Concepts and Applications of Human-Dependent Robots

Youssef Khaoula, Naoki Ohshima, P. Ravindra S. De Silva, and Michio Okada

Interactions and Communication Design Lab,
Toyohashi University of Technology, Toyohashi 441-8580, Japan
{youssef,ohshima}@icd.cs.tut.ac.jp,{ravi,okada}@tut.jp
http://www.icd.cs.tut.ac.jp/en/profile.html

Abstract. Relational artifacts (human-dependent) should have two aspects of subjective effects: Rorschach and evocative. During interaction, a robot has to anticipate the state (relationship) of the interactive person from the emotional to cognitive level to convey its Rorschach response. Consequently, the robot should behave as an evocative object to indicate the characteristic of animacy, which should accomplished using a potentially interactive architecture to coordinate the Rorschach and evocative effects. In this paper, we present two kinds of relational artifacts – a sociable trash box (STB) and a Talking-Ally.

Keywords: Relational artifacts, Rorschach and evocative effects, Hearership, Addressivity.

1 Introduction

As a newly emerging concept, previous studies have been concerned with developing a human-dependent robot [10][2]. Some research has defined a similar concept as either relational artifacts or a relatively operational robot. This concept can be defined as a robot being able to understand the situation of a human being (interactive user) in the relationship. The main attribute of the human-dependent robot (relational artifacts) is to convey its "attention" and "concern" of the interactive user by reading his/her state of mind to continue the interaction. A robot has to express attention and concern through stimulus interaction through its social cues (non-verbal behaviours, vocal interaction, etc.) by coordinating the cues of the interactive user [9].

Turkle [9] has discovered that relational artifacts (human-dependent robot) should have two aspects of subjective effects: Rorschach and evocative. During the interaction, if a robot can interpret/understand and express the user's relationship with the robot from an emotional to cognitive level then these can be defined as a Rorschach response. The second aspect addressed is that the robot should behave as an evocative object, which is concerned with how the robot reflects the aliveness (e.g., animacy), state of thought, and belief of defining status through its behaviours and interactions, while realizing the manners

S. Yamamoto (Ed.): HIMI 2014, Part II, LNCS 8522, pp. 435–444, 2014.

of the interactive person. This is due to the fact that a relational robot has to be proficient in behaving as an evocative object with the Rorschach effects.

Initially, few researchers were able to begin their work with relational artifacts (robots) from a different prospective. Okada [10][8] has been working on a child-dependent robot (relational artifacts) with a minimal designing concept to design a robotic platform to explore what are the essential factors in a child-dependent robot. Children are a better age group to explore the effectiveness of relational artifacts, because several studies have shown that children are able to naturally interact with robots as if they were interacting with their toys. Turkel [9] reported on a series of studies with a commercially available robotic (Paros, Furbys, etc.) platform by utilizing a group of the children to explore the essential factors in designing relational artifacts for medical and therapeutic vocations.

Fig. 1. Children interacting with the STB

Cynthia [3] developed an expressive anthropomorphic robot called Kismet that can engage with a natural facial expression, gaze direction, and vocal interactions. The study focused on developing a robot which is capable of engaging with human caregivers using natural social cues. The robot was capable of perceiving the state of human activities and appropriately responding to their social cues. Our focus here is on relational artifacts rather than on the frequencies and task outcome of human-robot interactions. In particular, we address the question of how to build up a novel active learning/interactive architecture to enhance the Rorschach and evocative capabilities of the robot. In addition, we explore how robots situated in a social relationship with the interactive user and how the interactive users are able to express their sense of relationship and their meaning through the interactions. In this paper, we introduce two kinds of relational artifacts (human dependent robot) called a social trash box (STB) robot and a Talking-Ally, with their concepts and applications.

