．

実験手法　オープンキャンパスで行った受付案内デモ中のロボットと被験者の対話状況を miniDV テープに撮影した．保存された記録画像から，目視で発話開始前後の非言語情報の現れ方を観察する．実験条件は，次のようになっている．対話が行われた場所は，本学情報科学研究科棟正面玄関ロビー，被験者とロボットの距離は約 1[m]，受付案内デモは音声対話とジェスチャによるデモを行った．音声入力は，受付案内ロボット ASKA の前にあるマイク台に固定されたハンドマイクで行っている．

対象となった被験者はオープンキャンパスに参加した一般人 71 人（男性 48 人，女性 23 人），年齢はおよそ 10 歳代〜50 歳代であった．非言語情報は，顔向き，視線，瞬きに注目して計測した．本研究では，これらの非言語情報を表 3.2 のように定義した．

予備実験の結果を表 3.3 に示す．この結果は，被験者 71 人中発話前後に顔向きをヒューマノイドロボットに向けて発話した被験者が 32 人，視線をロボットに向けて発話した被験者が 45 人，発話開始前後に瞬きをした被験者が 17 人いたことを表す．表 3.2 の非言語情報以外に，実験で観測した特徴を以下に示す．

表 3.2 非言語情報の定義

顔向き	顔方向をロボットに向ける
視線	視線をロボットに向ける
瞬き	まぶたの開閉

- 発話前までは視線が外れる
- 下方向を見る被験者が多い
- 発話時は顔方向・視線が安定する

表 3.3 予備実験結果

	顔向き	視線	瞬き
標本数 [人]	41	55	17
割合 [%]	57.7	77.5	23.9

考察 予備実験の考察として，発話開始時に視線を向ける割合が最も高く，人との対話と同様にロボットと対話する場合でも対話の相手に視線を向けることがわかった．視線と顔向きを比較すると，顔向きの割合のほうが低くなっているのは，マイクの位置と被験者の立ち位置の関係が原因と考えられる．瞬きと発話開始時期にはほとんど相関が無い．下方向を見る被験者が多かった結果に関しては，今回の実験はオープンキャンパスという実環境に近い環境で行ったため，人前のマイクや説明書きに対して注目したことが原因であると考えられる．発話前までは，視線が外れることに関しては，人間は物事を考えるときに視線を逸すという心理学的行動であると推測される．

この実験で最も興味深かった結果は，方向にばらつきはあるものの，ほぼ全員の被験者が発話の瞬間には顔方向と視線の方向が安定させていた事である．この結果は円滑な対話システムの実装に有効であるといえる．

3.4.5 非言語情報計測システムによる計測実験

前節で行った予備実験により，発話開始時期に被験者の顔方向と視線が安定している結果が得られた．しかし，予備実験での計測は目視で行ったため定量的な評価では無く，計測システムを用いた定量的に計測する必要がある．そこで，実験室環境下において，人とヒューマノイドロボットとの対話時における顔情報計測実験を行った．本実験の目的は，人とロボットの対話状況での人の非言語情報の現れ方を計測し，実験結果から対話での非言語情報の現れ方を定量的に評価することである．

実験手法

図 3.13 に示すような屋内の環境において，被験者が受付案内ロボットと対話を行い，その時の非言語情報（顔方向，視線，瞬き）を計測する実験を行った．発話区間を特定するためマイクから音声情報を録音し，ステレオカメラの多重化した画像と共に miniDV テープに保存した．

実験は被験者 10 人（本学学生）に対して行い，前準備として固定文 5 文・自由文 5 文程度用意してもらった．質問文を書いた紙を見ることを防ぐため，被験者には予め質問文を暗記してもらった．文章を暗記することの影響を考慮し，固定文には次に示すようなある程度の自由度を持たせた文章を用いた．実験中の受付案内ロボットは通常のジェスチャ付き音声対話のデモを行う．

- 開始の挨拶 (例：こんにちは，こんばんは)
- 研究室の場所を尋ねる (例：ロボティクス講座はどこにありますか?)
- 研究室の内線番号を尋ねる (例：ロボティクス講座の内線番号は?)
- 施設の場所を尋ねる (例：エレベータはどこにありますか?)
- 終了の挨拶 (例：ありがとう，さようなら)

実験結果

非言語情報の計測実験の結果を図 3.14，3.15，3.16 に示す．図 3.14，3.15，3.16 は，それぞれ被験者 A，B，C の計測結果中の 1 発話分を代表として示す．(a) はマイクに入力された音声データを表し，0 の区間は音声入力が無く 1 の区間は音声入力があったことを表す．(b) は頭部姿勢の pitch 角と yaw 角を示し，(c) は視線の pitch

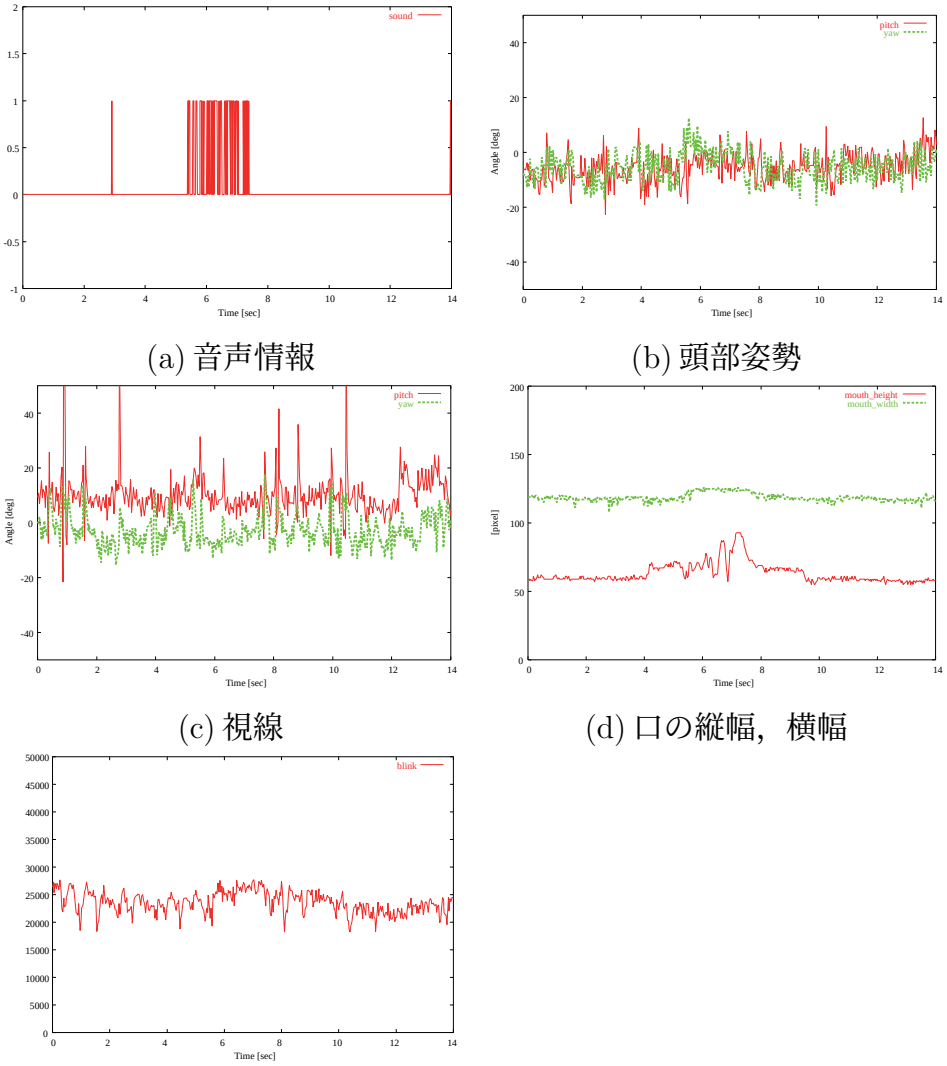
図 3.13 対話実験の様子

角と yaw 角を示す.（d）は口の縦幅と横幅を示す.（e）は瞬き検出処理によって検出した目領域の垂直成分の累積値となり，そのピーク箇所が瞬き示している.

実験結果の特徴

発話中の非言語情報の現れ方の特徴を以下に示す.

- 発話直前に ASKA に対して注目する
- 顔方向に関しては正確に ASKA の方向を向かない
- 発話直前に視線と顔方向は安定している
- 瞬きと発話には関連性がみられない
- 自由対話の時に発話前に視線，顔向きを対象から逸す傾向がある
- 発話直前に頭部の位置が動く
- ジェスチャ中はジェスチャに対して注目する



(a) 音声情報　(b) 頭部姿勢

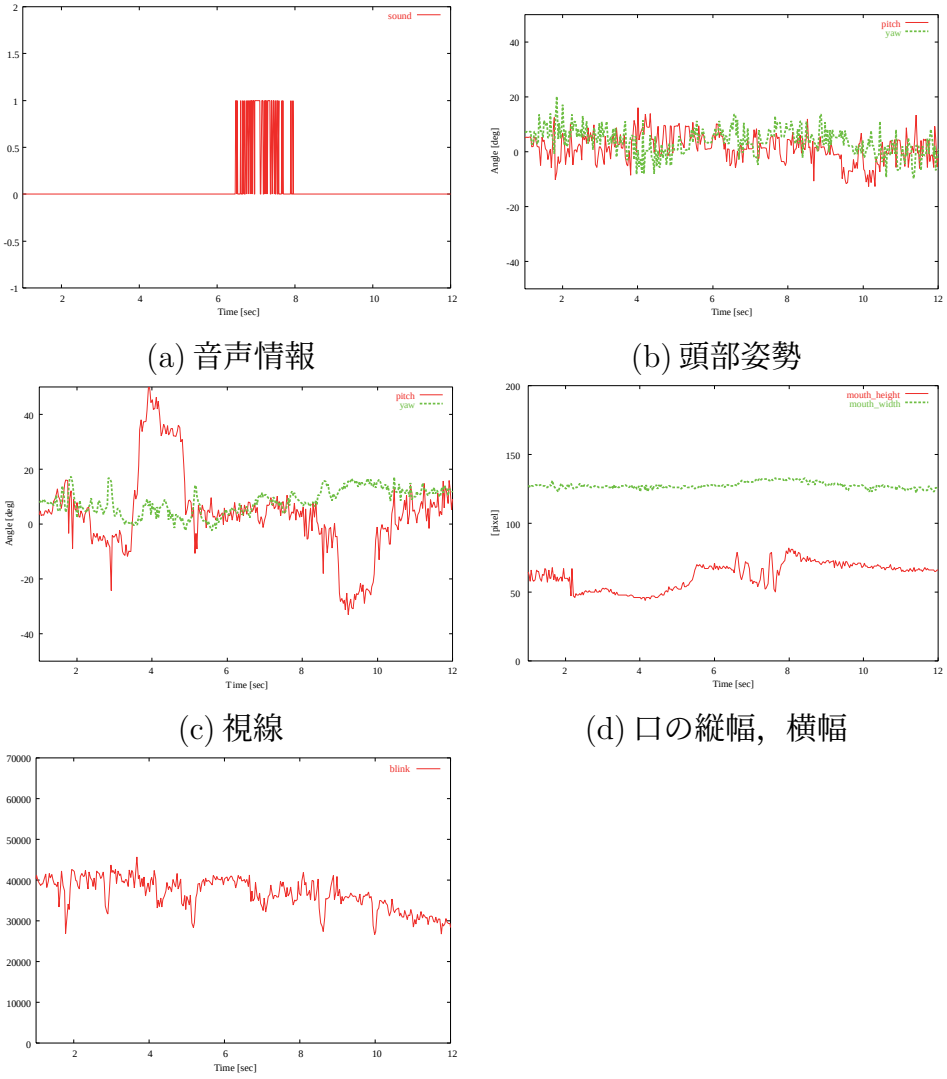
(c) 視線　(d) 口の縦幅，横幅

(e) 瞬き

図 3.14　被験者 A：実験結果（1 発話分）

3.4.6　考察

発話開始の直前には，視線・顔方向ともに角度が 0[度] に近づく傾向があり，このことから発話開始時には視線・顔方向を発話対象である ASKA に向けていることがわかる．また，顔方向に対して，視線のほうが変動が大きい．口唇動作に関しては，口の横幅は発話時にもあまり変動がなく，縦幅の方が変動が多いことから，自然な発音では横幅はあまり変動せず縦方向に動かして発音すると考えられる．瞬きに関



(a) 音声情報　(b) 頭部姿勢

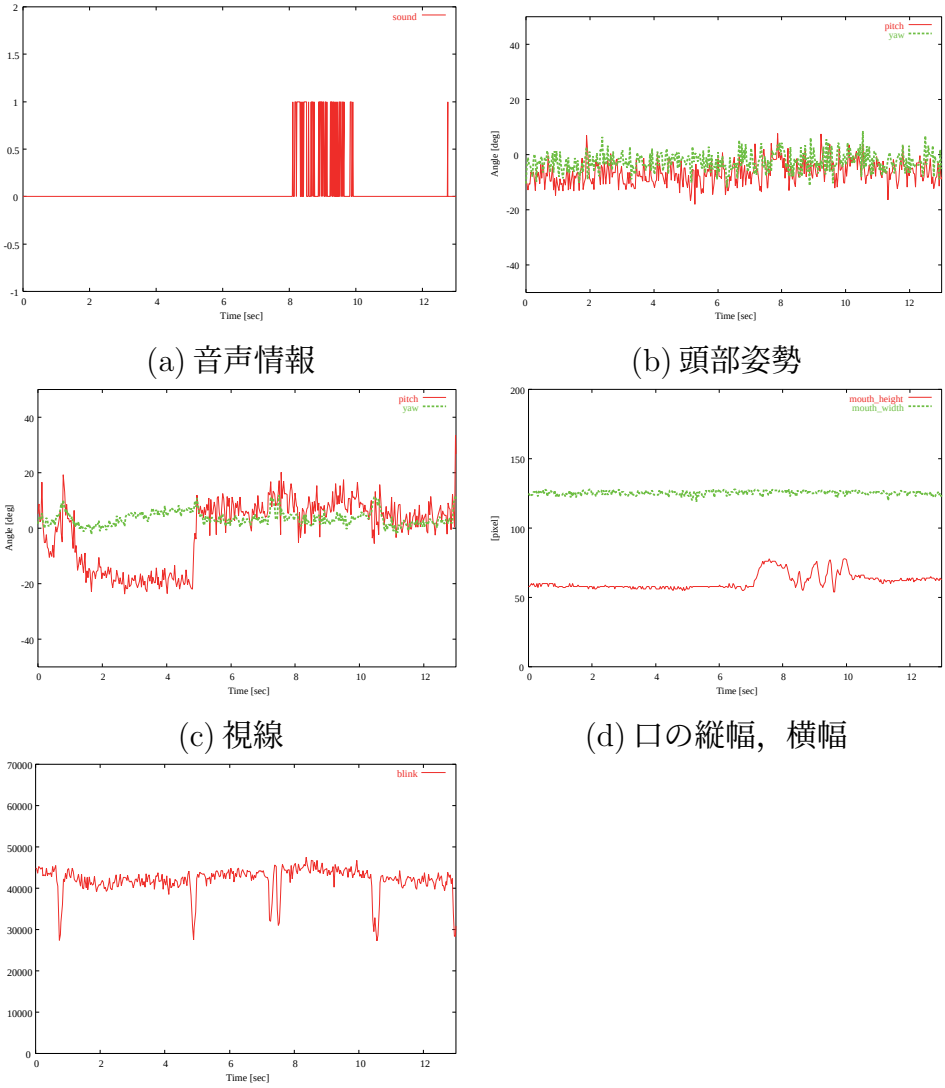
(c) 視線　(d) 口の縦幅，横幅

(e) 瞬き

図 3.15　被験者 B：実験結果（1 発話分）

しては，ほぼ定期的に現れ，発話区間という観点からは関連性は発見できない．

その他の注目すべき特徴として，発話開始時には頭部位置の変動，つまりマイクの間に顔を固定する動作があったこと，固定文による質問の場合は視線の変動が少なかったのに対して，自由文による質問の場合は視線の変動が多かったこと等が挙げられる．後者の固定文と自由文の違いは，固定文は暗記しているのに対して，自由文はその瞬間に文章を考えているため，考えるという行為が視線に現れていると



(a) 音声情報　(b) 頭部姿勢

(c) 視線　(d) 口の縦幅，横幅

(e) 瞬き

図 3.16　被験者 C：実験結果（1 発話分）

思われる．これは，先の予備実験の際にも見られた傾向であり，本ロボットとの対話に関して慣れていないためだと考えられる．

これらの結果から，予備実験での考察をほぼ実証できたと言える．予備実験と被検者による計測実験では，顔方向や視線の安定する方向に違いがみられた．予備実験では，受付カウンタ上の紙や ASKA 後方に設置したディスプレイ等，注目する対象物が多数存在したため，視線を向ける方向に違いがでたと考えられる．

表 3.4 発話認識に用いる非言語情報

	非言語情報	計測量
発話意志の検出	顔	位置・姿勢
	視線	方向
発話終了の特定	口	口の位置・縦横距離

3.4.7 非言語情報を用いた発話区間推定

対話において誤認識や雑音を防ぐためには，対話者の発話意志を読み取り，発話意志に合わせて音声認識をする必要がある．前節で行った対話時の非言語情報計測実験によって，顔向き・視線が発話開始時期の推定に，口唇動作が発話終了時期の推定に有効な情報であることがわかった．そこで本研究では，非言語情報として顔方向と視線を用いて発話意志を特定し，口唇動作を用いて発話終了を特定する発話区間認識システムを構築した．

非言語情報を手がかりに発話区間を認識するにあたって，発話開始点では音声認識が開始している必要があるため，発話者が発話が開始する前には音声認識も開始されなければならない．しかし，実環境で音声認識を行うには，環境の雑音やロボットに対する発話意志の無い音声情報等の問題があり，音声認識の精度が低下する．以上の問題の解決手法として，視覚的に計測できる非言語情報を手がかりに発話意志と発話終了を検出し，発話区間を特定する対話システムを構築した（表 3.4）．

人間の顔や目の動きは意図や興味を表す重要な情報であり，前節の計測結果からも発話時に人間は発話対象に向けて注意を示す傾向があることがわかっている．本研究では，顔と視線が発話対象に向けられている場合に，発話意志があると考え，顔向きと視線の注目点から発話意志を検出する．発話終了の特定に関しては，発話者は口唇動作をすることによって発話するので，口唇動作の終了を発話終了と判断する．この口唇動作の終了は，通常は音声の終了と同時であるが，背景雑音がマイクに入力される状況を考慮して計測を行った．

発話開始点の特定

発話開始時には音声認識を開始するために，人の注意や興味が顕著にあらわれる顔方向と視線を用いて発話意志の認識を行う．以下に処理手順を示す．

1. 顔方向ベクトル，視線ベクトルそれぞれのカメラ平面との交点を計算
2. ステレオカメラ中心からの距離を計算
3. カメラ中心からの距離に応じて注目度を設定
4. 顔方向，視線それぞれの注目度が閾値以上ならば発話開始と認識

注目度の決定

顔方向，視線情報をもとに，人がどれだけ対象に対して注目しているの度合いを注目度という指標で表す．注目度は顔・視線ベクトルとカメラ平面の交点がステレオカメラ中心からどれだけ離れているかで決定される．つまり注目点がカメラ中心位置であるロボットの顔付近であれば，注目度も高い値を示す．

顔，視線の計測データからカメラ中心位置と顔，目の三次元的な位置関係を求め，顔方向，視線方向からカメラ光軸と直交するカメラ平面上の注目点を求める（図 3.17）．カメラ平面上の点 $p_a(x_a, y_a, z_a)$ は三次元位置 $P_s(X_s, Y_s, Z_s)$ と，方向を示す単位ベクトル (V_x, V_y, V_z) から以下のように求める．

$$\begin{aligned} x_a &= X_s - Z_s(\frac{V_x}{V_z}) \\ y_a &= Y_s - Z_s(\frac{V_y}{V_z}) \\ z_a &= 0 \end{aligned} \tag{3.5}$$

ここで，顔方向と視線それぞれ注目点とカメラ中心からの距離を注目の度合い，として考えてみる．この距離が短いほど，ASKA のステレオカメラ中心部分，つまり目の部分を見ていることになるが，実際にどの程度の距離がどの部位の相当するのかを以下に述べる．その距離が 0〜100[mm] の場合は ASKA の顔中心部分，100〜300[mm] の場合は ASKA の顔全体，300〜500[mm] の場合は ASKA の上半身部分を注目している状態となる．また，これ以上の距離の場合は，ASKA を見ていないと考えられる．

発話終了点の特定

前述した用に余分な雑音が入力される可能性を排除するために，口唇動作をもとに発話の終了時期を特定する．人間が発話するときの口唇動作は口の開閉動作の繰

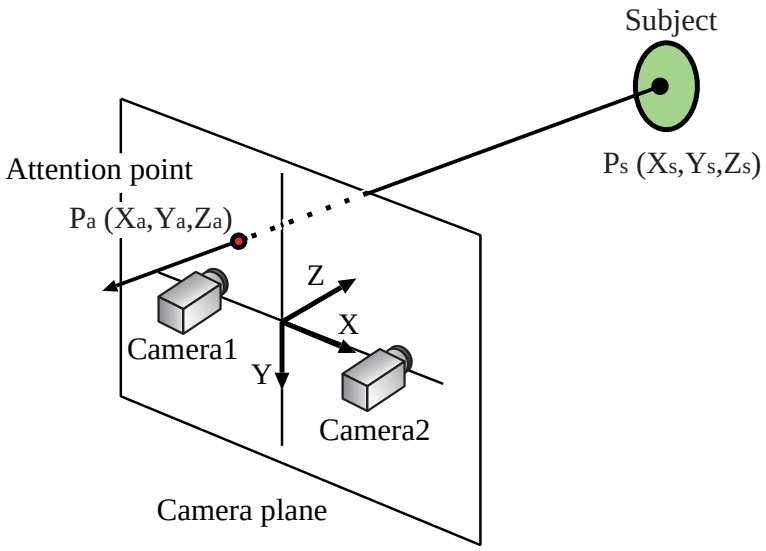


図 3.17 カメラ平面と注目点

り返しであり，それに伴って口の横幅と縦幅が変動する．しかし，口唇動作のうち口の横幅の変動はそれほど大きくなく，縦幅の変位のほうが顕著にあらわれる．また，発話には口を閉じるパターンも含まれており，単純に口を閉じた時点で発話終了という判断はできない．そこで，発話終了点の特定のために，口の縦幅情報を用いて口の変動を検出し，変動状態から終了時期を特定する．以下に発話終了特定の処理手順を示す.

1. 1 フレーム前の口の縦幅と現在の口の縦幅の差分を求める
2. フレーム間の差分値が一定値（計測誤差値）以上であれば，口唇動作と判断する
3. 一定時間縦幅の変動が無ければ発話終了と認識する

発話終了特定処理　口唇動作は口の縦幅の変動であり，フレーム間差分値として計測できる．フレーム間差分を求めることで，口唇動作を検出する．計測された口の縦幅を図 3.18 に示す．録音した音声データより，図中（A）,（C）は発話してない区間，図中（B）は発話区間であることがわかっている．このように計測データには高周波数の計測誤差が含まれているため，常にフレーム間差分値は一定値以上を示す．そこで，計測誤差以上の変位がある場合を口唇動作とみなす．また，発話終了区間 C を認識するために，20 フレーム分の差分の平均を考慮する．

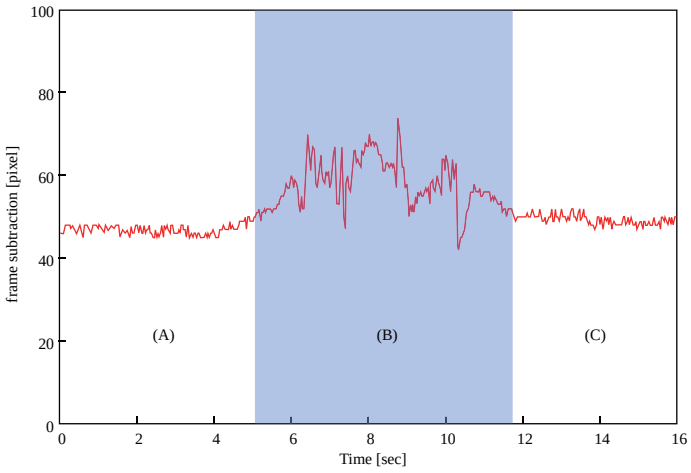


図 3.18 口唇動作

表 3.5 注目度の設定

注目点とカメラ中心の距離 [mm]	注目度
0 〜 100	5
100 〜 200	4
200 〜 300	3
300 〜 400	2
400 〜 500	1
500 〜	0

3.4.8 発話区間認識の評価

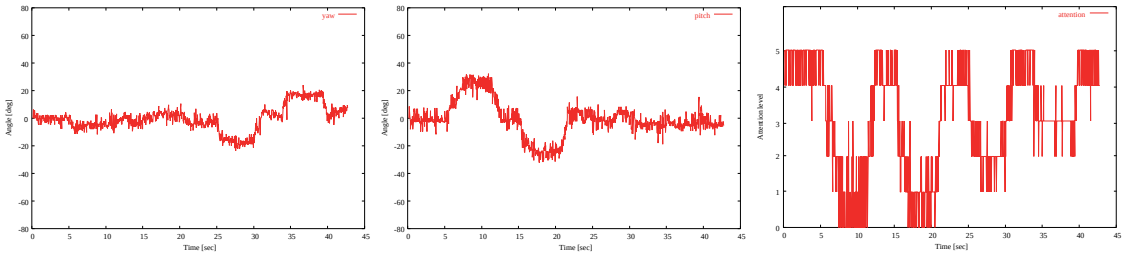
発話意志推定と発話終了特定の評価実験を行った.

