

観客を話に引き込むミュージアムガイドロボット - 言葉と身体的行動の連動 -

Museum Guide Robot Attracting Visitors in a Talk: Synchronously Coordination in Verbal and Bodily Behaviors

岡田真依^{1*} 星洋輔¹ 山崎敬一¹ 山崎晶子² 久野義徳¹ 小林貴訓¹

Mai Okada¹, Yosuke Hoshi¹, Keiichi Yamazaki¹,
Akiko Yamazaki², Yoshinori Kuno¹ and Yoshinori Kobayashi¹

¹ 埼玉大学

¹ Saitama University

² 東京工科大学

² Tokyo University of Technology

Abstract: In this paper, we propose a novel museum guide robot engaging visitors in its talk. We focused on the human guide's behavior specifically when s/he asks questions to engage visitors. Then we analyzed how the human guide organizes verbal and bodily behaviors from an ethnographical point of view. Based on the result of this analysis, a guide robot have been developed that can coordinate its utterances and body movement like a human guide. In our experiments that conducted in a real art museum, the effectiveness of synchronously coordinated verbal and bodily behaviors has been confirmed. In addition, we investigate a few cases of our experiments that the visitor lost their interest in the robot's talk and we draw on a strategy of the human guide gazing toward visitors in a multiparty setting.

1 はじめに

ロボットをミュージアムガイドとして用いる研究は様々なグループで行われている[10].

こうしたミュージアムガイドの研究の焦点は、単に安全にナビゲーションを行うことから、どのように観客の注意を引き付けて説明を行うか、さらにどのように観客と相互行為を行うかに関心が移ってきている。Sidner らは、ロボットが話の途中で観客を振り返ることが、観客のうなずき等のエンゲージメントを増やすということを報告している[11]。また、Kuzuoka らは、ロボットが説明の途中で一時停止や再スタートをおくことで、観客の関心を引き付けることができることを示している[5]。

人間と人間のコミュニケーション場面だけでなく、人間とロボットのコミュニケーションにおいても言語的行為と非言語的行為の協調が重要であることが明らかになりつつある[2]。それゆえ、いくつかの研究は、説明を行う際の人間の言語行為と非言語的行

為の関係を調べることで、人間に対して説明を行うロボットの開発を試みている。例えば、Mutlu らは、物語を語るロボットに、人間が行うのと同様な方法で、文法的に適切な場で聞き手に視線を向けるようにした[8]。

我々の研究でも、会話分析の移行関連場という考え方を用いて、ロボットが説明している最中に単に観客に振り返るだけでなく、文の切れ目などでの発話の移行が適切となる場所(transition relevance place TRP)で観客に振り返ることが、観客の適切な反応を促すことを示してきた[4][12]。しかし、これまでの研究では実験室的環境で有効性が示されるにとどまっていた。

本稿では、実験室的環境ではなく、実際の美術館に実験の場所を変えた。また、科学系のミュージアムではなく、美術系である大原美術館(倉敷)で実験を行うことにより、ロボットを見に来たわけではない観客にロボットの話を最後まで聞かせるという新たな視点を持ち込んだ。

そして、美術館で観客に説明を行うガイドロボットの分析と評価において、次の2つの点に着目した。

1. 自発的に参加した観客が、ロボットの説明

* 連絡先: 埼玉大学文化科学研究科
〒338-8570 埼玉県さいたま市桜区下大久保 255
E-mail: moma_asian_beat@yahoo.co.jp

(120秒)を最初から最後まで聞くかどうか。

2. 観客がロボットの説明に対して、積極的にかつ相互対話的に反応するかどうか。

以上のことを明らかにするため、最初に美術館において学芸員による絵の説明がどのようになされているかを調べた。

2 学芸員の解説行動の分析

我々の展示における解説者の行動の分析(1人対1人)によれば[4][12], 解説者が展示物から観客に視線を移すのは、キーワード、直示、質問および文の切れ目のような TRP ポイントであった。そこで、TRP ポイントを特定しそのタイミングで観客を振り向くロボット(Systematic mode)とランダムな場所で振り向くロボット(Unsystematic mode)への観客の反応を比較した。その結果、観客が Systematic mode ではロボットへ適切なタイミングでうなずきや視線を向けることが分かった。