2 Human-Dependent Robots

We used different perspectives in designing these two relational robots in a human-centric environment. A social trash box robot is a human-dependent robot that cannot collect the trash by itself, but conveys its intention to collect the trash from human assistance. Since the STB has to interpret/infer and express the user's relationship with the STB from at every stage of the interaction, it should demonstrate the characteristic of animacy to reflect aliveness in the interaction. In our second platform of Talking-Ally, we focus on developing the utterance generation mechanism to interact with the user in real time. Having the robot talk during the interaction is really challenging, as it strongly taps into the relational artifacts. By considering the aspects of Rorschach and evocative, we developed an utterance generation mechanism by considering hearership and addressivity.

3 Human-Dependent Sociable Trash Box (STB) Robot

Our main objective is to obtain child-assistance in collecting trash from a public space, while establishing a social interaction between the child and the robot (Figure 1). Our robot is capable of displaying manifold affiliation behaviors to build social rapport with the goal of collecting trash in and around an environment. In particular, the STB is a child-dependent robot that walks alone in a public space for the purpose of tracing humans and trash and collecting trash. In a crowded space, STBs move toward trash by engagement, using an attractive twisting motion (behaviors) and vocal interaction to convey its intentions to children. The robot is incapable of collecting trash by itself. In this sense, children have to infer the intentional stance of the robot or expectation for interaction with the STB. It is a novel concept to be able to collect trash while creating social rapport with children. The robot engages by using twisting and bowing motions as children place trash into an STB container. In order to collect the trash, each of the STBs communicates with one another to create a distance between one another to avoid collapsing.

3.1 Designing

We implemented the minimalism designing mechanism for STB robots (Figure 1). STB has two parts as its body (upper and lower), and the upper part contains three servomotors: one for twisting itself to the left and right, and the other two motors for bending forward and backward. The lower part has two servomotors for moving its entire body to the left and right directions. The STB contains three kinds of sensors and a single camera to obtain environmental informatics: a pyroelectric infrared sensor, an infrared ray sensor (IR sensor), and a distance sensor [10].

3.2 Experiment

To evaluate the effectiveness of the child-dependent robot (STB), we conducted an experiment using a natural setup of a child-centric environment (Developmental Center for Children) as a public space by using the five action scenarios; move toward trash (MT-I), communicate (electronically) with other STBs to move and create a distance with them (MT-G), move to a crowded space without communicating with other STBs (MC-WC), move to a crowded space to communicate with other STBs (MC-C), and STBs do not move and behave (NMB). When looking attentively at the videos, we categorized three main behaviors of the children: interest to the STB (Int-to-STB), indirect interaction (showing the interests far from the STB) with the STB (Ind-Int-STB), and state of collect the trash to the STB (St-Colt-Trash). The following Table 1 shows the behaviors of the children based on the STB action scenarios.

Table 1. Table depicted the child's behaviors are based on the STB's behaviors: PT (children place trash into the STB's container), and DPT (children have not placed trash into the container)

STB's Action Scenarios	Child's Behaviors		
	Int-to-STB	**Ind-Int-STB**	**St-Colt-Trash**
MT-I	PT=0, DPT= 9	PT=10, DPT= 0	PT=3, DPT=3
MT-G	PT=0, DPT= 6	PT=8, DPT= 5	PT=4, DPT=7
MC-WC	PT=0, DPT= 5	PT=3, DPT= 2	PT=6, DPT=8
MC-C	PT=0, DPT= 4	PT=2, DPT=13	PT=4, DPT=6
NMB	PT=0, DPT= 0	PT=1, DPT=3	PT=2, DPT=5

We attempt to link the trash box action scenarios with the children's behaviors (reactions or feedback) in the contexts of trash collection from the child assistance. In the present experiment, 108 children (between the ages of 4 and 11 years old) participated in naturally interacting with the STBs.

