ロボット漫才

—社会的受動メディアとしての二体のロボットの利用—

林 宏太郎^{*1*2} 神田 崇行^{*1} 宮下 敬宏^{*1}
石黒 浩^{*1*2} 萩田 紀博^{*1}

Robot Manzai

—Robots' Conversation as Passive Social Medium—

Koutarou Hayashi^{*1*2}, Takayuki Kanda^{*1}, Takahiro Miyashita^{*1},
Hiroshi Ishiguro^{*1*2} and Norihiro Hagita^{*1}

This paper reports on the development of a multi-robot cooperation system for human-robot communication. In the system, robots behave as if they are communicating by speech, while the system exchanges information through a network for keeping the timing of the communication natural. The exchanged information through the network is based on analysis of inter-human conversation. This system is based on a scripting language for coordinating multi-robot communication, which has a merit for developer of easy development. The developed system is used for Manzai, which is a Japanese comedy conversation mainly performed by two people. While the tempo and timing are particularly important in Manzai, the developed “robot Manzai” system was more highly evaluated than the Manzai shown in a video performed by humans. We believe that this system demonstrates the potentials of robots as a passive-social medium; like television and computers.

Key Words: Passive-social Medium, Robot-robot Communication, Robot-human Interaction

1. はじめに

近年、ロボット工学の進歩により、人間に近い身体表現ができるヒューマノイドロボットが開発されてきた[1]。我々は、ヒューマノイドロボットは人間に似た外見を利用して、人間との自然なコミュニケーションができるようになると思える。例えば、ヒューマノイドロボットが人間に似たゼスチャを行うと、人間はそのゼスチャの意味を直感的に理解できるだろう。これによりヒューマノイドロボットを、道案内などの人間とのコミュニケーションを中心とする役割に用いることが近い将来可能になると考えられる。

最近の HCI (human-computer interaction) 分野の研究からも情報を媒介するインタフェースとしてのロボットの有用性が明らかになりつつある。Nass らは、コンピュータが従来のテレビやラジオのように情報を媒介する新しいメディアであることを示した一連の研究の中で、テキストのみのシンプルなインタフェースであっても、人は他の人に対して振る舞うときと同様

にコンピュータに対しても対人的に振る舞うことを明らかにした[2]。さらに、身体性エージェントに関する Cassell らの一連の研究は、頭部や腕といった擬人的な表現を用いる擬人化エージェントはコンピュータエージェントと人との間の効率的な情報伝達を実現することを明らかにした[3]。そして、実ロボットとコンピュータ上のエージェントを比較した研究において、コンピュータ上のエージェントよりも実空間に存在するロボットのほうが実空間上の物体に関する対話に適していることが見出されている[4]。

これらの研究は、ヒューマノイドロボットが新しいメディアとして活用される一面を持つことを示唆している。この活用形態の一つとして、複数のロボットが会話をしてみせるという形のコミュニケーションに我々は注目した。これまでに神田らは、複数のロボットが会話をしてみせることで、この会話を観察している人が後にロボットと会話する際にロボットからの説明の理解が容易になり、またロボットに自然に振る舞うようになることを示した[5]。

このように複数のロボットによる会話を観察する場合、観察している人間はロボットと一対一で対話する場合と異なり、必ずしもロボットの会話に口をはさむ必要はない。この観察している人間は会話を行っているロボットに対して観察者の立場にあり、言語学者の Clark の言葉[6]を借りると、「観察者は傍観

原稿受付 2005 年 5 月 9 日

^{*1}ATR 知能ロボティクス研究所

^{*2}大阪大学

^{*1}ATR Intelligent Robotics and Communication Labs

^{*2}Osaka University

する立場であり、会話を行う責任は生じない」。つまり、既存メディアであるテレビの中の会話を見ている場合と同様に、複数ロボットの会話を観察する人間は、例えば挨拶や質問といったロボット同士の会話についても返答を行う必要はない。我々は、このような情報伝達の形態を社会的受動 (passive-social) メディアと名付けた。以下に社会的受動メディアの特徴について説明する。

Fig. 1 に、この社会的受動メディアと、対比として、従来からある他の二つのメディアの形態を示す。Fig. 1 (a) は、テレビにおけるニュースキャスターのように、一体のロボットがただ一方的に情報を提供するものであり、ここでは受動 (passive) メディアと呼ぶ。Fig. 1 (b) は、利用者から要望を受けその要望を満たす情報を提供するメディアであり、インタラクティブ (interactive) メディアと呼ぶ。Fig. 1 (b) の形式は、現在多くの研究者がロボットによる実現のために研究を続けており、展示会場での展示などが行われるようになってきている。しかし、主にセンシングの難しさが原因で、言葉を交わして相手の意に沿った情報提供を行うなどの、人間同士が行うような自然な対話レベルでの実用に耐えうるものはまだ実現に遠い。