まず，注視距離の評価実験として，屋内環境で ASKA に正対し，顔の動きと顔の注目度，視線と視線の注目度を計測する実験を行った．ここで，注目度とは，表 3.5 の用に，注視点から ASKA 顔中心までの距離を離散化したもののラベル値である．顔方向の実験では，最初 ASKA を見ている状態から，顔を右，左，上，下と動かし計測を行った．実験結果を図 3.19 に示す．実験結果から，被験者の左右の顔振りの計測結果を図 3.19（b）に，その後被験者が上下に顔を振った計測結果を図 3.19（a）に示す．そして，図 3.19（c）に顔振りに伴って注目度が 0〜5 の範囲で変動してい

ることがわかる．当然，正面を向いている時に注目度が高く求まっている．

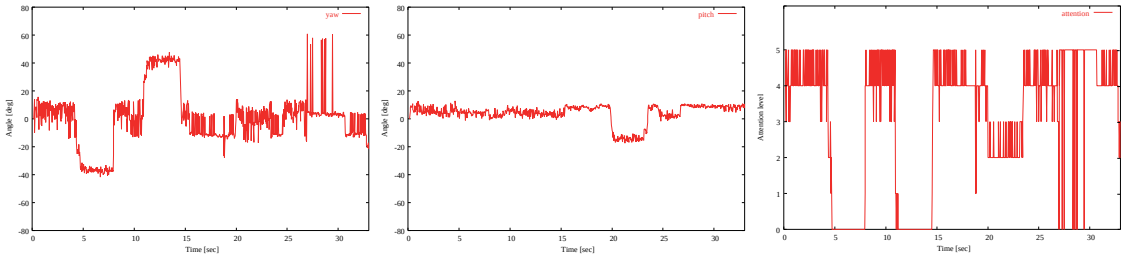
また，視線の実験も同様に，最初 ASKA を見ている状態から，視線を右，左，上，下と動かし計測を行った．実験結果を図 3.20 に示す．実験結果から，被験者の視線の左右の動きの計測結果を図 3.20（b）に，その後被験者が視線を上下に動かした計測結果を図 3.20（a）に示す．そして，図 3.20（c）に視線に伴って注目度が 0-5 の範囲で変動している．ここでも，視線を正面に向けている時に注目度が高く求まっていることが確かめられた．

図 3.21 に，顔向きと視線それぞれの注目度と，各注目度から判定した発話意志認識の結果を示す．図 3.21（a）,（b）にある区間の顔向きと視線により決定した注目度を示し，図 3.21（c）に顔向きと視線の注目度から求めた発話意志を 0-1 の値域で示す．顔向き・視線の注目度がそれぞれ閾値以上の場合に，発話意志が 1（発話意志が有る），顔向き・視線のどちらかの注目度が低い場合は，発話意志が 0(発話意志が無い) と判定できていることがわかる．



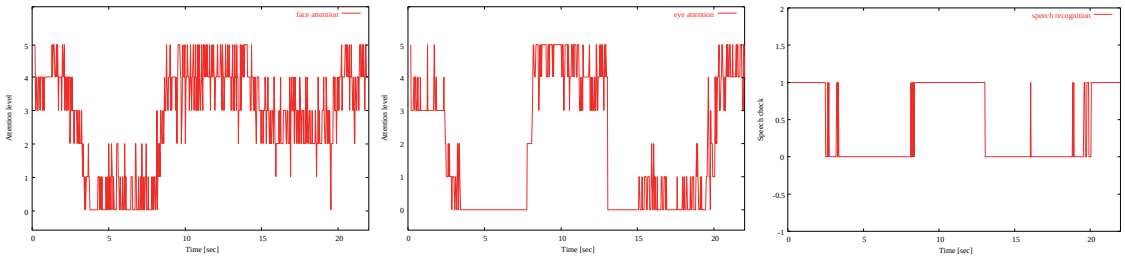
(a) 頭部姿勢 yaw　(b) 頭部姿勢 pitch　(c) 顔方向の注目度

図 3.19　頭部姿勢と注目度



(a) 視線 yaw　(b) 視線 pitch　(c) 視線の注目度

図 3.20　視線と注目度



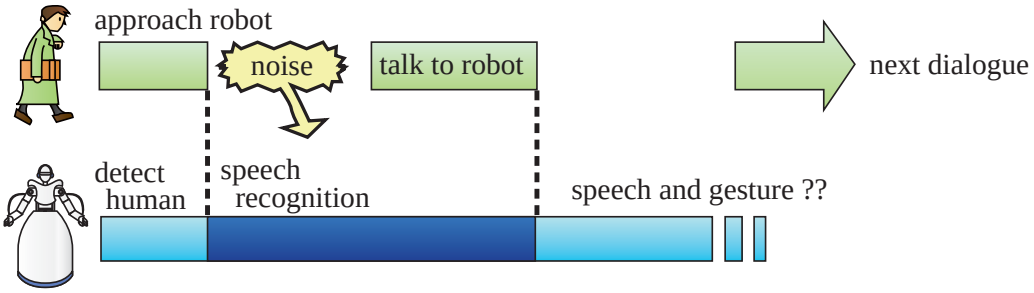
(a) 顔向き注目度　　(b) 視線注目度　　(c) 発話意志認識結果

図 3.21　顔向きと視線の注目度と発話意志認識結果

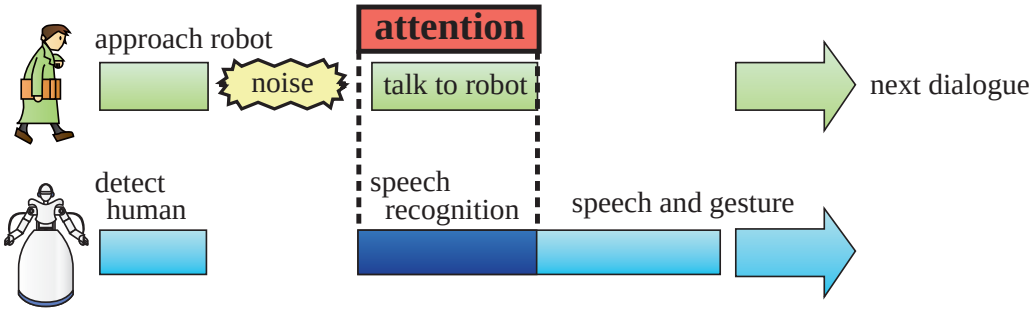
3.4.9　発話区間推定を利用した対話実験

前述のように，発話区間を推定することで，事前にユーザの発話意図を推定し，より適切な対話が可能になる．発話区間推定の有無に対するインタラクションの違いを図 3.22 に示す．発話区間推定を行わない場合を表す図 3.22(a) では，ユーザが受付近くに来た時点で音声認識が開始するため，発話時以外のノイズに対して誤った応答を返してしまう可能性がある．しかし，発話区間推定を行う場合をは，図 3.22(b) のように，ユーザがロボットの顔に注目している場合，つまりユーザが発話意図を持った場合のみ音声認識を行うので，発話以外のノイズを誤って認識してしまう可能性が低い．

また，図 3.23 は，本学受付において対話を行っている様子である．図中の（1），(4)，(7）では，マイクの前に立つ人物が，その隣の人間に対して発話を行っている．発話者の音声はマイクに入力されているが，発話者がロボットに注目していないため，ロボットは自らに向けられた発話ではないと判断して応答を返していない．



(a) without estimation of user's attention



(b) with estimation of user's attention

図 3.22 発話区間推定の有無に対するインタラクションフロー

図 3.23 本学受付での対話の様子

3.4.10 ジェスチャ認識と音声認識の統合

本研究では，対話時における人のジェスチャ情報を認識するために，ステレオ画像を用いたジェスチャ認識手法を提案する．本研究で提案する計測システムとして，人のジェスチャ認識機能を実装している．また，実装の際には ASKA に搭載されている 2 種類のカメラのうち広角カメラを用いている．

3.4.11 ジェスチャ認識手法

本手法は岡らによる連続 DP マッチングを用いたジェスチャのスポッティング認識手法 [48] をベースとして構成されており，ステレオ画像からソフトウェアによって生成した視差画像を用いて，その人物の 3 次元的なジェスチャを連続して認識している．本手法では，連続 DP マッチングに用いる特徴量として，視差画像中の近距離，中距離，遠距離領域の画素数，およびフレーム間差分画像中の変化領域の画素数と，分割された矩形領域との面積比を用いてジェスチャ認識を行う．特徴量のリストとジェスチャ認識の処理手順を以下に示す．

特徴量:

- 視差画像中の近距離領域の画素数と，分割された矩形領域の面積比
- 視差画像中の中距離領域の画素数と，分割された矩形領域の面積比
- 視差画像中の遠距離領域の画素数と，分割された矩形領域の面積比
- フレーム間差分画像中の変化領域の画素数と，分割された矩形領域の面積比

ジェスチャの認識:

1. 入力画像から人領域の設定，分割を行う
2. 特徴量を計算して標準パターンの登録，閾値の設定を行う
3. DP を計算し，認識を実行する

この手法は，フレーム間差分に加えて視差画像の近距離，中距離，遠距離の各成分を連続 DP マッチングの特徴量として採用しているため，距離情報の影響を受けることになり，3 次元的なジェスチャを認識することができるという利点がある．

人領域の設定と領域分割

人の存在する領域を分割し，分割された領域毎に特徴量の計算を行うことにより，それ認識器の入力とする．ここで，人の位置が左右や奥行き方向に変化すると，適切な領域の分割位置が変わる事に注意する必要がある．以下に各領域分割法の説明を述べる．

ここでの分割手法は，ステレオ画像から得られる視差画像について，X 軸における各点の明度値の累積を利用することによって人の中心位置を検索し，入力画像内においてジェスチャを行う人が横方向に移動した場合でも問題なくジェスチャ認識を可能とするための画像分割法である．さらに，視差画像全体の明度累計値の合計を計算し，最高点を示す X 軸の点から左右方向へ明度累計値を累算していき，視差画像全体の明度累計値に対して 70[%] を占めるまでの範囲を指定して，その指定された範囲を人領域とし，分割の対象とする．これにより，分割された領域矩形は正方形に固定されず，画像上の人物の大きさに対応して，必要な領域のみが分割の対象となる．以下に，領域の分割手順を述べる．

1. 入力画像の X 軸の各点における明度値の累計を求める
2. 明度値の累計が最も高い値を出力した X 軸の点を求める
3. 最高点から左右方向に一定割合を占めるまで明度累計値を累算する
4. 最も左側の点と最も右側の中点を計算し，この点を中心に人領域を等分割する

この分割範囲を指定する割合は，被験者数名を計測結果の平均において，肩幅が収まるサイズであった 70[%] を設定した．図 3.24 にヒストグラムと設定した分割範囲を示す．ASKA と対話者の距離はマイクの前にいる場合が 1.0[m]，遠くにいる場合が 2.0[m] である．この手法では，最高点からの割合で人領域を設定するので，人が多少遠ざかったり，近づいたりしてもジェスチャ認識時に対応することができる．また，例を挙げると，4×4 では図 3.25 のように分割が行われる．

特徴量の設定

マッチング手法で使用する特徴量について，また特徴検出において行われる処理について説明する．本手法では分割された領域ごとにフレーム間差分と視差画像から得られる距離成分の二種類を用い，各々の領域全体の面積とそれらの特徴が現れた面積との比を特徴量として認識を行う．

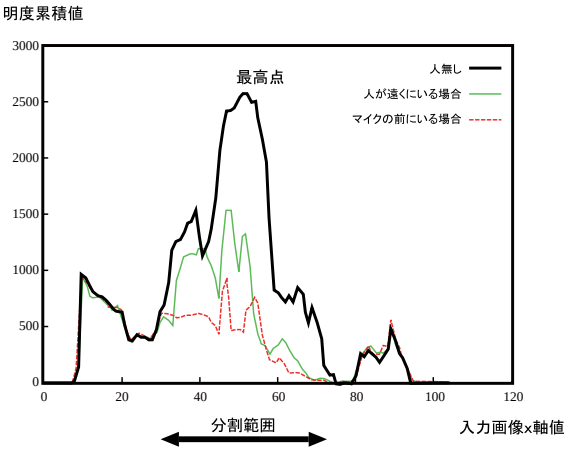
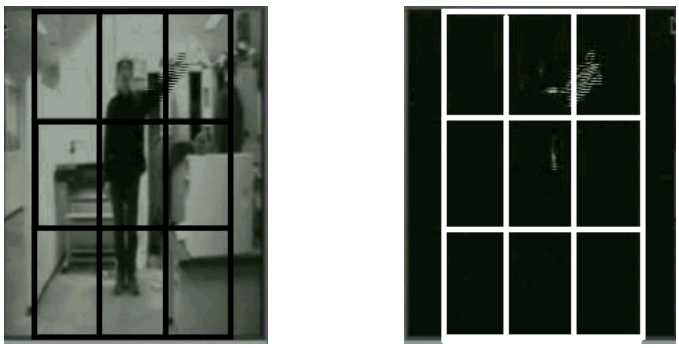


図 3.24　人の有無に関してのヒストグラムの変化

図 3.25　奥行き方向移動に対応した方法で分割された人領域 (4×4)

フレーム間差分　フレーム間差分とは現在の入力画像と 1 フレーム前の入力画像の異なる部分を表したもので，本手法では 256 階調で表現されている現フレームと前フレームにおいて，差分値を 40 以上を閾値として設定して二値化することにより使用している．この情報を用いることにより，人領域の中のどの部位で動作が起こっているかを知ることができる．生成されたフレーム間差分を図 3.26 に示す．



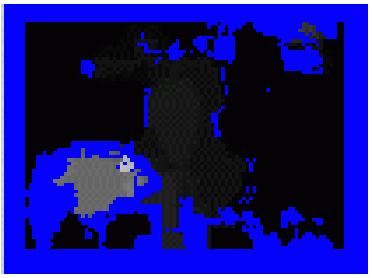
(a) 入力画像　　(b) 二値化したフレーム間差分画像

図 3.26　入力画像と生成されたフレーム間差分画像

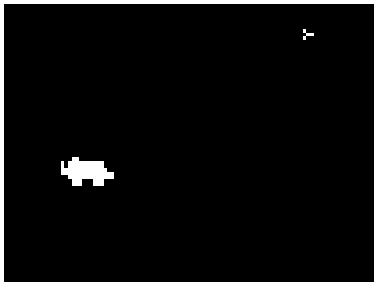
距離情報　本研究で用いる距離情報として生成されている視差画像は，人発見処理で使用したものがベースとなっている．その詳細は，3.4.2 で述べた．また，本手法

では，視差画像により得られた処理情報を近距離，中距離，遠距離の 3 つに分類しており，図 3.27(c),(d),(e) に入力画像に対するそれぞれの距離成分を二値化したものを示す．

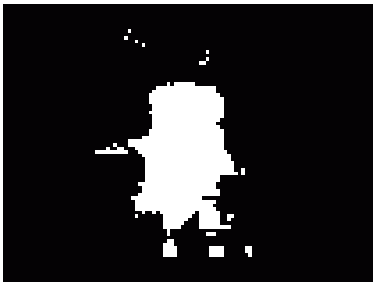
(a) 入力画像



(b) 視差画像



(c) 近距離成分



(d) 中距離成分

(e) 遠距離成分

図 3.27　入力画像，視差画像と各距離成分

マッチング手法

各領域で生成された特徴量で用いて行うマッチング手法について説明する．本手法では，まず入力画像系列に対し標準パターンを作成するのと同様の特徴抽出処理を施し，あらかじめ作成してある標準パターンとの距離をスポッティング整合方式 [48] により計算し，認識結果をフレームごとに出力する．この標準パターンは人間のジェスチャ動作を表現しているモデルであり，始点および終点の定まった特徴ベクトル系列として表現されている．また，標準のパターンはシステムに認識させたい動作の数だけ作成し，その長さはそれぞれのジェスチャで異なっている．以上の準備をもとに，ジェスチャ認識のためのスポッティング整合処理を行う．ジェスチャを

スポッティング認識するためのマッチングとは，始終点の定まっていない系列のある時点がもう一方の系列の終点に対応すると仮定し，それ以前の部分の最適適応を求める方法で，これは系列パターンとその区間の判定を同時に行うことができるということである．そして，このスポッティング認識を実現する具体的アルゴリズムとして連続 DP を用いたマッチングを用いる．連続 DP の出力は登録ジェスチャの個数の時系列となっている．例えば 2 つのジェスチャを登録したときには，2 つの時系列が得られる．そのとき，図 3.28 に見られるようにユーザが登録ジェスチャと類似のジェスチャをし終わったときのみ，該当するジェスチャに対応する連続 DP の値が下がり終わり上昇に転じる．そこで，この連続 DP 値の値が一時的に下落している箇所を検出すれば該当ジェスチャの認識ができることになる．連続 DP の値は常に出力しているので，ジェスチャはいつ始めても，またいつ終わってもよく，さらに登録ジェスチャ以外のものを入力しても反応，つまりへこみができないのでそれらは無視されるという利点がある．

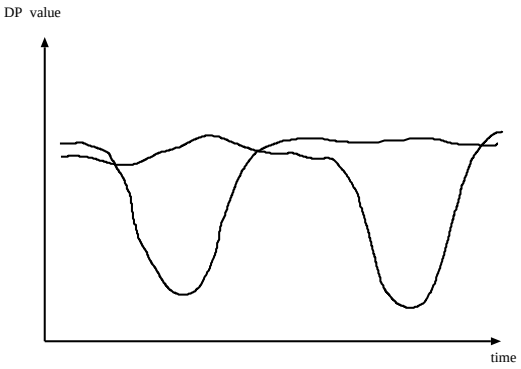


図 3.28 連続 DP 値の谷

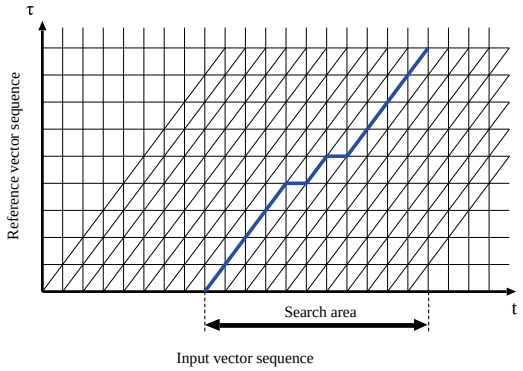


図 3.29 連続 DP の探索範囲

1 つの標準パターン Z を，標準動作をとらえた T フレームの動画像から得られる特徴ベクトル $\mathbf{z}_\tau$ の系列

$$Z = \{\mathbf{z}_\tau | \, 1 \leqq \tau \leqq T\} \tag{3.6}$$

で表す．ここで，特徴ベクトル z_τ はその次元数を N とすると

$$\mathbf{z}_\tau = (z_\tau(1), z_\tau(2), \cdots, z_\tau(N)) \tag{3.7}$$

である．入力画像からも同様な特徴ベクトル系列 $(\mathbf{u_t} 0 \leqq t < \infty)$ が連続的に得られ

る．このとき，$\mathbf{u_t}$ と $\mathbf{z}_\tau$ との局所距離を $d(t,\tau)$ と表記する．この $d(t,\tau)$ の定義の一例を以下に示す.

$$d(t,\tau) = \frac{1}{N}|\,\mathbf{u_k} - \mathbf{z}_\tau|^2. \tag{3.8}$$

ここで，入力，標準パターンの時間軸をそれぞれ t,τ と区別する．

さらに，点 (t,τ) を終点とした標準パターンを入力系列との累積距離を $S(t,\tau)$ で表す．連続 DP では $S(t,\tau)$ を以下のような式で更新する．

初期条件:

$$S(-1,\tau) = S(0,\tau) = \infty \tag{3.9}$$

以下繰り返し: $\tau = 1$ のとき

$$S(t,1) = 3d(t,1) \tag{3.10}$$

$\tau = 2$ のとき

$$S(t,2) = min\begin{cases} S(t-2,1) + 2d(t-1,2) + d(t,2) \\ S(t-1,1) + 3d(t,2) \\ S(t,1) + 3d(t,2) \end{cases} \tag{3.11}$$

$3 \leqq \tau \leqq T$ のとき

$$S(t,\tau) = min\begin{cases} S(t-2,\tau-1) + 2d(t-1,\tau) + d(t,\tau) \\ S(t-1,\tau-1) + 3d(t,\tau) \\ S(t,1,\tau-2) + 3d(t,\tau-1) + 3d(t,\tau) \end{cases} \tag{3.12}$$

となり，出力は,

$$A(t) = \frac{S(t,T)}{3T} \tag{3.13}$$

となる．

ここで，$S(t,T)$ は,

$$S(t,T) = \sum_{\tau=1}^{T} d(Z(\tau), f(t-\beta(\tau))) \tag{3.14}$$

となり，$1 \leqq \tau \leqq T, t \geqq \beta(\tau), \beta(\tau+1) \geqq \beta(\tau)$ の最小値，つまり連続 DP の最短距離であり，最適値に対応している．連続 DP では，各標準パターンが持つその出力 $A(t)$

群の中で閾値以下で連続 DP 値が一時的に下落している箇所をなす時刻のものがスポッティング認識されたジェスチャであるといえ，連続 DP 値が一時的に下落している箇所が検出されたときに，その標準パターンと対応するものが入力パターン系列中に区間として決まることになる．

また，各時刻 t において連続 DP の出力値を与える探索範囲を図 3.29 に示す．ここで縦軸は標準パターンの時間軸を示し，横軸は入力の動画像系列の時間軸を示し，$\tau = 1$ から $\tau = T$ に至る経路が標準パターンと入力パターンとの対応関係を示す．この経路上の局所距離の和が連続 DP の時刻 t における出力値となる．

3.4.12 ジェスチャ認識の評価実験

予備実験として人のジェスチャ認識の評価を行った．実験のプラットフォームは受付案内ロボット ASKA とした．

実験手法

本手法の有効性を確認するために，4 人の被験者によるジェスチャ認識実験を行った．被験者には予め登録されたジェスチャをそれぞれ 10 回試行してもらい，その中から正しいジェスチャが認識された回数を見ることによって評価を行った．

図 3.30，図 3.31 に登録されているジェスチャについてジェスチャを行っている様子と視差画像を表す．それぞれのジェスチャは ASKA に対して正面を向いて直立している状態を初期状態として，bow は上半身を前へ傾けるおじぎ動作，stop は両手を前に出す停止サイン動作，front は右手のみを正面に出す前方指示動作，side は右手を真横に出す右手指示動作を表す．

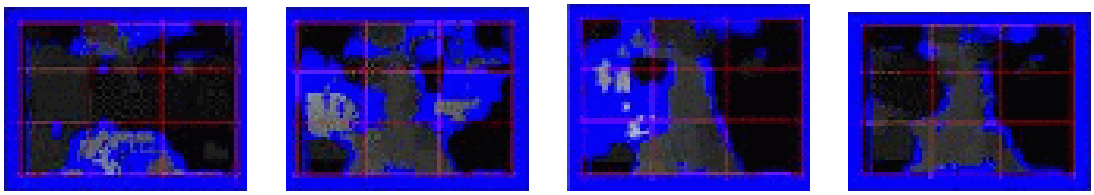
実験結果

各ジェスチャデータにおいては標準パターンに個人平均，画像の分割数は 4×4，画像分割法に 縦横移動に対応した手法を用いている．

認識処理におけるステップで生成される特徴量別の連続 DP 値の時系列データについて，認識を行った各々のジェスチャごとに図 3.32，図 3.33，図 3.34，図 3.35 に標準パターンの値の変化を示す．X 軸は時間 [フレーム]，Y 軸は DP 値を表してお

(a) おじぎ　(b) 停止サイン　(c) 前方指示　(d) 右手方向指示

図 3.30　登録されているジェスチャ



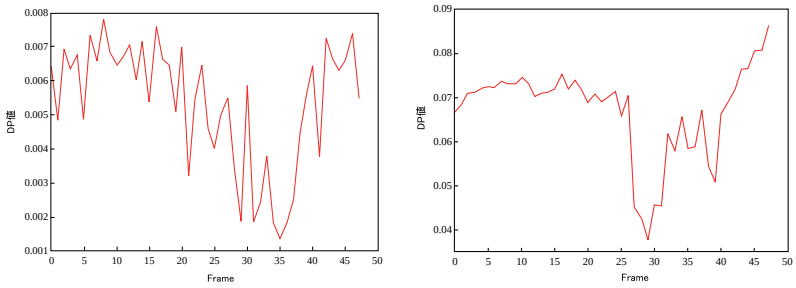
(a) おじぎ　(b) 停止サイン　(c) 前方指示　(d) 右手方向指示

図 3.31　登録ジェスチャの視差画像

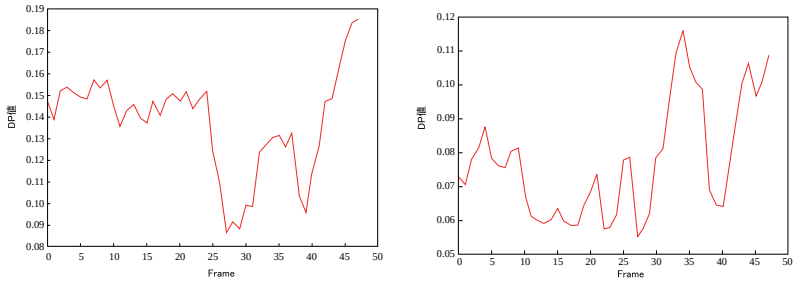
り，それぞれ対応するジェスチャについて，ちょうどグラフの真中辺りの時間でそのジェスチャが行われている．

おじぎ動作を表す図 3.32 については近距離成分の (d) が多少不安定であるが，(a),(b),(c) に関しては 30 フレーム付近において一時的な DP 値の減少が見られる．両手を正面に出す動作を表す図 3.33 については，すべてのグラフにおいて 40 フレーム付近に一時的な DP 値の減少が見られる．右手のみを前に出す動作を表している図 3.34 は，近距離成分 (d) の一時的な DP 値の減少量が少ないが (a),(b),(c) に関しては 40 フレーム付近において一時的な DP 値の減少が見られる．右手のみを真横に出す動作を表している図 3.35 は，ジェスチャの特性上横移動のみであるので近距離成分 (d) について DP 値の一時的な減少は見られない．(a),(b),(c) に関しては 50 フレーム付近において一時的な DP 値の減少が見られる．

連続 DP 値の時系列グラフ全般に言えることとして，図 3.33(a) や図 3.34(a) のような W 型のグラフが多いという点がある．これは指定したジェスチャを行った後にそのジェスチャ状態から初期状態に戻る際の動きが標準パターンと似ているために



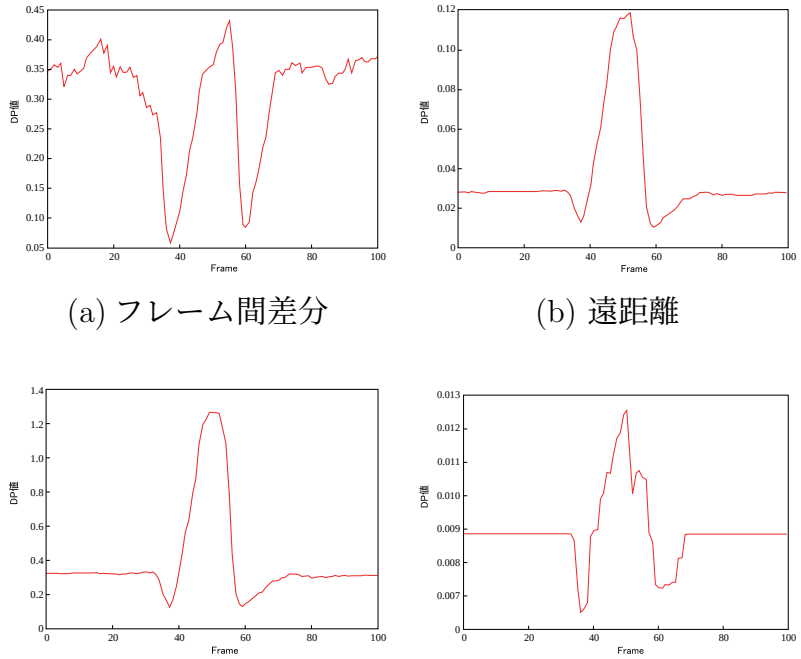
(a) フレーム間差分　　(b) 遠距離



(c) 中距離　　(d) 近距離

図 3.32　おじぎ動作における特徴量別の連続 DP 値の時系列データ

DP 値が減少することによって起こる．また，図 3.35(d) においてジェスチャが行われた際に発生する連続 DP 値が一時的に下落している箇所が見られないという結果が得られたが，これは右手を真横方向に出すというジェスチャの特性上，視差画像から近距離成分が得られないために連続 DP 値が一時的に下落している箇所が発生しなかったと考えられる．さらに近距離成分は他の距離成分と違って標準パターンで登録されてる面積比の値が小さいので，その場合，登録されているジェスチャとは異なる入力が得られた場合に非常に DP 値が大きくなる．



(a) フレーム間差分　(b) 遠距離

(c) 中距離　(d) 近距離

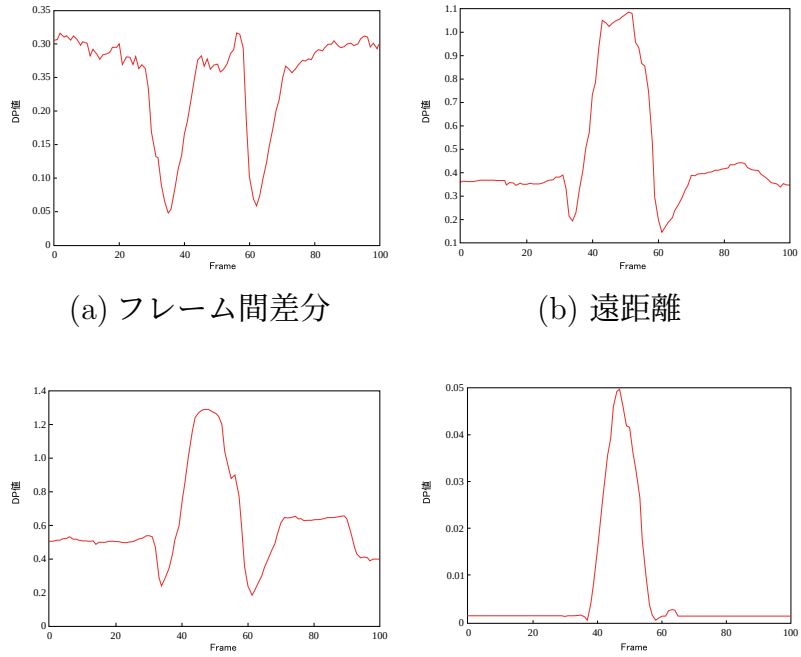
図 3.33　停止サイン動作における特徴量別の連続 DP 値の時系列データ

続いて実験の結果を表 3.6 に示す.