3.3 Moving vs. Immobile STBs

A chi-square test was employed to determine if either the STBs' movement or immobility (fixed as a typical trash box in the corner) was independent of the collection of trash. Each context was tested separately. The resulting p-value (χ^2=6.87, d.f.=1, p=0.009) was less than the critical p-value of 0.05 for the STB movement scenario, while in the case of the STB immobile scenario, the p-value (χ^2=1.31, df =1, p=0.252) was greater than the critical p-value of 0.05. The results therefore indicate that the null hypothesis can be rejected in the STB moving scenario and that a significant relationship exists between the two groups. However, in the STB immobile scenario, the null hypothesis cannot be rejected, indicating nonsignificant relationship between the two groups. The results of the above statistical method therefore reveal that the STB movement was essential in conveying its intentions toward collecting trash.

3.4 STBs Moving Direction toward to Trash vs. toward to People

In the former experiment, we discovered that the STBs' movements were most important in collecting the trash with child assistance. Accordingly, we have to reveal the direction of the STB, whether it is toward trash or toward people, in order to evoke its intentional stance in the children's minds. We therefore employed a chi-square method to evaluate the relationship between the trash (STB movement) and the trash collection from the children. The resulting p-value (χ^2=9.35, d.f.=1, p=0.002) was smaller than the critical p-value of 0.05. These results indicate that the null hypothesis can be rejected and that the two groups have a significant relationship. We applied a similar procedure for the context of STBs moving toward people, with a resulting p-value (χ^2=2.38, d.f.=1, p=0.123), which was greater than the critical p-value of 0.05, meaning the null hypothesis can't be rejected. This reveals that the STBs' movement toward the children did not correlate with the collection of the trash with child assistance.

3.5 STBs Moving (Interacting) as a Swarm vs. Individually

In this phase, we were interested in discovering whether the swarm behaviors (moving around public space as a group) or individual behaviors (moving around public space individually) were more effective in triggering the intentional stance of the STB in the children's minds. We considered the behaviors of the child subjects (Int-to-STB, Ind-Int-STB, and St-Colt-Trash) with the STB action scenarios of MT-I (individual behaviors) and MT-G (swarm behaviors). For this purpose, we employed a chi-square test to verify the relationship of the robot's demeanor (behavior as individual or group) and the trash collection via the child assistance. The results revealed that the p-value (χ^2=4.00, d.f.=1, p=0.046) was less than the critical p-value of 0.05, indicating the null hypothesis can be rejected and that the STBs' demeanor (individual or group) had a strong relationship with the trash collection. When carefully analyzing the contexts, we found that when the STBs moved in a group, many children (70%) interacted with the STBs compared with when the STBs moved individually (i.e., only 30% children participated to the interaction). In this sense, we believe that the group movement of the STBs more effectively helped to convey their intentions and to establish social rapport with the children than when the STBs moved around individually in the space.

4 Concept of Talking-Ally

The speaker refers to the hearer's behavioral information (nonverbal and vocal) to structure (organize) his/her utterance, and is also capable of dynamically aligning the structure of the utterances according to the resources (nonverbal and verbal) of the behavioral variation. Within a conversation in the interactions between hearer and speaker, the hearer is reacting to a speaker through nonverbal channels (e.g., attention coordinate, eye-gaze following etc) or a vocal response (e.g., back channel) toward prompting the interactions, which is defined as hearership in the conversation [4]. The concept of hearership is a resource (referring eye-gaze behaviors) for Talking-Ally to shape its utterance generation by considering the state of the hearership in dynamic interactions (Figure 2).

Bakthin [1] is suggested on the concept of addressivity, which can be defined as that through individual words can be directed toward someone, and then become completed utterances consisting "of one word or one sentence, and addressivity is inherent not in the unit of language, but in the utterance." The addressivity is a kind of never-ending communication that changes toward shaping the communication while adapting to the hearer's communication variations. The hearer influences the speaker's utterance generation mechanism, which is a prompt to adding/modifying sentences in order to coordinate a productive conversation [7]. Talking-Ally coordinates the addressee's eye-gaze behaviors (state of the hearership) to change the structure of the utterance generation (synchronized with bodily interactions) toward addressivity.