本研究において我々が注目するのは、ロボット同士が会話してみせることにより観察者に情報を伝達する Fig. 1 (c) の形式、つまり社会的受動 (passive-social) メディアである。受動メディアや社会的受動メディアはセンシングをほぼ行わず、利用者からの要望に反応することはない。我々は、「社会的受動メディアは受動メディアよりも自然で分かりやすい情報提供を行える」との仮説を持つ。例としてあげるなら、テレビのニュース番組において2人のアナウンサーがニュースについてコメントしあうといった状況がこれに当てはまる。ただし、本稿ではこの仮説は仮説のままにとどめ、むしろより基本的な問題である「社会的受動メディアとして、ロボットがテレビのような従来メディアに比べて効果を持つかどうか」の検証をまずは行うことを目的とする。

本稿では、この社会的受動メディアをロボットにより実現するロボット会話システムの開発と、社会的受動メディアとしてのロボットの、メディアとしての可能性の検証結果について報告する。具体的には、社会的受動メディアをロボットに実装するに当たって、ロボット同士の会話のタイミングを調節する機構が必要となる。複数の小型ヒューマノイドロボット SDR-3X (Sony 製) のダンス等、複数ロボットの協調動作にはいくつかの例が存在するものの、それらはロボットの内部クロックを連動させることにより同期させているものと思われる。これは、あ

らかじめ想定されている状況下において厳密に調整する必要があるが、一度環境が変われば、再調整には多くの手間と時間が必要となる。また、仮に外部刺激に対応する場合、ロボットとセンサ同士がネットワークを介してタイミング情報をやり取りすることになり、違う制御方式を併用するという複雑な状態になると考えられる。むしろ、我々は制御を、タイミング情報をやり取りするという方式に統一し、ロボット同士があたかもタイミングをとりあって会話するかのように、背後のネットワーク経由でタイミング情報を交換する方式により、複数ロボット協調対話システムを実現する。これにより、開発者が微妙なタイミングの調整や発話・ゼスチャの変更を容易に行えるようにするとともに、観察者の反応に応じてロボットが行動を変化させるといった外部刺激に対するシナリオ分岐メカニズムを実現する。

なお、ロボット同士のタイミング制御に関して、ローカルネットワークにより結合されたロボットにおいては通信遅延時間は十分小さく、ネットワーク経由のタイミング制御自体は難しいが、むしろ実際の会話に近い自然なタイミングを実現するための手法が重要となる。人-ロボット対話におけるタイミングに関連して、これまでに、エージェントの返答するタイミング [7]、挨拶における返答遅延 [8] や道案内のゼスチャに関する遅延 [9] といった自律的な応答タイミングに関する研究が行われてきたが、本研究では漫才などの高度なタイミング調整が必要で、かつ事前に開発者による調整が可能である題材を対象とし、自律的なタイミング調整よりもむしろ開発者による調整のしやすさを重視する。つまり、開発者が多様に編集できる環境が重要となる。

開発した複数ロボット協調対話システムは、社会的受動メディアとしてのロボットの有効性を検証するために実現するロボット漫才のアプリケーションとして用いる。有効性の検証において、我々はまず、社会的受動メディアとしてのロボットの利用が、既存のメディアに代わるものとしての能力を有するかどうかに着目した。その上で、この漫才アプリケーションを題材に、複数ロボットによるメディアと、漫才を流通させてきた既存メディアであるテレビとを比較する実験を行った。実験結果は、社会的受動メディアというロボットを用いた新たなメディア形態への利用方法の有効性を示すとともに、現時点で漫才を題材としたエンターテインメントへの実用性を示すものである。

2. 人間同士の協調対話の分析

2.1 協調対話の具体的な形式の選択

我々は協調対話の具体的な形式として、漫才を選択した。漫才では相方との適切なタイミングでの発話・動作 (例えば、つつこみ) が必要になるため、本システムでロボットに漫才をさせることが成功したとすれば、多種の協調対話への応用が期待できる。

漫才とは、2人ないし複数人の人間が、日常の出来事や世間の流行ごとなどを題材に、滑稽な問答のやり取りをするというものである。基本的に、「まけ」と言われる役がおかしなことを発言し、「つつこみ」と言われる役がそれをやや誇張気味に指摘するという舞台喜劇である (Table 1)。漫才では、「タイミン

Fig. 1 Robot(s) as medium

グの“良い”つつこみ」などと言われるように、通信遅延の短さよりもむしろ適切なきっかけによる自然なタイミングでの発話が観客の笑いにつながるため、協調対話の中でもタイミングの自然さが重要となる。

2.2 漫才の分析

ロボット漫才の実現に当たって、プロの漫才師の漫才を分析し、協調対話に必要な要素を調査した。芸歴も芸風も異なる4組の漫才師の漫才を対象とした。

漫才の基本となる流れは、1人が話を進行させ、もう1人が話の脱線、曲解、邪魔などを行って滑稽さを出すという会話である。その会話の話者間での発話タイミングの分析を行った。録画された漫才のビデオを筆者が視聴し、話し手が交代するときの交代のきっかけを分類した。結果、発話と発話の間の時間こそ芸人ごとに差があるものの、5種類に分類できることが判明した（Table 2）。ここで、Type1は相手の発話の終了後に発話が始まる場合である。このタイプの遷移がどの場合でも最も多かった。次に多く見られたのがType2の、相手の発話に割り込んで発話を開始する場合である。Type3は相手の動作終了後に発話が始まる場合、Type4は両者が同時に発話する場合、Type5は観客の拍手や笑い声がおさまるのを待って発話が始まる場合である。