今回、さらに広域的な視野の分析をするために、大原美術館で美術の専門家である学芸員が絵について解説を行う場面(1人対1人)を撮影し、分析した。ここで焦点を、学芸員がどのように絵を言葉と身体の動きによって解説を行うかということに当てたところ、観客への振り返り、キーワード直時の際の絵を指差しながらの観客の観察、文の切れ目での観客の反応の確認動作が観察された。

こうした美術館での解説行動の分析から、ロボットに次の機能を実装することにした。

1. ロボットは、絵を説明する際に、平叙文に加えて、絵を参照する質問をする。またロボットが答えを述べるようにする。
2. 従来の研究と同様に、ロボットは発話の切れ目で視線を絵から観客に向ける。
3. TRP の中で、質問の終わりの部分で視線を絵から観客に向ける。
4. TRP の中でも、ロボットが正解を言った後に、ロボットは絵から観客に視線を向ける。
5. キーワードを言う時に、ロボットは絵を指差しながら観客に視線を向ける。

ロボットに絵を参照させる質問をさせたのは、絵についての前提的な知識を問うような質問だと、間違った答えが観客から返された際に、ロボットがうまく対応できないためである。実際に、人間の観客を相手にした場合でも、観客の答えが想定と異なっていた場合には解説者が戸惑う場面があった。絵を参照する質問の場合には、観客からどのような答えが返された場合でも、ロボットは絵を見ながら正解を言うことができる。

今回の実験では TRP や質問、答えの際の観客の反

応を調べることを重視し、ロボットは複数の観客がいる場合でも、常に一人の対象者を選び、その人の方向に視線を向けることにした。

3 ガイドロボットシステム

本研究では、貴重な作品を展示する美術館という環境を考慮し、レーザー光を照射しない純粋なパッシブセンサであるカメラからの画像によって周囲の状況を観察し、適応的に動作するロボットを開発した。

作品の解説を行うロボットは、聞き手である観客を検出し、その聞き手に対して魅力的な説明を行わなくてはならない。そのため、カメラで観察された領域内に人物が出現した場合に、それを速やかに検出し、「説明いたしましょうか？」などの発話を行い、観客の興味を引き付ける必要がある。また、前節で述べたように、本ロボットは聞き手である観客に対して視線を投げかけたり、作品に視線を移したりすることで、より効果的な説明を試みる。そのため、カメラ画像から自己位置や聞き手の頭部位置・向きなどを計測し、得られた情報に基づいて適切なロボットの制御を行う。

3.1 ロボットシステムの概要

本システムの概要を図1に示す。ロボットは、コミュニケーション研究用ロボット開発プラットフォーム Robovie-R ver.2 を用いた。このロボットは2台の汎用 PC(Windows XP)を搭載し、シリアル通信により制御可能な関節(合計 29 自由度)を持つ人型のロボットである。また、ロボットの身長は約 120cm であり、下部に格納された2個の車輪を制御することで移動することができる。頭部は3軸(Yaw, Roll, Pitch)に対応する各関節の制御が可能であり、人間に近い頭部動作ができる。また、腕の稼働範囲も広く、指先の曲げ伸ばしが可能なことから、指差しなどのジェスチャーを行うことができる。さらに、ロボットの頭部の動きに影響されずに周囲の視覚情報を獲得するため、ロボットの背中に設置したボール上部に3台の USB カメラ(Logicool QCAM200V)を図1に示したように設置した。それぞれのカメラ映像は、解像度 320×240pixel, 30fps で取得できる。2台の内蔵 PC は有線ネットワークで接続され、同時に無線ネットワークを介して、内蔵 PC の操作画面を外部 PC で、閲覧、操作することができる。

本システムは相互に通信を行う2つの処理モジュール(人物検出・追跡モジュール、ロボット制御モジュール)で構成され、それぞれのモジュールは独立したプログラムとしてロボットに内蔵された2台の PC でそれぞれ実行される。人物検出・追跡モジュールが動作する PC には、正面向きの USB カメラ1台