4.1 Design of Talking-Ally

We followed the minimal-designing concept to develop the Talking-Ally as depicted in Figure 2, which has three flexible points (head, neck, and torso) to generate bodily interactions with the user. All of its external appearance (body) is made with artificial wood, and its eyes and head are designed on the iPod visualizer. The face-lab is located on the table to track the user's eye gaze-behaviors in real-time. Talking-Ally has a voice synthesizer to generate an interactive conversation (in Japanese) by obtaining a news source (through RSS) in real-time while synchronizing it's bodily interaction (nodding, leaning it's body to the left and right, and eye-gaze is able to follow and look around the environment) through servo-motors [6].

The Utterance Generation. Talking-Ally is interactively disseminating the news from the web (through RSS) to the participant and simultaneously some exciting (sport-based) TV-program is broadcasted behind the robot to obtain attention variation from the addressees (participants). The manipulation of attention variation is utilized to obtain a variety of utterance generation patterns which can evaluate the performance of Talking-Ally. We employed a simple method to track the addressee's attention-region which is a primal reference to generate/adjust the robot's speech interactions. The robot decides the addressee's attention-region according the frequency scores of the eye-gaze behaviors in each region by considering them at every 60 frames as a segmentation point (states of the hearership). Parallel to the robot, the virtual-plane is constructed. The virtual plane is divided into six regions: two regions for Talking-Ally, two regions for room-environment (away from robot), and two regions for TV. The robot synchronizes the position of the eye-gaze coordinate with virtual-plane to determine the attention-region of the user.

Fig. 2. Talking-Ally is interaction with children

Based on the addressee's attention-region, Talking-Ally decides the relevant news sources, turn-initial, and entrust behaviors to generate its utterances. Table 2 lists the relevant robot's behaviors (non-verbal and utterance) in each region and execute randomly (region-wise) for speech interaction while synchronizing the bodily-interaction (process of the addressivity). Any change (variation) of the addressee's behaviors has the influence of changing its bodily interactions, attention-coordination, and structure of utterance in the dynamic adaptation unit. The whole process continually concatenates toward getting-back/keeping the addressee's attention (influences) by changing the structure of utterances to enhance the degree of communicative persuasion of Talking-Ally.

This study mainly focuses on exploring the performance of an utterance generation mechanism in order to enhance the persuasive power of the robot's communication and the effectiveness of the communication (naturalism of robot's communication) while Talking-Ally interactively disseminates exciting news from the web. Our study is mainly concerned with the dynamic interactive history of the robot (utterance generation/adaptation and non-verbal communication) and addressee (attention behaviors/adaptation through eye gaze behaviors) to evaluate the above performance.

4.2 Experimental Protocol

A total of 14 participants (aged between 20 and 24 years) were involved in the experiment in four separate sessions in which the conditions of the robot (interactions) were changed as follows: *A*-(attention-coordination (−), turn-initial and entrust behavior (−)), *B*-(attention-coordination (−), turn-initial and entrust behavior (+, random)), *C*-(attention-coordination (+), turn-initial and entrust behavior (+, random)), and *D*-(attention-coordination (+), turn-initial and entrust behavior (+)). The (−) indicates that robot did not considered these

Table 2. The virtual plane is divided into 6 regions (AG1, AG2, AG3, AG4, AG5, and AG6); at that particular time the robot modifies its utterance by considering the eleven types of turn-initials (TI) or six types of entrust-behaviors (EB) while synchronizing its six kinds of bodily interactions (BHV)