表 3.6　ジェスチャ認識システムの評価実験結果

標準パターン	領域分割法	領域分割数	平均認識率
A) 全員平均	等分割	3×3	44.20%
B) 個人平均	等分割	3×3	63.70%
C) 個人平均	縦横移動に対応	3×3	78.75%
D) 個人平均	縦横移動に対応	4×4	76.88%
E) 個人平均	等分割	3×3	49.38%

標準パターンによる認識率の変化　ジェスチャ認識時に登録する標準パターンとは DP を計算するときの基準となるデータである．まずはじめに，標準パターンをすべての人の平均データを用いる場合と，実験を行う人ごとの平均データを用いる場合とで認識率の変化を比較した．表 3.6（A），(B）に実験結果を示す．この結果から，



(a) フレーム間差分　(b) 遠距離

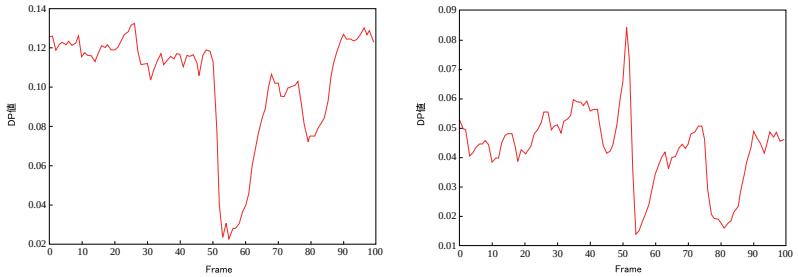
(c) 中距離　(d) 近距離

図 3.34　前方指示動作における特徴量別の連続 DP 値の時系列データ

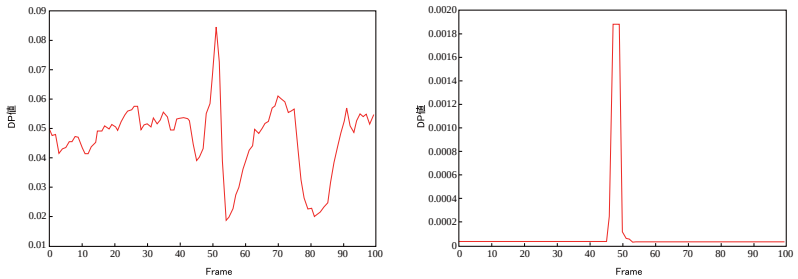
全員の平均を標準パターンに用いると認識率が大幅に低下することがわかった．これは，ジェスチャの個人差が大きいことに起因する．例えば，右側を指す動作でも，前腕を傾ける動作だけで済ます場合もあれば，上腕と前腕を両方動かして大きなジェスチャを行う場合もある．以後，ジェスチャの認識を行う際には個人別の標準パターンを使用する．

人中心基準の領域分割による認識率の変化　画面の分割を行ってジェスチャ認識を行う場合，人の立ち位置によって得られる各領域の情報が大きく変化することが予想される．そこで次に，人の領域を等分割した方法と，横移動に対応した方法，そして縦横移動に対応した方法の認識率の比較を行った．表 3.6（B）,（C）に実験結果を示す．この結果から，縦横の移動に対応した方法で領域分割を行い，ジェスチャ認識を行うと認識率が向上することわかった．

領域分割数と認識率の関係　本研究では，ジェスチャ認識を行う際に，人領域をいくつかの領域に分割している．そこで，ジェスチャ認識を行うのに適した領域分割



(a) フレーム間差分　　(b) 遠距離



(c) 中距離　　(d) 近距離

図 3.35　右方向指示動作における特徴量別の連続 DP 値の時系列データ

数を調べるために，3 × 3 の 9 領域に分割した手法と 4 × 4 の 16 領域に分割した手法との比較実験を行った．表 3.6（D）,（E）に実験結果を示す．その結果から，3 × 3 の領域分割数を採用した場合の手法が 4 × 4 の領域分割数を採用した結果よりも平均認識率が良いことがわかった．しかし，図 3.36 に示すジェスチャの種類別の認識率を見ると，ジェスチャの種類によって認識率に差があることがわかる．このことから，領域分割数は，登録されているジェスチャの種類によって認識率が変化するので，一概にどちらかの優位性を示すことはできないと言える．

従来手法との比較　本手法は岡らによる連続 DP マッチングを用いたジェスチャ認識手法 [48] をベースとしているが，この手法と提案している手法との比較を行った．従来手法では，入力画面全体を人領域を設定し，その人領域を 3 × 3 の等分割している．また，DP マッチングに用いる特徴量として，フレーム間差分画像中の変化領域の画素数と，分割された矩形領域との面積比を用いてジェスチャ認識を行っている．表 3.6（E）に実験結果を示す．この結果から，同様の条件である表 3.6（B）と比較すると，本手法を用いることにより，従来手法の認識率を改善することができたと

言える.

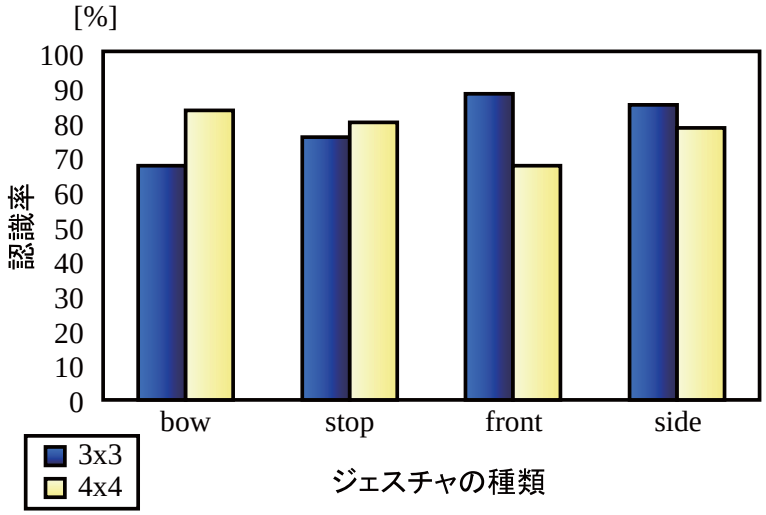


図 3.36 ジェスチャ別の認識率

3.5. 音声とジェスチャを用いた対話

ジェスチャ認識を用いて円滑な対話システムを構築するためには，人の発した音声の内容と行ったジェスチャの内容とを同時に認識し，対応できなければならない．そこで，音声認識のみを用いた対話実験と，ジェスチャ認識と音声認識を用いた対話実験を行い，ジェスチャ認識を含めた対話機能の評価を行った，

3.5.1 音声とジェスチャ情報の処理

対話実験を行うにあたって，対話者の質問に対して適切な受け答えを行うために，入力される音声とジェスチャの情報をどのような手順で処理するのかを決定する必要がある．以下に音声情報のみを用いて対話を行う場合の処理手順と，音声とジェスチャの情報を用いて対話を行う場合の処理手順について示す.

音声情報のみを用いた対話の処理手順

1. 人発見モジュールによってマイクの前の人を検知すると，音声認識モジュールが音声の入力を開始する．
2. ユーザが ASKA に質問する．
3. 音声認識モジュールの結果を見て入力音声に対する応答文を作成，サーバに結果を送信する．
4. 音声合成部は，応答文から合成音声を作成，発話待ちの状態で待機する．
5. 胴体と頭部のジュスチャ部は，応答文などの必要パラメータがサーバに入力されたのを検知し，ジェスチャの動作パターンに基づいて動作を開始する．
6. 音声合成部は，ジュスチャと同時に発話を開始する．
7. ユーザからの発話待ち状態に戻る．

音声情報とジェスチャ情報を用いた対話の処理手順

1. 人発見モジュールによってマイクの前の人を検知すると，音声認識モジュールが音声の入力を開始する．
2. ユーザが ASKA に質問する．
3. ジェスチャ認識モジュールが質問時のユーザのジェスチャを認識する．
4. 音声認識モジュールとジェスチャ認識モジュールの結果を見て入力音声に対する応答文を作成，サーバに結果を送信する．
5. 音声合成部は，応答文から合成音声を作成，発話待ちの状態で待機する．
6. 胴体と頭部のジュスチャ部は，応答文などの必要パラメータがサーバに入力されたのを検知し，ジェスチャの動作パターンに基づいて動作を開始する．
7. 音声合成部は，ジュスチャと同時に発話を開始する．
8. ユーザからの発話待ち状態に戻る．

ここで，音声情報とジェスチャ情報を用いた対話の処理手順の 4. における，音声認識モジュールとジェスチャ認識モジュールの結果を見て入力音声に対する応答文を作成について詳細を述べる．以下におじぎや手振りのような挨拶を意味するジェスチャが行われた場合と，指さしのジェスチャが行われた場合について，音声認識から応答文作成までの処理について示す．

挨拶のジェスチャが行われた場合

1. 認識した音声から得られる音声認識モジュールの応答候補を参照する．

2. 応答候補の中に挨拶をされた場合に応答する候補 (こんにちは，さようなら等) があるかを検索する．
3. 応答候補の中に挨拶をされた場合に応答する候補があったときは，その候補の優先順位が一位でなくとも正式な応答文と設定する．なかった場合は元の応答候補を正式な応答文と設定する．

指さしのジェスチャが行われた場合

1. 認識した音声から得られる音声認識モジュールの認識結果を参照する．
2. 認識結果が方向を表す指示代名詞を含む質問 (あれは何ですか?等) であるかどうかを調べる．
3. 認識結果の中に方向を表す指示代名詞があった場合，ジェスチャが指し示す方向に何があるのか情報を読み出して応答文を作成する．なかった場合は，音声認識モジュールから得られる応答文を設定する．

例えば，右方向に掲示板があるときに，右方向を指さして “あれは何ですか?” と質問したとき，右方向に掲示板があるという情報を ASKA が保持していて，右方向を指さしているというジェスチャを認識し，“あれは何ですか?” という質問を音声認識できた場合，“あれは掲示板です” というような適切な応答を返すことができる．

3.5.2 音声認識のみを用いた対話実験

音声認識のみを用いた対話実験の評価を行った．実験のプラットフォームは，受付案内ロボット ASKA である．

実験手法

本手法の有効性を確認するために，4 人の被験者による対話計測実験を行った．被験者には予め定められた内容の質問文で対話をそれぞれ 10 回行ってもらい，その対話からキーワードを認識できた回数と，正しい受け答えができた回数を見ることによって評価を行った．図 3.37 に対話実験の様子を示す．

以下に ASKA との対話に用いる質問文を示す．

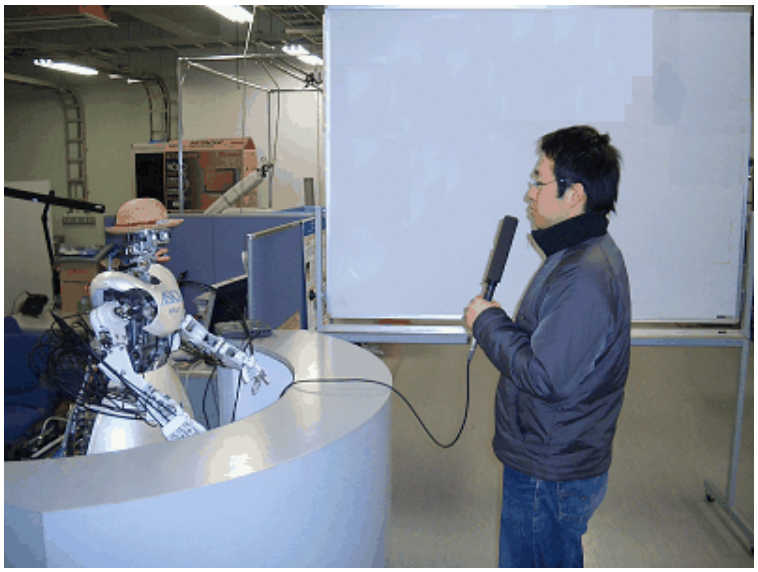


図 3.37 対話実験の様子

質問文

1. こんにちは
2. あれは何ですか
3. 向こうには何がありますか
4. では，向こうには何がありますか
5. ありがとう

実験結果

実験の結果を図 3.38 に示す．Y 軸の番号は実験手法で説明した説明文の番号に対応している．また，音声認識率は話かけた言葉を認識できてきるかどうかを表し，対話成立率は，ASKA が問いかけられた言葉に対して正しい受け答えをすることができたかどうかを表している．音声認識率に関しては，認識率 84[%] という結果を得ることができたが，対話成立率に関しては，質問文の 2,3,4 における「あれ」等が何を指し示すのかを ASKA は判断することができなかったので，当然 0[%] という結果になっている．

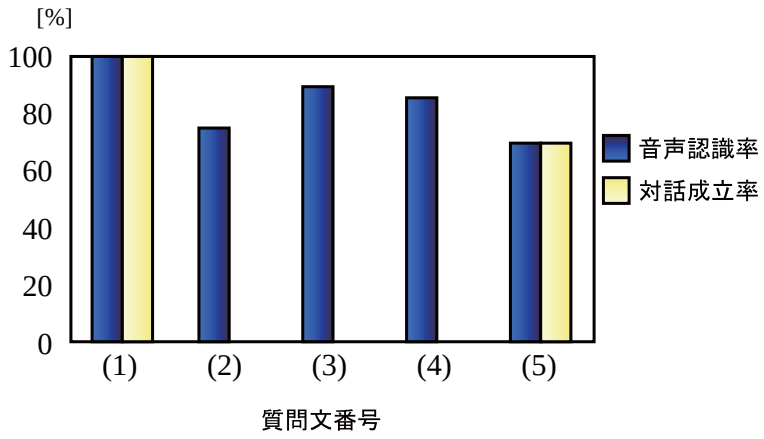


図 3.38 音声認識率と対話成立率

3.5.3 音声とジェスチャを用いた対話実験

次に，音声認識とジェスチャ認識を用いた対話実験の評価を行った．実験のプラットフォームは同様に受付案内ロボット ASKA である．

実験手法

実験の手法は前節で行われた実験に図 3.39 に表すジェスチャを付随するものである．以下に ASKA との対話時に行われたジェスチャを示す．

対話時に行うジェスチャ

1. おじぎ
2. 前方の指さし
3. 右横方向の指さし
4. 左横方向の指さし
5. 手振り

実験結果

実験の結果を図 3.40 に示す．Y 軸の番号は実験手法で説明した説明文の番号に対応している．また，音声認識率は話かけた言葉を認識できてきるかどうかを表し，ジェ

(a) おじぎ　(b) 前方の指さし　(c) 右横方向の指さし

(d) 左横方向の指さし　(e) 手振り

図 3.39　実験で行われたジェスチャ

スチャ認識率は被験者のジェスチャを正しく認識することができた割合を示している．そして，対話成立率は，ASKA が問いかけられた言葉に対して正しい受け答えをすることができたかどうかを表している．

このグラフの質問文 1,5 に注目すると，質問文 1 についてはジェスチャ認識率 60[%]，音声認識率 100[%] に対して対話成立率は 100[%]，質問文 5 についてはジェスチャ認識率 100[%]，音声認識率 65[%] に対して対話成立率は 95[%] という結果を得ることができた．これは，“こんにちは” のジェスチャを認識できていなくても，音声を認識できていれば正しい応答を返すことができ，逆に “ありがとう” については，音声が正しく認識できていなくても，ジェスチャを認識できていれば正しい応答を返すことができた．

また，グラフの 2,3,4 に注目すると，質問文 2 についてはジェスチャ認識率 65[%]，音声認識率 75[%] に対して対話成立率は 65[%]，質問文 3 についてはジェスチャ認識率 90[%]，音声認識率 90[%] に対して対話成立率は 85[%]，質問文 4 についてはジェスチャ認識率 85[%]，音声認識率 85[%] に対して対話成立率は 85[%] という結果を得ることができた．傾向としては，対話成立率はジェスチャ認識率，音声認識率と同様か，やや低い値を示した．

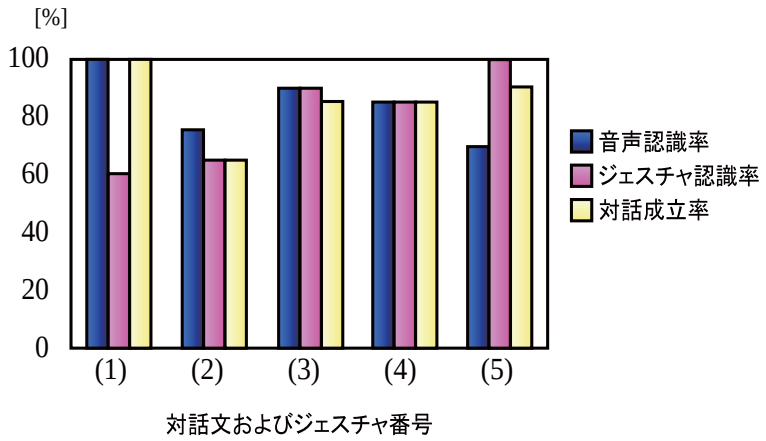


図 3.40 音声認識率，ジェスチャ認識率と対話成立率

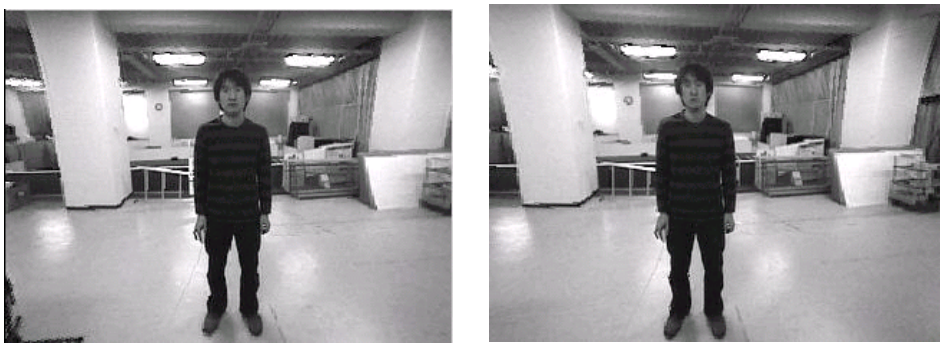
3.5.4 考察

図 3.40 に示した実験結果から，質問文 1,5 のような挨拶を受ける場合について，音声認識とジェスチャ認識は互いの機能を補完し合って応対することができていることがわかる．質問文 2,3,4 について考えると，これは挨拶を受ける場合と異なり，音声認識とジェスチャ認識が同時に正しく行われた時にのみ正しい応答を返すことができる対話内容のため，対話成立率が低めになっていると考えられる．

これまでの ASKA の対話システムでは，ステレオカメラから入力される視覚情報を処理するシステムと，音声認識，音声発話を行うシステムの連係が行われていなかったため，“あれ”，“むこう” というような方向や物を指し示す指示代名詞が含まれる対話内容に対して適切な応答をすることができなかった．この問題を解決するため本システムでは，対話の応答を行う際にジェスチャ認識の認識結果と音声認識の認識結果を参照した．

3.6. 本章のまとめ

本章では，自然なヒューマンロボットインタラクションの実現を目的として開発された，受付案内ロボット ASKA について述べた．ASKA のシステムは，互いに独立して動作する音声認識，音声合成，人発見，人計測，胴体ジェスチャおよび頭部ジェスチャ部のモジュールと，サーバから構成される．本システムは，各要素技術の



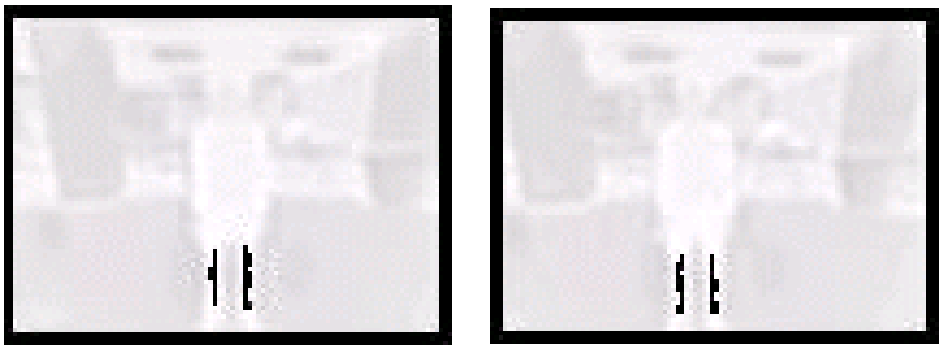
(a) 左カメラ画像　　(b) 右カメラ画像

図 3.41　キャプチャ画像 (312×234)

(a) 左カメラ画像　　(b) 右カメラ画像

図 3.42　平滑化画像 (104×78)

実証実験のためのプラットフォームとして構築され，人とコミュニケーションを行うための視覚計測機能と対話機能が実装された．視覚を用いた計測機能には，ステレオ視による距離計測および顔情報計測機能がある．視差画像により人発見を行い，その距離情報を特徴量とする DP マッチングによって，三次元ジェスチャを認識し，より自然な対話を実現した．また，顔情報を計測する事により発話者の発話区間を推定し，従来認識が困難であった状況でも音声認識を行うことが可能になった．音声対話機能は，大語彙連続音声認識エンジン Julius をベースとしたキーワードマッチによる音声認識理解部と音声合成部で構成され，自然発話の認識を実現している．また，事前に規定された文章だけでなく，ネットワーク上から情報を取得して応答を返すことが可能になっている．



(a) 左カメラ画像　　(b) 右カメラ画像

図 3.43　LoG フィルタ後の画像 (104×78)

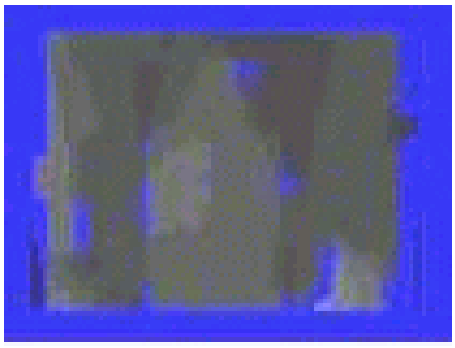


図 3.44　視差画像 (104×78)

第4章　顔ロボットを用いたインタラクション

4.1. 遠隔コミュニケーション

従来，工場などでの限定された環境での作業を想定した産業用のロボット研究・開発は，近年のPCの発達にともない，エンターテインメントやコミュニケーションなどを目的とした共存型ロボットの研究・開発へと広がりを見せている．人との深い関わりが求められる共存型ロボットにおいて，自然で円滑なコミュニケーションの実現は重要な課題となっており，顔ロボットにおける表情の生成など，対話の円滑化を目的とした研究も盛んに行われている．

一方で，近年のネットワークインフラの急速な普及に伴い，遠隔地にいる人との情報伝達や情報共有の方法において，電子メールやチャットを利用した新しい遠隔コミュニケーションシステムが広がりを見せている．たとえばチャットや電話など，従来文字や音声に頼っていたコミュニケーション手段に対して，ビデオ画像やアバターなどの顔情報伝送機能が加えられるようになってきた．コミュニケーションにおいて，非言語情報はときに音声情報以上の情報を伝達する [49] と言われ，メッセージの6割〜9割を非言語情報が占めるという研究結果もある [50]．中でも，円滑なコミュニケーションには，表情から得られる情報が特に重要であると言われている [51]．

従来の遠隔コミュニケーションシステムにおいて，表情を伝達するためにオペレータをカメラで撮影しその映像を直接転送する方法がある．これは最もシンプルな方法であり，古くから映像の圧縮技術についての研究が主に行われてきた．その結果，現在ではMicrosoft Windows MessengerのようにOSに標準的に搭載されるほど普及している．この方法では，自然なコミュニケーションが実現できる反面，顔や背景がそのまま相手に送信されるためプライバシーが守られないという欠点を持っている．

そのため，オペレータの表情を認識し本人以外のもので再現し遠隔コミュニケーションを行う方法が提案されてきた．大場らは，オペレータの表情を認識し，あらか

じめ用意しておいた別の人間やキャラクターに同じ表情を投影している [52]．Marc らは表情のみでなく全身をアバターで再現している [53]．また，R^3（アールキューブ）構想 [54] のようにロボットを遠隔コミュニケーションのツールとして利用する研究も行われており，オペレータの動作をロボットに投影しコミュニケーションを試みる研究も進められている [55, 56]．これらの方法はあらかじめモデルを用意し，そのモデルを動かすパラメータのみを伝送すればよいため，映像をそのまま伝送する方法と比べ大幅に情報伝送量を減らすことができるという利点がある．

多くのシステムは基本 6 表情（驚き，恐怖，嫌悪，怒り，喜び，悲しみ）を認識して，その結果を伝送し相手側の端末で表情の再現を行うものが多い．そのため，人間の顔の動きを離散的に認識することになり自然な顔の動きを伝達できなかった．表情を認識するのではなく，顔の動きそのものを計測する研究も行われている．Elagin[57] らは単眼カメラを用いて顔を含むビデオ画像から顔の動きをトラッキングしているが，顔情報を 2 次元データとして計測しているため，顔位置の並進・回転を正確にトラッキングすることが不可能である．

本章では，これまでのアバターやビデオチャットの問題点を解消し，さらに実存感のある遠隔コミュニケーションを実現するために，ロボットを遠隔コミュニケーションのメディアとして利用することを提案する．提案する遠隔コミュニケーションシステムは，ステレオカメラシステムを用いてユーザの頭部・眼球・眉毛・口唇の各器官の運動をそれぞれ 3 次元計測する．計測されたユーザの顔器官の運動を，表情生成可能なロボットに投影することにより，頭部の並進・回転に対応した表情の再現・伝送を行う．以下の章では，表情提示ロボットを介した遠隔コミュニケーションシステムの詳細について延べた後，印象評価実験に基づき，提案システムのコミュニケーション能力について評価する．

4.2. ステレオカメラを用いた顔情報計測

Ekman らは表情を測定する方法として顔面行動分類法 FACS(Facial Action Coding System) を提案した [58]．また Birdwhistell は FACS を用いて，人間の表情を外から観察できる表情筋の動きを 44 の顔面動作単位である AU(Action Unit) によって定義し，これらの組み合わせで様々な表情を表せることを示した [59]．

