Talking Ally: Toward Persuasive Communication in Everyday Life

Yuki Odahara, Naoki Ohshima, P. Ravindra S. De Silva, and Michio Okada

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

Abstract. Exploiting a social intelligent agent (robot) for the embodiment of communication and interaction in a social and engaging manner is more exceptionally challenging than developing the basic capabilities of embedded robots (moving and acting). A robot should mainly have the capability of executing its communication capabilities within a social manner that is sufficient for establishing interaction with humans. The field of social robotics is mainly concerned with exploring desirable conventions embedded in social robots, which must contemplate and incorporate non-verbal communication. As such, we are developing a social robot (Talking-Ally) that is capable of liking the state of the person (addressee) through an utterance-generation mechanism (addressivity) that refers to the hearer's resources (hearership) in order to persuade the user through dynamic interactions.

Keywords: Mutually influences, Persuasive communication, Hearership, Addressivity.

1 Introduction

Humans are experts in adapting to different contexts through their social interaction. If we examine a human's daily conversation, we find that they mostly lengthen their conversation while adjusting to the hearer's behavioral information and to environmental changes. But how do they adjust their conversation? As we experience, humans employ entrust behaviors and turn-initial, etc., to change the direction of the conversation, and error-handling in the conversation, etc., is mostly utilized to continue the conversation by referring to the partner's behavioral changes [5].

As we thought, such a strategy or mechanism should be embedded into a social robot toward enhancing the persuasiveness of robot communication [4]. Many previous achievements [8][10][9] have addressed the persuasiveness of robotic communication by considering the design of powerful natural social cues in order to reduce the communication barriers in human-robot interactions. There is a great concern to adhere to important guidelines in the designing of persuasive robots [16][15]. A recent study by [2] reports on an experiment in which they

C. Stephanidis and M. Antona (Eds.): UAHCI/HCII 2013, Part II, LNCS 8010, pp. 394–403, 2013.

Fig. 1. Children interacting with Talking-Ally

explored how a robot persuades users by vocal and bodily interactions in the context of robotic suggestions within a Desert Survival Task. The study mainly focused on exploring the effectiveness of the above channels to measure the degree of the persuasiveness of robot, e.g., non-verbal communication.

Vocal interaction might play an important role in forming such persuasive robotics, but most previous work has considered establishing the vocal interaction through a sequence of utterances instead of dynamically generating a robot's utterance in real-time by referring to the hearer's behaviors [3]. Consequently, the robot is capable of changing its conversational direction by utilizing a conversational strategy as humans do to continue the conversation in any irregular time of the dynamic conversation. As such, we were motivated to integrate a similar utterance generation mechanism for Talking-Ally.

Therefore, in this study we developed Talking-Ally, which is capable of producing utterances (toward addressivity) by utilizing the state of the hearer's resources (eye-gaze information) to persuade the user (toward the hearership) through dynamic interactions. Talking-Ally stretches the conversation by connecting utterances by utilizing fillers and entrust behaviors, etc., while referring to the variation of the hearer's attentions. Our main focus is to explore how the utterance-generation mechanism enhances the persuasive power of the robot's communication and the effectiveness of the communication (naturalism of robot's communication) in human-robot interactions.

2 Organizing of the Talk-in-Interactions

The human talk-in-interaction mechanism should be profoundly explored in order to extract the essential components which can enhance a human's engagement. Human engagement and talk-in-interaction have a strong relationship in

their long-term interactions, but a simple rule is followed in their communication. What are these simple rules? Initially, a human indicates who is being addressed in the conversation by mentioning the person (name) or the contents of the conversation. The conversation is stretched while referring to the hearer's behaviors and responses. Also, the addresser uses attractive body gestures, facial expressions, and eye gaze behaviors while adjusting his/her vocal interaction. The addresser is continually synchronizing of all the affective channels to enhance the engagement while switching the direction of vocal interactions — changing the contents of the conversation, switching the direction of the conversation, and error handling of the conversation, etc. We believe the above talk-in-interaction mechanism should be embedded in the communication structure of social robotics.

