Robot *Manzai*

– Robots' conversation as a passive social medium –

Koutarou Hayashi[1&2], Takayuki Kanda[1], Takahiro Miyashita[1], Hiroshi Ishiguro[1&2], Norihiro Hagita[1]

[1]ATR Intelligent Robotics and Communication Labs.
Kyoto, Japan

[2]Osaka University
Osaka, Japan

E-mail: {hayashik, kanda, miyasita}@atr.jp, ishiguro@ams.eng.osaka-u.ac.jp

***Abstract* - 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, but the system is designed based on communication through a network. Actual information exchanged 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 the advantage for developer's being easy to develop. The developed system is applied to "robot *Manzai*," which is a Japanese comedy conversation usually performed by two people. Although tempo and timing are particularly important in *Manzai*, the robot *Manzai* system was more highly evaluated than the *Manzai* shown in a video performed by humans. We believe that this system demonstrates the potential of robots as a passive-social medium, similar to television and computers.**

Index Terms - inter-robot communication; robot cooperation.

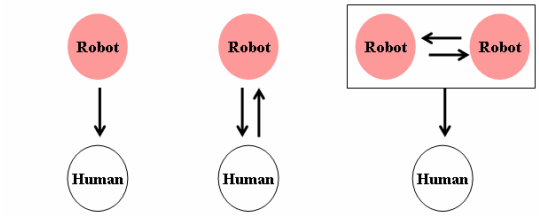
I. Introduction

Over the past several years, many humanoid robots have been developed that can typically make sophisticated human-like expressions. We believe that humanoid robots will be suitable for communicating with humans. The human-like bodies of humanoid robots enable humans to intuitively understand their gestures and cause them to unconsciously behave as if they were communicating with their peers. That is, if a humanoid robot effectively uses its body, people will communicate naturally with it. This could allow robots to perform communicative tasks in human society such as route guidance.

Recent research in HCI (human-computer interaction) has highlighted the importance of robots as a new interface media. Reaves and Nass researched the role of computers as a new interface medium in the manner of previous media, such as television and radio, and they proved that humans act toward computer interfaces (even a simple text-based interface) as if they were communicating with other humans [1]. Cassell et al. demonstrated the importance of anthropomorphic expressions, such as arms and heads on embodied agents, for effective communication with humans [2]. Cory and Cynthia compared a robot and a computer-graphic agent and found that the robot was suitable for communication about real-world objects [3].

We believe that humanoid robots will be used as an interface medium, particularly by showing conversation among multiple robots. For example, Kanda et al. proved that users understand a robot's speech more easily and actively respond to it after observing conversation between two robots [4].

Moreover, users are under no obligation to take part in the robots' conversation when they see it, which is contrary to the situation where a user talks with a person or a single robot. In other words, borrowing the words from linguistic literature of Clark [5], the user is placed in a *bystander* position (free from responsibility for the conversation), where robots are speaking and listening as *participants* (responsible for it). This characteristic is also similar to previous media: people do not have to respond to what the medium (the person in the medium) says, or to return the greeting when the medium greets, or feel uncomfortable about leaving in front of the medium.



(a) *passive* (b) *interactive* (c) *passive-social*

Fig.1: Robot(s) as medium

Figure 1 shows the difference of this type of medium compared to other forms of human-robot interaction. At times robots have been used for merely presenting information to people, which we named a *passive* medium. This is the same as a news program on TV where one announcer reads the news. On the other hand, many researchers have been struggling to realize robots that act as an *interactive* medium, which intends to accept requests from people as well as present information to people. The robot-conversation-type medium, on which we focus in this paper, is named a *passive-social* medium. It does not accept requests from people, which is as same as the *passive* medium; however, it does present more information than the *passive* medium, through its *social* ability: expression of conversation. It is rather similar to a news program on TV where announcers discuss comments spoken by others.

To develop a robot system as a *passive-social* medium, one technical difficulty must be overcome: a method for timing adjustment must be developed. To achieve natural human-like robot conversation, it is crucial to consider the timing between robots, such as intervals of speech and motion. Although a few research works have reported on the automatic adjustment of timing, such as a computer interface that naturally gives back-channel responses [6], it still lacks naturalness. We believe that it is currently realistic enough to let developers prepare and adjust the timing of speech and motion for the robots.