これらの AU のうち，ステレオ画像から 3 次元計測可能な眉毛，口唇の運動に着目する．また，表情だけでなく頭部運動によるジェスチャーも計測する．頭部運動に

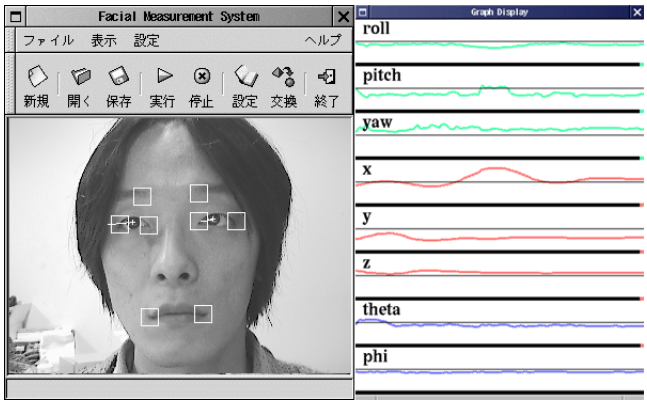


図 4.1 顔情報の計測

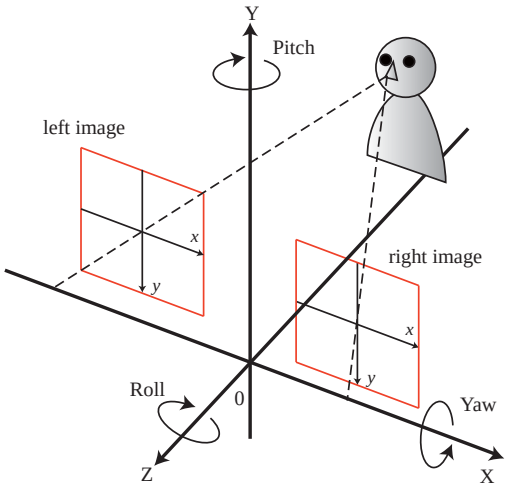


図 4.2 顔情報に関する座標系

よるジェスチャーの代表的なものとして首をかしげる “疑問”，うなずく “同意”，横に振る “否定” がある．さらに，アイコンタクトなど視線情報はコミュニケーション時には重要な因子であることから，視線情報の計測も行う．

我々は，これまでにステレオビジョンを用いることで顔と視線の方向を検出することができ，非接触，受動的，リアルタイム，ロバスト，コンパクトという利点をもつ顔トラッキング・視線計測システム [60] を提案してきた．このシステムは，ユーザに非接触であるため，センサの接触によりユーザへ負担を与えることなく，自然な動作を検出することができる．また，一定照明条件下の実験において大きな変形や隠れがない場合，顔の位置について ± 1[mm]，姿勢について ± 1[deg]，視線については ± 3[deg] 程度の計測精度を実現している．図 4.1 に，計測結果の一例を示す．矩形は視覚追跡するための顔の特徴の位置を示し，2 つの直線は検出された視線方向を示している．

本研究では，表情を提示するために，従来の顔トラッキング・視線計測システムに眉毛の位置と口唇の形状の計測機能を追加する．以下に眉毛，口唇形状計測方法の詳細について述べる．

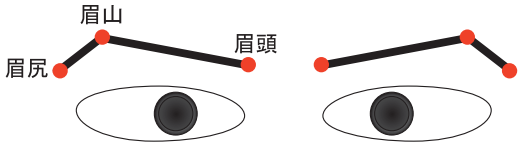
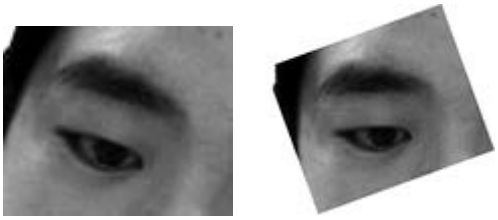


図 4.3 眉モデル



(a) カメラ画像　(b) 回転後の画像

図 4.4 眉画像のアフィン変換

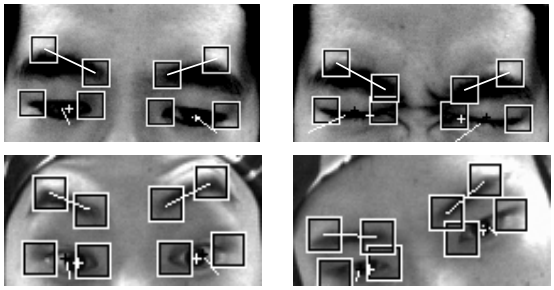


図 4.5 眉位置検出結果

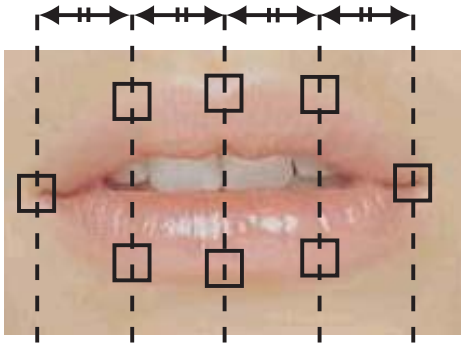
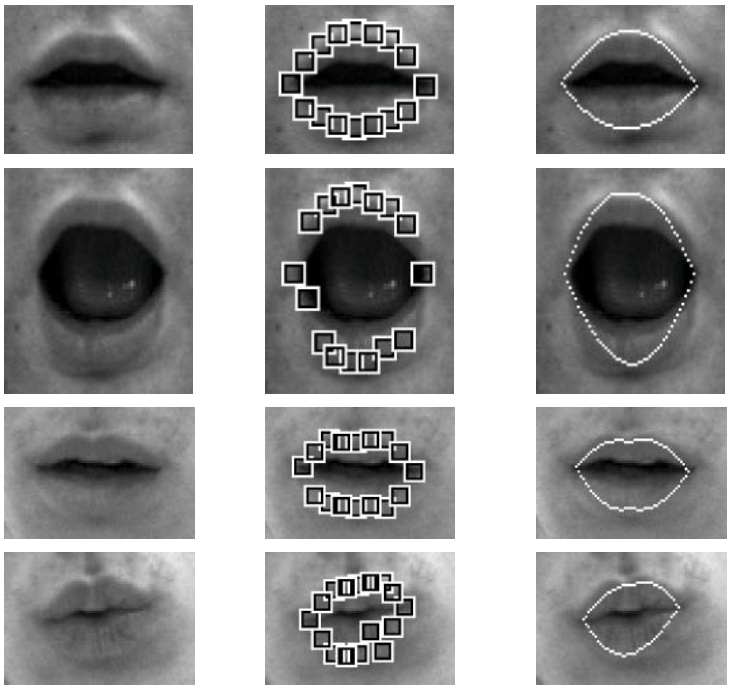


図 4.6 口唇形状計測のためのテンプレート

4.2.1 眉毛の位置計測

本研究では，眉毛を図 4.3 のように眉頭・眉山・眉尻をつないだ折れ線として定義する．ただし，眉尻は眉山との相対位置がほとんど動かないことから眉頭・眉山を計測することで眉毛の位置の計測とする．眉山と眉頭の画像を，あらかじめテンプレートとして登録しておき，左右のカメラ画像からテンプレートマッチングによりその位置を検出する．ステレオ画像上でのそれぞれのテンプレート位置を用い眉頭と眉山の 3 次元位置を求める．眉毛位置の計測手順を以下に示す．

1. 左右の眉毛の眉頭と眉山をテンプレートとして登録する．
2. 顔画像の傾きを補正する．
3. 目の両端位置から眉頭・眉山の探索範囲を設定する．
4. 各探索範囲内でマッチングを行う．
5. 1.〜4. を左右のカメラ画像両方で行い，眉頭と眉山の 3 次元位置を計測する．



(a) 入力画像　(b) トラッキング結果　(c) 曲線補完結果

図 4.7 口唇形状の検出

図 4.2 に示すような顔情報計測座標系において，顔の Yaw 回転 (頭部うなずき方向の回転）や Pitch 回転（頭部鉛直軸回りの回転）は，特徴点がカメラに撮影されている場合マッチング結果にほとんど影響を与えない．しかし，頭部が Roll 回転（カメラ光軸回りの回転）している場合，図 4.4（a）のように，カメラ画像にも頭部は回転されて投影される．そのためテンプレート登録時の頭部の姿勢と大きく異なる場合には，相関値が低下して，テンプレートマッチングが失敗してしまう．そこで，顔トラッキング・視線計測システムから得られた頭部の姿勢を用いて，眉毛周辺の画像に対してアフィン変換を行う（図 4.4（b)）．図 4.5 に眉毛検出結果を示す．頭部を回転した場合や眉毛を上下させたときも，正しく追跡できていることがわかる．

4.2.2 口唇形状計測

本研究では，唇輪郭上の特徴点をテンプレートマッチングによりトラッキングし，トラッキングした特徴点に曲線をフィッティングすることにより，口唇形状を検出す

表 4.1 頭部位置および顔部品の可動範囲

		頭部位置	眉頭	眉山	上唇点	下唇点	右口角点	左口角点
自由対話	平均	0.7	0.8	0.7	0.7	0.8	0.7	0.7
	最大	13.1	8.3	7.4	14.5	10.4	10.1	9.0
誇張表情	平均	0.4	0.6	0.5	0.5	1.1	0.7	0.7
	最大	13.1	12.7	6.8	9.7	27.2	8.2	5.7

単位 [mm]

る．通常のテンプレートマッチングでは，特徴の少ない唇輪郭上の点を安定して検出することは困難である．そこで，まず特徴的で位置変化の少ない唇の両端を見つけ，図 4.6 に示すように両端位置から探索範囲を細かく分割することで残りの特徴点の安定検出を可能にした．口唇動作の検出の処理手順を以下に示す．

1. 唇輪郭上の特徴点をテンプレートとして登録
2. 口の両端の探索範囲を設定
3. 口の両端をマッチング
4. 口の両端位置から特徴点の探索範囲を設定
5. 各探索範囲内でマッチング
6. 特徴点の間を補間

口の両端位置から他の特徴点の探索範囲を設定する場合，図 4.6 に示したように唇の両端の間を均等に分割している．しかし，顔が Pitch 回転をしていると，射影的なゆがみが発生するため，唇の両端を均等に分割しても唇の中心を求めることはできない．そこで，ゆがみを含む分割割合を計算し分割距離を決定している．顔の Roll 回転に対しても眉毛位置の計測の時と同様に，顔トラッキングシステムから得られた値を用いて，傾きを補正する為のアフィン変換を行う．図 4.7 に唇の運動検出結果を示す．

3 名の被験者に対して，以下の 2 条件下での顔特徴の 33[msec] 毎の変位を光学式モーションキャプチャシステムにより計測した．

- 2 分間の自由対話
- 眉毛および口唇を大きく動かす誇張表情動作

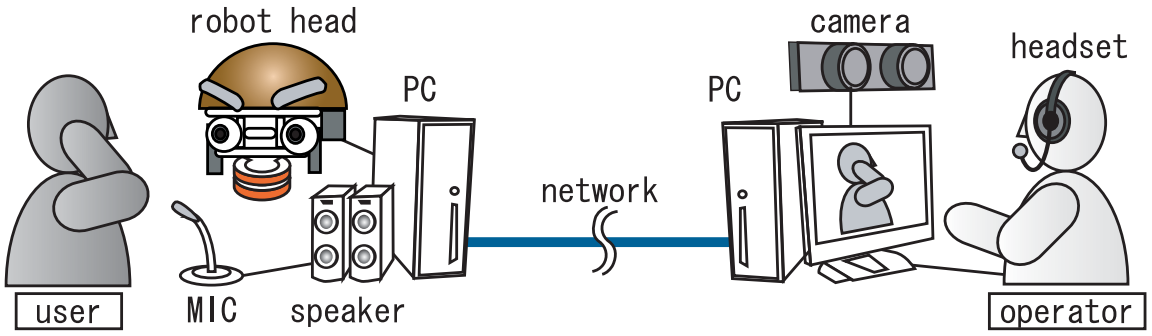


図 4.8 遠隔対話システム概要

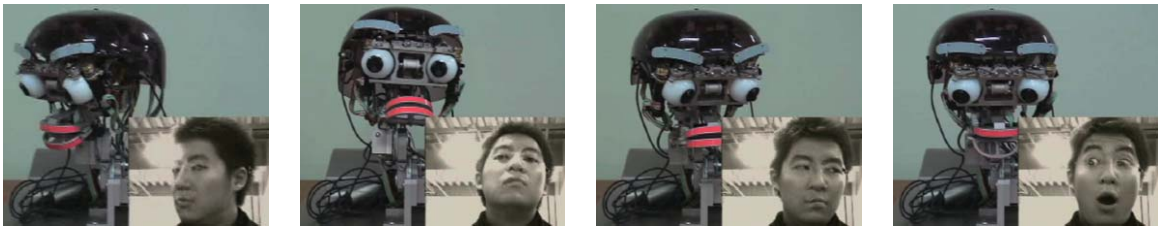


図 4.9 ロボットへの顔情報付加

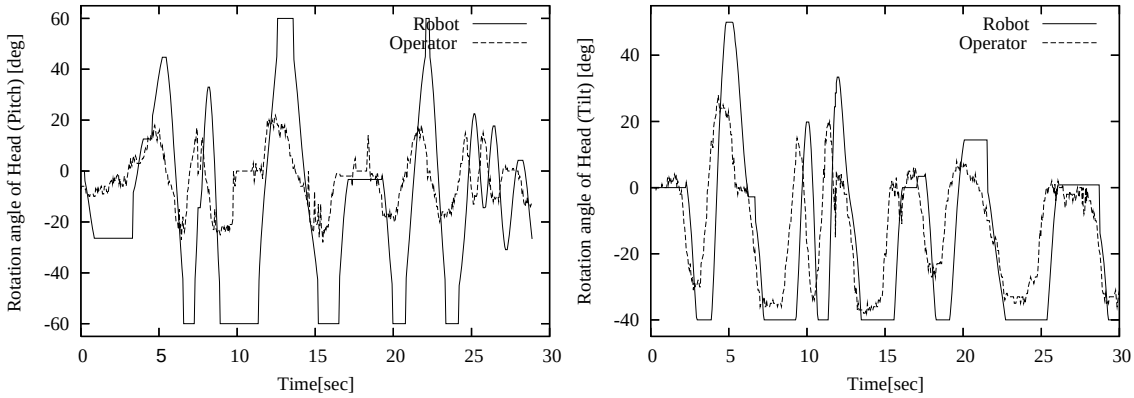
計測結果を表 4.1 に示す．この表より，人間の自然な対話時にはどの部品も平均で 0.7[mm] 程度，最大では上唇点の 14.5[mm] の変位となっている．誇張表情の場合でも，変位の最大値は下唇点の 27.2[mm] である．今回の実装では，焦点距離 12[mm]，画角 25.5[deg] のレンズを使用しており，カメラから 600[mm] 離れた位置に顔があるとすると，下唇点の最大移動量 27.2[mm] に対して必要な探索範囲は，24.3[pixel] となる．今回，提案実装した顔情報計測システムでのテンプレート探索範囲は最小でも 32[pixel]×44[pixel] であることより，照明条件の急激な変化やテンプレートの隠蔽等が生じない限り，対話時の顔部品をロバストに探索するには十分な探索範囲であると言える．

4.3. 顔ロボットの表情生成

前節で述べた手法により，各画像キャプチャフレーム毎に顔部品の 3 次元位置が計測される．この結果をネットワークを介してロボット側に同時に伝送し（図 4.8），

表 4.2 計測結果のロボットへのマッピング方法

ロボットの動作	オペレータの顔情報
口唇の開閉	上唇点・下唇点間距離
口唇の傾き	口角点間の距離
眉毛位置	眉毛特徴点位置 （ローパスフィルタによるノイズ除去済み）
顔向き	計測された頭部回転角度（上下・左右方向） （提示角度は計測角度の 2 倍）
視線方向	計測された視線角度

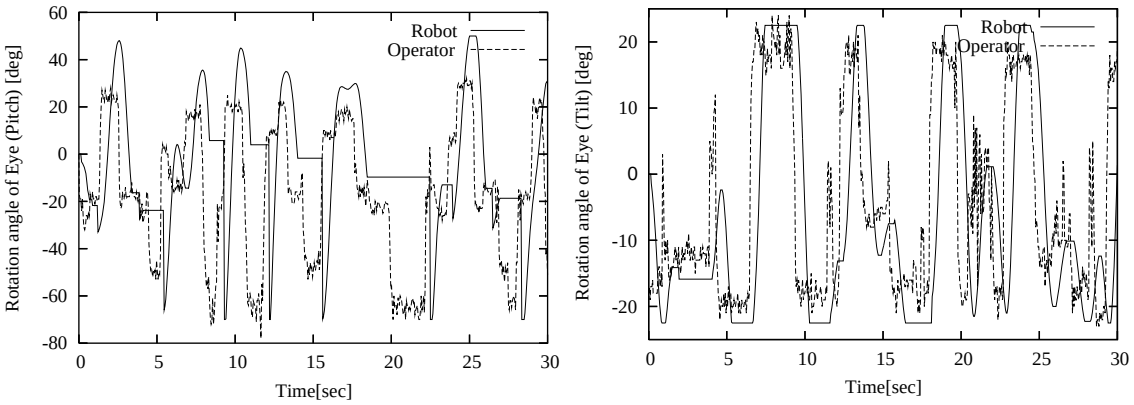


(a) 頭部姿勢（Pitch）　　(b) 頭部姿勢（Tilt）

図 4.10 頭部姿勢の計測値および，ロボットに対する指令値

直接ロボットの顔部品の位置として再現することで，顔情報を転送する．計測されたオペレータの顔情報はソケット通信によりロボット側に伝送され，ロボットによりオペレータの表情を再現し対話者に提示する．今回の実験においてオペレータの表情をを再現するメディアには，顔ロボット Infanoid2[61] を使用した．計測した顔情報のマッピング方法を表 4.2 に示す．また，計測結果のロボット提示例を図 4.9 に示す．

オペレータの運動の計測値に対するロボットの動きを図 4.10-4.12 に示す．また，オペレータの顔情報を得てからロボットの各顔パーツが動くまでの遅延時間は唇開閉で約 0.1[sec]，それ以外のパーツではノイズ除去フィルタの処理時間を含むため約 0.6[sec] となっている．



(a) 眼球姿勢（Pitch）　　(b) 眼球姿勢（Tilt）

図 4.11 眼球姿勢の計測値および，ロボットに対する指令値

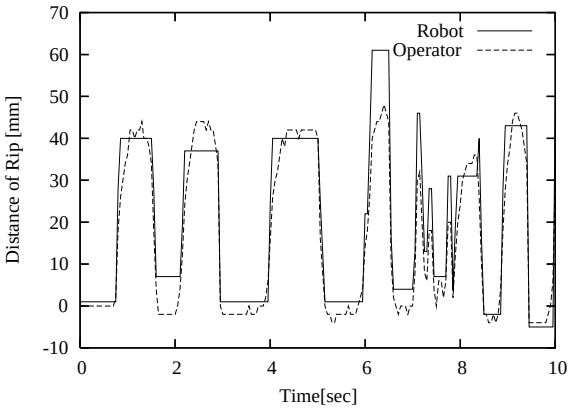


図 4.12 口唇開閉動作の計測値と指令値

4.4. 実験結果

開発した遠隔コミュニケーションシステムを用い（1）オペレータの表情がどの程度正しく伝わるかを調べる表情伝達実験,（2）各メディアに対する印象を調べる印象評価実験，の 2 種類の実験を行った.

4.4.1 感情伝達実験

構築したシステムでどの程度オペレータの感情が伝わるか実験を行った．被験者数は 5 人，提示する感情は図 4.13 に示すとおり，“怒り”，“幸福”，“驚き”，“悲し

(a) 無表情　(b) 怒り　(c) 幸福　(d) 驚き　(e) 悲しみ

図 4.13　ロボットによる表情提示

表 4.3　提示表情の認識率

提示した感情	認識率 [%]		
	人間 (静止画)	ロボット (静止画)	ロボット
怒り	64	82	100
悲しみ	96	96	100
幸福	100	80	87
悲しみ	100	96	93
平均	90	88	95

み” の 4 種類である．比較のためにオペレータ，ロボットがそれぞれ 4 種類の表情を提示した写真を，合計 3 回ずつランダム表示し，被験者にどの表情を提示しているか判定させる実験も行った．実験結果を 表 4.3 に示す．

表から，表情識別においては静止画に比べ，全体的に識別率が高くなっていることが分かる．この原因は静止画に比べてロボットを介した表情提示では，目の前で実際にロボットが動き表情を提示するまでの動作過程を持つことで，提示表情を識別する情報が増えたためだと考えられる．この結果は，構築したシステムが基本的なオペレータの 4 種類の表情を伝えられる事を示している．

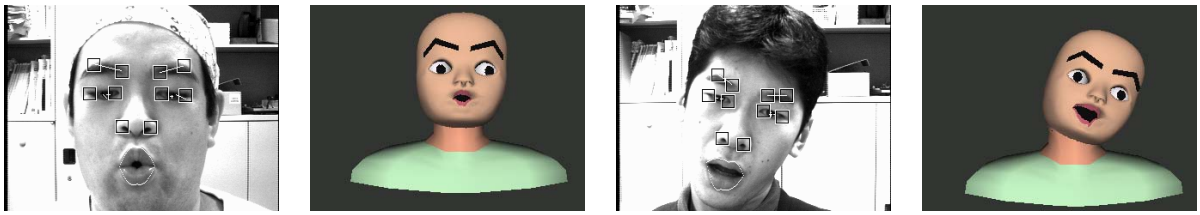


図 4.14 アバターチャットシステム

4.4.2 印象評価実験

これまでに，遠隔操作ロボットとのインタラクションに関しては，山岡らがロボットの背後にいる人の存在の有無によるコミュニケーションの印象評価を行い，ロボットの身体性により相互作用を楽しむ傾向が大きくなることを示唆している [62]．また坂本らは，アンドロイドロボットを介したコミュニケーションシステムとビデオ会議，音声会議システムとの印象比較を行い，ビデオ会議システムと同程度に自然な会話ができたことを示している [63]．しかし，これらの印象評価においては，一部の顔情報と全身動作によるインタラクションが評価対象となっており，頭部動作に特化したロボットインタラクションに対する印象評価はされていない．森田らは，頭部動作を転送する遠隔ロボットコミュニケーションシステムを構築しロボット会議，ビデオ会議，対面会議の 3 条件で発表者が感じる注目状態を比較し，ロボット会議がビデオ会議より優れていることを示している [64]．しかし，このシステムもロボットに転送・再現しているのは，頭部姿勢のみであり我々の提案のように顔部品の位置情報は転送していない．

本論文で提案する遠隔コミュニケーションシステムを評価するために，従来の遠隔コミュニケーション手法との印象比較実験を行った．比較の対象は，音声のみ，ビデオチャット，アバタチャット，ロボットの遠隔コミュニケーション手法に，人対人の直接対話を加えた，合計 5 種類とした．各手法によるコミュニケーション実験を 10 人の被験者に対して行った．また，各コミュニケーション手法の実験順序は，被験者毎にランダムに設定した．対話内容は特に限定せず，被験者の好みに任せたが，話題が無い場合にはオペレータが，“今日の仕事内容”，“旅先での出来事”，“家族のこと”，等のテーマを提示して対話を行った．

ビデオチャットは画像と音声をソケット通信により相手側に送る手法をとり，アバタチャットは，本論文で述べた顔情報計測システムを用いて独自に開発したチャット

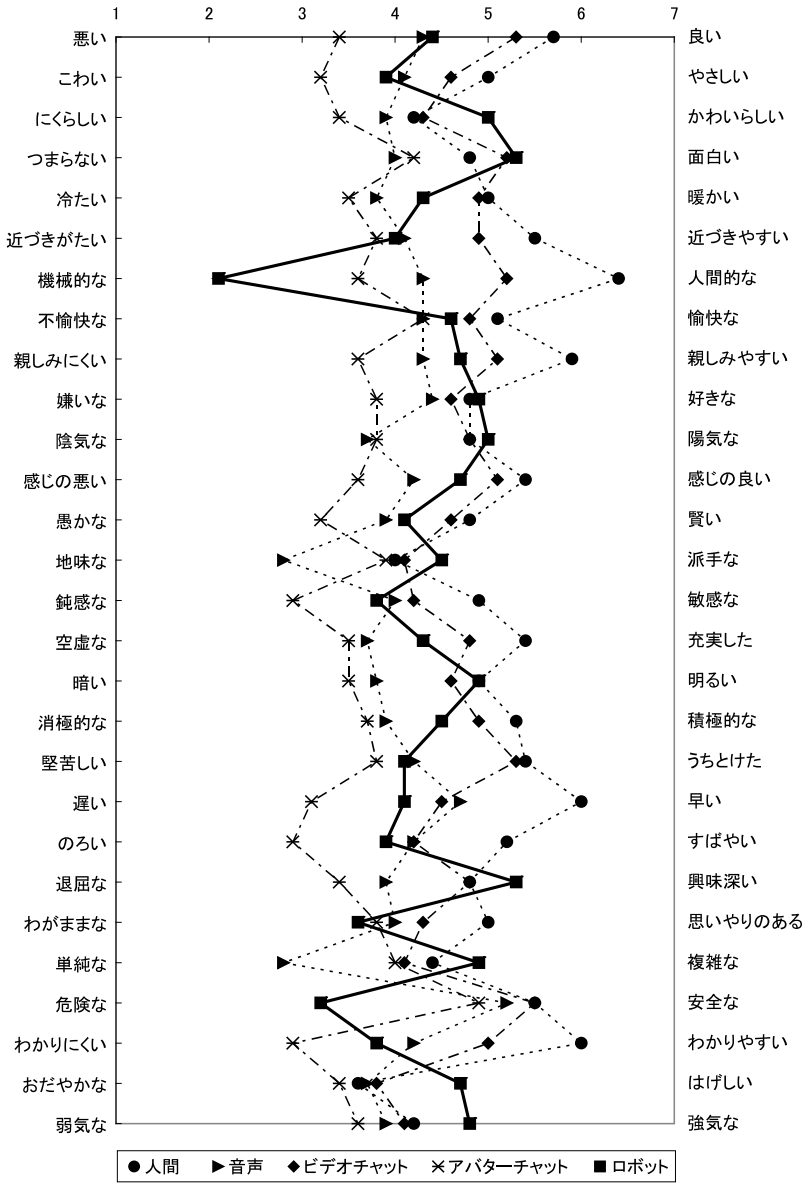


図 4.15 印象評価（SD プロフィール）

システム [65] を用いている．実験は被験者がオペレータと自由に3分間の対話を行った後，被験者が受けた印象を印象評価用紙に形容詞対のスコアとして記述するSD法を，各コミュニケーション方法ごとに繰り返す方法で行った．オペレータの性格により評価にバラつきがでないように，オペレータは常に同一人物としている．