を接続し、人物の検出、及び頭部の位置と向きを追跡を行う。検出・追跡の結果は、ロボット制御モジュールに送信され、必要に応じて利用される。モジュール間の通信には Socket を用い、非同期でデータの送受信を行っている。ロボット制御モジュールでは、シリアル通信によりロボットの各関節のモータのステップ値を設定することで、ロボットの動きを制御する。また、ロボット制御モジュールが動作する PC には、左または右向きの USB カメラ 1 台を接続し、側方の画像を取得し、移動時の自己位置の同定に用いている。

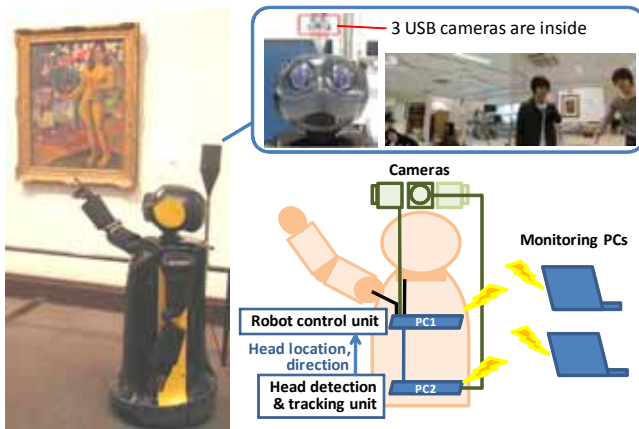


図1. ガイドロボットシステムの概要。

3.2 センシング手法とロボットの動作

カメラ画像に基づいたセンシングについて述べる。我々は画像処理ライブラリOpenCVに付属する顔検出器を用いて観客の頭部を観察した。顔検出後の頭部の追跡にはパーティクルフィルタを用いた[7]。頭部領域中の肌色画素と黒色画素を水平軸上に射影することで人物のおおよその顔向きを取得する。このシステムはロボットから3m以内の顔位置を8方向に判定することができる。ロボットの自己位置の計測には説明対象の絵作品の四隅をテンプレートマッチングにより検出し、画像中での四隅の位置関係からおおよその自己位置を推定する。

次に、作品説明時のロボットの動作について述べる。本システムは、説明内容を女性の声、男性の声、合成音声で記録した音声データを持っており、設定により音声を変更して、説明することができる。

説明の途中で聞き手を振り向く動作や、作品を指差す動作などは、そのタイミングを含めて事前に記録しており、ロボットは記録された動作を指定されたタイミングで再生する。ただし、ロボットが聞き手を振り向く方向は、追跡された聞き手の頭部の方向に応じて変更し、身長や立ち位置の異なる聞き手に対しても正しく聞き手の頭部の方向へ振り向くことができる。具体的には、聞き手頭部の画像中での

位置と大きさに基づいてどの方向へ振り向けば良いかを事前に記録しておき、説明時には、この情報に基づいて振り向く方向を適応的に切り替える。これにより、例えば、聞き手として人物が選ばれるとロボットは頭部を聞き手の顔の方向へ向け、「説明いたしましょうか？」などと発話し、観客の興味を引き付ける動作を行うことができる。また、説明をしている間も、ロボットは常に聞き手の頭部の向きを観察し、聞き手が作品やロボットの方向を向いていない場合は、振る舞いを適応的に変化させることもできる。例えばロボットが絵の一部を指差し、「この女性がイヴですね。」と説明した時点で聞き手が絵の方向を向いていない場合には、「この女性ですよ。」と説明を繰り返すこともできる。

4 大原美術館での実験

これまでに述べた特徴を持つロボットを実際の美術館に設置し、一般の観客に向けた大規模な実験を行った。実験は、大原美術館にて2007年11月17日に実施し、ロボットが説明を行う作品は、「かぐわしき大地」(ゴッガン作)とした。

ロボットは待機状態では説明対象である絵の前で身体を左右に向ける動作を連続して行い、観客に対して自身が対応可能であることを示している。

観客が近づいてきて顔が検出され、聞き手として選ばれると、ロボットは「説明いたしましょうか？」などと発話し、観客の興味を引き付ける。この発話の後、追跡されている人物が絵の方向、又はカメラの方向を向いている場合には、あらかじめ準備された説明を行う。説明の終了後は、再度待機状態に戻り、次の観客が現れるのを待つ。