Although there exists multiple-robot cooperation, such as

synchronized dancing of SDR-3Xs, to the best of our knowledge, previous methods seem to be based on synchronized time clocks. That is, by synchronizing all robots' internal clocks, every robot can behave exactly according to pre-recorded sequences, carefully prepared by the developers. Once the sequences are prepared, however, it is difficult to re-adjust them and to adapt to external stimuli, such as audience applause.

This paper reports the development of a multi-robot conversation system based on communication through a network. The exchanged information among robots is designed based on an analysis of inter-human conversation. The system also features the advantages that it is easy to re-adjust and to adapt to external stimuli. Based on our developed system, we have implemented a robot *Manzai*, which is a Japanese comedy conversation mainly performed by two people. The effect of the developed system and Robot *Manzai* are verified experimentally. We believe this is the first application of multi-robot cooperation for human-robot communication, indicating a positive perspective for robots as a *passive-social* medium.

II. Analysis of Inter-Human Communication

A. Why did we choose Manzai as a first application?

First, we analyzed *Manzai* as one form of inter-human communication because it requires precise control of timing and coordination. That is, if we can achieve *Manzai* in robots, we can probably apply the same mechanism to various other forms of conversation.

Manzai is Japanese stand-up comedy in which two or three performers carry on a funny dialogue. *Manzai*-performers try to make the audience laugh by their utterances and gestures. Each performer has a different part in the dialogue: the *Boke* says the funny lines and the *Tsukommi* responds to the *Boke* in the "straight-man" fashion (**TABLE 1**). It is necessary to accurately regulate the *Manzai* dialogue's timing, which as we said above, makes it an appropriate and very demanding first test application for the inter-robot communication system we are developing.

B. Analysis of Manzai to classify types of timing

We analyzed four *Manzai* (M1…M4) performed by professional *Manzai*-performers in order to investigate the necessary conditions for the performance. **TABLE 2** shows the result of the analysis. During the fundamental progression of a *Manzai*, one performer usually respects the speech and movement timings that are tightly tied to his partner's. The analysis led to classification of these timings into five types.

The most common timing was the performer starting his speech after the partner's speech had ended (shown as **Type 1** in the table). This is also a very common aspect of our daily conversation and is known as turn-taking. Similarly, **Type 2** represents the timing of the partner's end-of-action. **Type 3** is common in Manzai but quite unusual in daily conversation; that is, starting speech simultaneously with the partner's speech. This type is often observed at the beginning of a *Manzai* routine, such as both members speaking "hello everyone" together. **Type 4** is barging into the partner's speech, which is observed in daily conversation, but occurs quite often in *Manzai*. **Type 5** is very special timing for *Manzai*: starting speech or actions after the audience has finish clapping or laughing. In contrast to TV news, *Manzai* sometimes involves the audience by making fun of them and taking account of their responses in the scenario. We believe that this matches with the merits of robots as a *passive-social* medium, because robots will not obtrusively require responses from people due to their insufficient recognition ability, but will implicitly react to listeners to promote interaction with people.

TABLE 1: Example of *Manzai* (role of *Boke* and *Tsukommi*)

Boke (the part of speaking funny lines)***:***
There we go! The bride and the groom made it through the staff entrance!
Tsukommi (the part of responding to funny lines)***:***
Staff entrance? Why?
Boke:
A truck carried the bride and the groom into here!
Tsukommi:
Truck? Why were they carried by a truck?

TABLE 2: Occurrence of speech timing for the Manzai scenario

Speech timing	Manzai scenario				Rate
	M1	M2	M3	M4	
Type 1 (speech end)	238	561	173	175	82.2%
Type 2 (action end)	5	1	2	3	0.8%
Type 3 (simultaneous speech)	2	21	3	2	2.0%
Type 4 (barge in)	31	29	35	15	7.9%
Type 5 (wait audience)	18	17	34	30	7.1%

C. Necessary conditions for the system

We have analyzed four different *Manzai* and found that in all cases the starts of speech are classified into five types of timing. Thus, this analysis suggests the importance of building a system that can achieve these five types of timing for a human-like Robot-*Manzai* system. This is easily realized in **Types 1** and **2** by sending a message from the first robot to the second one when the first is finished acting or speaking. Upon receiving the message, the second robot starts speaking. For **Type 3**, similarly, we can consider sending a message from the first robot to the second one, as soon as the first starts speaking. For **Type4**, we split the voice file at the point where we want to have the second robot barge in. It is sufficient to send a message from the first to the second when the first is done playing the first part of the voice file. For **Type 5,** we should check the reaction of the audience (clapping and laughter) at important points of the *Manzai* routine. For example, it could be realized by measuring the noise produced by the audience with a sound-level meter.

