Appearance and Physical Presence of Anthropomorphic Media in Parallel with Non-face-to-face Communication

Analyses of Non-verbal and Underconscious Expressions

Tomoko Yonezawa[1] Noriko Suzuki[2] Kenji Mase[3] Kiyoshi Kogure[4]

[1] Kansai University
[2] Doshisha University
[3] Nagoya University
[4] Kanazawa Institute of Technology

Abstract: Puppets could become a new tool for expressive communication, in parallel with traditional communication channels in human-human interaction. This research aimed to verify the effectiveness of appearance and embodied presence of anthropomorphic media. In this paper, we focused on a usage of the anthropomorphic medium and the user's conscious or under-conscious behaviors in parallel to non-face-to-face conversation. We conducted a non-face-to-face conversational experiment by adopting a stuffed-toy robot, which allowed expression via motion and vocal cues. The bare-robot condition was prepared to compare the appearace, and the monitor which showed the stuffed-toy was adopted to compare the embodied presence. The analyses of the results showed that the appearance affects on the unconscious behaviors of the user and that the embodied presence affects on the conversational utterances. We conclude the physical embodied presence and appearance of the stuffed-toy robot play an important role in non-verbal communication in non-face-to-face conversation with the robot system.

1 Introduction

Recent media for synchronous communication have been developed various types of expressions even in the case of non-face-to-face communication. One of the most richest human-human communications is a face-to-face interaction in a real space. We underconsciously or consciously use various expressions to tell contents of the communication in detail.

On the otherhand, many researches have tried to enrich our communication by new methods for sharing information parallel to on-line communications. There have been a research of communication sharing a real virtual space [1]. The usage of avatars instead of the users is also discussed for enriched nonverbal communication in virtual space [2]. These are new methods for virtual communication and tele-presence depending on the emersive feelings of individuals. The modalities of the communication channels are expected to help the virtual communication.

Focusing on the nonverbal and multimodal channels that helps verbal communication, we have proposed an expressive method using a musical puppet system [3, 4] as a new channel in parallel with a conventional face-to-face communication. When we observed the communications between the holder of the system (player) and the participants who do not have the system (listener), the balance of the communication irregularly varied by the expression of the puppet system.

In order to estimate the factors for this irregular change, we suppose that the characteristics of the anthropomorphic media consist of *a) embodied presence* and *b) appearance*. Presence of the anthropomorphic media may cause the contribute the reality of anthropomorphism. Appearance with a face, body, or some other bodily parts like a human has an ability to make people to treat the presence as though the media were living matters. In this research, we focus on a non-face-to-face communication using anthropomorphic media.

IP RobotPhone [5] is a tele-communication medium for gestural expressions of a stuffed-toy robot. This system was designed as communication assistance parallel to conventional speech communication in IP phone. Here, we aim to clarify the effects of both factors, presence and appearance, of anthropomorphic media in human-human non-face-to-face communications using an anthropomorphic device. This paper introduces our experiment of the difference effects of a) the appearace and b) the embodied presence of anthropomorphic media in parallel to non-face-to-face communication with the other person. Based on the comparisons using the analyses of both subjective evaluations and behavioral observations, we discuss the different types of the effect on the user's expression.

2 Related Research

There have been many researches on social robots in these two decades. Robovie system [6] was designed to research social reactions of the humanoid robot. Several pet robots [7, etc.] were developed for com-

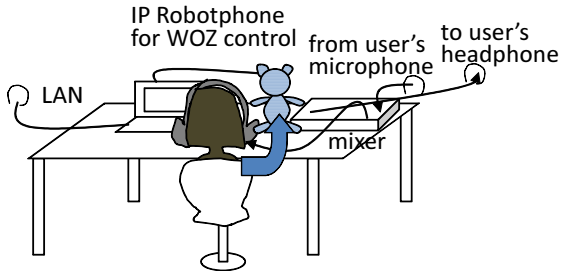

Figure 1: Participant-side Experimental System

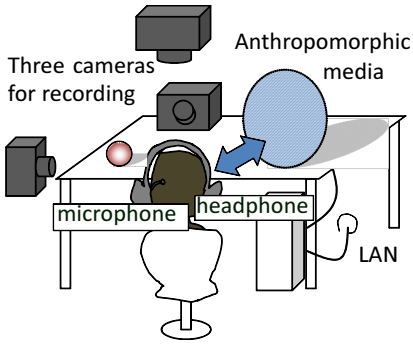

Figure 2: Experimenter-side System

Figure 3: View of Experiment (puppet)

fortable and relieved situations in our lives. These robot systems are focusing on basic tactile interactions in social relationships.