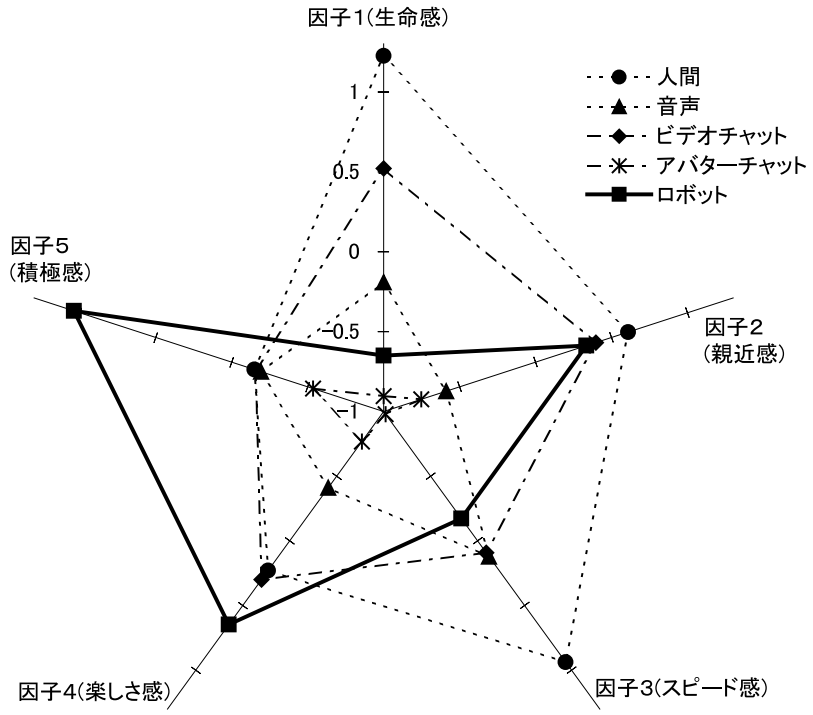


図 4.16 各コミュニケーションメディアの違いによる因子得点の平均値

図 4.15 に示す SD プロフィールでは，人対人の直接対話は，多くの項目で印象が良い方向に支持が偏っている．しかし，それ以外のコミュニケーション方法については印象評価に大差ないこともわかる．ロボットを介した遠隔コミュニケーションとそれ以外のメディアとを比較すると，“人間的な” と “安全な” の項目でロボットを介したコミュニケーション方法が顕著に値が低い．これは見た目に基づく直接的な印象であるが，それ以外のコミュニケーション自体に対する印象では，他のメディアを用いたコミュニケーションに比べて特に悪印象のものは見られず，むしろアバタチャットや音声対話に比べては好印象の項目が多いことがわかる．

比較する変数を減らし，直感的に各コミュニケーション方法が与える印象の傾向を評価するために，因子分析を行った．因子分析の結果，その固有値が 1 以上である 5 因子を抽出した．各評価項目の因子負荷量を，表 4.4 に示す．ここで，0.5 以上の値をボールド体で表している．また各因子の解釈を表 4.5 に示す．それぞれの因子負荷量を用いて，各サンプルの因子得点を計算し，コミュニケーション方法別の平均値を比較した．その結果を図 4.16 に示す．

表 4.4 因子負荷量

評価項目	因子 1	因子 2	因子 3	因子 4	因子 5
積極的な	**0.816**	**-0.450**	0.172	0.265	0.125
強気な	**0.794**	0.099	0.136	-0.184	0.047
派手な	**0.748**	0.058	-0.122	-0.177	0.116
充実した	**0.745**	0.142	0.163	0.060	-0.250
明るい	**0.555**	0.216	-0.113	0.249	0.104
うちとけた	**0.503**	0.348	0.022	0.116	-0.073
感じの良い	-0.124	**0.885**	0.062	0.175	-0.153
暖かい	0.179	**0.723**	-0.317	0.028	0.062
陽気な	0.182	**0.623**	-0.010	-0.111	0.248
親しみやすい	0.150	**0.607**	0.128	0.161	-0.119
好きな	-0.160	**0.497**	0.368	-0.099	0.307
敏感な	-0.130	**0.490**	**0.477**	-0.305	0.140
人間的な	0.243	**0.442**	-0.022	0.256	-0.229
良い	0.203	0.345	0.010	0.215	0.027
早い	0.119	-0.072	**1.019**	-0.190	-0.085
すばやい	0.081	0.010	**0.961**	0.001	-0.187
複雑な	0.194	-0.278	**0.418**	0.327	0.189
わかりやすい	0.262	0.239	0.285	0.064	-0.010
思いやりのある	0.146	0.148	-0.359	**0.861**	0.011
やさしい	-0.287	0.139	0.174	**0.617**	0.091
近づきやすい	-0.044	0.301	0.298	**0.416**	-0.053
面白い	0.074	0.022	-0.193	0.018	**0.957**
興味深い	0.055	-0.120	-0.018	0.310	**0.577**
愉快な	0.313	0.354	-0.100	0.061	0.358

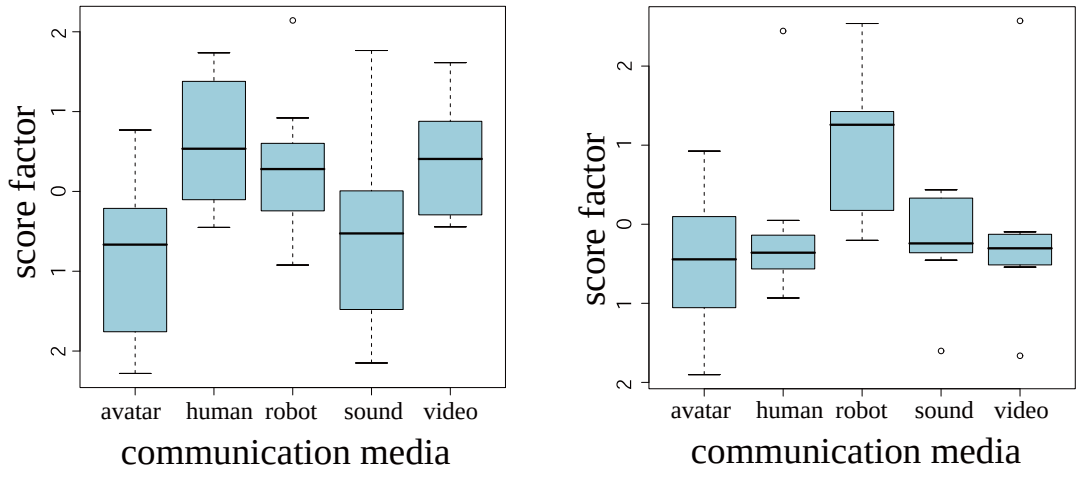
図 4.16 を見ると，実際の人とのコミュニケーションは，多くの因子において高い値を示している．特に，生命感およびスピード感に関する因子では，他のコミュニケーション方法をとの差が大きい．この結果は，従来の遠隔コミュニケーション手法には，生命感やスピード感を感じさせる機能が欠けており，それらを追加するこ

表 4.5 各因子の解釈

	解釈	影響が大きい主な評価項目
因子 1	能動感	積極的な・強気な・派手な
因子 2	生命感	感じの良い・暖かい・陽気な
因子 3	スピード感	早い・すばやい
因子 4	親近感	思いやりのある・やさしい
因子 5	楽しさ感	面白い・興味深い

とで自然な直接対話に近づく可能性があることを示している．次に，実際の人の動画像を用いるビデオチャットを見てみると，生命感と親近感に関する因子で高い値を示しており，人とのコミュニケーションと似た傾向が伺える．反対に，本論文で提案したロボットを介した遠隔コミュニケーションは，特に積極感に関する因子で高い値を示しており，親近感に関する因子では人間やビデオチャットによるコミュニケーションと同様の高い値を示している．図 4.17 に，これらの因子得点の分布の一例を示す．実際に，因子 5（積極感）の因子得点に対して有意水準 5%で分散分析を行ったところ，コミュニケーション手法間で有意差が確認できた．さらに，Tukey の方法を用いて多重比較検定を行ったところ，ロボットを用いた手法とそれ以外のすべての手法との間に有意水準 5%で有意差が認められた．これは，ロボットという新しいコミュニケーション媒体であることで，被験者の興味を引いていること，物理的に存在していること，また，全体がダイナミックに動くこと等に起因していると考えられる．また，因子 2（親近感）について同様に検定を行ったところ，人間，ビデオチャット，ロボットの各手法間では有意差が認められず，同等の親近感をユーザに与えている事が確認できた．最後に，アバタチャットは，アバタの作り込み度合いによって印象が大きく変化する可能性が高いが，今回の単純なアバタでは，あまり芳しい評価が得られなかった．

これらの結果より，提案システムによるコミュニケーションは，他のメディアと比較して生命感が低いという特徴があるものの，ビデオチャットや直接対話と同等かそれ以上の親近感や積極的印象を与えるものであると評価できる．



(a) 因子 2　　(b) 因子 5

図 4.17 因子得点の比較

4.5. 本章のまとめ

本章では，自然で円滑な遠隔コミュニケーションを実現するために，音声と表情を伝達できるロボットを介した遠隔コミュニケーションシステムを提案し，評価を行った．最初に，これまでに開発してきた視覚センサを用いた顔情報計測システムに，基本表情の伝達を可能とするための眉毛・唇の位置情報計測機能を追加した．次に，計測された人の顔情報と音声情報をロボットに投影する遠隔コミュニケーションシステムを構築した．最後に，構築した遠隔コミュニケーションシステムの表情伝達とコミュニケーション時に対話者が受ける印象に関する，評価実験を行った．評価実験の結果より，ロボットを介した遠隔コミュニケーションシステムが音声対話・アバターチャットを用いた遠隔コミュニケーションシステムに比べて物理的な存在感からの優位性を示すことが出来た．今後は

- 音声，表情を再現するタイミングの同期
- 全身動作を計測・再現することで，より存在感のあるコミュニケーションシステムへの改良

を目指し，より自然な遠隔コミュニケーションシステムとしての改良を図っていく必要がある．

第5章　ヒューマノイドロボット HRP-2を用いたインタラクション

3章では，受付案内ロボットASKAを構築およびその評価を行った．その結果明らかになった幾つかの問題点を解決するべく，また現在までの研究成果の実装としてヒューマノイドロボットHRP-2を用いたシステムを構築した．本章では，そのシステムの詳細について述べる．

5.1. システム構成

5.1.1 ハードウェア構成

本システムは，図 5.1に示す様に，ヒューマノイド本体と2台の外部PC，およびカメラ，マイクロフォン，スピーカから構成されている．ロボット本体部分には，HRP-2（川田工業）を利用し，内蔵PCに対して，IEEE1394，USBのポートの追加，およびIEEE1394カメラへの換装を行った．HRP-2頭部には，8chのマイクロフォンとA/Dボードが内蔵され，各入力を同時にサンプリングする事が可能になっている．また，ロボット本体頭以外にハンドマイクを装備し，入力を切り替えることで遠隔から音声も入力可能である．ロボット本体以外に，HRP-2のセンサ情報の監視を行うAuditor動作用のPC1台，音声認識および音声合成を行うPC1台を外部PCとして使用している．

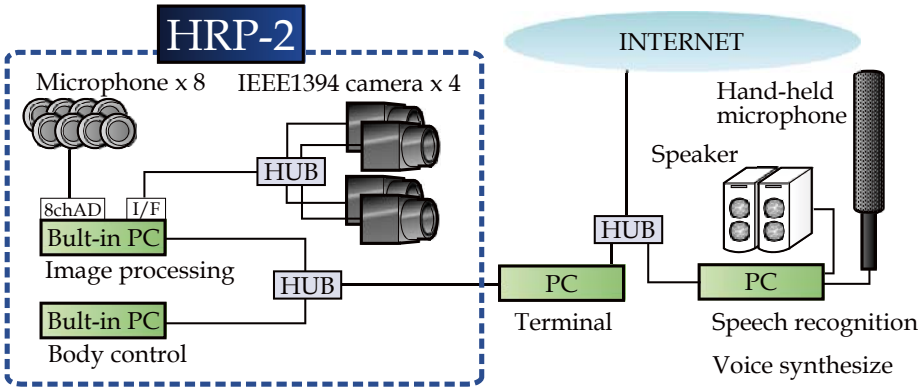


図 5.1 HRP-2 を利用したシステムのハードウェア構成

5.1.2 ソフトウェア構成

本システムを構成する各モジュールは，ソケット通信によって互いの状態を伝えながら動作している．この様子を，図 5.2 に示す．この状態伝達には，単純な黒板モデルが用いられている．サーバには全モジュールの状態（動作中，待機中，出力値など）が常に蓄積されており，各モジュールは，それを参照して次の状態を決定する．各モジュールにおいて，これらの状態を取得・蓄積するためのインタフェースが定義されているが，モジュールの実装自体は比較的自由である．Speech Recognition, Speech Synthesize は，文字通り音声認識および音声合成を担当する．Gesture Script は，Jython で記述された，HRP-2 をコントロールするスクリプト．Face Tracking, Gesture Recognition は，それぞれ対話者の顔情報および方向指示ジェスチャを認識するモジュール．Draw-Data Generation は，“似顔絵データ” と呼ばれる，対話者の顔を描画するためのベクトルデータを作成する．またこれらの視覚情報を利用したモジュールは，Vision SubServer で統括され，さらに全モジュールの状態を保持する Server に接続されている．

5.1.3 インタラクションシナリオ

HRP-2 は上述の様なシステムにより，対話者とのインタラクションを行う．その基本的なシナリオは，以下のようになる．

1. HRP-2 がビジョンにより人の顔向きを計測する

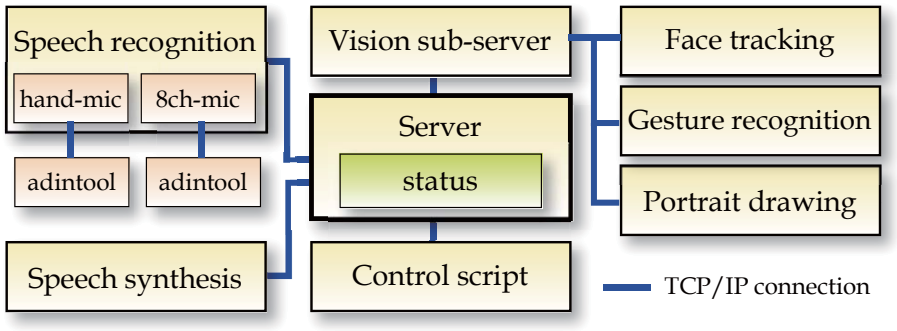


図 5.2 HPR-2 を利用したシステムのソフトウェア構成

2. 対話者が音声とジェスチャにより質問等を行う
3. HRP-2 が発話者の音声を認識し，合成音とジェスチャにより応答するか，又は命令されたタスクをこなす

また，応答に用いる定義済みのモーションは，MotionCreator[66] を利用して作成し，予め応答文とのマッピングを行ってある．

5.2. インタラクション機能

5.2.1 音声および非定常雑音認識

本システムの音声対話は，受付案内ロボット ASKA で利用したオープンソースの音声認識エンジン Julius[37] に加え，Julian を用いた並列音声認識システムを利用している．Julius は主にハンドマイク音声認識時に利用しており，語彙数 4 万の 3-gram 言語モデルと PTM triphone HMM 音響モデルを持つ．モデルの学習には，奈良県生駒市の音声情報案内システム「たけまるくん」[67] のフィールドテストで収集した大人及び子供の実発話を利用しており，子供に対する認識精度の向上を得た．Julian は頭部マイクのような録音環境が悪い状況下のための音声認識を担っており，言語モデルには約 400 の単語数の記述文法を利用する．Julius 使用時と異なり単語数を制限することで認識精度の低下を補っている．

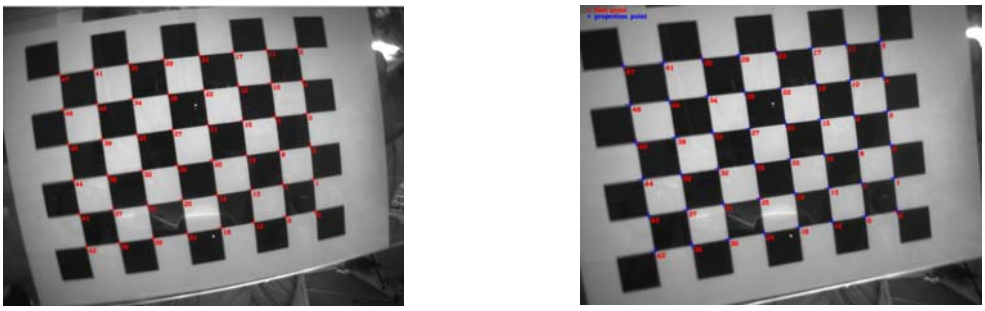
表 5.1 GMMs の訓練条件

Class	# of training data
Adult voice	7,497
Child voice	7,503
Laughter	849
Coughing	321
Beating by hand	101
Beating by soft hammer	104
Background noise	5,000
Other noise	6,380
Sampling rate/bit	16 kHz, 16 bit
Window width/shift	25/19 msec
Parameter	MFCC (12 dim.), ΔMFCC, ΔPower
Mixtures of Gaussian	64

非定常雑音認識

通常の対話システムにおいて，ノイズは不要な入力と見なされていた[68]．本システムの特徴として，人とロボットのインタラクションに役立つであろう咳やくしゃみ，自己の頭部を叩かれる音など非定常雑音を認識することができる．ロボットを実環境に設置した際に発生する音声対話に不必要な入力音を棄却すると共に，雑音の内容自体を識別することで柔軟なインタラクションの実現を目指した．例えば，利用者が咳きをすると「風邪ですか?」と応答することができる．この機構は，上記の音声認識部と平行するように，64 混合の混合正規分布モデル（GMM）を音響モデルとして持つ雑音認識用の Julian を並列動作させることで実装した．なお，GMM 音響モデルには大人と子供の収集発話も含めている．入力音の音声と非音声の識別の際に必要となり，また，大人と子供の発話も識別することで利用者が大人であるか，子供であるかもシステムは判断することが可能になっている．

表 5.1 は，GMM のトレーニング条件を表す．HRP-2 が実際に音声認識に用いている内蔵マイクを用いて収録された．



(A) 歪み補正前画像　　(B) 歪み補正後画像

図 5.3　画像の歪補正

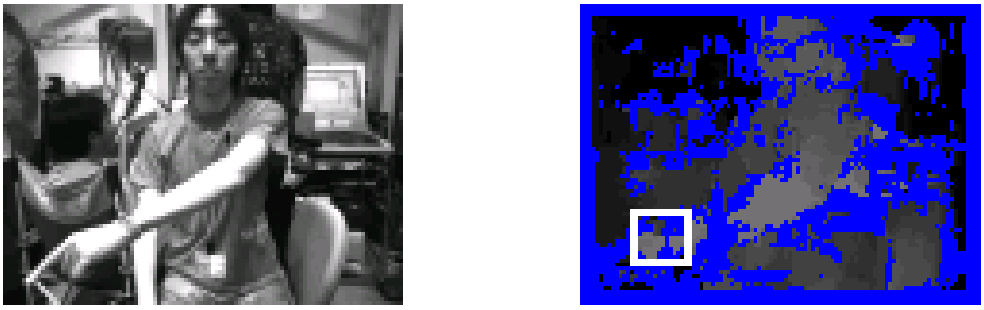


図 5.4　指示動作認識

5.2.2　アイコンタクトと発話区間推定

5.2.3　ジェスチャ認識

本研究では，SAD（Sum of Absolute Difference）形式の相関演算と再帰相関演算を利用した高速局所相関演算アルゴリズムを利用し，簡便な方向指示ジェスチャの認識をリアルタイムに行っている（図 5.4）．以下に，その処理手順を示す．

1. 広角カメラのレンズ歪を補正し，視差画像を生成する
2. カメラまでの距離が最も短い領域を検出し，そこを手先と仮定して指示方向を認識する

本機能により指示方向を認識することで，“あっち” や “こっち” といった指示語が入った発話内容に対して適切な応答や，例えば指定した場所にある新聞を取ってもらう把持行動等を行うことが出来る．

5.2.4 似顔絵作成

似顔絵データの作成

顔情報計測システムを使い HRP-2 は人との対話を行ないながら似顔絵の生成を行なう．あらかじめ登録された対話者の顔の目じり口元などの 8 つの特徴点の画像と，それぞれの 3 次元位置を併せ持つ 3 次元顔モデルを作成する．次に顔情報計測システムで，3 次元顔モデルを使用し，入力画像と 3 次元トラッキングを行うことで対話者の顔方向や視線方向の検出を行なう．3 次元顔モデルと入力画像との相関値をとり，信頼度を求める．信頼度が閾値以上になったときのみ，ステレオカメラの左右の入力画像と，そのときに計測される 8 つの特徴点の画像平面上の (x,y) 座標を取得する．

まず，入力画像になるべく線と線をつなげるようなエッジ画像を出す Canny のエッジフィルタをかけ，エッジ画像を作成する．しかし，似顔絵を描くためには入力画像から顔部分だけを抜き出さなければいけない．そこで入力画像から背景を除去し，対話者の顔だけを抜き出すため depth フィルタと顔の形を近似した楕円フィルタの作成を行なう．depth フィルタは SAD 法を用い作成し，HRP-2 から遠くにある背景と対話者より前にあるもの除去する．次に顔情報計測システムによって得られる顔の 8 つの特徴点の座標を用い，楕円フィルタの作成を行なう．楕円は左右の目の目元と目尻の計 4 つの座標の重心を楕円の中心とし，長径は目の重心から 2 つの口元座標の重心位置までの距離に係数をかけたもの，短径は左右の目の目じり間の距離に係数をかけたものとしている．以上により作成されたフィルタをエッジ画像にかけ，顔領域のみのエッジ画像を作成する．また，顔の 8 つの特徴点重心位置が画像の中心に来るように得られたエッジ画像を移動させる．

HRP-2 に絵を描かせるためには，一つの線を画像データではなく，線データとして保持する必要がある．そこでエッジ画像にラベリングを行い，ラベル付けされた順に画像圧縮などに広く用いられている，chain 法を使い画像データを線データ（x, y）として変換する．また，線データにはそれぞれの点ごとに開始点，中間点，終点の 3 つの状態フラグをつける．線データと状態フラグを合わせて似顔絵データとして保持する．次に HRP-2 がなるべく少ない腕の移動量で絵を描くことができるように，線データをソートする．すべての開始点と終点を比較し終点にもっとも近くにある開始点を持つ線データを次の線データとする．比較の際，すでに選ばれた線データが出た場合は，次に近いデータを選ぶ．また，HRP-2 が 3 分間程度で絵を描き終えることができるようデータ数に応じて，中間点のデータ数をなるべく一定になる

ように間引く．以上によって作成したデータを制御用 PC に転送し HRP-2 に実際に絵を描かせる．これらのデータ生成のフローチャートを図 5.5 に示す．

似顔絵描き動作の生成

はじめに，HRP-2 の手先を描きやすい位置，高さに調節を行い，手先の初期位置と姿勢として登録する．似顔絵データの開始点でペンを下ろし，終点でペンを上げる．開始点から終点の間は座標点を直線で補完し線をつなげ似顔絵を描く．似顔絵を描く際には，蓄積誤差を減らすために，手先座標を相対座標ではなく，登録した初期位置を原点とした絶対座標で表し，逆運動学を解くことにより手先を移動させる．移動時間は移動距離に比例して変化させる．手先が目標座標となるような目標関節角を求めるため，腰 2 自由度，肩 3 自由度，肘 1 自由度，手首 2 自由度，計 8 自由度の逆運動学を解く．ただし，手先の姿勢角は初期姿勢に固定する．解いた目標関節角と現在関節角との間を HRP-2 の制御周期である $5[ms]$ 間隔で移動時間までの間を補完し，各時間ごとに補完した関節角を読み込み HRP-2 を動かす．これを似顔絵データがなくなるまで繰り返し似顔絵を描く．似顔絵データがなくなると手先を初期位置に戻し，似顔絵を終了する．

5.3. 対話実験

5.3.1 展示会における一般来場者対話実験

前述した受付案内ロボット ASKA を用いた実験では，発話時にユーザがロボットの顔を見る傾向が確かめられている．しかし，これはロボットの外観に大きく影響される可能性が考えられる．HRP-2 におけるアイコンタクトの有効性を確認するために，“2005 国際ロボット展” において，一般来場者が HRP-2 との対話を行い，その際のユーザの顔情報を，ヒューマノイドロボットの頭部カメラを用いて記録した．ここでの被験者は全て来場者である一般人，男性 5 名，女性 5 名，子供 5 名の計 15 名である．対話のトピックは自由であるが，どのような質問が可能であるのかの例を被験者には示してある．また，音声入力にはハンドマイクを使用し，似顔絵作成などの作業を行っている部分のデータは含んでいない．

実験結果からユーザが HRP-2 に対して発話行う際に，ユーザの顔，視線，あるいはその両方が HPR-2 に向けられている事が確認できた．また，ユーザが HRP-2 以

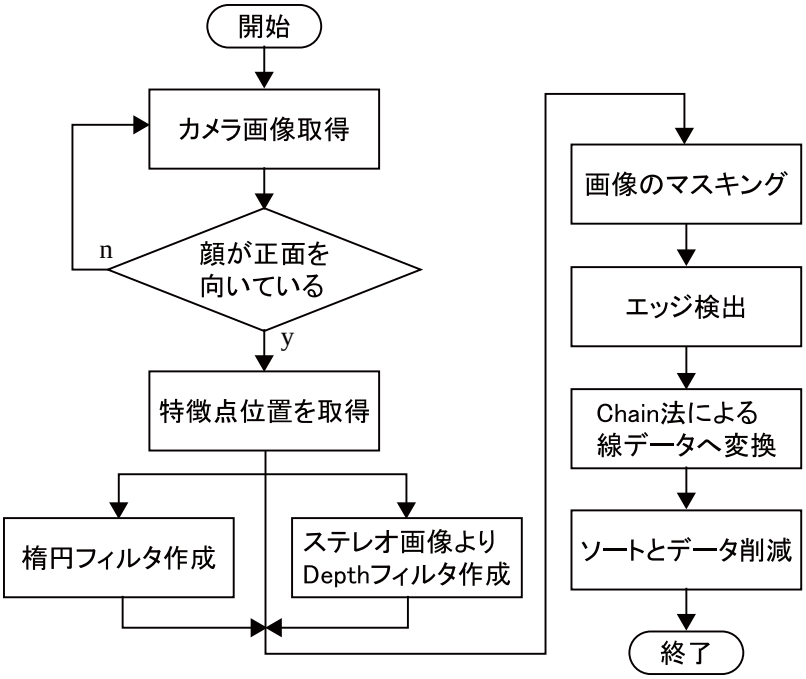
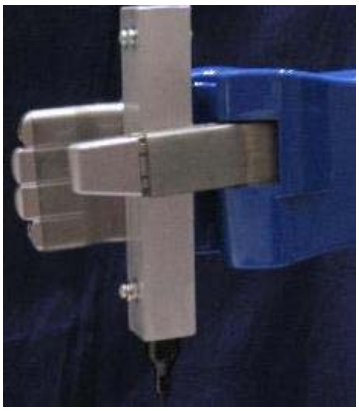


図 5.5 似顔絵データ生成のフローチャート

(A) HRP-2 in portrait drawing



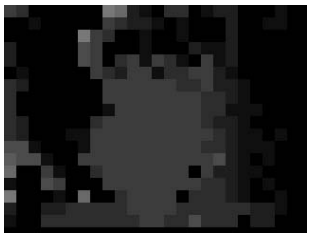
(B) The pen and holder

図 5.6 似顔絵作成の概要

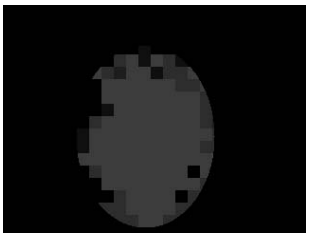
外に向けての発話（多くの場合は，実験をサポートしているスタッフやユーザの同伴者に対する発話）を行う場合は，やはり HRP-2 から顔向きと視線を外している事が確認された．ここから，ユーザの顔情報を利用することで，実際の対話においても不要な発話を棄却できている事が確認できる．