III. Multi-Robot Communication System

This system consists of language for multi-robot communication, a set of robots that interpret and execute scripts written in the language, and sensors that recognize the state of the audience (**Fig. 2**).

A. Language for multi-robot communication

We have created a scripting language that developers can use to prepare scripts for each robot. Each robot reads the script file and interprets its commands. This language is com-

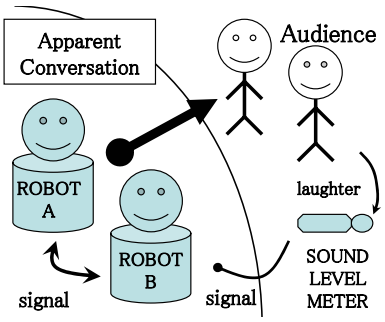


Fig. 2: Outline of a Multi-robot communication system

posed of the following commands:

signal(*S*)
This command is used for sending an "*S*" signal. The similar command **signalwait(*S*)** is used for keeping the robot in a waiting state until it receives an "*S*" signal. There is another way to send a signal: placing ":*S*" at the end of the following commands.

motion "*motion_file_name*" : *S*
This command is used to make the robot read the motion file "*motion_file_name*", and interpret it into a motion sequence. This **motion** command does not block the script interpreter until the end of the motion.

motionwait "*motion_file_name*": *S*
This is a similar to the **motion** command except that the interpreter does not move on to the next command until the motion sequence is over. Like **motion** and **motionwait**, there are other coupled commands that have the same syntax, **speak** and **speakwait**, which make the robot play (speak) the content of a voice file.

move(*x, y, θ*)
This command makes the robot move to the specified spot represented by (*x, y, θ*). There is a **movewait** command that waits until the robot reaches the new position before allowing the interpreter to continue.

check_reaction()
When this command is executed, a connected sensor returns a signal that reflects the state of the audience. Developers can program behaviors for the robots by using the "**if"** sentence syntax to process the sensor response.

wait (*t*)
This command is used to make the robot wait for a time interval equal to *t [milliseconds]*. This is useful *w*hen developers want to set intervals between actions in a scenario.

exit()
This command is used to exit the interpretation of the script. When this command is executed, the robot sends a signal informing others that it has finished interpreting the script.

B. System components

- Humanoid robot "Robovie"

We use a humanoid robot, Robovie [7], for this system. **Figure 4** shows Robovie having arms with four degrees of freedom and a head with three degrees of freedom. Consequently this robot is capable of human-like body expression. The robot interprets a script written in the previously mentioned language and performs accordingly.

TABLE 3: Example of scenario

robot A	*robot B*
speakwait "helloB.wav" :S1	signalwait(S1) speakwait "helloA.wav" :S2

signalwait(S2)
check_reaction()
if(reaction=COOL_DOWN)
  wait(100)
endif	

signal(S3) speak "hey.wav" :	signalwait(S3) speak "hey.wav"

- Sensor for recognizing the state of the audience

This sensor features a sound-level meter and is continually receives the volume data [dB] from it. When a command comes from a robot, the sensor returns an answer with a signal corresponding to the level of the audiences' reactions: *BURST_OUT*, *LAUGH*, *COOL_DOWN*. Developers can set an appropriate volume level that corresponds to each signal.

C. An example of implementation

TABLE 3 shows an example of the script written in this language. This example is progresses as follows (also shown in **Fig. 3**):

1. Robot *A* plays (speaks) the voice file "*hellowB.wav*." When robot *A* finishes speaking, it sends the signal "*S1*" to robot *B*.
2. When robot *B* receives signal "*S1*", it plays the voice file "*hellowA.wav*." When robot *B* finishes speaking, it sends the signal "*S2*" to robot *A*.
3. When robot α receives signal "*S2*," it receives information on the state of the audience from the sensor.
4. If robot *A* receives *COOL_DOWN*, it moves on to step **4'**. If robot *A* receives something else, it moves on to **5**.

4'. At **4'**, robot *A* waits for *100* milliseconds.