Human behaviors (Attention regions)	Robot's bodily interaction	Resources for utterance generation	
		Turn-initials (indirect request)	Entrust behaviors (direct request)
AG1, AG2 (Space of Talking-Ally)	BHV1(Initial-position), BHV3(Nodding)	TI1: "a-a", TI2: "ano-", TI3: "anone", TI4: "anosa", TI5: "e-tto", TI6: "e-ttone", TI7: "etto", TI8: "etto-", TI9: "ne-ne", TI10: "ntto", TI11: "nttone"	–
AG3, AG4 (Looking around the environment)	BHV4(Trun left-side), BHV5(Trun right-side), BHV6(Look around)	TI1: "a-a", TI2: "ano-", TI3: "anone", TI4: "anosa", TI5: "e-tto", TI6: "e-ttone", TI7: "etto", TI8: "etto-", TI9: "ne-ne", TI10: "ntto", TI11: "nttone"	–
AG5, AG6 (Attention to the TV)	BHV2(Bending forward), BHV3(Nodding)	–	EB1: "kite ne"—get attention, EB2: "kite yo"—get attention, EB3: "kite yone"—get attention, EB4: "kotti mite ne"—get gaze-attention, EB5: "kotti mite yo"—get gaze-attention, EB6: "kotti-mite yone"—get gaze-attention

channels in the condition and the (+) sign indicates that the robot considered these channels in the interactions. All participants participated in four sessions (A, B, C, and D), and each of the sessions took approximately three minutes to complete.

4.3 Results

We might consider the experimental conditions of B and D because within condition B, the robot did not trace the addressee's attention (tracking the eye-gaze behaviors) but randomly executed the utterance generation (mixing with turn-initial and entrust-behaviors). The condition did plainly not consider the state of the hearership. But in condition D, the robot traced the addressee's attention (state of the hearership) to generate the utterances (mixing with turn-initial or entrust-behaviors) while synchronizing its bodily interaction (whole process of addressivity). By comparing B and D, we can extract the persuasiveness power of the robot when integrating both hearership and addressivity.

We have gathered the turn-initials or entrust-behavior of Talking-Ally and relevant addressee's attention behaviors during the interactions for all of the participants (number of times) for both conditions (Figure 3(left-side)). The robot used a turn-initial or entrust behavior which was quite higher than the number of times in condition B, and also proportionally increased the obtaining of the attention of the addressees with a percentage of 68%; but in condition D, the usage of filler or entrust-behavior was reduced and also started to increase the acquisition of attention of addressees with 73%.

The response time of the addressee was another worthwhile parameter to use in evaluating the power of the robot's communication, because a lower responsive-time significantly indicated the persuasiveness of the robot's communication – both the clearness of the communication and the degree of influence

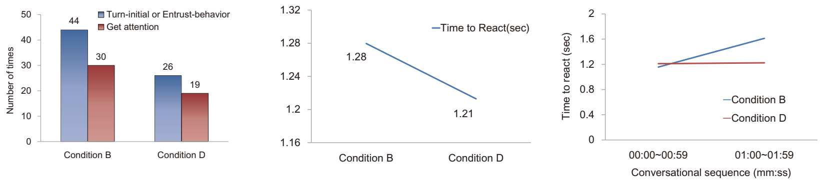

Fig. 3. Figure showed the selected addressee's responses (attention) within the selected segment for condition B and condition D. Here the addressees' attention-behaviors (red-color) when the robot utilized the turn-initial or entrust-behaviors (blue-color) during the interactive condition of B and D (left-side figure). The center figure shows the average responsive time within the segment and right-side of the figure depicted the responsive time within the selected segment according to the time interval.

of the communication, etc. [5]. Figure 3 (center and right-side) shows one of addressee's response (attention) times according to the robot's turn-initial or entrust-behaviors in the middle of the interaction (one of segmentations) that might be a perfect manifestation in comparing conditions B and D. The center figure shows the mean value of the responsive time within the selected segment. This indicates that the responsive time in D was lower than that of condition B, indicating that when we integrate hearership and addressivity, the addressee's response time begins to decrease. The right-side of Figure 3 shows an interesting pattern of responsive time. At some point in the segment, the addressee's response (attention) time in condition B suddenly increases, but the attention response time in interactive condition of D starts to decrease.