On the other hand, there were studies to adopt the anthropomorphic media for education and guidance. Actimate-burney [8] is an interactive education robot connected to video contents. There was a museum guide robot for intuitive and natural guidance [9]. Johnson et al. [10] introduced an interface robot for virtual space. We regard the anthorpomophic media help our communication in various shape and proposed a robot system which helps human-human conversation in videophone [11]. Katagiri et al mentioned [12] that different behaviors of an agent differently affect the user's knowledge and performance. Thus, the anthropomorphic agent systems are expected to develop our communication and knowledges.

Not only as intuitive media, but also as a motion medium, RobotPhone [5] had been introduced for gestural and motion communication in parallel to IP phone. The motion of the robot was sent to the motion of the other robot to show various motions during conversation. The anthropomorphic robot was able to regard as both the avatar of the user and the agent of the other person.

We had proposed Com-music system [3], a musical communication puppet, and observed user's behaviors using the system in parallel to face-to-face communications [4]. The results showed irregular effects just by holding a small puppet. In this paper, we focus on appearance and embodied presence of a stuffed-toy robot to clarify the effects by each factor. Ono et al. [13] and their experimental trials have shown some quantitative observations of human and robot interaction by using motion capture systems. Consequently, we observed the number and duration of each participant's behavior and her/his utterances.

3 Experiment on Parallel Communication using Robot

To clarify the effects of appearance and presence of anthropomorphic media, we conducted an experiment in parallel with speech communication in IP Phone. We adopted IPRobotPhone with two degree of freedom for the head, left and right hand the robot.

Hypotheses: I) The appearance of the stuffed-toy robot affects on the non-face-to-face speech communication. II) The embodied presence of the stuffed-toy robot affects on the non-face-to-face speech communication.

Participants: Twenty-one people (six females and fifteen males) aged from twenties to early thirties.

Conditions: To verify the anthropomorphic presence, we prepared a presence in a monitor to remove the factor of the embodied presence. To verify the appearance, we prepared a bare robot to remove the factor of the anthropomorphic appearance of the stuffed-toy robot. The three conditions as follows were designed as between-subjects factor. Condition 1) `puppet`: The stuffed-toy robot was used as an anthropomorphic medium. Condition 2) `monitor`: A monitor displaying a stuffed-toy robot was used as an anthropomorphic medium. Condition 3) `robot`: A bare robot without any stuffed-toy cover was used as an anthropomorphic medium.

Procedures: The participants were instructed to communicate with the other person through IP phone. They were also informed that the motions of a robot were controlled by the other person. We adopted a Wizard-of-Oz method to refrain from the effects of the interaction in human-human communication. In other words, "the other person" for the participants indicated some other person, but "the other person" for the experimenter indicated himself.

A prepared scenario with voices (see Figure 1) and recorded reactions (see Figure 2) of the robot was playbacked with precise timings.

Each participant sit in front of the anthropomorphic medium in each conditions. Figure 1 shows the settings of the experiment for the participants. The experimenter controlled both timings of the motion

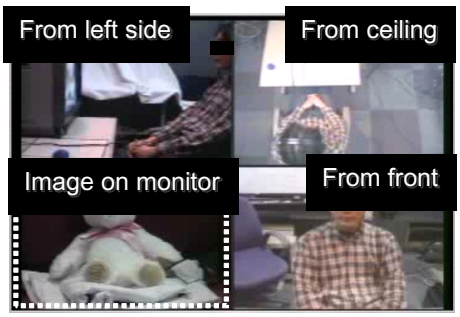

Figure 4: View of Experiment (monitor)

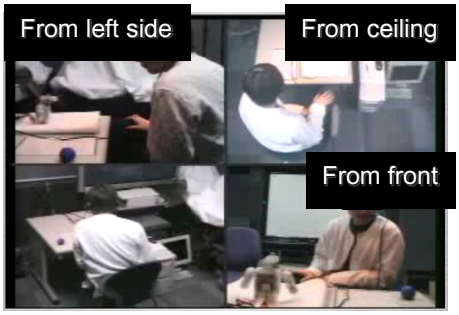

Figure 5: View of Experiment (robot)

and the speech of the robot in the experimental system as shown in Figure 2. The robot's speech (RS) was created by the change of F_0 from real speeches. The contents of the scenario were displayed to the participants in advance to roughly understand the topic. The participants were instructed to reply to the other person in the communication without any limitation.