なお、条件を一定とするため、顔向きによる説明内容の適応的な変更はせず、一旦説明を開始した後は、観客が途中で立ち去っても説明を打ち切らずに最後まで説明を行うこととした。

4.1 ロボットガイド実験の分析

4.1.1 観客がロボットの話を聞いていたか

こうしたロボットの説明は95回行われた。また1回の説明における観客の最大数の平均は4.8人であり、500人近い観客がこのロボットのデモを見た。この説明のうち、実験スタッフによる調整、実験スタッフが観客に加わったデモ、マスメディアの取材が多数あった。それらを除いて、観客が自発的に参加した有効な実験回数は63回であった。ここでは、観客が自発的に参加した63回のデータを分析する。

この63回の内容を図2に示す。最初にロボットが話しかけた観客がロボットの話を最後まで聞いてい

た場合(本稿では「FULL」と呼ぶ)は、38 回あった。また、前回の説明に途中から加わった観客に次回でもロボットが話しかけ、2 回合わせて観客がロボットの話を最後まで聞いた場合(本稿では「併せて FULL」と呼ぶ)が 6 回あった。さらに、FULL の観客が次回まで居残ってしまい前回と同一の観客を認識してしまった場合や、ロボットが通りすがりの人を認識した場合が 6 回、別の場所にいた観客に呼びかけられてその場を離れた場合が 4 回あった。他の要因なしに途中退出したのは 9 回だった。

このような外的要因や、ロボットの認識失敗を除いた 53 回の場合で、FULL で聞いたのは 71.7%，併せて FULL を加えると 83% の観客(ロボットが最初に認識した観客)がロボットの 120 秒の説明を最後まで自発的に聞いたことになる。以上より、ロボットを見に来たわけではない美術館の観客にロボットの話を最後まで聞かせるという目標は十分達成されたと思われる。

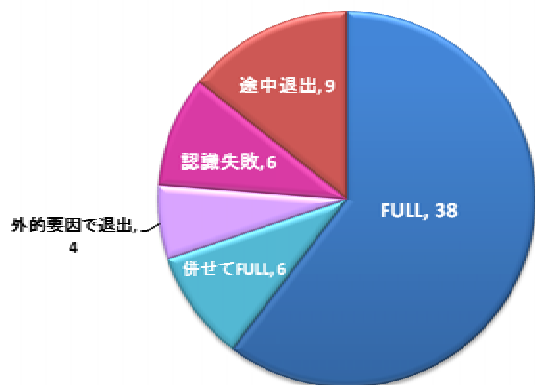


図 2. 全デモ数 63 回中の観客の内容.

4.1.2 観客の全体の反応

次に、TRP でのロボットの視線の動きに対する観客の反応を分析する。観客がロボットの方を見る場合(R)、うなづく場合(N)、ほほえんだり、絵に視線をさらに向けたり、絵を指し示す場合(P)、何らかの発話を行う場合(U)のそれぞれに分けて、観客の反応は分析した。ここでは、絵をそのまま見る等の観客の動きや発話がない場合はカウントされない。

表 1 は観客が最初から最後まで見た 38 場面において、TRP で観客が何らかの反応をした割合を示したものである。ここで、R+N+P+U は、観客がこのどれかの反応をしたことを示している。この表は 50% 以上の確率で、観客がロボットの視線の動きに何らかの反応をしたことを示している。

表 1. TRP における観客の反応.

	R	N	P	U	R+N+P+U
平均	30.3%	21.0%	11.2%	13.1%	53.1%

図 3 は、FULL における観客が TRP において何らかの反応をした回数である。この図から、観客の反応には大きなばらつきがあることが分かる。実際 1 回から 2 回しか反応しない場合が 7 場面あった。逆に、11 回の TRP のうち、半数の 6 回以上の反応を示す場合が 20 場合と半分以上を占めていた。

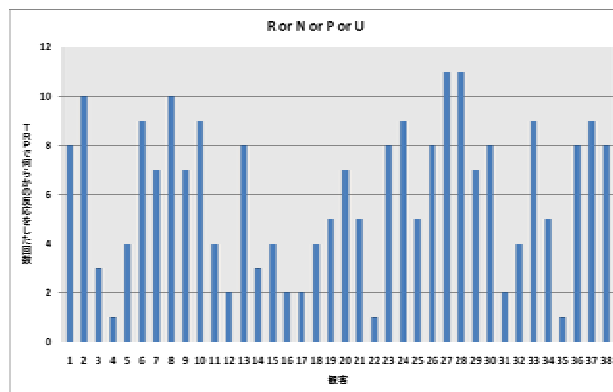


図 3. 観客の TRP における反応回数.