5. Robot *A* sends the signal "*S3*" and immediately plays the voice-file "*hey.wav*." At the same time, robot *B* receives signal "*S3*" and plays "*hey.wav*" simultaneously with *A*.

This system includes two major merits as follows:

- It is easy to change the lines of the scenario

In systems where the user must prepare time schedules for each robot, if a developer wants to change something in the script, he/she must change the time schedule of every other robot too. However, with our system a developer needs to change what he/she desires, such as the name of the voice file, to a new one.

- It is easy to change intervals between speech lines

In a conventional system, if a developer wants to change the time interval between speech lines at a certain point, he/she must change all the time schedules starting from that point. However, in our system, developers need only insert the command ***wait(t)*** where intervals are required, thus they merely adjust the argument *t*.

For example, if a developer wants to make the scenario in **TABLE 3** slower for elderly people, he/she can go through these two steps: First, in the script file the developer can change the played voice files to low-speed voice files that

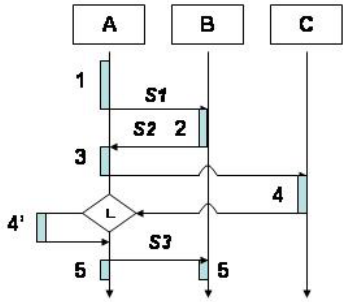


Fig. 3: Example of signal exchange

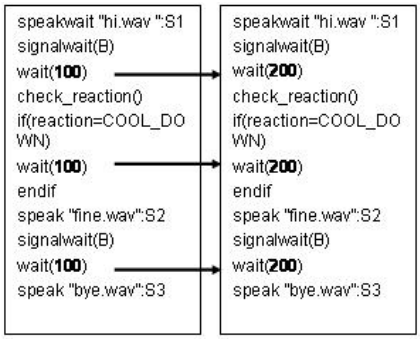


Fig. 4: Example of making slower intervals

Fig. 5: Robot condition

Fig. 6: Human condition

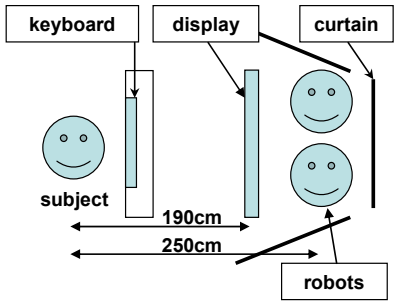


Fig. 7: The settings of the experiment

We placed the display and robots at different distances from subjects, so that the apparent sizes of humans in the display and the robots were the same.

he/she had slowed down in one instance, using a system command, or had ready beforehand. Second, insert a "**wait**(*t*)" command at necessary places such as after a "**signalwait**" command (**Fig. 4**). In our experience, it took almost 30 seconds to change a script file for a one-minute long scenario using this method.

IV. Experiment

We conducted an experiment to investigate the possibility of robots as a *passive social* medium in the *Manzai* application by comparing it with television, since we believe that in the future robots will be utilized as a *passive-social* medium in a similar way as to how television are used today. This is why we tried to compare television and robots utilization for the same task: showing a *Manzai* performance.

A. Method

Thirty-two university students participated in the experiment as subjects. The subjects are in not only engineering but also various departments. Their average age was 19.6. Thirteen of them were females and were nineteen males. Each of subjects watched the *Manzai* performance in a large room one by one.

In this experiment we could not use the sensor that recognizes the state of the audience because audience size was reduced to one, thus we could not expect large volume of laughter sound. Instead, we used a "*Virtual laughing system*," which is a simple system that generates the sound of laughter when the subject presses a key and sends the reaction signal *LAUGH* or *COOL_DOWN* to the robot.

B. Condition

Each subject watched a *Manzai* performed by either robots (**Fig. 5**) or humans (**Fig. 6**) in front of them (**Fig. 7**):

Robot condition:

Two robots perform *Manzai* in front of the subject. Please refer to the movie in proceeds CD for a scene of the robot Manzai.

Human condition:

Two amateurs, who have practiced *Manzai* for three years, perform *Manzai* on a TV screen in front of the subject.

C. Measurement

Once the experiment was finished, we distributed questionnaires to the subjects who rated the *Manzai* on a 1-to-7 scale for each question where 7 is the most positive. There were five questions: naturalness of motion, naturalness of voice, naturalness of timing, presence, and overall impression. We retrieved the subject's laughing time duration from the virtual laughing system.