To avoid the actuation sound from the robot, the participant used a headset to listen and talk to the other person. The experimental views for each condition are shown in Figures 3 to 5.

Instructions: At the beginng of the experiment, we instructed the participants as follows. "You are going to talk with your friend whose voice is changed. She/he will control the anthropomorphic object in front of you during the conversation." After the first instruction, the experimenter showed the dialogs of the other person in the scenario (Figure 1) and instructed as follows. "The conversation is limited to this scenario, however, you can ad-libs the dialog in your turn."

Table 1: Dialogs in the scenario

01: Now I came back.
02: What's up recently?
03: Yeah-yeah.
04: Well, is that blue ball soft?
05: That ball is this size, but I wanna get *this* size.
06: I'm enjoying internet life!!
07: I found that [Website name] is not good for brain.
08: Yup yup.
09: Thank you for talking with me.
10: Let's talk again in this style.

Table 2: Motions of each dialog in the scenario

01-M: put up its hands
02-M: incline its head
03-M: make a nod
04-M: indicate by its left hand
05-M: show different sizes by itshands
06-M: make a nod
07-M: swing its arms between front and back
08-M: make two nods
09-M: put up its right hand
10-M: swing its arms between front and back

Table 3: Labeled behaviors

B-1) Gaze of the object	
[look]	at the anthropomorphic medium
[look-ball]	looking at the ball on the desk
[switch]	counts of the gaze turns
B-2) Facial expressions	
[smiling]	smiling or laughing
B-3) Gestures and motions	
[hands-wave]	waving or putting up her/his hand
[mimicry]	mimic the robot's motion
[nod]	make a nod
[head-incline]	head motions except nods
[make-a- bow]	making a bow
[body-incline]	body motions except bows
[touch-ball]	touching the ball

Observation of for Labelings: We recorded the behaviors of the participants by three cameras and microphone as shown in Figure 1. The recorded movie data were checked and labeled from the viewpoint of the behaviors as follows by an observer using WaveSurfer[14] as shown in Figure 6.

A): Labeling of Utterances The timings of the participants' utterances were labeled for each step of the scenario. The speech of the utterances were dictated in text files. From the labeled data, we calculated 1) the duration of the utterance, 2) the delay of the utterance from the robot's dialog, 3) the moras of the utterance for each utterance. We did not give a different label for filler utterances to treat them as short answers.

B): Labeling of Behaviors The behaviors of the participants were marked the begin and end time with counting the times of the behaviors for each utterance as showin in Table 3.

C): Subjective Evaluations We asked the participants to answer to a questionnaire for three items: QA) whether the participant imagined the other person, QB) whether the participant felt affection to the anthropomorphic object, and QC) whether the participant strongly imagined the other person rather than the object.