さらに、ロボットに対して、人がほとんど反応しない場合を見てみた。その内訳は次のものである。ロボットが正確な相手の場所を見ていない (No31)、観客が動くことによって観客の視界からロボットがはずれる (No17)、最初の観客とロボットの間に別の観客入り込み、その人がロボットに反応する (No4, No35)、観客とロボットの距離がある (No16)、絵をじっと見ていた (No12, No.22)。

38 場面全体の構成は、成人女性 22 人、成人男性 12 人、男児 2 人、女児 2 人だった。2 回以下の場面における属性別は、成人男性 3 場面、成人女性 4 場合だった。逆に 6 回以上反応する場面では、成人女性 13 回、成人男性 4 回、女児 1 回、男児 2 回だった。このように全体的には、女性やこどものほうがロボットに反応する傾向が見られた。なおロボットの声による差はほとんどなかった。

5 考察

実際の美術館で学芸員の行動を分析し、それに基づいてロボットの行動を設計した。本稿では、以上のようにして設計されたロボットの話を観客は最初から最後まで聞いていたか、観客がロボットの説明に対して積極的かつ相互対話的に反応するかということについて分析した。結果として、人間はロボットに対して学芸員に対する反応と同じように相互対話的な行動をとった。これは、ロボットの言語的・非言語的行動に対する観客の反応から、ロボットに肯定的な反応が示され、ミュージアムガイドロボットとしての有効性が確かめられたといえよう。

しかし、本システムには次のような問題点が明らかになった。

5.1 ロボットシステムの改善点

図2に示したように実験回数63回中6回の認識失敗, 9回の途中退出があった。この失敗例を分析すると2つの改善すべき点が挙げられる。

1点目は, 前回に説明していた人が説明終了後に居残り, 一度説明が終了した人に対して次回も説明を開始してしまう点である。これが, 観客の途中退出の原因である。これは, ロボットが誰か人を発見したら説明を開始することとしていたためであり, 誰に対して説明をしていたかということは検討されていなかったからである。一般に, ガイドは同じ人に対して同じ説明をすることはない。この場合を改善するため, 追跡している人が判別される必要がある。具体的には, 説明対象として追跡している人の服装などの特徴を覚えておき, その観客が居残った場合には説明を開始しない。または説明内容を変化させたりするなどの工夫をすることで, このような状況においても観客を飽きさせないガイドが可能になると考えている。同様に, 現在追跡している人は通りすがりの人であるかどうかを判別する必要もあると考えられる。

2点目は, 説明の途中で追跡の対象者が変化してしまい, 説明開始時に注目していた人と違う人に注視の方向も変化し, 途中でロボットがTRPごとに向ける視線の方向も変化してしまった点である。これが, 途中からロボットに注視されなくなった観客の興味が薄れてしまうことにつながっていると思われる。ロボットシステムの次段階としては, 複数人の追跡を実現し, 1人対N人のコミュニケーション場面においても円滑なガイドをするロボットシステムへの拡張が求められる。この際, 観客のうなずきや首振りなど検出することで, 質問する人を選択したり, 視線配布したりするといった観客の興味を引き付けるガイドが目標とされる。そこで, 複数人ガイドシステムを導入する前に学芸員と複数人の観客間の相互行為を分析することが重要となる。

5.2 学芸員の複数人に対する解説行動の分析

本節では, 大原美術館における解説者と複数人の観客間の相互行為に焦点化し, 分析を行う。

図4は, ピカソの「頭蓋骨のある静物」を解説している場面である。解説者と観客が1人対1人の場合と異なり観客が複数人いることで解説者の視線配布の対象者が増えた。それにより解説者はすべての観客に視線配布を行っている。この視線配布には3つの機能があると推測される。1点目は, 観客の反