(A) Original image

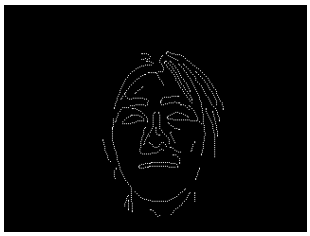
(B) Canny edge image



(C) Depth image



(D) Mask for elimination



(E) Line segment data

(F) Result of portrait drawing

図 5.7 似顔絵作成の処理プロセス

表 5.2 発話時間とアイコンタクト成立率

class	talk to HRP[s]	talk to other[s]	face & talk [%]	gaze & talk [%]
male	13.0	5.1	94.6	64.1
female	20.2	4.3	96.4	68.7
child	8.1	2.3	89.3	74.5

表 5.2 に，発話時間および，発話中に顔向き・視線によるアイコンタクトを行う割合の平均値を示す．HRP-2 に対する発話中に，顔向きや視線を HRP-2 方向に向けていた割合をアイコンタクトの成立率とすると，顔向きによるアイコンタクトの成立率は，男女ともに約 95[%] と高い割合を示している．また，子供の場合は，約 89[%] と若干低い割合を示している．これは，子供の場合は，大人に比べて対話に集中し

図 5.8 展示会デモにおける一般来場者との対話（HRP-2 のカメラ画像）

ている時間が短いことが原因であると考えられる．また，視線によるアイコンタクトの場合は成立率が大きく下がる．HRP-2 は，目や口などの “顔” の機構を有しておらず，また応答時以外においては，ロボットが動いたり話したりすることがないため，ユーザの注目がヒューマノイドロボットの顔に集まりにくい事が原因として考えられる．

5.3.2 アイコンタクトを利用した対話実験

次に，実験室環境において，アイコンタクトを利用した対話実験を行った．2 人 1 組として計 10 組の被験者が．予め提示されたシナリオに沿って対話を行う．対話は，二人の被験者および一台のロボットとの間で行われ，人対人，人対ロボットの対話が繰り返される構成になっている．図 5.9 に実験の様子を示す．図 5.10 に，対話の誤応答率を示す．ここで誤応答率とは，ロボットに向けられていない発話に対してロボットが誤って応答を返してしまう確率を表す．図 5.10（a）は，アイコンタクトによる有効発話推定を行わない場合の誤応答率を，図 5.10（b）は，行った場合の誤応答率を示している．平均 75.7[%] から，4.3[%] に減少してることが確かめられる．

5.3.3 愛知万博でのデモンストレーション

実環境下におけるシステムの実現性を確認するために，愛知万博 2005 のプロトタイプロボット展においてデモンストレーションを行った．その様子を 5.11 に示す．多くの観客による話し声や隣接するブースのデモに伴う背景雑音を考慮して，実際の展示エリアではハンドマイクを用いて音声入力を行った．万博時のインタラクションシナリオでは，非定常雑音認識によってユーザの咳を認識することで風邪に関す

図 5.9 アイコンタクトを利用した対話実験の様子

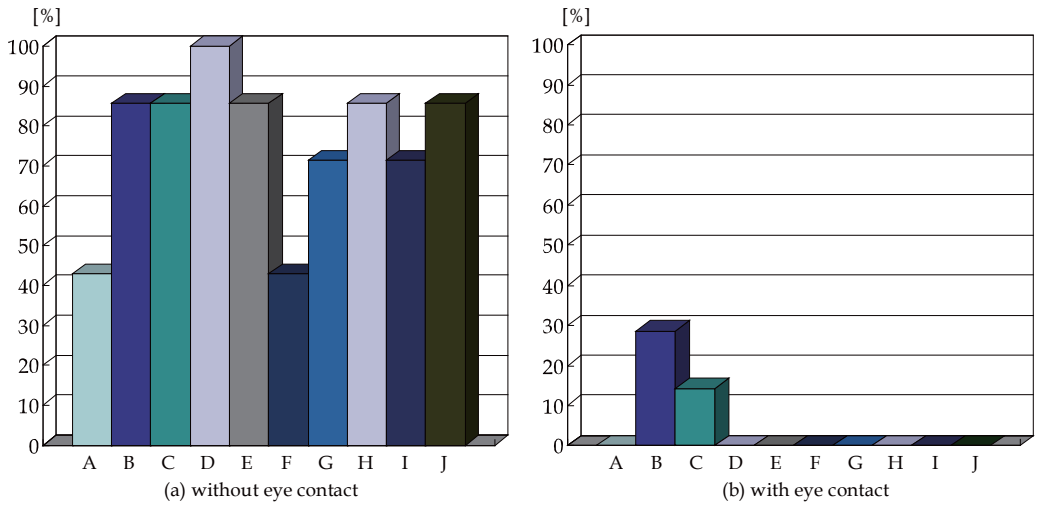


図 5.10 有効発話推定による誤応答率の低下

る対話を行い，また，Web から情報を取得する事で，天気予報やニュースヘッドラインなどの質問に答えた．

このような照明条件を制御できない状況は，画像処理を行う上で厳しい条件である．しかし，我々の手法は，照明条件の変化に影響を受けやすい肌色検出などを利用していないため，このような環境下においても，顔情報計測やジェスチャ認識などの処理がロバストに動作した．

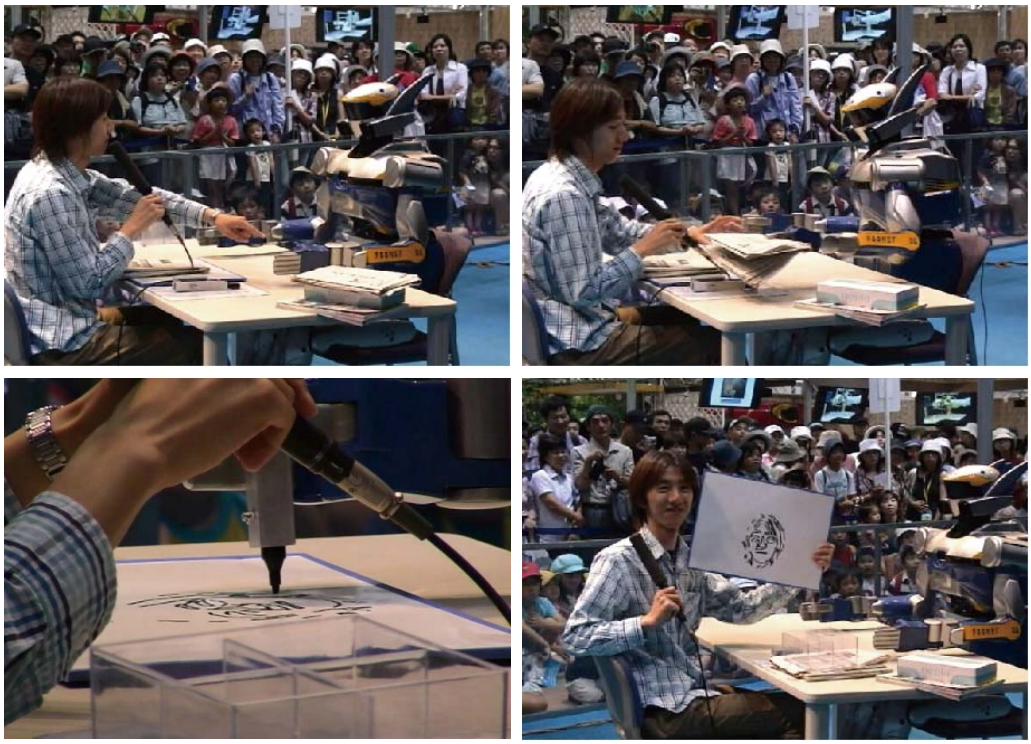


図 5.11 愛知万博におけるステージデモンストレーションの様子

5.4. ヒューマノイドロボットのナビゲーション

これまでは，二足歩行可能なヒューマノイドロボットを敢えて座らせる形で，対話やオブジェクトの受 け渡し等のインタラクションを行ってきた．そのため，ロボットに実行可能なタスクは，音声やジェスチャ，あるいは机上で完結するものに限られた．しかし，歩行による移動機能を利用したタスク，例えば移動による案内やオブジェクトの移動などを実行可能とすることで，人とロボットのインタラクションが，より自由なものとなる事は間違いない．一方，移動ロボットのナビゲーションは，

ここでは，ビューシーケンスを利用した，“teach and play” に近いナビゲーション手法をヒューマノイドロボットに適用するすることで，屋内環境下でのナビゲーションを実現する．

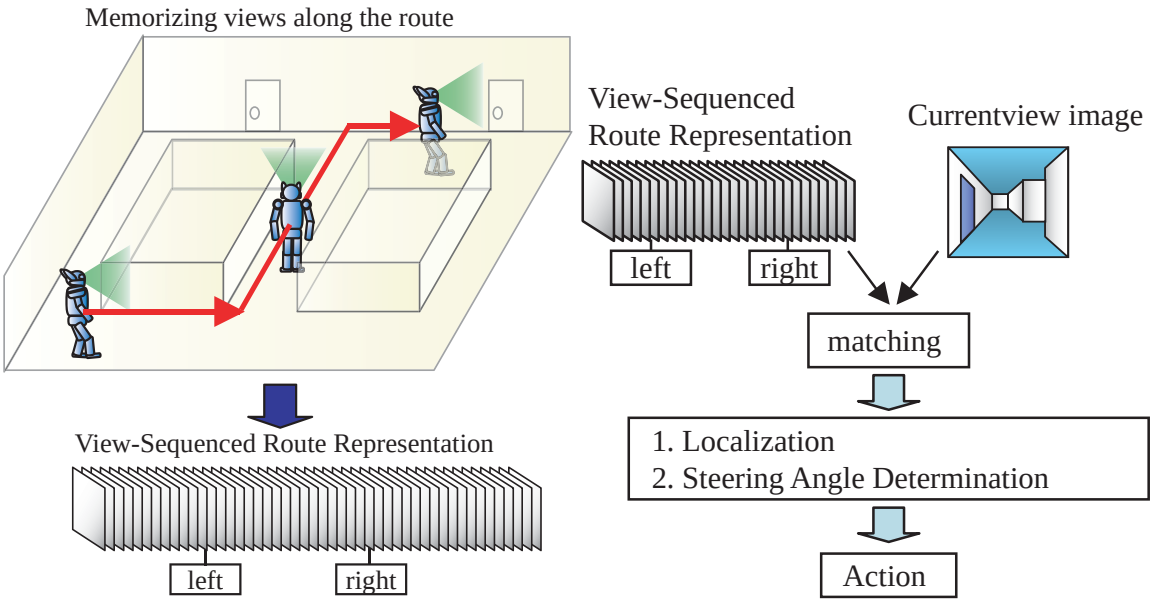


図 5.12 ビューシーケンスを用いたナビゲーションの概要

5.4.1 ビューベーストナビゲーション

“ビューシーケンス” を利用するナビゲーションは，二つのステップから構成される．記録走行のステップでは，実際の移動経路に沿ってロボットを走行させ，そのときの画像を時系列に沿って記憶する．次の自律走行のステップでは，記録走行によって得られた画像とリアルタイムに取得した画像のマッチングを行い，どの記録画像に最もマッチするかを計算することで自己位置を推定する．また，記憶した画像とのずれ量を計算しロボットの駆動系にフィードバックすることで経路からのずれを修正する．図 5.4.1 に，ビューシーケンスを利用したナビゲーションの概略図を示す．

ビューシーケンスの作成

ビューシーケンスを用いたナビゲーションでは，前方を向いたカメラから得られる画像が記録される．このとき，計算コストを削減するために，低解像度の画像が用いられる．画像を平滑化および縮小することで，データサイズを小さくできるだけで

なく，元の画像のノイズの影響を小さくすることができ，ロバストで高速なマッチングが可能になる．記録される画像$\mathbf{I}$の配列である画像列$\{M_i|0 \leq i < N\}$が，ビューシーケンスとなる．

純粋な画像列の情報だけでは，隣接する画像間の関係が不明であり，走行経路をモデル化するには不十分である．これは，次の画像（ビュー）に移るために，ロボットがどのような行動をすれば良いのかが分からないという事を意味する．従って，あるビューI_iに対して，次の隣接画像との関係がタグT_iとして付加される．例えば，直線上に前進して右に曲がるような経路でのビューには，“Foward” および”Right” というタグが付加される．ここでは，タグ$\{T_i\}$が付加されたビュー$\{M_i\}$の列を”ビューシーケンス” と呼ぶ．

ビューシーケンスを用いたナビゲーション

ビューシーケンスを利用する場合の，ロボットの基本的な行動は次ぎようになる：

- ビューシーケンスに従って前進する
- ビューシーケンスのコーナー部で回転する

ナビゲーションは，これら二つの行動の繰り返しにより実現される．

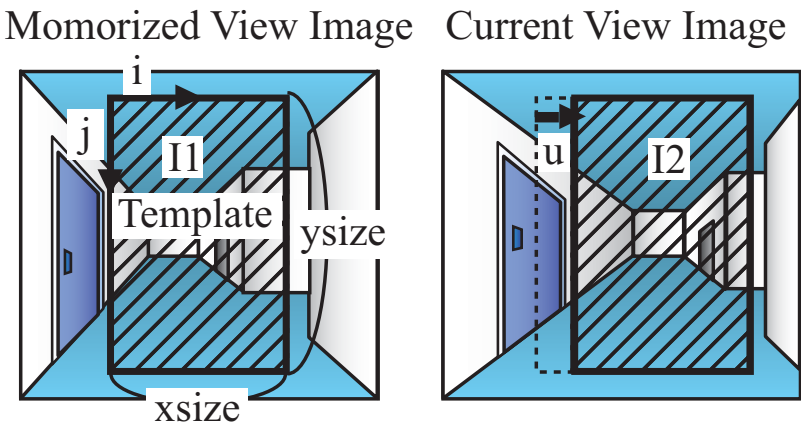
二つの画像を比較し，相関値を計算するプロセスをマッチングと呼ぶ．ここでは，テンプレートマッチングが利用され，画像中央の矩形領域がテンプレートとして利用される（図 5.13 を参照）．マッチングプロセスの結果，水平方向のずれ量uと相関値$corr$が得られる．

このずれ量uは，ロボットの横方向の位置と姿勢のによって生じる．単一の画像からだけではこれらの影響を分離できないので，ロボットの姿勢を求めることはできない．しかし，ロボットを教示走行時の位置と姿勢に近付くように制御することは可能である．この場合の具体的な行動については，次のセクションで述べる．

5.4.2 ビューシーケンスのヒューマノイドへの適用

二足歩行ロボット特有の問題

ビューシーケンスを図 5.14 のようなヒューマノイドロボットに適用するためには，ロボットの視点が三次元移動することを考慮する必要がある．ヒューマノイドロボッ



$$corr = \frac{\sum_{i,j} I1_{i,j} \times I2_{i,j}}{\sqrt{\sum_{i,j} I1_{i,j}^2} \times \sqrt{\sum_{i,j} I2_{i,j}^2}}$$

図 5.13　二枚の画像間のマッチング

トが二足歩行を行うことにより発生する画像のぶれと揺れは，画像のマッチングにおいて重大な問題となる場合がある．その結果，自律走行時に，次のような問題が発生する：

- 画像の揺れ幅がテンプレートマッチングの探索範囲以上になると，マッチングプロセスが失敗する．
- 探索対象となる画像にモーションブラーが発生すると，相関値およびマッチング位置が不正確な結果となる．
- 二足歩行ロボットは，左右に振れながら歩行を行うため，直進時のビューシーケンスでも左右に振れた画像が含まれてしまうと，自律走行での直進ができない．

まず，画像の揺れの影響を低減するために，歩行中の画像を調べる．図 5.15 は，ぶれた画像の一例である．左が歩行中に取得されたぶれた画像，右が同じ場所で静止した状態で取得されたぶれていない画像である．このようなぶれた画像は，歩行のパターンに合わせて毎回発生する．図 5.16 は，歩行時の揺れの影響を示す画像である．このように，ロボット自体が直進している場合でも，頭部カメラから得られる画像は左右に揺れている事が分かる．図 5.17 は，水平（上）および垂直（下）方向のオプティカルフローを表す．歩行の 1 サイクル，つまり左右一歩づつの合計二歩，における水平，垂直方向のフローの最大値はそれぞれ，9 [pixel]，および，12 [pixel] である．

これらの画像のぶれが発生する原因を調べるために，ロボットが歩行中，カメラが搭載された頭部の位置と姿勢がどのように変化するのかを計測した．計測には，光学式モーションキャプチャシステム Vicon を用いた．図 5.18 に，計測環境を示す．ロボットは，平坦なリノリウム床を 0.1 [m/s] という低速で前進を行い，そのときのマーカ位置を計測した．ここで計測された頭部の位置と姿勢を，図 5.19 に示す．図 5.19（a）から，ロボットが一定の速度で前進している事が分かる．また，図 5.19（b）から，ロボットが最大で 150 [mm] ほど左右に揺れながら歩行している事が分かる．さらに，図 5.19（c）は，頭部の上下の揺れが最大で，約 30 [mm/frame] である事を示している．

カメラのシャッタースピードを 10 [ms]，テンプレート画像に写る環境の奥行きの平均値が 5[m]，画像の解像度を 320×240 [pixel] とすると，横揺れ（図 5.19（b））は，

図 5.14 HRP-2 の起立姿勢

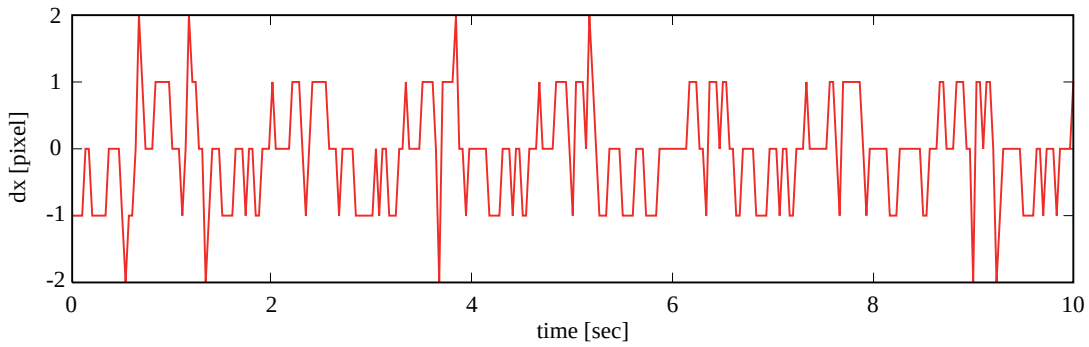
図 5.15 歩行動作中の取得画像（左）と，静止時の取得画像（右）

図 5.16 歩行時の揺れによる画像の左右変位

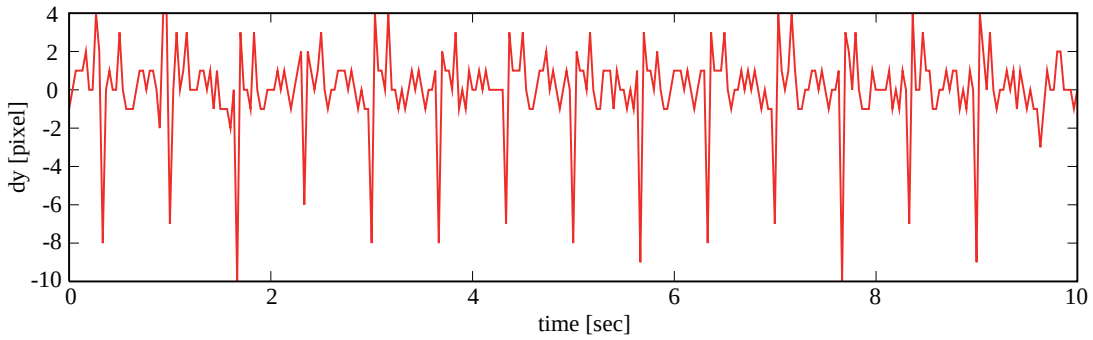
水平方向の変位 7.4 [pixel] に相当する．また，縦揺れ（図 5.19（c））は，垂直方向の変位 0.5 [pixel/frame] に相当する．図 5.19（d）は，ロボット頭部のピッチ変位を表す．この最大値は，約 2 [deg/frame] であり，これは，垂直方向の変位 2.8 [pixel/frame] に相当する．p つまり，画像の垂直方向のオプティカルフローのほとんどは，カメラが搭載されたロボット頭部のピッチ角変位による事が分かる．

ヒューマノイドロボット特有の問題に対する解法

ぶれや揺れのない安定した画像列を得るために，記録された画像から画像の選択を行う．図 5.17（a）に示すように，一歩移動する度に垂直方向の大きな変位が発生している．つまり，ロボットの遊脚接地直後に大きなぶれが発生すると思われるので，これらの変位が起こる中間の画像を選択することで，常に安定した画像が取得



(a) 横方向の変位



(b) 縦方向の変位

図 5.17 画像の変位

できる．また，ビューシーケンス内の隣接する画像は，十分に異なる画像が選択されるべきである．言い換えると，隣接する画像間の相関値が, ある閾値を下回った場合に，その画像がビューシーケンスに加えられる．さらに，二足歩行による左右の揺れの影響を除くため，同じ側の脚が前に出ている姿勢においてのみ，画像を取得する．図 5.20 に，このビューシーケンス作成プロセスの概要を示す．

5.4.3 教示走行

経路教示は，通行人などの動的な障害物が存在しない環境において行われる必要がある．オペレータは，端末コンピュータからヒューマノイドロボットを遠隔操作し，この場合の歩行速度は一定値に設定される．経路の教示は，直進指示と回転指示により行われる．各直線経路では，ロボットは，ビューシーケンスを記録しなが

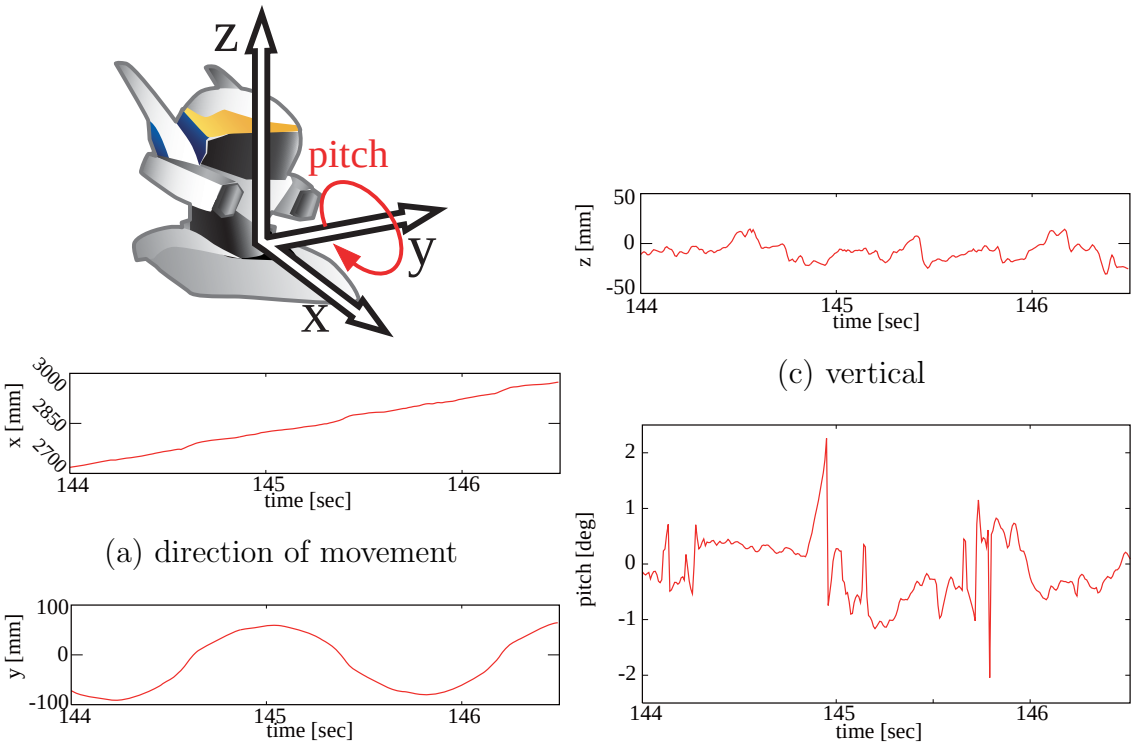
図 5.18 モーションキャプチャシステムによる歩行時の頭部位置，姿勢計測

ら，指定された直進歩行を行う．曲がり角では，“Right”（右回転），“Left”（左回転）の何れかのコマンドが端末から送られ，次の直進経路の方向を向くまでその場回転を行う．経路の終端では，“Stop”（停止）コマンドが送られる．この教示走行では，30[frame/s] でキャプチャが行われ，これらの画像は，$640 \times 480 \times 8$bit から $320 \times 240 \times 8$bit のサイズに縮小される．

5.4.4 自律走行

自律走行では，そこで使用されるビューシーケンスを作成した教示走行時と同じような照明条件である必要がある．画像の比較に用いられる正規化相関演算は，画像全体の大局的な照明変化に対しては頑健であるが，局所的な照明変化に対してはその影響を避けられないからである．

ロボットは，教示走行時と同じ場所からナビゲーションを開始し，前述したテンプレートマッチングの結果を基に姿勢を修正しながら，記録された経路に沿って歩



(a) direction of movement

(b) horizontal

(c) vertical

(d) pitch

図 5.19 歩行時のヒューマノイドロボットの頭部の変位と揺れ

く．また，ロボットは，ビューシーケンス内における自己位置を，相関値に基づいて離散的に認識する．言い換えれば，ロボットは，比較対象となるビューシーケンス N，$N+1$ のうち現在の画像に似ているを方を，現在位置を示すものとみなす．

ナビゲーションが終了する場所にくると，記録されたゴール位置の「数歩手前」で，ゴールの画像との相関値が一番高くなる．そこで，ここでは，通常の相関値比較とは異なる手法を用いる．図 5.21 は，教示走行時の相関値の変化を表す．まず，教示走行において，ゴール地点における相関値 M_{N-1} が記録される（図 5.21 の (p)）．自律走行時には，相関値 M_{N-1} が，記録された値 (p) に下がるまで歩行を続ける．この処理により，ロボットの停止位置がより教示時のゴールに近くなる．

自律走行時に歩行者などの移動障害物がロボットの視界に現れ，算出される相関値が低下すると，ロボットは，相関値が回復するまで一時的に停止する．つまり，このナビゲーション手法は,「恒常的な」環境変化においては無力だが,「一時的な」環境変化ならば，その変化が終了するのを待つことでナビゲーションを再開すること

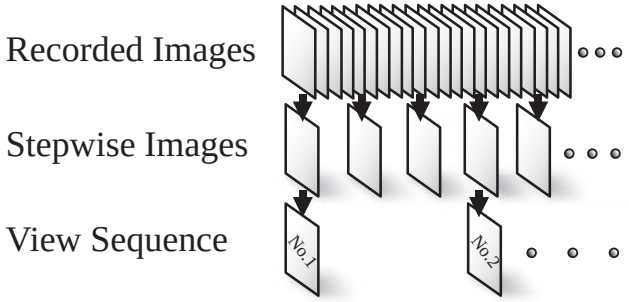


図 5.20　ビューシーケンスの作成

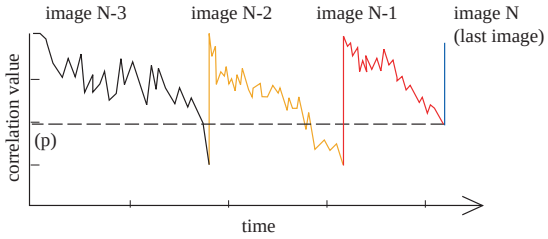


図 5.21　教示走行時の相関値の移り変わり

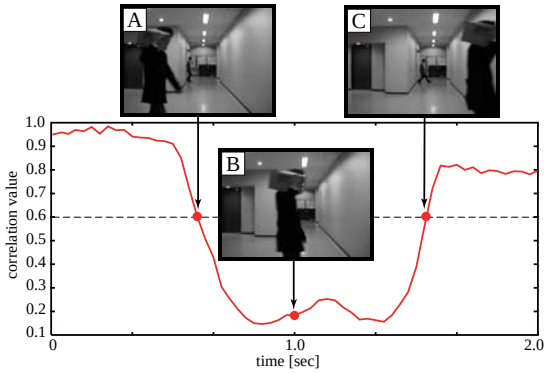


図 5.22　人が横切った場合の相関値への影響

ができる．図 5.22 は，ロボットの眼前を歩行者が横切った場合の相関値の変化を表している．この図中の点線で示される閾値まで相関値が低下するような状況では，ロボットは相関演算を続けつつ，その場に停止し足踏みを続ける．通行人がロボットの視界から消えると，相関値は閾値まで回復し，ロボットは再びナビゲーションを再開する．この場合は，閾値を 0.6 に設定している．ロボットは，障害物までの正確な距離が分かるわけではないので，障害物を避けたり移動させたりするためには，ステレオカメラなどの他のセンシング手法が必要になる．