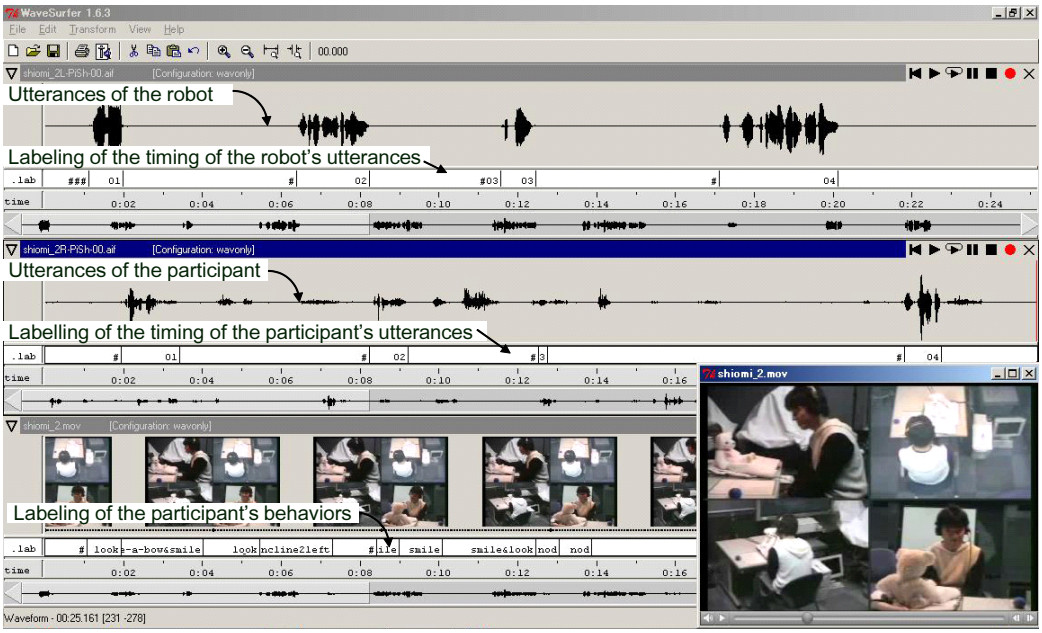

Figure 6: Labeling of User's Behaviors

Table 4: Answer patterns in subjective evaluations

	QA&QB	only QA	only QB	QC	all
puppet	3	2	3	1	8
monitor	0	1	5	3	7
robot	0	3	3	1	6

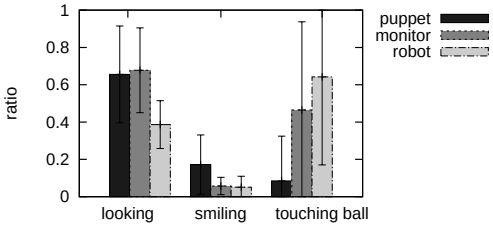

Figure 7: User's Behaviors during Experiment

4 Analyses of Results

4.1 Examples of Observed Behaviors

We could observe some participants talked with waving the hand or looking at the robot. It is especially remarkable that the hand gestures were observed although the gestures are frequently appeared in face-to-face communications. The observation indicated that the possibility of the effect of the anthropomorphic media on the communication.

4.2 Results of Subjective Evaluations

All the evaluations were yes-no decisions. QA and QB are independent questions each other, and QC is the impressive comparison between the robot and the other person. The summary of the combination of QA and QB is shown in Table 4. There were three participants who voted "Yes" to both QA and QB in the puppet condition, although there was no participant in the other conditions. In the monitor condition, almost all participants answered "Yes" to QB. It is conjectured that the participant felt affection to both the other person and the robot in the puppet condition, and that the appearance of the robot even in the monitor could draw the participant's interest.

The correlation coefficient betwen QA and QB was -0.61. From the result, it is presumed that the participants had a consciousness to either the robot or the other person. In the robot condition, three participants had a consciousness to the other person rather than the robot. All the three participants answerd "No" to QA and "Yes" to QB. It is assumed that the robot condition made these participants concentrate on the object.

The free descriptions of the experiment are shown in Appendix A. There were the positive descriptions of nods and back-channel feedbacks in all conditions. There were two descriptions of the other person's voice for the monitor condition, and two for the monitor condition. There were some participants who preferred a real presense in the monitor condition.

4.3 Analyses of User's Behaviors

The labels of [look], [smiling], and [touch-ball] were used to analyse the participant's interest. The avarages of the normalized durations for each label are shown in Figure 7. We held analyses of variance (ANOVA) among three conditions with each duration. The results are shown in Table 5. There were significant tendencies of ANOVAs for each label with significant tendencies of the post-hoc tests using Tukey-Kramer.

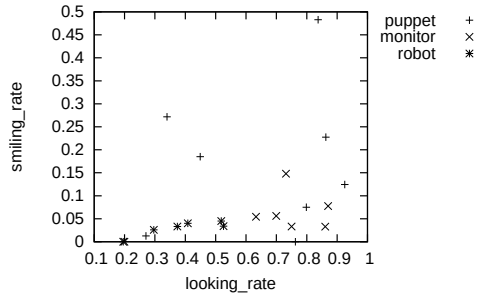

Figure 8: Propotions between "looking" and "smiling"

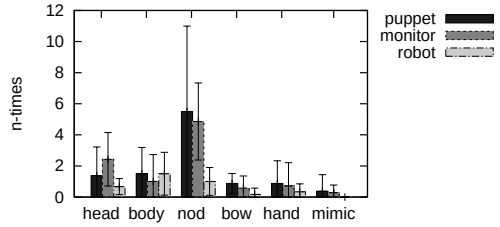

Figure 9: User's Motions during Experiment

The analyses show possibility of the different behaviors of the participants during the scenario especially between the puppet and robot conditions. Thus, Hypothesis I) was confirmed.