2.3 システムの必要条件

人間に近いロボット漫才を現実にするには、前節で述べた5種類に分類した発話タイミングを実現できるシステム開発が必要である。本節では、これらを実現するために複数のロボット間で交換されるべき情報の必要条件を検討する。

Type1, Type3については、相手側が発話もしくは動作を終了したことを相手に送信する必要がある。受け取った側はこの相手側からの合図を待って発話を開始すればよい。Type4はどちらかが発話を開始したことを相手に送信する必要がある。受け取った側もこれに応じて直ちに発話を開始する。

やや複雑になるのはType2の場合であるが、この場合必要

Table 1 Example of Boke & Tsukkomi

ばけ:	さあ新郎新婦、従業員入り口より出てまいりました
つつこみ:	どっからきとんねん（どこから来ているのですか）
ばけ:	搬入トラックに乗っての入場でございます
つつこみ:	何でのとんねん（一体何故乗っているのですか）

Table 2 Example of speech timing

Speech Timing	Occurrence			
	Manzai1	Manzai2	Manzai3	Manzai4
Type1	238	561	173	175
Type2	31	29	35	15
Type3	5	1	2	3
Type4	2	21	3	2
Type5	18	17	34	30

(Type1: On partner's speech end, Type2: Barging in the partner's speech, Type3: On partner's end of action, Type4: simultaneously starting with partner's speech, Type5: After audiences finish clapping or laughing)

となるのは、割り込まれる側のロボットが割り込みを期待する部分での合図の発信である。つまり、割り込まれる側はここで発話を一時停止し、相手に合図を送信した後に発話を再開するという方法をとる。具体的には、Table 1の「搬入トラックに乗っての入場でございます」の途中でつつこみを割り込ませようとする場合、ばけの音声ファイルを「搬入トラックに乗っての」と「入場でございます」の二つに分割する。そして「搬入トラックに乗っての」という音声ファイルが終了した段階で相手のロボットに終了情報を送信すればよい。Type5については、漫才のシナリオ上、笑いや拍手など観客の反応が予想される部分において笑い声や、拍手などの音を把握する必要がある。これは、センサで観客の発する音を測定し、観客が音を発し終わるのを待つというシステムを用いて実現した。

3. 複数ロボットによる協調対話システム

2.3節において述べた必要条件を満たし、また発話・動作や話者間のタイミングの調整が容易な複数ロボット対話システムを実現した。このシステムは、複数ロボット間協調対話言語によって書かれたスクリプトと、それを実行することで発話や動作を行う複数のロボット、観客の状況を認識する観客反応センサから構成される（Fig. 2）。以下に、各構成要素について説明する。

3.1 複数ロボット間協調対話の記述言語

協調対話システムの構築に当たり、開発者が簡便に記述できるように独自のシンプルな記述言語を作成した。この記述言語によって書かれたスクリプトは、ロボット内にファイルとして置かれ、それを各ロボットが読み取ることで協調対話を実現する。以下に、この記述言語に含まれる命令を説明する。

• signal (シグナル)

括弧内のシグナルを他のロボットに送る命令文である。

• signalwait (シグナル)

他のロボットから括弧内の指定されたシグナルを受信するまで待機する命令文である。

• motion “動作テキストファイル名”:シグナル

あらかじめ作成した動作テキストファイルをロボットに実行させ、動作開始と同時にシグナルを送り、次の命令にすぐ移行するという命令文である。末尾の“:”以降は省略可能であり、省略した場合シグナルは送らない。

Fig. 2 Outline of multi-robot communication

● motionwait “動作テキストファイル名”:シグナル

あらかじめ作成した動作テキストファイルをロボットに実行させ、動作終了の際にシグナルを送り、次の命令には動作完了の後に移行する命令文である。末尾の“:”以降はこれも省略可能であり、省略した場合はシグナルを送らない。

この、motion、motionwait と同じ記述方式をとる命令文として、speak、speakwait (発話命令)、move、movewait (移動命令) が存在する。

● move(x, y, θ):シグナル

協調対話が始まった地点の初期値を (0,0,0) とし、 x, y で指定された座標に移動。 θ で指定された角度に体の向きを変える。move は動作開始と同時に、movewait は動作終了と同時に、シグナルを送る。末尾の“:”以降は省略可能であり、省略した場合シグナルは送らない。

● check_reaction()

騒音計を利用した観客反応センサに状況確認命令を送り、観客反応センサから返答として送られてくる現在の状況を取得する。シナリオ上で観客の反応を参照したい箇所にきた場合に、この命令を用いて観客反応センサに対して問い合わせ信号を送る。問い合わせ信号を受けた観客反応センサは、現在の観客の状況として騒音量を測定し、そのレベルに応じて設定された信号を送り返す。if 文を用いることで観客の状態 (例えば、「受けた」か「受けなかったか」など) に応じて異なった振る舞いをさせることができる。

● if(reaction=check_reactionの戻り値)()