5 Conclusion

We have presented two relational artifacts which were developed using different perspectives. An STB is a human-dependent robot that collects the trash from a public space by conveying its intention. Movement is a main characteristic used in gaining the attention of people or animals. We believe from the above results that characteristic of attention-grabbing is gaining more consideration as an effective means to infer an object (e.g. robot) or people's behaviours. Another important aspect is the use of rich social cues (e.g., vocal interaction or twisting behaviours) based on the contexts, as these factors help to map various perceptions to infer someone's behaviours, e.g., intentional stance about an STB. We can examine similar perspectives by way of inferring a robot's behaviour in a child's mind. Accordingly, the results suggest that the STB movements and social cues directly correlated with the trash collection from the child assistance.

The results of the Talking-Ally showed that the resource of the hearer (state of hearership by tracing the addressee attention) was significant in generating/adjusting to the structure of the utterance generation mechanism (toward addressivity) to persuade the addressees. Additionally, the analysis of dynamic interaction showed that both the human and robot influenced each other's behaviors: the robot influenced the addressees' attention, and the humans influenced the robot in changing its utterance generation mechanism. The results of the subjective rating indicated that the robot recognized the participants as the hearer (life-likeness of robot), and the robot was capable of utterance generation and moving autonomously, which was vital in enhancing the characteristic of relational artifacts.

Acknowledgement. This research has been supported by both Grant-in-Aid for scientific research of KIBAN-B (21300083) and Grant-in-Aid for scientific research for HOUGA (24650053) from the Japan Society for the Promotion of Science (JSPS).

References

1. Bakhtin, M.: The problem of speech genres, pp. 60–102. University of Texas Press, Austin (1986)
2. Breazeal, C., Scassellati, B.: A context-dependent attention system for a social robot. In: IJCAI, pp. 1146–1153 (1999)
3. Brooks, R.A., Breazeal, C., Irie, R., Kemp, C.C., Marjanovic, M., Scassellati, B., Williamson, M.M.: Alternative essences of intelligence. In: AAAI/IAAI, pp. 961–968 (1998)
4. Goodwin, C.: Embodied hearers and speakers constructing talk and action in interaction. Belmont, CA: Wadsworth 16(1), 51–64 (2009)
5. Lutz, R., Swasy, J.L.: Integrating Cognitive Structure And Cognitive Response Approaches To Monitoring Communications Effects. Association for Consumer Research (1977)
6. Ohshima, N., Ohyama, Y., Odahara, Y., De Silva, P.R.S., Okada, M.: Talking-ally: Intended persuasiveness by utilizing hearership and addressivity. In: Ge, S.S., Khatib, O., Cabibihan, J.-J., Simmons, R., Williams, M.-A. (eds.) ICSR 2012. LNCS, vol. 7621, pp. 317–326. Springer, Heidelberg (2012)
7. Okada, M., Kurihara, S., Nakatsu, R.: Incremental elaboration in generating and interpreting spontaneous speech. In: Proc. of 3rd International Conference on Spoken Language Processing, pp. 103–106 (1994)
8. Okada, M., Sakamoto, S., Suzuki, N.: Muu: Artificial creatures as an embodied interface. In: 27th International Conference on Computer Graphics and Interactive Techniques (SIGGRAPH 2000), The Emerging Technologies: Point of Departure (2000)
9. Turkle, S., Taggart, W., Kidd, C.D., Dasté, O.: Relational artifacts with children and elders: the complexities of cybercompanionship. Connect. Sci. 18(4), 347–361 (2006)
10. Yamaji, Y., Miyake, T., Yoshiike, Y., Silva, P.R.S.D., Okada, M.: Stb: Child-dependent sociable trash box. I. J. Social Robotics 3(4), 359–370 (2011)