In the ratio of [touch-ball] and [smiling], there were significant differences. The number of the participants who touched the ball was different by the conditions. In the puppet condition, only one person in eight people touched the ball, and four in the seven and five in the six were also touched the ball after the dialog 04. After the dialog 04, the participants touched the ball regardless of the scenario. From the results, it is conjectured that the real presence decresed meaningless touches. Differently from the significant tendencyt in [look], the duration of [smiling] was significantly different between the puppet and robot conditions. The smiling expressions of the user seems to be appeared when the user faces to the real presence.

Figure 8 shows a schatter graph between [looking] and [smiling]. As can be seen, we could confirm that the results of "robot" gathered at the left-bottom and that the results of "monitor gathered at the right-bottom. The result indicates some possibilities of the different affects by appearance and presence.

Figure 9 shows the number of each behavior during the scenario. We could not find any significant result by large variances, but the puppet condition seemed to lead larger number of the user's gestures than the robot condition.

Figure 10 shows the number of the gaze turns during or without utterances either the other person's voice or the participant. The approximating expression and R^2 value for all samples were different from the expression and R^2 for the robot condition. The result shows that appearance of the puppet leads some stable rhythm of the conversation.

Table 5: ANOVA results for behavioral durations

	F-value	p-value	post-hoc test
looking	3.51	.052	none
smiling	3.94	.038	sig. in *puppet-robot*
touching	3.71	.045	sig. in *puppet-robot*

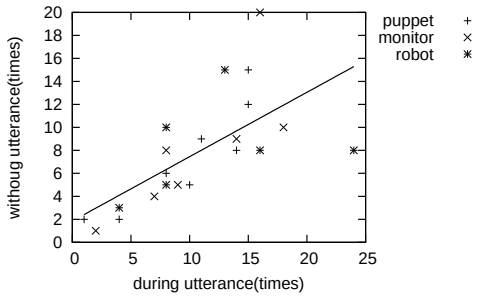

Figure 10: Switching of Gaze Target

4.4 Analyses of Utterances

We observed the duration, delay, and mora-number of the participants' utterances. The replied voices after 01, 09, 10 in Apendix A were summarized as the utterances for "*greeting*(`gre`)," the voices after 03 and 08 were summarized as "*back-channel feedbacks*(`fbak`)," the utterances after 02, 04, 07 were summarized as "*replies for question*(`que`)," and the utterances after 05, 06 were summarized as "*replies for insistence*(`insi`)." Figure 11 to 13 show the results for each category, and Table 6 to 8 show the results of ANOVA and the post-hoc tests.

From Figure 12 and Table 7, there was no significant differences among three conditions. The delays from the utterance from the other person expresses hesitation, discretion, etc [15]. It is conjectured that appearance and presence did not affect on the internal thinking of the participant.

The duration of utterances for "`que`" showed significance between monitor and robot. The number of mora in `que` converation was significantly different comparing puppet-robot and monitor-robot. The highest value was found in the monitor condition, and the presence could increase the participants' utterances. Thus Hypothesis II), the effect of the embodied presence, was confirmed.

Figure 14 shows a scatter graph of the duration between `que` and `insi` that are comparatively large amount. When we observed the data comparing with the puppet condition as a standard, the monitor condition frequently appears in right side. That means the participants in the condition did not replied to insistences of the other person.

From the analyses, we could confirm Hypotheses of the experiment with several findings.

5 Discussions

The subjective evaluations and the analyses of the observed data showed different effects of a) apppearance