観客反応センサからの戻り値が括弧内の等式と一致した場合、ロボットは if() の後に続くスクリプトを endif が書かれる部分まで実行する。

● wait(時間 [ms])

括弧内に記述されている時間分だけ待機する。例えば、あるロボットが発話後に少し間を置いて再度発話する場合などに用いられる。1 体のロボットが自ら少し発話や動作に間を置くなどの調整にも使用する。

● exit()

スクリプトの実行を終了し、終了命令を他のロボットおよび観客反応センサにも送る。

3.2 構成要素

3.2.1 ヒューマノイドロボット

Fig. 5 に本研究で用いたヒューマノイドロボット Robovie [10] を示す。4 自由度の腕、3 自由度の頭部、各 2 自由度の眼球部 (カメラ内部の稼動部) により人間とのコミュニケーションに用いる様々なゼスチャが可能である。マイクロフォン、接触センサ、超音波センサ、ステレオカメラ、全方位視覚センサ、といった多様なセンサや、コンピュータやモータ制御ボードといった制御機器をすべて内蔵している。

3.2.2 観客反応センサ

具体的には騒音計を利用しており、騒音計から常時騒音量 (dB) を受け取り、他のロボットから状況確認命令がきた場合、客席の騒音レベルに合わせて BURST_OUT (大爆笑)、LOUGH (普通の笑い)、COOL_DOWN (受けなかった) の 3 種類のシ

グナルを送る。この判定基準 (dB) は設定ファイルを書き換えることにより変更可能である。

3.3 実装例

実装例を示す。Table 3 はこの言語によって書かれたスクリプトであり、Fig. 3 はその進行を図で示したものである (ϵ : 観客反応センサ)。

流れとしては、以下のようになる。

- (1) 1 でロボット α が発話し、終了時にシグナル A を送る。
- (2) 2 でシグナルを受けロボット β が発話し、終了時にシグナル B を送る。
- (3) 3 で A がシグナルを受信し、観客反応センサに状況認識命令を送る。
- (4) 4 (Fig. 3 の L) において COOL_DOWN を受け取れば 4'、他のシグナルなら 5 に移動。
- (5) 4' でロボット α が 100 [ms] 停止する。
- (6) 5 でロボット α が発話し、開始時にシグナル A を送る。ロボット β がそれを受けて発話。

3.4 スクリプト開発における本手法の有用性

本稿で提案する手法は、ロボット同士を対話させるスクリプトを開発するうえで、開発者が開発を行いやすくなる利点がある。本節ではこの利点を具体的に示し、簡単な比較実験により客観的にその有用性を示す。

3.4.1 スクリプト開発上の利点

本システムは、以下のような利点を持っている

● シナリオ内容の変更の容易さ

すべてのロボットが同期した内部クロックを持ち、あらかじめ決められたタイムスケジュールで動作するような内部クロック同期方式のシステムを考えると、あるロボット

Table 3 Example of scenario

no.	Robota	Robot β
1	speakwait “ γ .wav”: A	signalwait(A)
2	signalwait(B)	speakwait “ σ .wav”: B
3	check_reaction()	signalwait(A)
4	if(reaction=COOL_DOWN)	
4'	wait(100)	
	endif	
5	speak “ θ .wav”: A	
	signalwait(B)	speakwait “ ν .wav”: B
	speak “ ϵ .wav”: A

Fig. 3 Example of signal exchange

の台詞を変更する場合に変更後の台詞の発話時間にあわせて他の全ロボットのタイムスケジュールの変更が必要となり、非常に煩雑である。しかし、本システムでは個別のロボットの台詞（具体的には、発話音声ファイル）を変えるだけでよく、他ロボットのシナリオ変更の必要性はない。

● 発話間合いの調整

スクリプトを作る際には、ロボット同士の発話間合いを微調整して、適切な間合いで発話するように調整する必要がある。また、我々は協調対話の発話タイミングは聴衆（観察者）に応じて変更される必要があると考える。例えば、老人を対象とする場合には、協調対話の速度やタイミングを遅くする必要性が生じる可能性がある。

システムでは発話と発話の間にある wait 命令の引数を変更するだけで発話間合いを容易に調整可能である。

例えば、Table 3 のシナリオを年配の観客向けにすることを想定してゆっくりとした速度・タイミングに変更した場合、以下の二つの手順のみで完了した (Fig. 4)。

- (1) その漫才の音声ファイルをあらかじめ所有していたスクリプトで速度を一括して変換する。
- (2) シナリオ上の wait() のカッコ内の指定待機時間を、一括してテキストエディタで置換する。

3.4.2 比較方法

本提案方式の上記二つの利点を検証するために、簡単な比較実験を行った。実験では、2 台のロボットが簡単な会話を行うためのスクリプトを、提案手法と、その対比として内部クロック方式の 2 通りの方法で用意し、開発者が「タスク 1) シナリオ内容の変更」と「タスク 2) 発話間合いの調整」のためにスクリプトの修正を試みた。これらのスクリプトを修正するのにかかった時間をそれぞれ計測した。