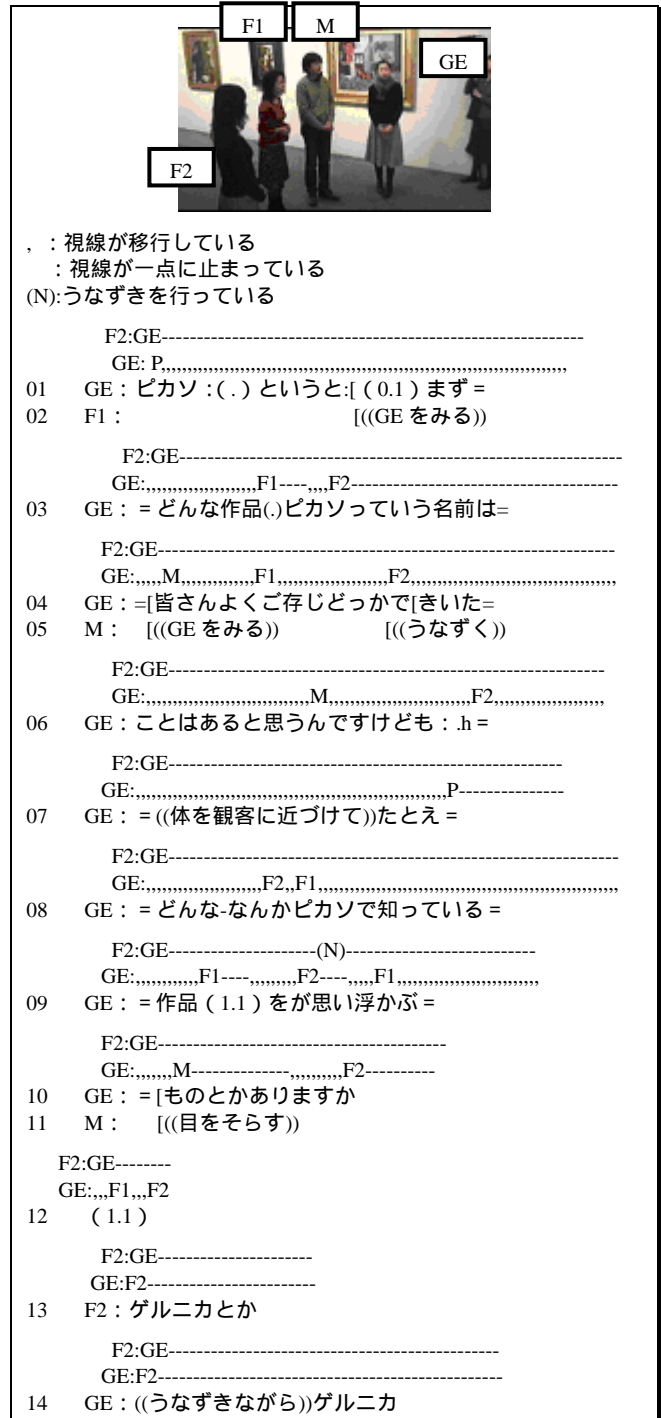


図4. 解説者と観客の相互行為場面.

応を確かめる機能である。これは Kendon が「モニター機能(monitoring function)」として指摘しているものである[3]。つまり, 観客は解説者が投げかけた質問の内容を理解しているかどうか, また, 観客の注意を引くことができているかなどの確認作業である。2点目は, 観察をしていることを観客に示すことで答えや反応を促す機能である。そして, 3点目は Lerner や Sacks らが指摘しているように, 次の話

し手である応え手を選ぶという機能である[6][9]。解説者はすべての観客に対し均等に視線配布を行うことで、均等に応答の機会を与える環境を作っているといえる。

これまで、解説者の視線配布にのみ焦点を絞り言及してきたが、相互行為とは、参加者が相互に「話し手」と「受け手」という役割を継的に交換しながら織りなすものである。従って、受け手となる観客の行為も詳細にみていく必要がある。ここで、Heathの受け手性の視点に基づいて、観客の行為を分析する[1]。すると、最も解説者に対して志向の明示化を積極的に行った観客(F2)が、最終的に応答の機会を得ていることが分かる。F2は解説者が質問のシークエンスを話し始めた時点から、ずっと解説者に対し視線を送り続けている。一方、図4の11行目のMは、わざと視線を外すこと、つまり意図的な受け手性の非提示により、応答の回避を行っている。解説者はMからの受け手性の提示を得られなかったために、他の観客に視線を配ることを余儀なくされ、最終的に視線を獲得できたF2からの応答を得ることができたのである。また、10行目から14行目にかけての解説者の視線配布をみると、F2に対する視線配布を頻繁に行っているため、解説者はF2からの強い受け手性の提示に気付いていたものと思われる。つまり、ここから、受け手性の提示度合を的確に読み取るということが、円滑な相互行為を進める上で一つの重要な要素となっていると考えられる。