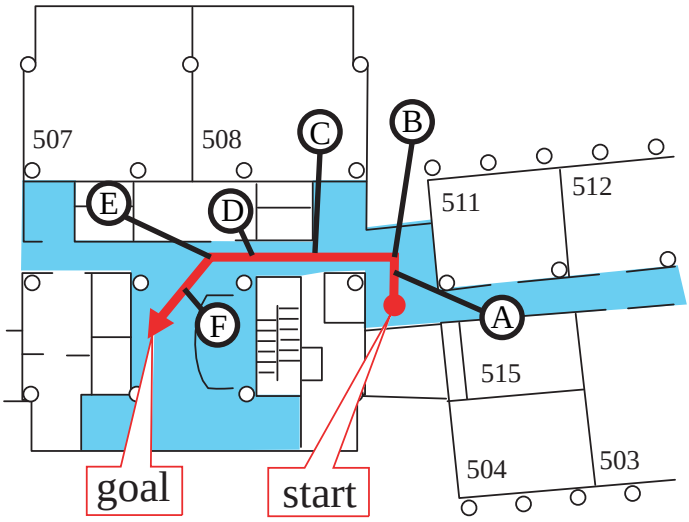


図 5.23　ビューシーケンスを用いた屋内ナビゲーションの実験環境

図 5.24　ビューシーケンスのサンプル画像（合計 41 画像）

5.4.5　ナビゲーション実験

教示走行

提案手法の有効性を実証するために，図 5.23 に示すような環境で実験を行った．図中の陰付き領域は，歩行可能な廊下を表す．

教示走行では，実験室入り口からエレベータまで，二カ所のコーナーを含むおよそ 16[m] の経路を歩行した．歩行速度は，0.1[m/s]，コーナーでの回転速度は 0.2[rad/s] である．この教示走行では，合計で 3427 枚の画像が記憶された．

記憶された画像から，前述ような画像選択を行い，41 枚の画像からなるビューシーケンスが作成された．図 5.24 に，ビューシーケンス内の画像のサンプルを示す．

自律走行

自律走行時のスナップショットを図 5.25 に示す．右上と右下の画像はそれぞれ，自律走行時のカメラ画像とそれに対応するビューシーケンス内の画像を示す．例えば，図 5.25（c）にあるように，教示走行時と自律走行時では，ドアの開閉状態が変化するような環境の変化が起こっているが，問題なく歩行している．

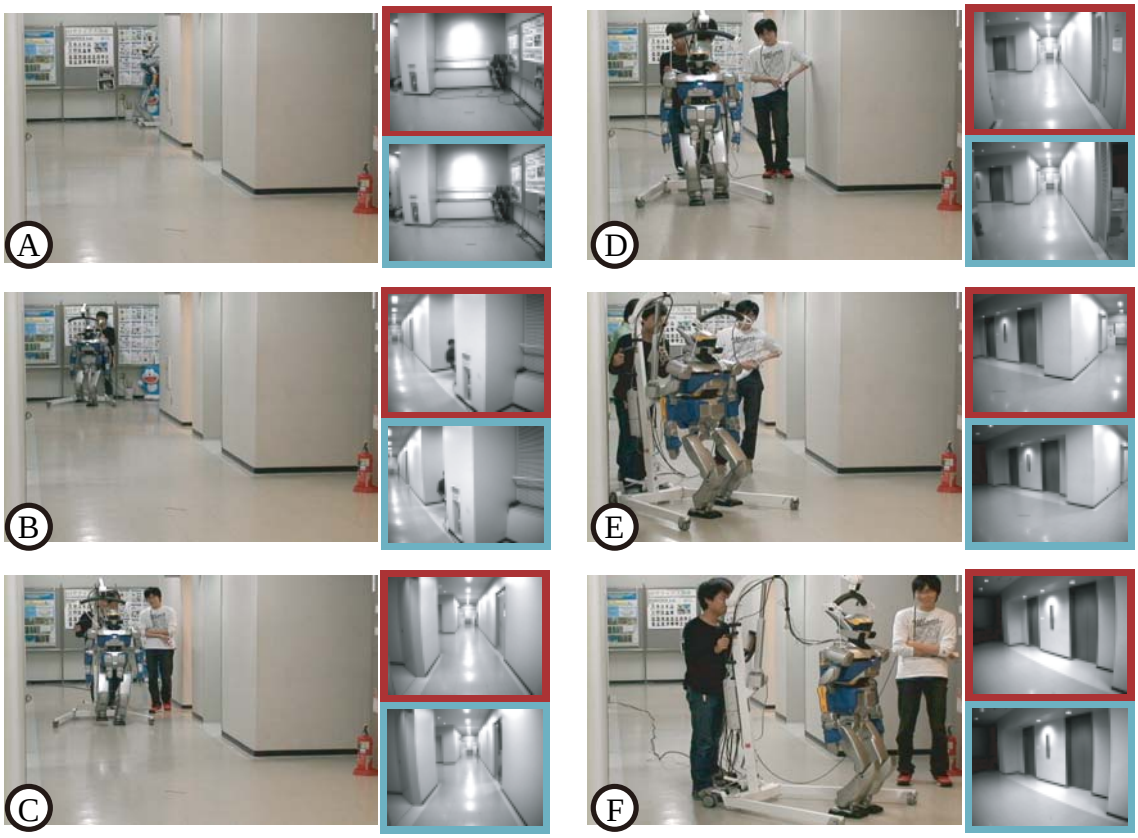


図 5.25　自律走行時のスナップショット

図 5.26 は，マッチングプロセスで計算された相関値の変化を表す．図中の二カ所の陰付きの領域，つまり（a）と（b）は，ロボットが回転している部分である．ビューシーケンスは，回転中の画像を保持していないので，ここでは一時的に相関値が下がっている．直進している期間の相関値の平均は 0.91 であり，ナビゲーションが安定して実現されていることが分かる．このナビゲーションには，およそ 180[s] の時間を要した．

提案手法の評価

提案手法の評価を行うため，モーションキャプチャシステムを利用して，教示走行および自律走行時の経路を計測した．モーションキャプチャシステムの計測可能範囲は，図 5.23 の（D）地点からゴールまでである．

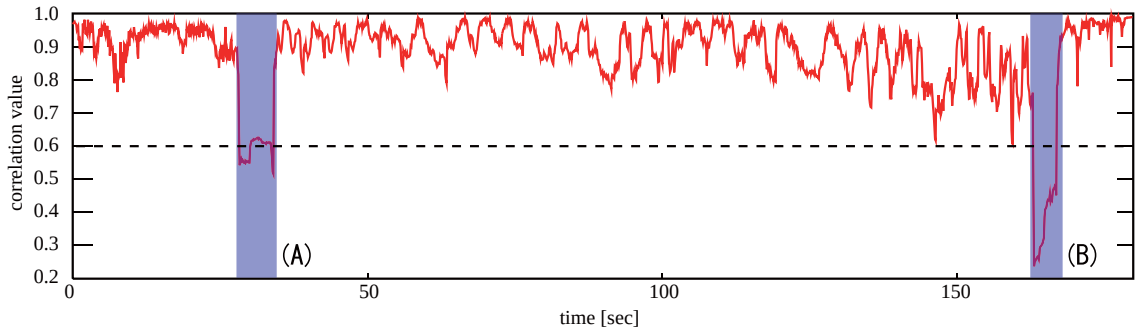
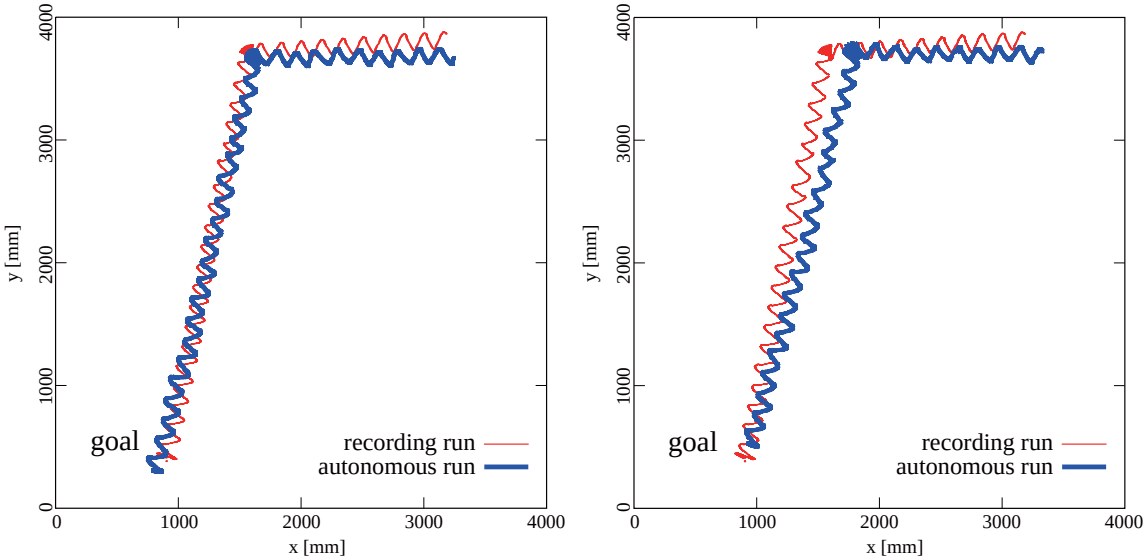


図 5.26　自律走行時のカメラ画像の相関値の変化

同一のビューシーケンスを用いて，5 回の自律走行を行った．そのうち，二試行分の結果を図 5.27 に示す．ナビゲーションのゴールは，図に示す軌跡の右下の点である．図 5.27（a）では，自律走行時の経路が，教示走行時とほぼ同じ軌跡を描いている．図 5.27（b）では，ロボットは，教示走行時よりも早くコーナーを曲がってしまっているが，ゴールに近づくにつれ教示走行時の経路に漸近している．経路からどれほど外れた場合にマッチングが失敗するかは，カメラの画角と環境に依存する．図 5.25 に示す実験環境では，廊下の幅は 2.0[m] であり，ロボットが廊下の端までずれて移動した場合でもマッチングは正常に行われ，ロボットの経路は補正される．図 5.28 は，自律走行時にロボットが停止した位置を示す．図の中心にある十字マークは，教示走行時の停止位置，つまりゴール位置を示す．平均誤差は，101[mm] であった．

次に，ロボットの位置誤差に対する頑健性を評価するために，ロボットの移動開始位置をずらしての自律移動実験を行った．図 5.29 は，実験経路および複数のスタート位置を示している．教示走行では，図中の（A）地点からゴールまでの歩行を行った．次に，自律走行では，(A)，(B)，(C) の異なる三地点から歩行を開始し，同一のビューシーケンスを利用して教示時のゴールを目指した．(B) 地点は（A）地点の 0.9[m] 後方，(C) 地点は（B）地点の 0.75[m] 横側に位置している．図 5.30 は，それぞれの自律走行時におけるロボットの取得画像を示している．自律走行開始地点でのカメラ画像は，その位置の違いにより，それぞれ大きく異なっている．しかし，自律走行が正しく行われるに従い歩行経路が収束して，カメラ画像が似たものになっている様子が分かる．各スタート地点からの自律走行をそれぞれ 5 回行い，それぞれのゴール付近における停止位置を記録したものを図 5.31 に示す．停止位置誤差は最大で，0.3[m] であった．



(a) 試行 1　　(b) 試行 2

図 5.27　教示走行時および自律走行時の移動軌跡

5.5. 本章のまとめ

本章では，自然なヒューマンロボットインタラクションの実現を目的として開発された，ヒューマノイドロボットシステムについて述べた．本システムには，人とコミュニケーションを行うための視覚計測機能と対話機能が実装されている．従来の ASKA のシステムと比較すると，その身体機能を生かしたより適切なジェスチャ応答や把持，似顔絵作成などのタスク実行が可能となった．また，二足歩行ロボットの移動機能を生かしたタスクを実行するために，屋内ナビゲーション機能を実装し，その評価を行った．音声処理に関しても，本体内蔵マイクを用いた音声認識や，非定常雑音の識別などの機能が新たに追加されている．

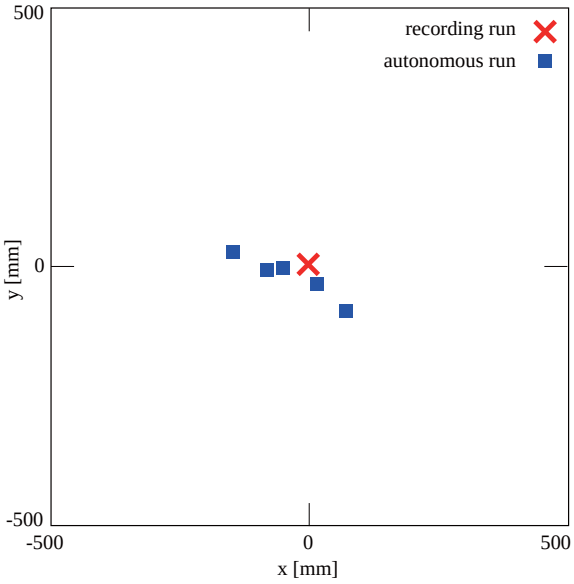


図 5.28　教示走行時および自律走行時の停止位置

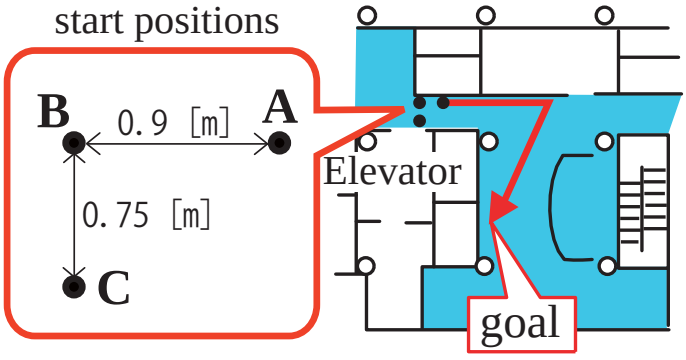


図 5.29　実験環境および，三つのスタート地点

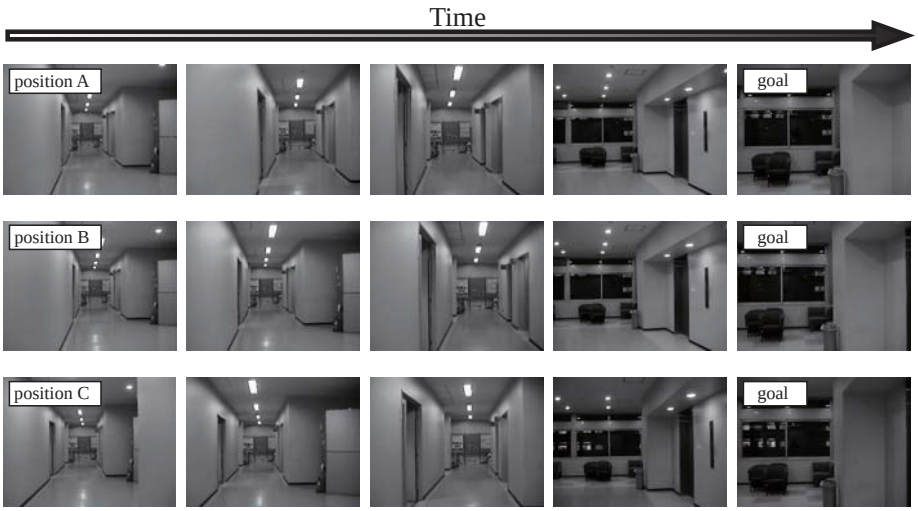


図 5.30　自律走行時のカメラ画像：異なる三地点からのスタート

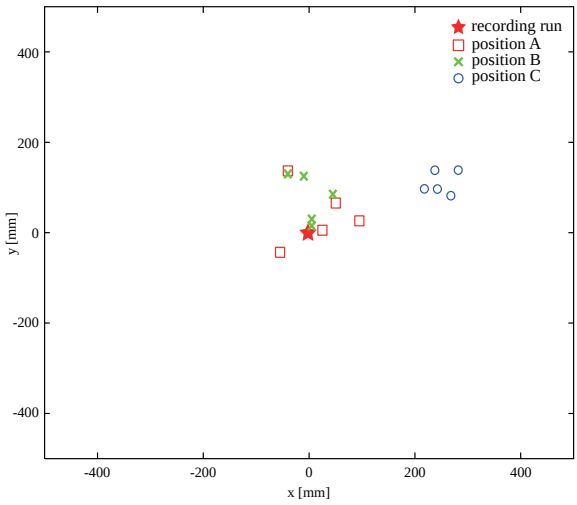


図 5.31　異なるスタート地点からの自律走行時の停止位置

第6章　結論

6.1. 本論文のまとめ

本研究では，人とヒューマノイドロボットとのコミュニケーション円滑にするために必要な情報を検討し，その情報を基に実装されたインタラクション機能を実環境下で運用，評価を行った．ここにその概要をまとめる．

第 1 章では，序論として拡大するロボット市場を背景とする人とロボットの共存の重要性，そしてその中での本研究の目的を述べた．

第 2 章では，はじめに人間同士のインタラクションについて考察することによって，人間とロボット間で成立し得るインタラクションの種類を整理し，インタフェースの観点から 4 種類に大別した．また，それぞれの情報を用いた人間とロボットのコミュニケーションに関する先行研究について言及し，本研究のアプローチについて述べた．

第 3 章では，自然なヒューマンロボットインタラクションの実現を目的として開発された，受付案内ロボット ASKA について述べた．学外からの来場者向けのオープンキャンパス時に，受付案内ロボット ASKA を用いた実験を行い，ヒューマノイドロボットに対して対話行動を行う場合においても，発話時は顔方向・視線が安定しロボットの頭部を注視するといった傾向を確認した．ASKA のシステムは，互いに独立して動作する音声認識，音声合成，人発見，人計測，胴体ジェスチャおよび頭部ジェスチャのモジュール群と，サーバから構成され，これにより人とコミュニケーションを行うための視覚計測機能と対話機能が実装した．音声対話機能は，大語彙連続音声認識エンジン Julius をベースとしたキーワードマッチによる音声認識理解部と音声合成部で構成され，視覚を用いた計測機能は，視差画像生成，DP マッチングおよび顔情報計測技術をベースに構成した．開発は，奈良先端科学技術大学院大学の複数の研究室によって行われ，様々な技術を実環境下で評価するための研究プラットフォームとして活用された．また，顔情報計測機能と音声認識機能を統

合した「アイコンタクト」による，有効発話区間推定機能や，ジェスチャ認識と音声認識を統合した，指示語を含む発話の認識機能，などのマルチモーダルな認識機能が実装・評価された．

第 4 章では，顔ロボットを用いたインタラクションについて述べた．前章で構築した ASKA システムの顔部分の情報伝達に着目し，これを用いて遠隔対話システムに応用したシステムを構築した．前章では，人間と自律システムであるヒューマノイドロボットとの対話を前提としていたが，遠隔操作型のヒューマノイドロボットシステムとの対話を考慮することで，ロボットシステムはより柔軟なものとなった．本システムを，従来の電話やビデオチャットなどの遠隔対話システムと比較することによって，ロボット遠隔対話の特徴を明らかにした．

第 5 章では，自然なヒューマンロボットインタラクションの実現を目的として開発された，ヒューマノイドロボットシステムについて述べた．HRP-2 を利用することで，従来の ASKA のシステムと比較すると，その身体機能を生かしたより適切なジェスチャ応答や把持，似顔絵作成などのタスク実行が可能となった．また，音声対話機能に関しても，本体内蔵マイクを用いた音声認識や，非定常雑音の識別などの機能を追加することで，対面インタラクションにハンドマイクを用いる不自然さが解消された．さらに，タスクの一つとして HRP-2 の二足歩行機能の利用した屋内ナビゲーション機能を実装した．従来，車輪型移動ロボットに利用されていたビューシーケンスを，ヒューマノイドロボットに適用することで，屋内での移動を可能にした．

これらのシステムは共に，ROBODEX や愛・地球博をはじめとした様々な環境下でデモンストレーションが行われ，実環境下での有用性が確認されている．以上の事により，アイコンタクトをはじめとする非言語情報と音声情報を人とヒューマノイドロボットとのコミュニケーションに利用する事の有効性が確認されたといえる．

6.2. 今後の課題と展開

以下，に本研究で開発したインタラクションシステムにおける現状での問題点と今後の課題を簡単にまとめる．今回開発したシステムでは，ASKA，HRP-2 共に単純な黒板モデルを用いてシステム全体のコントロールを行っている．これは，移植性やモジュール実装のシンプルさという利点がある反面，モジュール間における応答などメッセージのやり取りが煩雑になりがちである．また，それに付随してメッ

セージの取りこぼしなども起こりやすい．今までのような一問一答型のタスクにおいては，それがあまり問題にはならなかったが，今後要求されるタスクが複雑化するに従い，新たなアーキテクチャの採用に迫られる可能性がある．また，研究・開発の長期化に伴い，筆者の専門ではない音声認識技術分野においては，先端技術との乖離が見られた．例えば，我々のシステムでは，基本的にキーワードベースの認識を行っているが，同じ Julius を利用する「たけまるくん」などのシステムでは，より柔軟な例文ベースのシステムを利用している．また，音源分離などの技術は，実環境での動作に非常に適していることから，これからロボット分野にも多く取り入れられていくと考えられる．

また，最後に，ヒューマノイドロボットとのインタラクションに関する研究の今後展開について，本論文では扱う事ができなかった観点から簡潔に述べたいと思う．一つは，ヒューマノイドロボットとの外見とインタラクション機能に関してである．本論文執筆時においても，精巧な外見をもつヒューマノイドロボットが次々と開発されている．より人間的な外見を持つヒューマノイドロボットは，これからも積極的に研究，開発されていくと考えられる．しかし，ロボットの外見が人間に近づくほど，人間は無意識のうちにロボットに人間的な，つまり高機能で自然な応答を期待してしまう．今後，敢えてロボットの外見をその機能に見合ったものに制限することで，ユーザの期待の壁を低くするという考え方も重要になると思われる．例えば，ロボットが認識しやすいようにハッキリと話す，あるいは，複雑な対話を避けるといったような人間側の適応を引き出しやすくする，という効果が期待できる．もちろん，これはユーザが感じるストレスとトレードオフの関係になることが予想されるので，ロボットの外見と機能との適切な対応関係を探る必要がある．

もう一つは，人間同士の対話に縛られないマルチモーダル性に関してである．本論文では，人間同士の対話を規範としてロボットのマルチモーダルなインタラクション機能を研究してきた．例えば，アイコンタクトやジェスチャ認識がその例にあたる．しかし，この手法だけでは，人間と同等の認識能力が実現されない限り，真に自然なインタラクションを成すことはできない．それはロボットであることの利点を生かしきれて為であるとも考えられる．もちろん，ヒューマノイドロボットとの自然で人間的なインタラクションが実現できれば，それはすばらしい成果になるが，それとは別のアプローチも同時に探られるべきである．つまり，ロボットとのインタラクションの自然さを人間同士のそれを基準に決定する必要はなく，ヒューマノイドロボットとのインタラクションにおける自然さ新たに提唱できれば良い．例え

ば，コントローラと自然対話入力をシームレスに利用するような，人間同士の対話とは微妙に異なるインタラクション手法が，人間とロボットとの間においては自然なものになる可能性もある．ヒューマノイドロボットとの真に自然なインタラクションとは何か，それはどのように評価されるべきものなのか，残念ながら現時点では明らかではない．しかし，序章に述べたようにこの分野の研究はまだ若く，大いに発展の余地がある分野である．今後の展開に期待したい．

謝辞

本論文に書き記した一連の研究は，筆者が奈良先端科学技術大学院大学 情報科学研究科在学中および，研究員として本学に在職中，ロボティクス講座（小笠原研究室）において執り行われたものです．この研究は実に多くの方々のお力添えにより実現することができました．ここで御礼を申し上げます．本学 情報科学研究科 小笠原 司 教授は，修士課程の学生だった時分より長きに渡り，未熟な筆者を懇切な御指導と適切な御助言によりここまでお導き下さいました．深く感謝致します．本学 情報科学研究科 鹿野 清宏 教授には，本研究をまとめるにあたり，ご多忙にもかかわらず博士論文の審査委員を引き受けて下さり，数々の御助言，御指摘を頂きました．深く感謝致します．本学 情報科学研究科 高松 淳 准教授には，短い間でしたが研究に関する有用な助言や，新たな視点からのご指摘を頂きました．深く感謝致します．本学 情報科学研究科 松本 吉央 客員准教授には，お忙しい中，筆者の拙い英語論文や発表を入念にチェックして頂き，丁寧な御指導，御助言を頂きました．また，学内のプロジェクトに始まり，ROBODEX や愛知万博をはじめとする大規模なイベントへの参加など多数の貴重な経験をさせて頂きました．深く御礼申し上げます．本学 情報科学研究科 栗田 雄一 助教，および 竹村 憲太郎 助教には，教員として赴任されてからはもちろん，御二方共に学生の時分より，私の研究や私生活において様々なご助力を頂きました．ありがとうございました．本講座秘書 金岡恵様には，物品発注や出張資料作成などで多々お世話になりました．おかげで研究を円滑に進めることができました．また適切な資産運用法について御教授頂き，多くの貴重な経験をさせて頂きました．ありがとうございました．本学 情報科学研究科 上田 淳 助教（現：Georgia Tech Assistant Professor）には，研究会で様々な質問・提言を頂きました．また，アメリカに移られてからも，学生の指導を通して，多くの助言を頂きました．ありがとうございました．上田悦子 氏（現：奈良産業大学 情報学部 准教授）には，本学の助教を勤めていた時分から様々な御指導，御意見を頂きました．博士論文の執筆の際には適切なアドバイスを頂き，精神的に窮迫していた筆者を助けて

くださいました．西村 竜一氏（現：和歌山大学 システム工学部 デザイン情報学科 助教）には，氏が学生の時分よりROBODEXや愛知万博などのプロジェクトの遂行にあたって音声認識など多くの場面で力を貸していただきました．心より感謝いたします．小枝正直氏（現：大阪電気通信大学　総合情報学部　メディアコンピュータシステム学科 准教授）にも，栗田，竹村の両助教と同様，学生時代よりお世話になりました．特に研究室の環境やネットワークに関して多くのことを学ばせていただきました．本学ロボティクス講座の学生の皆様には，研究の一部を様々な形で手助けして頂き，感謝しています．特に卒業生である，塚谷浩司氏，森永英文氏，根津猛氏，志水芳直氏，高松渉氏には，学生時代にプログラムや実験などに関して多くの貢献をしていただきました．感謝しています．川田工業株式会社および，General Robotixの皆様には，HRP-2のメンテナンスや基本操作の指導などでご助力いただきました．

本研究の一部は，21世紀COEプログラム「ユビキタス統合メディアコンピューティング」，NEDO戦略的先端ロボット要素技術開発プロジェクト「高齢者対応コミュニケーションRTシステム（サービスロボット分野)、快適生活支援RTシステムの開発」の支援により，奈良先端科学技術大学院大学で実施されました．誠に遺憾ながら，本論文は推敲を重ねた現在も，未だ洗練されているとは言い難い状況にありますが，わずかなりとも誰かの助けになればと思います．最後に，学生生活，それも他人よりも長時間，をおくるにあたり全面的に協力・応援してくれた家族に心から感謝します．

著者研究業績

学術雑誌論文

1. 松本 吉央, 怡土 順一, 竹村 憲太郎, 小笠原 司 : ”リアルタイム顔・視線計測システムの開発と知的インタフェースへの応用”, 情報処理学会論文誌 コンピュータビジョンとイメージメディア, vol.47, No.SIG 15(CVIM16), pp.10-21, 2006
2. 怡土 順一, 上田 悦子, 松本 吉央, 小笠原 司 : ”3 次元顔情報計測に基づく対話ロボットを介した遠隔コミュニケーション”, 日本顔学会誌 vol.8, No.1, 2008
3. J. Ido, Y. Shimizu, Y. Matsumoto, T. Ogasawara : ”Indoor Navigation for Humanoid Robot Using View Sequence”, International Journal of Robotics Research

国際会議議事録 (査読あり)

1. J. Ido, K. Takemura, Y. Matsumoto, T. Ogasawara : ”Robotic Receptionist ASKA : A Research Platform for Human-Robot Interaction”, Proceedings of the 2002 IEEE Int. Workshop on Robot and Human Interactive Communication (ROMAN2002) , pp.306-311, Berlin, 2002.
2. J. Ido, Y. Myouga, Y. Matsumoto, T. Ogasawara : ”Interaction of Receptionist ASKA Using Vision and Speech Information”, Proceedings of International Conference on Multisensor Fusion and Integration for Intelligent Systems, pp.335-340, 2003.
3. Y. Matsumoto, J. Ido, K. Takemura, M. Koeda, T. Ogasawara : ”Portable Facial Information Measurement System and Its Application to Human Modeling and Human Interfaces”, The 6th International Conference on Automatic Face and Gesture Recognition (FG2004), pp.475-480, 2004.