- (1) 提案手法：本提案手法に基づき用意されたスクリプト (Table 4) を修正する。
- (2) 内部クロック同期方式：2 台のロボットが同期した内部クロックを持つとの前提で、対話開始時刻からの経過時刻とそれの際のロボットの振る舞いを記述する方式のスクリプト (Table 5) を修正する。

いずれのスクリプトでも、片方のロボットの発話終了時から 100 [ms] 後にもう片方のロボットが発話するように設定してある。

修正の具体的な内容は以下のとおりである。また、修正の際に、各音声ファイルの長さを開発者に示した。

Fig. 4 Example of making slower the intervals

タスク 1) シナリオ内容の変更

音声ファイルの内容が変更されたことを想定し、その際に必要となるスクリプトの修正を行う。Table 4, Table 5 に示されるスクリプトそれぞれにおいて、各音声ファイル (γ .wav, σ .wav, ...) の長さがそれぞれ 2 倍になったとして、音声発話のタイミングを調整する。発話間合いは、修正前と同様に片方のロボットの発話終了時から 100 [ms] 後にもう片方のロボットが発話するものとする。

タスク 2) 発話間合いの調整

2 台のロボットの発話間合いの調整を行うことを想定し、その際に必要となるスクリプトの修正を行う。片方のロボットの発話終了時から 200 [ms] 後にもう片方のロボットが発話するようにそれぞれのスクリプトを修正する。

3.4.3 比較結果

3 人の、いずれも対話ロボット開発に携わる学生が、開発者としてシナリオの修正を試みた。タスク 1 とタスク 2、それぞれにかかった平均時間を Table 6 に示す。タスク 1 に関しては、本方式ではスクリプトの変更が不要であるため時間がかからなかったが、内部クロック同期方式では平均 66.7 [s] かかった。また、タスク 2 に関しては、本方式のスクリプトでは wait 命令により明示されている間合いの時間を変更するだけでよく 12.0 [s] で作業が終わったが、内部クロック同期方式では各動作開始時刻を 100 [ms] 遅くする、200 [ms] 遅くする、など順々に開始時刻を遅らせていく必要があり、47.3 [s] とより多くの作業時間がかかった。結果的に、タスク 1、タスク 2 ともに、提案方式によるスクリプトの方が短時間で修正できたことが分かる。

Table 4 Scenario script written in the proposed method

Robot α	Robot β
speakwait " γ .wav": A1	signalwait(A1)
	wait(100)
signalwait(B1)	speakwait " σ .wav": B1
wait(100)	
speakwait " θ .wav": A2	signalwait(A2)
	wait(100)
signalwait(B2)	speakwait " ϵ .wav": B2
wait(100)	
speak " ν .wav"	

Table 5 Scenario script written in a synchronized clock based method

Robot α	
Time	Action
0	speak " γ .wav"
3400	speak " θ .wav"
10300	speak " ϵ .wav"

Robot β	
Time	Action
2000	speak " σ .wav"
5300	speak " ν .wav"

Table 6 Result for the modifying time

	task 1 [sec]	task 2 [sec]	reaction
proposed method	0.0	12.0	possible
clock-based method	66.7	47.3	impossible

実際に後述の実験で用いた漫才スクリプトを作成した際には、何度もタスク 1, 2 のような手順を繰り返し、スクリプトを動作させて実際の動きを見てはタイミングを変更する必要がある、またスクリプトの長さも長いので、これらの手法の差はより大きいものとなる。また、本提案手法は観客の反応に応じてロボットの振る舞いを切り替えることが可能であるが、内部クロック方式では実現が困難である。

4. 実験

本実験の目的は、構築したシステムの能力と、社会的受動メディアとしてのロボットの可能性を従来メディアとの性能比較によって検証することにある。題材としては、2 章で述べたように、複数の人の間でタイミングよくやり取りが行われる漫才を取りあげ、テレビモニタ上の映像として行われる人間の漫才と比較する。これにより、開発したシステムのタイミングなどの自然さに関する基本能力が従来メディアと同程度あることを確認するとともに、ロボットが実際に観察者の目前に存在することによる臨場感がどの程度総合的な性能に影響するのかを評価する。

4.1 被験者

本実験の被験者は、平均年齢 19.6 歳の大学生 32 名（男性 13 名、女性 19 名）である。

4.2 実験方法

Fig. 7 に示すように、被験者は幅 120 [cm] のディスプレイまたは 2 体のロボットの前に用意された席に着席し、漫才を見る。なお、今回の実験において観客は 1 人であり、大きな笑い声は期待できないため、むしろ 1 人でも笑い声をあげやすいように「仮想笑いシステム」を使用した。これは、キーボードのキーを押すことにより、押している間はスピーカから、笑い声が流れるシステムである。また、この実験ではこの仮想笑いシステムからの出力を、観客反応センサとして用いた。

さらに、このボタンを押している時間をデータとして抽出した。なお、ロボット側から観客反応の認識命令が来ている場合は、キーから手を離すと終了のシグナルが送られ、押されていない場合は 1.5 [s] で終了のシグナルが送られる機構を実験のために追加し、本システムの状況認識機能の実証を行った。