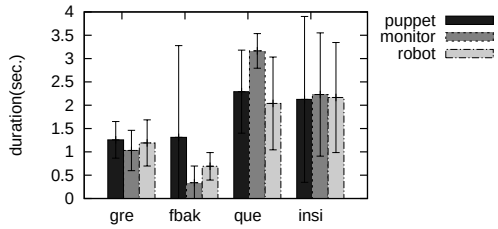

Figure 11: Durations of Utterances

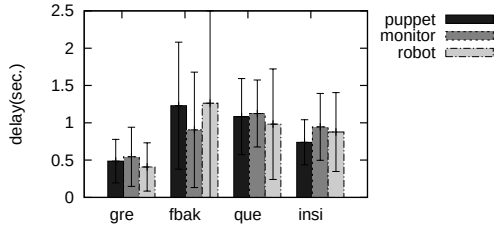

Figure 12: Delays of Utterances

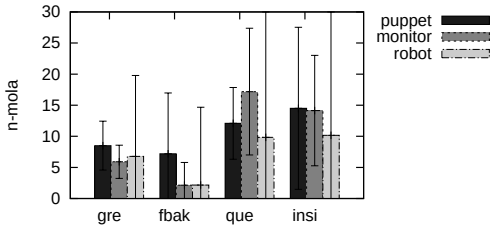

Figure 13: Moras of Utterances

Table 6: ANOVA results for durations of the participants' utterance

	F-value	p-value	post-hoc test
`gre`	0.52	0.60	none
`fbak`	1.16	0.34	none
`que`	3.75	0.044	sig. in *monitor-robot*
`insi`	0.078	0.93	none

and b) presence.

From the analyses of the subjective evaluations for the puppet condition, the participants had both familiar impressions to the stuffed-toy's appearance and the consciousness to the other person at the same time. The participants were significantly smiling in front of the stuffed-toy and did not touch the ball at their side. It is conjectured that the appearance of the stuffed-toy evlked the impresson as though they were in face-to-face communication with the other person. The bare robot reduced moras in the participants' utterances. From these results, it is supposed that the appearance of the stuffed-toy makes the user to enjoy talking with the other person.

On the other hand, the results of the subjective evaluations show that there could some possilibity to make the user to concentrate on the other person rather than the robot medium. Contrary to the result in the robot conditions, the participants' utterances were incresed for answering to the questions in the monitor condition. It is also presumed that the user proactively talked without the embodied presence of the anthropomorphic media.

Consequently, we could verify the effets of the appearace and embodied presence of the anthrpomorphic media on the under-conscious behaviors and utterances in the conversation. It should be important to emphasize and take advantage of the embodied presence and to carefully design the appearance when we build communicatoin systems using anthropomorphic media.

6 Conclusion

This research aimed to verify the effectiveness of appearance and embodied presence of anthropomorphic media. We conducted a conversational experiment by adopting IP RobotPhone. In this paper, we focused on a usage of the anthropomorphic medium and the user's conscious or under-conscious behaviors in parallel to non-face-to-face conversation. The analyses of the results showed that the appearance affects on the unconscious behaviors of the user and that the embodied presence affects on the conversational utterances. As future works, we investigate to give compositive labels among multiple behaviors of the participants. The importance of the auditory communication was indicated by some participants. We should discuss the design of the parallel communication with the other person and the anthropomorphic medium. In our experiment in this paper, we did not allow any chance to touch the robot during the conversation. The possibility of the embodied anthropomorphic presence should be discussed with tactile interactions.

Acknowledgements

This research was supported in part by Subsidy for Supporting Young Scholars in Kansai University and KAKENHI 25700021. The basis of the research achievement was supported in part by KAKENHI 20700106 and the National Institute of Information and Communications Technology of Japan when the author was in ATR IRC.

References

[1] Keiji Hirata, Toshihiro Takada, Yasunori Harada, Shigemi Aoyagi, Yasuhiro Yamamoto, and Kumiyo Nakakoji, An Approach to Expressing and Sharing Value in Collaboratively Producing Creative Knowledge Artifacts, *In Proc. of the 11th International Conference on Human-Computer Interaction (HCII)*, 2005.

[2] Daisaku Arita, Rin-ichiro Taniguchi, Presentation of Human Action Information via Avatar: From the Viewpoint of Avatar-Based Communication, *Knowledge-Based Intelligent Information*

Table 7: ANOVA results for delays of the participants' utterance

	F-value	p-value	post-hoc test
`gre`	0.26	0.77	none
`fbak`	0.17	0.85	none
`que`	0.11	0.90	none
`insi`	0.38	0.68	none

Table 8: ANOVA results for mora-number in the participants' utterance

	F-value	p-value	post-hoc test
`gre`	2.38	0.12	none
`fbak`	1.60	0.23	none
`que`	6.38	0.01	sig. in *puppet-robot* and *monitor-robot*
`insi`	0.48	0.62	none

and Engineering Systems Lecture Notes in Computer Science Vol 3683, pp. 883–889, 2005.