D. Results

- Naturalness of motion, voice, timing

Figure 8 shows a comparison of questionnaire results for naturalness of motion. From the results of a one-way between-groups design ANOVA (analysis of variance), there is no significant difference between robot condition human condition in the average score ($F(1,31) = 0.06$, *n.s.*). **Figure 9** shows the results for the naturalness of voice while **Fig. 10** shows those for the naturalness of timing. A one-way between-groups design ANOVA also indicated no significant difference between them (Naturalness of voice: $F(1,31) = 1.21$, *n.s.*, naturalness of timing: $F(1,31) = 0.96$, *n.s.*).

- Presence and overall impression

Figure 11 shows the results for presence and **Fig. 12** shows them for overall impressions. A one-way between-groups design ANOVA shows a significant difference between

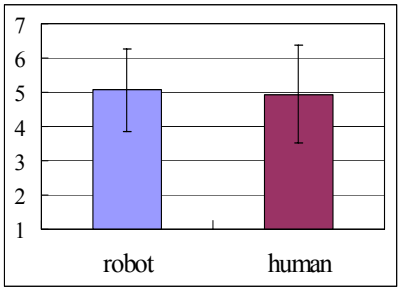


Fig. 8: Naturalness of motion

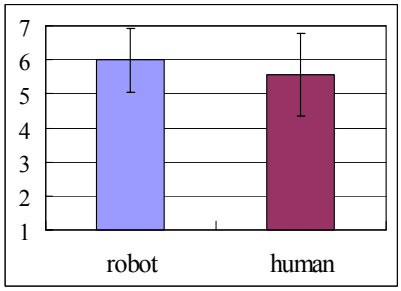


Fig. 9: Naturalness of voice

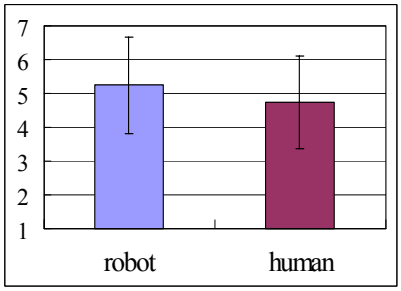


Fig. 10: Naturalness of timing

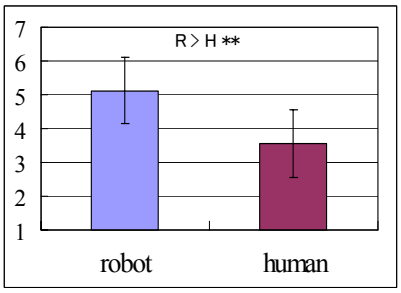


Fig. 11: Presence

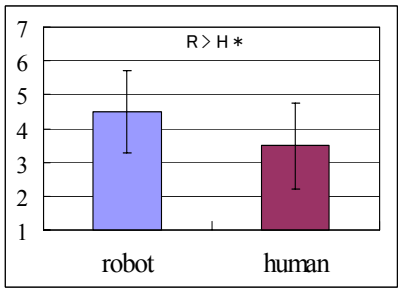


Fig. 12: Overall impression

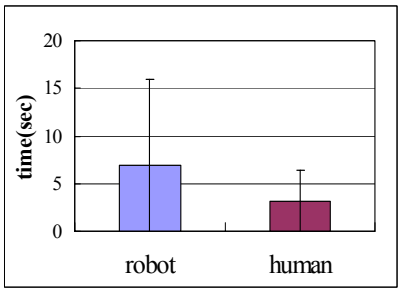


Fig. 13: Laughter duration

the robot condition and the human condition in the average score when we consider the presence and the overall impression (Presence: $F(1,31) = 18.49$, $p<.01$; Overall impression: $F(1,31) = 5.71$, $p<.05$). That is, the Manzai performed by two robots gave higher presence and overall impressions to subjects than the one performed by humans shown in the video.

- The average duration of laughter

Figure 13 shows the average duration of subjects' laughter. From the results of the one-way between-groups design ANOVA, we see there is no significant difference between the robot and human conditions as long as the average time of laughter is considered ($F(1,31) = 0.21$, *n.s.*).