4.3 実験条件

実験は被験者間比較により行った。被験者は約 1 分間のロボット漫才 (Fig. 5)、もしくは人間漫才 (Fig. 6)、のどちらか一方のみを Fig. 7 で示す位置で観察した。どちらの条件でも、漫才のシナリオおよび音声は同一である。

- (1) ロボット条件 (robot): 人間型ロボット (Robovie) 2 体により漫才を行うもの。
- (2) 人間条件 (human) 人間が行った漫才をビデオ録画したもの (ビデオに録画された漫才を行ったのは、大学の落語研究会に 4 年間在籍して落語・漫才の練習を行ってきたアマチュア 2 名である)。

ロボット条件にはロボットの実物を用いたのに対し、人間条件ではビデオ録画したものをを用いた理由は、我々の興味は社会的受動メディアとしてのロボットの可能性の検証にあるためである。つまり、従来は人間が行った漫才をビデオに録画しテレ

Fig. 5 A scene of the experiment at Robot condition

Fig. 6 A scene of the experiment at Human condition

Fig. 7 View of experiment.

(We placed display and robots in front of subject with different distance, so that apparent sizes of humans in display and robots were same)

ビというメディアが繰り返し放送するといった利用が行われていたが、この代わりに複数のロボットが繰り返し再現してみせるという将来的な利用を考えている。したがって比較する対象は、従来の漫才を流通させてきたメディアとしてテレビを選び、ディスプレイに写される人間同士の漫才とした。

また、被験者から見えるロボット／人間のサイズが類似のものとなるようにディスプレイの位置をロボットの位置よりも手前に設置した。音声は、ロボット条件では各ロボットの頭部に取り付けられたスピーカから発話させた。人間条件では、Fig. 6 に示すように、ディスプレイの手前に小型ステレオスピーカを置き、そこから発話させた。

4.4 ロボットの動作

ロボットの動作に関しては、人間条件のビデオを参考に、だいたい同じような動きをロボットがするように作成した。特に、漫才特有の「つつこみ」に関しては、シナリオ上の同じ箇所

human condition

robot condition

Fig. 8 An example of "Tsukkomi" motion

可能な範囲で同じような動き（軌道，速度，加速度）で行った（Fig. 8 の人間条件・ロボット条件ともに左側が「つつこみ」を行っている）。ただし，現状の人間型ロボットが，人間と完全に同じ動作を行うことには無理があり，一方人間がロボットの真似をすることも不自然である [11]。このため，人間条件，ロボット条件ともに，細かい動作に関してはそれぞれ最善をつくしたモーションを行うようにした。つまり，ロボット条件の開発者によるある程度の作りこみ，デザインが含まれてしまうことはやむを得ないと考え，開発者にゆだねた。その上で，動作，声，タイミングの自然さに関して，人間条件と差が生じないように留意してロボット条件の動作などを作成した。

4.5 評価方法

被験者が漫才を見た後，以下の五つの評価項目に関する 7 段階評価の主観評価と自由にコメントを回答する自由記述を行うアンケートに記入を求めた。具体的な評価項目は，(a) 動作の自然さ，(b) 声の自然さ，(c) タイミングの自然さ，(d) 臨場感，(e) 漫才全体の評価，である。これと，仮想笑いシステムのボタンが押されていた時間（笑っていた時間）のデータ，以上 6 項目により評価を行った。

5. 実験結果と仮説検証

5.1 仮説

本実験では，以下の仮説を検証する。

仮説 1：動作，声，タイミングの自然さに関して，人間条件と差が生じないように留意してロボット条件の動作などを作成したため，評価において条件間に差がない（つまり，開発したロボットシステムにより，人間同士の会話と同程度の自然さが実現される）。

仮説 2：臨場感に関しては，ビデオよりも実物のほうが大きい。

Fig. 9 Naturalness of motion

Fig. 10 Naturalness of voice

Fig. 11 Naturalness of timing

つまり，ロボット条件の方が人間条件より臨場感に関する評価が高い。

仮説 3：漫才全体の評価に関して，ロボット条件も人間条件も同じシナリオの漫才を行うため漫才のストーリーとしては条件間に差はないが，臨場感が影響することにより，ロボット条件のほうが人間条件よりも評価が高い

5.2 実験結果

各評価項目に関して Mann-Whitney の U 検定により 2 条件を比較することで，仮説の検証を試みた。

● 仮説 1 の検証

Fig. 9, Fig. 10, Fig. 11 はそれぞれ，動作，声，タイミングの自然さの評価を，ロボット条件と人間条件の間で比較したグラフである。図中の塗りつぶした棒は評価の平均値を，棒上の直線は平均値±標準偏差の範囲を示す。

Mann-Whitney の U 検定の結果，動作 [$U=120.0$, $p=.756$, n.s.]，声 [$U=104.5$, $p=.339$, n.s.]，タイミング [$U=105.0$, $p=.370$, n.s.]，以上 3 項目に関して，人間条件とロボット条件の間に有意差は存在しないことが示された。また，各比較における p 値も，有意傾向を示すような値よりも大きいことから，被験者数を増やすことで有意差が生じる可能性も少ないと考えられる。つまり仮説 1 が検証されたといえる。