これらのことから、複数人に向けてガイドを行うロボットには、観客に対して均等に視線配布を行い、ロボットへの観客の積極的反応を見分けることが望まれることが分かった。

6 おわりに

本稿では、美術館での学芸員の行動を観察することで、ロボットにその言語と身体の協調を付与することを行った。これにより、人間と円滑な相互行為をするロボットの設計をすることができ、美術館で観客の注意を十分に引き付ける解説も行うことができた。ロボットからの知識提供に理解を示す観客の反応さえ見られた。また本稿では、実験からの改善点を踏まえ、複数人観客に対する有効な解説方法を検討し、視線配布と観客の積極的反応の判別の重要性について明らかにした。

こうしたことがまた、会話分析の成果をロボット研究に取り入れることの重要な意義の一つであると思われる。

謝辞

本研究の一部は総務省・SCOPE-Sの助成による。

参考文献

- [1] Heath, C. Talk and reciprocity: sequential organization in speech and body movement, in J. Maxwell Atkinson & John Heritage eds., *Structures of Social Action*, Cambridge University Press, (1984)
- [2] 神田崇行, 石黒浩, 今井倫太, 小野哲雄. 人-ロボット相互作用における身体動作の数値解析. 情報処理学会論文誌, Vol. 44, No. 11, (2003), 2699-2709.
- [3] Kendon, A. *Conducting interaction: patterns of behavior in focused encounter*, Studies in International Sociolinguistics 7, Cambridge University Press, (1990), 210-213
- [4] Kuno, Y., Sadazuka, K., Kawashima, M., Yamazaki, K., Yamazaki, A. and Kuzuoka, H. Museum guide robot based on sociological interaction analysis. In *Proc. CHI 2007*, ACM press (2007), 1191-1194.
- [5] Kuzuoka, H., Pitsch, K., Suzuki, Y., Kawaguchi, I., Yamazaki, K., Kuno, K., Yamazaki, A., Heath, C. and Luff, P. Effect of restarts and pauses on achieving a state of mutual gaze between a human and a robot. In *Proc. CSCW 2008*, (2008).
- [6] Lerner, G.H. Selecting next speaker: the context-sensitive operation of a context-free organization. *Language in Society*, 32(2), (2003), 177-201.
- [7] Matsumoto, Y., Kato, T. and Wada, T. An occlusion robust likelihood integration method for multi-camera people head tracking. In *Proc. INSS 2007*, (2007), 235-242.
- [8] Mutlu, B., Hodgins, J.K. and Forlizzi, J. A storytelling robot: modeling and evaluation of human-like gaze behavior. In *Proc. HUMANOIDS 2006*, (2006), 518-523.
- [9] Sacks, H., Schegloff, E. A., and Jefferson, G. A simplest systematics for the organization of turn-taking for conversation, in Jim Schenkein ed., *Studies in the Organization of Conversational Interaction*, (1974), 25-28,54
- [10] Shiomi, M., Kanda, T., Koizumi, S., Ishiguro, H. and Hagita, N. Group attention control for communication robots with wizard of OZ approach. In *Proc. HRI 2007*, (2007), 121-128.
- [11] Sidner, C.L., Lee, C., Kidd, C.D., and Rich, C. Explorations in engagement for humans and robots. *Artificial Intelligence*, 166 (2005), 140-164.
- [12] Yamazaki, A., Yamazaki, K., Kuno, Y., Burdelski, M., Kawashima, M. and Kuzuoka, H. Precision timing in human-robot interaction: coordination of head movement and utterance. In *Proc. CHI 2008*, ACM press (2008), 131-140.