[3] Yonezawa, T., Clarkson, B., Yasumura, M., and Mase, K., Context-aware Sensor-doll as a Music Expression Device, *CHI2001 Extended Abstracts*, pp. 307–308, ACM SIGCHI, 2001.

[4] Yonezawa, T. and Mase, K., Musically Expressive Doll in Face-to-face Communication, *IEEE International Conference of Multimodal Interfaces Proceedings*, pp. 417-422, 2002.

[5] D. Sekiguchi, M.Inami, S. Tachi: RobotPHONE: RUI for Interpersonal Communication, *CHI2001 Extended Abstracts*, pp. 277–278, 2001.

[6] H. Ishiguro, T. Ono, M. Imai, T. Maeda, T. Kanda, R. Nakatsu, Robovie: an interactive humanoid robot, Int. J. Industrial Robot, Vol. 28, No. 6, pp.498–503, 2001.

[7] Takanori Shibata, Kazuyoshi Inoue, and Robert Irie, Emotional robot for intelligent system-artificial emotional creature project, *5th IEEE International Workshop on Robot and Human Communication*, pp.466–471, 1996.

[8] Strommen, E., When the Interface is a Talking Dinosaur: Lerning across Media with ActiMates Barney, *CHI98 Proceedings*, pp. 288-295, 1998.

[9] Wolfram Burgard, Armin B. Cremers, Dieter Fox, Dirk Hahnel, Gerhard Lakemeyer, Dirk Schulz, Walter Steiner, and Sebastian Thrun. Experiences with an interactive museum tour-guide robot. *Artificial Intelligence*, 114(1-2):3–55, 1999.

[10] Lewis Johnson, Jeff Rickel, Randy Stiles, and Allen Munro. Integrating pedagogical agents into virtual environments. *Presence: Teleoperators and Virtual Environments*, 7(Issue 6):523–546, 1998.

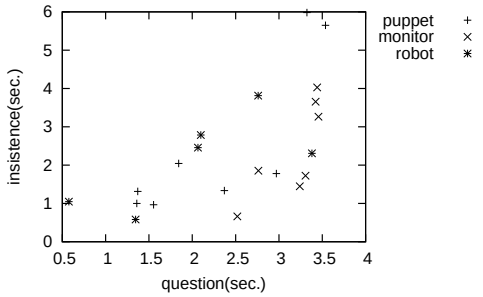

Figure 14: Utterance durations between question and insistence

[11] Tomoko Yonezawa, Hirotake Yamazoe, Akira Utsumi, and Shinji Abe, Assisting video communication by an intermediating robot system corresponding to each user's attitude, *Human Interface Society Journal*, Vol.13, No.3, pp.5–18, 2011.

[12] Y. Katagiri, T. Takahashi, and Y. Takeuchi. Social persuasion in human-agent interaction. *Second IJCAI Workshop on Knowledge and Reasoning in Practical Dialogue Systems*, IJCAI-2001:64–69, 2001.

[13] T. Ono, M. Imai, and H. Ishiguro, Embodied communications between humans and robots emerging from entrained gestures, *In Proc. CogSci 2001*, pp. 732–737, 2001.

[14] Sjlander, K., and Beskow, J., WaveSurfer, http://www.speech.kth.se/wavesurfer/

[15] Pomerantz, A., Pursuing a response. *In J. M. Atkinson and J. Heritage (Eds.), Structures of social action, Cambridge University Press*, pp. 152-164, 1984.

Appendix A: free descriptions

–puppet–
The participant unintentionally looked at the robot when the voice came.
Back channel feedbacks are important.
The puppet made familiar feeling to the communication.
The communication was easier than IP phones by the presence.
The participant preferred familiar friend or unknown person.
When the user naturalized to the system, it would be more enjoyable.
It was difficult to understand emotional gestures.
The participant wanted to see more gestures.
The participant hesitated to talk because she/he felt like a soliloquist.

–monitor–
The robot's nods were important.
Disconnectedness feeling decresed even he/she felt an artificial conversation.
The conversation style was sort of vague through a

monitor.
The participant preferred real presence rather than the image in a monitor.
The voice was like an animation character's one.
The participant became curious of the voice.

–robot–
The back-channel feedbacks were important.
The nods and gestures for showing the ball size was good.
The participant was not conscious to the other person.
The participant wanted to touch the robot and show to the other person.
The motions of the robot were scary.
Bigger motions were preferred.
The participant was conscious to the voice as a different presence of the robot.
The voice was cute.