V. Discussions

A. Effect of the multi-robot communication system

There is no significant difference in the average score for naturalness of motion, voice, and timing between the robot condition and the human condition. Concerning naturalness of the voice, this result is what we had expected because we used recorded voices of the human condition for the robot condition; thus the voices subjects heard were identical in both conditions. On the contrary, the results for motion and timing were quite surprising because even the current humanoid robot (Robovie) was evaluated as being a natural as humans in the *Manzai* application. We believe that these results indicate the system's capability of achieving human-like natural timing to a certain extent.

On the questionnaire, we also allowed subjects to freely describe their experience. Ten subjects reported: "sometimes, the robots stopped unnaturally." This is probably due to the virtual laughing system. This system sends the *COOL_DOWN* signal to the robot after a 1.5-second delay, which is the likely reason for the feeling of unnaturalness. In this experiment, we cannot verify the efficiency of the sensor that is meant to recognize the state of the audience.

B. Robot Manzai v.s. human Manzai

- Naturalness of motion, voice, and timing

There was no significant difference between the robot condition and human condition in the average score for naturalness of motion, voice, and timing. Thus, the robots reached some degree of human-like naturalness in the *Manzai* application. Of course, we should not overestimate the meaning of results, because the people performing *Manzai* in the video were amateurs who have practiced for only three years. If we compare the robot *Manzai performance* with one by professional *Manzai* performers, the average score of the professional performers would be higher than that of robots. Thus, the results only indicate that the realized quality of naturalness is equal to a three-year amateurs' level.

On the other hand, the virtual laughing system had an unexpected influence, and thus it might have been possible for robots to score higher than humans to a significant level.

- Presence and overall impression

For both presence and overall impression, the average score attained by the robots was significantly higher than that of humans'. We believe that the higher overall impression is produced by the more impressive presence of the robots, since the *Manzai* scenario was same and the naturalness of motion, voice, and timing was evaluated as similar in both conditions. In other words, the presence score reflects the added value of robots as a medium, in comparison with television.

- The average duration of laughter

There was no significant difference between the robot condition and human condition in the average score, but the

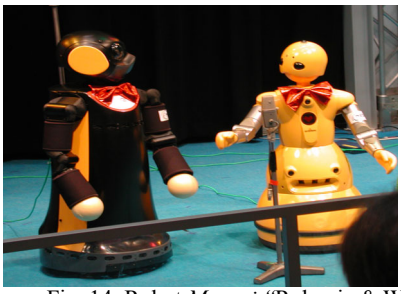



Fig. 14: Robot *Manzai* "Robovie & Wakamaru" and audiences at Expo 2005.

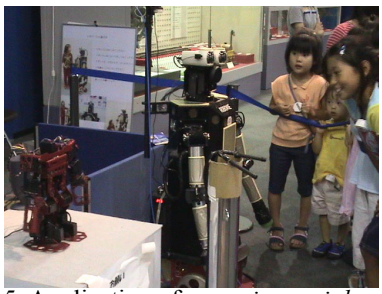


Fig. 15: Application of a *passive social* medium

standard deviation for the robot condition was relatively higher than that for the human condition. That is, the duration of subjects' laughter was more varied among subjects in the robot condition. We believe that some subjects did not laugh because of the robots' unnaturalness, while some laughed because they rather preferred the novelty of robots performing *Manzai*. For example, one subject answered that "the lack of facial expression makes the robots unfamiliar," while another one answered that "I was impressed that the non-human existences (robots) can perform *Manzai*"

- Reliability of comparison

Presence is the principal factor affecting the results in this experiment comparing humans in a display with humanoid robots. On the other hand, social factors affect subjects' behavior and the subjectivity of their evaluations. It can not be ignored that the novelty of the robots affected the evaluation. For example, when robots were novel, people placed high value on basic functions like walking. News of the first robot that could do bipedal locomotion moved people profoundly. Now they see it as usual. As humanoid robots become more popular, subjects are expected to become more critical in their evaluations.

On the other hand, social factors affect human feelings, too. Some subjects had a stronger sense of affinity to human than robot. For example, on the questionnaire, we allowed subjects to freely describe their experience. One subject reported: "It made me feel ashamed not to laugh at the human." Accordingly, it is sure that social factors affected these results. However, we think they are side issues and the presence of robots gave subjects the strongest impact.