4. J. Ido, R. Nisimura, Y. Matsumoto, T. Ogasawara : "Humanoid with Interaction Ability Using Vision and Speech Information", Proceedings of International Conference on Intelligent Robots and Systems, pp.1316-1321, 2006.
5. J. Ido, W. Takamatsu, Y. Matumoto, T. Ogasawara : "Indoor Navigation for Humanoid Robot Using View Sequence", Proceedings of 9th International Conference on Climbing and Walking Robots, pp216-220, Sep. 2006.
6. J. Ido, E. Ueda, Y. Matumoto, T. Ogasawara : "Robotic telecommunication system based on facial information measurement", Proceedings of the 12th international conference on Intelligent user interfaces, pp.266-269, 2007.

国内発表

1. 西村 竜一, 怡土 順一, 李 晃伸, 松本 吉央: "情報科学研究の実環境プラットフォームとしての受付案内ロボット ASKA", 情報処理学会第 64 回全国大会講演論文集, Vol.4, pp.565-570, 4B-2-01, 2002-3.
2. 明賀 陽平, 怡土 順一, 松本 吉央, 小笠原 司: "共存型ロボットのための視覚に基づく対話行動の認識 ," 第 21 回日本ロボット学会学術講演会予稿集 CD-ROM, 3H13, 2003.9.
3. 塚谷 浩司, 怡土 順一, 松本 吉央, 小笠原 司："受付案内ロボット ASKA におけるステレオ画像を用いたジェスチャ認識", 第 4 回 SICE システムインテグレーション部会講演会講演論文集 (SI2003), 1I2-7, pp.305-306, 2003.12.
4. 根津 猛, 怡土 順一 , 上田 悦子, 松本 吉央, 小笠原 司, "顔ロボットを用いた双方向遠隔コミュニケーションシステム", 第 5 回計測自動制御学会システムインテグレーション部門講演会 (SI2004), 2D4-2,2004.12.
5. Florian Schorradt, 怡土 順一, 松本 吉央, 小笠原 司: "ヒューマノイドロボットのための視覚に基づくインタラクションシステム", ロボティクス・メカトロニクス講演会'04 講演論文集, 2P1-H-69,2004.6.
6. 根津 猛, 怡土 順一, 松本 吉央, 小笠原 司: "受付ロボット ASKA を用いた遠隔コミュニケーションシステム", ロボティクス・メカトロニクス講演会'04 講演論文集, 2A1-H-33, 2004.6.
7. 怡土 順一, 西村 竜一, 末永 剛, 佐々尾 直樹, 近藤 理, 松本 吉央, 小笠原司："HRP-2 によるマルチモーダルインタラクション ～視覚と音声を利用した対話システム

の構築〜", 第 23 回日本ロボット学会学術講演会予稿集, 1H12, 2005.9.15-17.

8. 佐々尾 直樹 近藤 理 末永 剛 怡土 順一 松本 吉央 小笠原司："HRP-2 によるマルチモーダルインタラクション〜似顔絵描きの実現〜", 第２３回日本ロボット学会学術講演会予稿集, 1H13, 2005.9.15-17.
9. 高松 渉, 怡土 順一, 松本 吉央, 小笠原 司 : "ビューシーケンス を用いたヒューマノイドロボットの屋内ナビゲーション", ロボティクスメカトロニクス講演会 2006 (ROBOMEC2006), 1A1-D23, 2006.5.26-28.
10. 怡土 順一, 上田 悦子, 松本 吉央, 小笠原 司 : "顔情報計測に基 づく表情提示ロボットを用いた遠隔コミュニケーションシステム", ロボティクスメカトロニクス講演会 2006 (ROBOMEC2006), 2P2-C20, 2006.5.26-28.
11. 末永 剛, 怡土 順一, 上田 悦子, 松本 吉央, 小笠原 司 : "顔情報計測に基づくヒューマノイドロボットを介した遠隔コミュニケーショ ン", ロボティ クスメカトロニクス講演会 2007 (ROBOMEC2007), 1A2-O09, 2007.5.11-12.
12. 志水 芳直, 湯浅 卓也, 怡土 順一, 松本 吉央, 小笠原 司 : "ビューシーケンスによるヒューマノイドロボットの屋内ナビゲーション", ロボティ クスメカトロニクス講演会 2007 (ROBOMEC2007), 2P1-C11, 2007.5.11-12.

書籍

1. 奈良先端科学技術大学院大学 OpenCV プログラミングブック制作チーム, "OpenCV プログラミングブック", 毎日コミュニケーションズ, 2007.9.26.
2. J. Ido, R. Nisimura, Y. Matsumoto, T. Ogasawara : "Multi-Modal Interaction with Humanoid Robot based on Eye Contact", I-Tech Education and Publishing

その他の発表

1. K. Takemura, J. Ido, Y. Matsumoto, T. Ogasawara : "Development of Non-Contact Drive Monitoring System for Advanced Safety Vehicle",IEEE/ASME International Conference on Advanced Intelligent Mechatronics(AIM2004),pp.1119-1122,Kobe, Japan, 2003.7.

2. K. Takemura, J. Ido, Y. Matsumoto, T. Ogasawara : "Drive Monitoring System Based on Non-Contact Measurement System of Driver's Focus of Visual Attention," Proceedings IEEE Intelligent Vehicles Symposium (IV2003), pp.581-586, 2003.6.
3. 坪田 智子, 怡土 順一, 松本 吉央, 小笠原 司. "顔情報の計測に基づくユーザの心理状態推定", 第 18 回日本ロボット学会学術講演会予稿集, pp.803-804, 2001.9.
4. 竹村 憲太郎, 怡土 順一, 松本 吉央, 小笠原 司 : "ドライバ注視点計測システム", 日本ロボット学会創立 20 周年記念学術講演会予稿集, 3C17, 2002.10.

5. 怡土 順一, 松本 吉央, 小笠原 司: "顔情報に基づくドライバーの状態計測", ロボティクス・メカトロニクス講演会'02 講演論文集,1P1-C05, 2002.6.
6. 近藤 理, 怡土 順一, 松本 吉央, 小笠原 司 : "ヒューマノイドによるレーザスキャナを用いた三次元環境認識", 第 6 回計測自動制御学会システムインテグレーション部門講演会 (SI2005), 2005.12.
7. 湯浅 卓也, 怡土 順一, 栗田 雄一, 松本 吉央, 小笠原 司： "ヒューマノイドによるレーザレンジファインダを用いた三次元地図作成と障害物回避", 計測自動制御学会システムインテグレーション部門講演会, pp.539-540, 2007.12.
8. 湯浅 卓也, 怡土 順一, 栗田 雄一, 松本 吉央, 小笠 原司："ヒューマノイドによるレーザレンジファインダを用いた三次元環境地図作成", 第 25 回日本ロボット学会学術講演会予稿集, 3D21, 2007.9.13-15.
9. 荒木 天外, 怡土 順一, 竹村 憲太郎, 栗田 雄一, 松本 吉央, 小笠原 司: "画像に基づくナビゲーションのための 3 次元環境地図の構築", ロボティクスメカトロニクス講演会 2008 (ROBOMEC2008), 2P2-C15, 2008.6.6-7.
10. 河村 雅人, 怡土 順一 , 栗田 雄一, 松本 吉央, 小笠原 司; "ロボットによる情報提示を目指した関心発生源マップの作成", ロボティクスメカトロニクス講演会 2008 (ROBOMEC2008), 2P1-H03, June 6-7, 2008.

メディア掲載など

1. 2000.12.01, 日本テレビ「ズームイン !! 朝 ! 」, 情報トレイン「21 世紀目前びっくりマシーン最前線」

2. 2001.10.25, 奈良テレビ, ロボティクス研究室の研究内容を紹介
3. 2001.11.05, 朝日新聞　朝刊, 「奈良先端科学技術大学院大学 最新研究を公開 市民ら感心　ロボットと話も」
4. 2002.03.20, フジテレビ　情報プロジェクト, 「お天気コーナー」にて ASKA が紹介
5. 2002.04.04, アメリカ　CNN テレビ, ROBODEX2002 特集　ASKA とレポーターのコミュニケーション
6. 2002.05, Newton 2002 年 6 月号 Vol.22, No.7, P127, コミュニケーション能力を得た－ＡＳＫＡ－
7. 2002.06.02, NHKBS2　「おーい日本　今日はとことん奈良県」10：00〜19：00 放映, 「おもしろ博物館」コーナー
8. 2002.09.11, NHK ニュースかんさい発 2002 奈良ウィーク「新・まほろば宣言」テーマ：産業 研究室紹介他, 受付ロボット「ASKA」が番組アシスタントとして登場
9. 2003.04.07, Weekly ぴあ　関東版 ROBODEX2003 関連記事で ASKA の紹介
10. 2004.10.08, 奈良テレビ「ざっくばらん」, 受付案内ロボット ASKA
11. 2004.12.29, 讀賣新聞　朝刊, 次世代ロボットお目見え ¡人型¿ アイ・コンタクトで対話－視線を合わせると対話するロボットと松本助教授
12. 2005.05.28, 奈良テレビ放送　街角ウォッチャー, [調査]　日本の技術は世界一！最先端のロボット研究とは？館内案内ロボット たけまるくん・受付案内ロボット ASKA・ヒューマノイド HRP-2・顔情報計測について紹介（小笠原教授出演）
13. 2005.06.11, NHK 総合　「生中継 万博で未来がわかる！最新ロボット大研究」20：00〜20：45, HRP-2 のデモの様子が放送
14. 2005.11.30, NHK ニュース　おはよう日本　5:30〜5:50, 人と話す最新ロボット
15. 2005.12.01, PC WATCH, 2005 年国際ロボット展開催 〜産業用ロボットからパックマンロボットまで子どもの顔を見て似顔絵を描く HRP-2〜
16. 2005.12, ロボコンマガジン No.43（Ohmsha), 地域 ICT　未来フェスタ in かがわ　HRP-2 のデモの様子が掲載
17. 2006.01.20, 毎日放送　ちちんぷいぷい, ＜中継＞奈良県生駒市－学生が作る次世代ロボット
18. 2006.03.06, 日本経済新聞（夕刊）, NAIST 宣伝広告（文部科学省飴些慾穎　繁

槁　梁魑　僂　鼎里　里薀察砲 HRP-2 のデモ

19. 2006.04, よくわかる国語の学習 3, 生活を便利にするロボット　ヒューマノイドロボット「HRP-2」
20. 2006.04, ワイド＆ビジュアル　最新国語資料集, 最先端技術がやってきた　ヒューマノイドロボット「HRP-2」
21. 2006.06, 東京エレクトロンデバイス『インレビアムラボ』第 3 号, 人とインタラクション可能なヒューマノイドロボット HRP-2

参考文献

[1] 新エネルギー・産業技術総合開発機構. 技術戦略ロードマップ. http://www.nedo.go.jp/roadmap/data/manu_rm1.pdf.

[2] M.Fujita. Digital creatures for future entertainment robotics. In *Proc. of IEEE International Conference on Robotics and Automation*, pp. 801–806, 2000.

[3] 藤田喜弘. パーソナルロボット R100. 日本ロボット学会誌, Vol. 18, No. 2, pp. 198–199, 2000.

[4] 横山 真男, 青山 一美, 菊池 英明, 帆足 啓一郎, 白井 克彦. 人間型ロボットの対話インターフェイスにおける発話交替時の非言語情報の制御. 情報処理学会論文誌, Vol. 40, No. 2, pp. 487–496, 1999.

[5] A.Kendon. Some functions of gaze-direction in social interaction. *Acta Psychologica*, Vol. 26, pp. 22–63, 1967.

[6] 深山篤, 大野健彦, 武川直樹, 澤木美奈子, 萩田紀博. 擬人化エージェントの印象操作のための視線制御方法. 情報処理学会論文誌, Vol. 43, No. 12, pp. 3596–3606, 2002.

[7] 三吉秀夫綿貫啓子. ヒューマンインターフェイスのための人間の振舞いの解析-マルチモーダル対話データの解析-. Technical Report SIG-SLUD-9902, 人工知能学会研究会資料, 1999.

[8] J.Cassell, D.McNeill, and K.E.McCullough. Evidence for one underlying representation of linguistic and non-linguistic information. *Pragmatics and Cognition*, Vol. 7, No. 1, pp. 1–33, 1999.

[9] 松井 俊浩, 麻生 英樹, John Fry, 浅野 太, 本村 陽一, 原 功, 栗田 多喜夫, 速水悟, 山崎 信行. オフィス移動ロボット jijo-2 の音声対話システム. 日本ロボット学会誌, Vol. 18, No. 2, pp. 300–307, 2000.

[10] R.Stiefelhagen, C.Fuegen, P.Gieselmann, H.Holzapfel, K.Nickel, and A.Waibel. Natural human-robot interaction using speech, gaze and gestures. In *Proc. of*

IEEE/RSJ International Conference on Intelligent Robots and Systems, 2004.

[11] M.Ehrenmann, R.Becher, B.Giesler, R.Zöllner, O.Rogalla, and R. Dillmann. Interaction with robot assistants: Commanding albert. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, 2002.

[12] 神田 崇行, 石黒 浩, 小野 哲雄, 今井 倫太, 前田 武志, 中津 良平. 研究用プラットフォームとしての日常活動型ボット “robovie” の開発. 電子情報通信学会論文誌, Vol. J84-D-I, No. 3, pp. 1–10, Mar. 2001.

[13] Shuji Hashimoto Kenji Suzuki, Riku Hikiji. Development of an autonomous humanoid robot, isha, for harmonized human-machine environment. *Journal of Robotics and Mechatronics*, Vol. 14, No. 5, pp. 324–332, 2002.

[14] 小林 哲則, 白井 克彦. ヒューマノイドロボットにおけるマルチモーダル会話インタフェース. Technical report, 電子情報通信学会技術報告, Dec. 1999.

[15] H.Takuya, S.Masaru, S.Taichi, K.Hiroshi. Development of the interactive receptionist system by the face robot. In *Proc. of SICE Annual Conference*, pp. 1404–1408, 2004.

[16] M.Montemerlo, J.Pineau, N.Roy, S.Thrun, and V.Verma. Experiences with a mobile robotic guide for the elderly. In *Proc. of National Conference on Artificial Intelligence*, pp. 587–592, 2002.

[17] H.Hüttenrauch, A.Green, M.Norman, L.Oestreicher, and K.Severinson-Eklundh. Involving users in the design of a mobile office robot. *IEEE transactions on systems, man and cybernetics. Part C*, Vol. 34, No. 2, pp. 113–124, May. 2004.

[18] H.Okuno, K.Nakadai, H.Kitano. Realizing personality in audio-visually triggered non-verbal behaviors. In *Proc. of IEEE International Conference on Robotics and Automation*, pp. 392–397, Sep. 2003.

[19] M. Castrillón, J.Cabrera, D.Hernández, A.C.Domínguez, J.Lorenzo, J.Isern, C.Guerra, I.Pérez, A.Falcón, and M.Hernández. Eldi’s activities in a museum. In *Proc. of WAF-2001*, pp. 61–73, 2001.

[20] R.Simmons, J.Fernandez, R.Goodwin, S.Koenig, and J. O’Sullivan. Xavier: An autonomous mobile robot on the web. Robotics and Automation Magazine, 1999.

[21] R.Simmons V.Matellán. Implementing human-acceptable navigational behavior

and a fuzzy controller for an autonomous robot. In *Proc. of WAF-2002*, pp. 113–120, 2002.

[22] R.Siegwart, K.O.Arras, S.Bouabdallah, D.Burnier, G.Froidevaux, X.Greppin, B.Jensen, A.Lorotte, L.Mayor, M.Meisser, R.Philippsen, R.Piguet, G.Ramel, G.Terrien, and N.Tomatis. Robox at expo.02: A large scale installation of personal robots. *Robotics and Autonomous Systems*, Vol. 42(3-4), pp. 203–222, Mar. 2003.

[23] V.Graefe R.Bischoff. Dependable multimodal communication and interaction with robotic assistants. In *Proc. of IEEE International Workshop on Robot and Human Communication*, 2002.

[24] 株式会社ココロ. The press release of actroid. `http://www.kokoro-dreams.co.jp/ng/actroid/pdf/press_english.pdf`.

[25] H.Asoh, S.Hayamizu, I.Hara, Y.Motomura, S.Akaho, and T.Matsui. Socially embedded lerning of the office-conversant mobile robot jijo-2. In *Proc. of International Joint Conference on Artificial Intelligence*, 1997.

[26] T.Asfour, A. Ude, K.Berns, and R.Dillmann. Control of armar for the realization of anthropomorphic motion patterns. In *Proc. of IEEE-RAS International Conference on Humanoid Robots*, 2001.

[27] R.Stiefelhagen K.Nickel, E.Seemann. 3d-tracking of heads and hands for pointing gesture recognition in a human-robot interaction scenario. In *Proc. of Sixth International Conference on Face and Gesture Recognition*, May. 2004.

[28] S.Thrun, M.Beetz, M.Bennewitz, W.Burgard, A.B.Creemers, F.Dellaert, D.Fox, D.Hahnel, C.Rosenberg, N.Roy, J.Schulte, and D.Schulz. Probabilistic algorithms and the interactive museum tour-guide robot minerva. *International Journal of Robotics Research*, Vol. 19, No. 11, pp. 972–999, Nov. 2000.

[29] M.Scheeff, J.Pinto, K.Rahardja, S.Snibbe, and R.Tow. Experiences with sparky, a social robot. In *Proc. of the Workshop on interactive robotics and entertainment (WIRE-2000)*, 2000.

[30] H.Miwa, K.Itoh, M.Matsumoto, M.Zecca, H.Takanobu, S.Roccella, M.C.Carrozza, P.Dario, A.Takanishi. Effective emotional expressions with emotion expression humanoid robot we-4rii. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 2203–2208, 2004.

[31] R.simmons A.Bruce, I.Nourbakhsh. The role of expressiveness and attention in human-robot interaction. In *Proc. of IEEE International Conference on Robotics and Automation*, 2002.

[32] R.Gockley, A.Bruce, J.Forlizzi, M.Michalowski, A.Mundell, S.Rosenthal, B.Sellner, R.Simmons, K.Snipes, A.Shultz, and J.Wang. Designing robots for long-term social interaction. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, Apr. 2005.

[33] 松坂 要佐, 東條 剛史, 小林 哲則. グループ会話に参与する対話ロボットの構築. 電子情報通信学会論文誌, Vol. J84-D-II, No. 6, pp. 898–908, 2001.

[34] 藤江 真也, 江尻 康, 菊池 英明, 小林 哲則. 肯定的/否定的発話態度の認識とその音声対話システムへの応用. 電子情報通信学会論文誌, Vol. J88-D-II, No. 3, pp. 489–498, 2005.

[35] R.A.Brooks. The cog project: Building a humanoid robot. *Computation for Metaphors, Analogy and Agents*, Vol. 1562, , 1999.

[36] S.Kotosaka. Humanoid robot 'db'. In *Proc. of International Conference on Mach Automat*, pp. 21–26, 2000.

[37] A. Lee, T. Kawahara, and K. Shikano. Julius–an open source real-time large vocabulary recognition engine. In *Proc. ISCA-EUROSPEECH2001*, pp. 1691–1694, 2001.

[38] 株式会社テムザック. `http://www.tmsuk.co.jp/`.

[39] H.Yano H.Kozima. A robot that learns to communicate with human caregivers. In *Proc. of The First International Workshop on Epigenetic Robotics*, 2001.

[40] H. P. Blackboard Systems Nii. The blackboard model of problem solving and the evolution of blackboard architectures. *AI Magazine*, pp. 38–53, Summer 1986.

[41] K.Shikano A.Lee, T.Kawahara. Julius–an open source real-time large vocabulary recognition engine. In *Proc. of 7th European Conference on Speech Communication and Technology*, pp. 1691–1694, 2001.

[42] R.Nisimura, T.Uchida, A.Lee, H.Saruwatari, K.Shikano, Y.Matsumoto. Aska: Receptionist robot with speech dialogue system. In *Proc. of IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1341–1317, 2002.

[43] 河原達也, 住吉貴志, 李晃伸, 武田一哉, 三村正人, 伊藤彰則, 伊藤克亘, 鹿野清宏. 連続音声認識コンソーシアム2000年度版ソフトウェアの概要と評価. Technical Report 2001-SLP-38-6, 情処学研報, 2001.

[44] 岡田慧, 加賀美聡, 稲葉雅幸, 井上博允. PCによる高速対応点探索に基づくロボット搭載可能実時間視差画像・フロー生成方法と実現. 日本ロボット学会誌, Vol. 6, No. 18, pp. 896–901, 2000.

[45] R.Y.Tsai. A versatile camera calibration technique for high-accuracy 3d machine vision metrology using off-the-shelf tv camera and lenses. *IEEE Journal of Robotics and Automation*, Vol. RA-3, No. 4, pp. 323–344, 1987.

[46] K.Konolige L.Iocci. A multiresolution stereo vision system for mobile robots. In *Proc. of the AIIA98 Workshop on New Trend in Robotics Research*, 1998.

[47] デビッドマー著, 乾敏郎, 安藤広志訳. ビジョン-視覚の計算理論と脳内表現-. 産業図書, 1990.

[48] 岡 隆一, 西村 拓一, 矢部 博明. ジェスチャ動作の動画像からのスポッティング認識について. 情報処理学会論文誌, Vol. 43, No. SIG4, pp. 54–68, 2002.

[49] 黒川隆生. ノンバーバルインターフェイス. オーム社, 1994.

[50] Kinesics R.L.Birdwhistell and Context. *Essays on Body Motion*. Communication. Univ. of Pennsylvania Press, 1970.

[51] 松尾太加志. コミュニケーションの心理学. ナカニシヤ出版, 1999.

[52] Kohtaro Ohba, Takehito Tsukada, Tetsuo Kotoku, Kazuo Tanie. “Facial Expression Space for Smooth Tele-Communications”. In *Proc. of Third IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 378–383, 1998.

[53] Marc Fabri, David J. Moore, Dave J. Hobbs. The emotional avatar: Non-verbal communication between inhabitants of collaborative virtual environments. In *Proc. of Gesture Workshop*, pp. 269–273, 1999.

[54] Susumu Tachi. “Telexistence and R-Cubed”. *Industrial Robot*, Vol. 26, No. 3, pp. 188–193, 1999.

[55] 葛岡 英明, 上坂 純一, 山崎 敬一, 山崎 晶子. “コミュニケーションメディアとしてのロボットの開発 - コミュニケーションにおける予期の支援 -”. 計測自動制御学会システムインテグレーション部門2003予稿集, pp. 886–887, 2003.

[56] 内田 誠一，森 明慧，倉爪 亮，谷口 倫一郎，長谷川 勉，迫江 博昭. “動作の早期認識およびその予測への応用に関する検討”. 電子情報通信学会技術研究報告, 第 104 巻, pp. 7–12, 2004.

[57] Egor Elagin, Johannes Steffens, Hartmut Neven. Automatic pose estimation system for human faces based on bunch graph matching technology. In *Proc. of Third IEEE International Conference on Automatic Face and Gesture Recognition(FG'98)*, pp. 136–141, 1998.

[58] P.Ekman，W.V.Friesen. *Facial Action Cording System*. Consulting Psychologist Press, 1978.

[59] P.Ekman，W.V.Friesen，工藤 力. 表情分析入門. 誠信書房, 1987.

[60] Yoshio Matsumoto, Alexander Zelinsky. An algorithm for real-time stereo vision implementation of head pose and gaze direction measurement. In *Proc. of Fourth IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 499–505, 2000.

[61] Hideki Kozima. Infanoid: A babybot that explores the social environment. In K. Dautenhahn, A. H. Bond, L. Canamero, B. Edmonds, editor, *Socially Intelligent Agents: Creating Relationships with Computers and Robots*, pp. 157–164. Kluwer Academic Publishers, 2002.

[62] 山岡 史享, 神田 崇行, 石黒 浩, 萩田 紀博. 遠隔操作型コミュニケーションロボットとのインタラクションにおける印象評価, 2007.

[63] 坂本 大介, 神田 崇行, 小野 哲雄, 石黒 浩，萩田 紀博. 遠隔存在感メディア としてのアンドロイド・ロボットの可能性, 2007.

[64] 森田 友幸，間瀬 健二，平野 靖，梶田 将司，岡留 剛. ヒューマノイドロボットを用いた遠隔コミュニケーションにおける注目伝達, 2007.

[65] 森永 英文，松本 吉央，小笠原 司. 表情伝達を目的としたアバターチャットシステム. 情報処理学会第 66 回全国大会講演論文集, 第 4 巻, pp. 239–240, 2004.

[66] A.Nakazawa S. Nakaoka and K.Ikeuchi. An efficient method for composing whole body motions of a humanoid robot. In *Proc. of the Tenth International Conference on VIRTUAL SYSTEMS and MULTIMEDIA (VSMM2004)*, pp. 1142–1151, November 2004.

[67] R. Nisimura, A. Lee, M. Yamada, and K. Shikano. Operating a public spoken guidance system in real environment. In *Proc. INTERSPEECH2005*, 2005.

[68] A. Lee, K. Nakamura, R. Nisimura, S. Hiroshi, and K. Shikano. Noise robust real world spoken dialogue system using gmm based rejection of unintended inputs. In *Proc. INTERSPEECH2004*, Vol. 1, pp. 173–176, 2004.