● 仮説 2 の検証

Fig. 12 は，臨場感に関する主観評価をロボット条件と

Fig. 12 Presence

Fig. 13 Overall impression

Fig. 14 Laughter duration

人間条件の間で比較したグラフである。

Mann-Whitney の U 検定の結果、臨場感において、ロボット条件への評価の方が人間条件より有意に高かった [$U=38.0$, $p=.000$]。つまり仮説 2 のとおり、ロボットの方がディスプレイの中の人間よりも大きな臨場感を与えることを示している。

● 仮説 3 の検証

Fig. 13 は、漫才全体の評価をロボット条件と人間条件の間で比較したグラフである。Mann-Whitney の U 検定の結果、漫才全体の評価に関しては、ロボット条件のほうが人間条件より有意に高いことが示された [$U=75.5$, $p=.038$]。よって、仮説 3 は検証された。

一方、Fig. 14 に、仮想笑いシステムのボタンを押していた平均時間（つまり、笑い時間）をロボット条件と人間条件の間で比較したグラフを示す。同様に漫才の評価にかかわると考えられる笑い時間に関しては平均点に有意差は存在しなかった [$U=102.0$, $p=.325$]。笑い量に関しては、グラフでの標準偏差にも示されるように、個人差が大きいことも有意差が生じなかったことに影響したと考えられる。

6. 考 察

6.1 システムの有効性

実験の結果、仮説 1 が検証されたように、人間条件とロボット条件との間で動作・発話・タイミングの自然さに関する差はみられなかった。つまり、本システムによって、社会的受動と

しての複数ロボット協調対話が、人間に近い自然さのタイミングで実現可能であるといえる。

一方、実験のアンケートの自由記述部分において、“不自然に停止することがある”という報告が多く寄せられた。これはおそらく、本実験のために用いた「仮想笑いシステム」に要因があると考えられる。仮想笑いシステムにおいて、観客がキーを押し続ける間は、ロボットが動作、音声等、完全に静止するという実装を行った。これが結果として不適切なタイミングを作成してしまったものと考えられる。仮想笑いシステムの使用回数自体は、今回使用した 2 種類の漫才シナリオにおいてそれぞれ 7 回と 2 回という少ないものであったが、仮想笑いシステムが働くよう設定したところは、漫才においてボケつつこみが終了、またはボケつつこみの間といった、観客が最も意識を集中しているポイントであった。次は何を言うのか、つつこみはまだなのか、そのような場面で仮想笑いシステムの反応待ちの停止が際立って見えたという効果も影響を与えたのではないかと考える。同時に多くの観客の利用を想定した観客反応センサの有効性を適切に検証するためには、実際に漫才が演じられている演芸場などでの実験が有効であろう。今後多くの観客の前で本システムを利用する機会にこの有効性の検証を行いたい。

6.2 人間の漫才とロボットの漫才の比較

● 自然さに関する考察

動作、声、タイミングの自然さにおいて、両条件の間に有意差は存在しなかった。声に関しては、元より同じ人物の音声を使用しているため、この結果は予想されたものである。動作、タイミングに関しては人間条件がアマチュアのものであるものの、ある程度人間同士の対話に近い自然さがロボットによっても実現されたといえる。つまり、我々はこのシステムが少なくとも観客に不自然さを感じさせないレベルに到達しているものと考えられる。

● 臨場感、漫才全体の評価

臨場感、漫才全体の評価は共に、ロボット条件が人間条件と比較して有意に高かった。ビデオに比べて実在するロボットの方が臨場感が高くなるということは予想されていた結果であるといえる。漫才全体に関する評価にも臨場感の影響は大きいようであるが、これもまた新たな社会的受動メディアとしてのロボットの有用性を示しているものと考えられる。

● 笑い時間に関する比較

笑い時間に関しては、グラフからは一見ロボット条件の方が大きいようにも見えるが、有意差は存在しない。ただし、極端に多くボタンを押す被験者とまったくボタンを押さない被験者がいたことが標準偏差を大きくしたことで、また、本実験で用いた「仮想笑いシステム」の悪影響も考えられるため、実際のロボット漫才が観客の集団に笑いを誘発させることができるかという問題に関してはさらなる検証を行う必要があると考えられる。

7. 結 論

本論文では、人-ロボット相互作用への新たなロボットの利用法として、社会的受動メディアとしての複数ロボット協調対話

システムの開発を報告した。本システムは複数のロボットをネットワーク通信によって、あたかも本当に会話しているかのような会話を行わせるものである。実験の結果、開発したシステムを用いて実現された2台のロボットによる漫才は、人間が行った漫才のビデオと比べて遜色のないものであることが分かった。動作やタイミングの自然さが人間と同等に評価され、臨場感や漫才の全体的な評価はロボットの行ったものの方が高く評価された。この結果は社会的受動メディアとしてのロボットに肯定的な見通しと、エンターテインメントとしてのロボット漫才の有用性を示しているといえる。

謝 辞 本システム開発にご協力いただいた三菱重工の塘中氏、吉本興業の上田氏、ATRの古城戸氏、寺内氏、そして筆者の林とともに人間漫才のビデオ撮影にご出演いただいた大阪大学の利川氏に深く感謝いたします。本研究は、総務省の研究委託により実施したものです。