C. Usefulness in entertainment applications

The technology reported in the paper is being employed at the 2005 World Expo in Aichi, Japan. A robot *Manzai* system named "Robovie & Wakamaru" has been exhibited at the Expo's prototype robot exhibition by Yoshimoto, an entertainment company that usually employs human *Manzai* performers, under the financial support of NEDO, Japan.

The prototype robot exhibition, which operated between June 9th and 19th, was one of the most popular exhibits at the Expo with more than a hundred thousand people visiting. Robovie & Wakamaru performed the *Manzai* twice per day on a stage (**Fig. 14**). The routine entertained many people and won second prize for the popular robots as voted by children from among all 65 robots in the prototype robot exhibition. We believe that this result also demonstrates the usefulness of the robot *Manzai* in entertainment application, as well as the potential usefulness of the developed system for easy preparation of contents.

D. Perspective for passive social medium

- Example as a medium

The experimental results indicate that the developed system is good enough to express conversation between robots in terms of naturalness and presence. Particularly for the *Manzai* application, these robots are ready to use. We believe that this positive perspective for an entertainment application is a good starting point. Moreover, the results show the fundamental potential of robots as a *passive social* medium. We believe that there are many possible applications. For example, **Fig. 15** shows a scene of an experiment at a science museum where two robots exhibited a conversation about exhibition guide and attracted many people. These robots were successfully used to motivate visitors' interest in science [8]. Similarly, robots in the future might be also used in show windows to talk about displays, like a television-based advertisement.

- Compatibility of this system for passive social medium

We will use this system as common software for other passive-social medium applications. Its flexibility helps us to effectively prepare scenarios.
It was helpful for the preparation of *Manzai* experiment, although we did not fully utilized its capability in our experiment, since the virtual laughing system was not so effective. Moreover, when humanoid robots will work in our daily life, even passive social medium applications requires more dynamic interaction capability, where our system will be more helpful.

VI. Conclusion

This paper reported the development of a multi-robot cooperation system for human-robot communication. In the system, robots behaved as if they were communicating by speech, but they were actually communicating through a network. The exchanged information among robots was designed based on an analysis of inter-human conversation. The developed system features a major merit: it is easy to re-adjust and adapt to external stimuli. We believe that this framework allows the system to autonomously adjust timing depending on the type of audience. It will thus be beneficial in our future works.

Based on the developed system, we have implemented a robot *Manzai*. We compared its performance with a *Manzai* performed by humans shown in a video. Experimental results

revealed that the robots' naturalness of motion and timing were similar to those of humans. Moreover, the overall impressions given by the robot *Manzai* was higher than that of the humans'. We believe that this system demonstrates the potential of the robot *Manzai* and presence of real robots. That is, it shows a positive perspective of robots as a *passive-social* medium, and that it illustrates the usefulness of the robot *Manzai* in entertainment applications.

ACKNOWLEDGEMENTS

We wish to thank Shingo Furukido and Hideaki Terauchi at ATR for their support. This research was supported by the Ministry of Internal Affairs and Communications of Japan.

REFERENCES

[1] B. Reeves and C. Nass, *The media equation.* 1996.

[2] J. Cassell, T. Bickmore, M. Billinghurst, L. Campbell, K. Chang, H. Vilhjalmsson, and H. Yan, Embodiment in Conversational Interfaces: Rea. *Conf. on Human Factors in Computing Systems (CHI'99),* pp. 520-527, 1999.

[3] C. Kidd and C. Breazeal, Effect of a Robot on User Perceptions. *Int. Conf. on Intelligent Robots and Systems (IROS'04),* 2004.

[4] T. Kanda, H. Ishiguro, T. Ono, M. Imai, and K. Mase, Multi-robot Cooperation for Human-Robot Communication, *IEEE Int. Workshop on Robot and Human Communication (ROMAN2002),* pp. 271-276, 2002.

[5] H. H. Clark, *Using Language,* Cambridge University Press, 1996.

[6] T. M. Takeuchi, N. Kitaoka and S. Nakagawa, Timing Detection for Realtime Dialog Systems Using Prosodic and Linguistic Information, *Int. Conf. on Speech Prosody,* pp. 529-532, 2004.

[7] 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.

[8] H. Ishiguro, M. Shiomi, T. Kanda, D. Eaton, N. Hagita, Field experiment in a science museum with communication robots and a ubiquitous sensor network, Proceedings of ICRA'05 Workshop on Network Robot Systems, 2005.