参 考 文 献

- [1] Y. Sakagami, R. Watanabe, C. Aoyama, S. Matsunaga, N. Higaki and K. Fujimura: "The intelligent ASIMO; System overview and intergration," IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS'02), pp.2478-2483, 2002.
- [2] B. Reeves and C. Nass: "The media equation. How people treat computers, television, and new media like real people and place?," CSLI Publications, 1996.
- [3] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson and H. Yan: "Embodiment in Conversational Interfaces: Rea," Conference on Human Factors in Computing Systems (CHI'99), pp.520-527, 1999.
- [4] C. Kidd and C. Breazeal: "Effect of a Robot on User Perceptions," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS'04), 2004.
- [5] 神田崇行, 石黒浩, 小野哲雄, 今井倫太, 中津良平: "人-ロボットの対話におけるロボット同士の対話観察の効果", 電子情報通信学会論文誌 D-I, vol.J85-D-I, no.7, pp.691-700, 2002.
- [6] H.H. Clark: Using Language, Cambridge University Press, 1996.
- [7] 竹内真士, 北岡教英, 中川聖一: "韻律・言語情報を素性とした決定木による自然な応答生成タイミングの検出", 日本音響学会講演集, 2-4-9, pp.75-76, 2003.
- [8] 山本倫也, 渡辺富夫: "ロボットとのあいさつインタラクションにおける動作に対する発声遅延の効果", ヒューマンインタフェース学会論文誌, vol.6, no.3, pp.87-94, 2004.
- [9] M. Kamasima, T. Kanda, M. Imai, T. Ono, D. Sakamoto, H. Ishiguro and Y. Anzai: "Embodied Cooperative Behaviors by an Autonomous Humanoid Robot," IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS2004), pp.2506-2513, 2004.
- [10] T. Kanda, H. Ishiguro, M. Imai and T. Ono: "Development and Evaluation of Interactive Humanoid Robots," Proceedings of the IEEE, vol.92, no.11, pp.1839-1850, 2004.
- [11] 神田崇行, 宮下敬宏, 長田拓, 配川有二, 石黒浩: "人ロボット相互作用における人型ロボットの外見の影響", 日本ロボット学会誌, vol.24, no.4, pp.497-505, 2006.

林宏太郎 (Koutarou Hayashi)

1981年9月23日生。2005年3月まで、大阪大学工学部応用理工学科在学。同学部を卒業し、2005年4月より奈良先端大学院大学情報科学研究科に入学。ATR 知能ロボティクス研究所実習生。ロボットインタラクション、人間-ロボット間のコミュニケーションに興味を持つ。(日本ロボット学会学生会員)

宮下敬宏 (Takahiro Miyashita)

1970年7月30日生。1993年大阪大学基礎工学部制御工学科卒業。1995年同大学大学院基礎工学研究科物理系制御工学分野博士前期課程修了。2000年同大学大学院工学研究科電子制御機械工学専攻博士後期課程単位取得退学。博士(工学)。1998年より日本学術振興会特別研究員。2000年ERATO北野共生システムプロジェクト研究員、和歌山大学システム工学部助手を経て、2002年7月よりATR 知能ロボティクス研究所研究員となり現在に至る。視覚と全身触覚を持つ多自由度ロボットの研究に従事。人工知能学会会員。(日本ロボット学会正会員)

萩田紀博 (Norihiro Hagita)

1954年2月11日生。1978年慶應義塾大学大学院工学研究科電気工学専攻修士課程修了。同年日本電信電話公社(現:NTT)武蔵野電気通信研究所に入所。文字認識や画像認識などの研究に従事。NTT基礎研究所などを経て、現在ATR 知能ロボティクス研究所所長。工学博士。IEEE、電子情報通信学会、情報処理学会、人工知能学会、各会員。(日本ロボット学会正会員)

神田崇行 (Takayuki Kanda)

1975年12月7日生。1998年京都大学工学部情報工学科卒業。2000年同大学大学院情報学研究科社会情報学専攻修士課程修了。2003年同専攻博士課程修了。博士(情報学)。現在、ATR 知能ロボティクス研究所上級研究員。ヒューマンロボットインタラクション、特にロボットの自律対話機構や社会的能力、人間型ロボットの身体を利用した対話に興味を持つ。(日本ロボット学会正会員)

石黒 浩 (Hiroshi Ishiguro)

1963年10月23日生。1991年大阪大学大学院基礎工学研究科物理系専攻博士課程修了。工学博士。同年山梨大学工学部情報工学科助手。1992年大阪大学基礎工学部システム工学科助手。1994年京都大学大学院情報学研究科社会情報学専攻助教授。2001年、和歌山大学システム工学部情報通信システム学科教授。現在、大阪大学大学院工学研究科知能・機能創成工学専攻教授。ATR 知能ロボティクス研究所第二研究室客員室長。視覚移動ロボット、能動視覚、パノラマ視覚、分散視覚に興味を持つ。人工知能学会、電子情報通信学会、情報処理学会、IEEE、AAAI 各会員。(日本ロボット学会正会員)