3 Design of Persuasive Communication of Talking-Ally

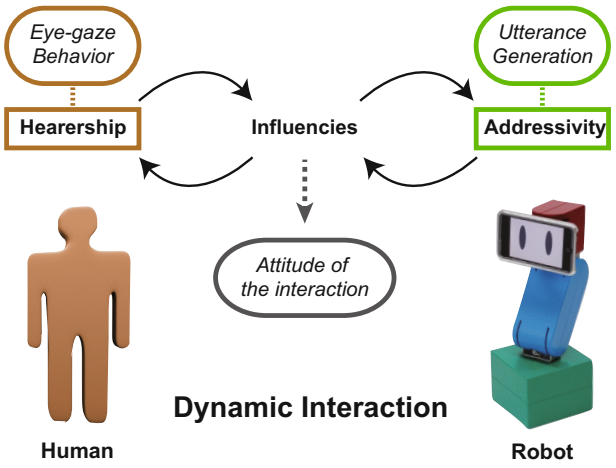


Fig. 2. Design of persuasive communication

Figure 2 depicts our proposed approach for persuasive communication of Talking-Ally by embedding it with the talk-in-interaction strategies as described in the previous section. Several studies have found that synchronizing of bodily gestures (non-verbal behaviors) in vocal interactions is mostly a potent strategy to transfer a speaker's message in an effective way toward a higher degree of persuasion. In addition, the addressee has to maintain eye contact with the hearer, which is important for obtaining his/her feedback and understanding how the hearer complies with speaker's message. These indicate the persuasiveness of the speaker's communication [12], as the Talking-Ally preserves vocal interaction while blending bodily interactions and attention coordination (eye-gaze behaviors) in dynamic interactions that are essential for talk-in-interaction.

The other important properties that influence each other's behaviors (persuasiveness) might depend on the experience during a significant interactive period, and influences are totally based on the contents of the conversation and method of delivering the message. Some degree in the strength of persuasiveness (both robot and human communication) is dependent on the series of communication chains (in dynamic interactions) to influence each other [6]. Also, the communication strategy of Talking-Ally might be capable of switching the hearer's attitude of interaction through a series of communication chains. Therefore, our main concern was to explore how robot and human behaviors (cues) mutually influence each other in different attributes of communication in dynamic interactions by implementing the utterance-generation mechanism (considering the addressivity and hearership) toward the intent of communicating persuasively as described in Figure 2.

3.1 Hearership and Addressivity

According to the Figure 2, 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 [13]. 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 [7]. 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.

Bakthin [1] is suggested on the concept of addressivity, which can be defined as that through which 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 [14]. 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 in the context of interactively disseminating exciting news from the web (through RSS).

4 The Design of Talking-Ally

Talking-Ally has been designed by following the minimal-standards to establish the interaction with the user. Moreover, all of its external appearance (body) is made with artificial wood, and its eyes and head are designed on the iPod

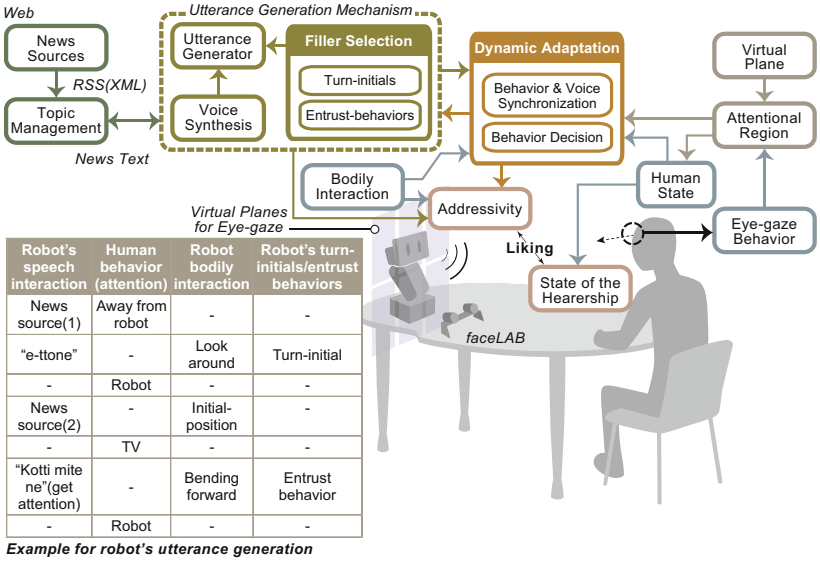
Fig. 3. The Design of Talking-Ally

visualizer Figure 3. 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 its bodily interaction (i.e., nodding, leaning the body to the left and right, and eye-gaze are able to follow and look around the environment) through servo-motors. Our main focus is to develop a robot that has embedded the social norm and skills of natural communication toward everyday life in human-robot interaction.

5 Generation of Smooth Flow of the Speech

To explore the performance of the robot's utterance generation, we have selected the context of Talking-Ally interactively disseminating the news from the web (through RSS). In addition to the robot, some exciting news (sport-based) is broadcasted through the display to obtain attention variation of the addressees. Talking-Ally detects the addressee's attention region according to the frequency score of eye-gaze behaviors by considering them at every 60 frames as segmentation while combining the defined virtual plane behind the robot. The virtual plane are divided into six regions: two regions for Talking-Ally (upper side (head), lower side (body)), two regions for room environment (away from robot), and two regions for TV. Talking-Ally estimates the position of addressees' attention by mapping the eye-gaze behaviors into a virtual plane (Figure 4).

The robot uses entrust behaviors and turn-initial based on the addressee's attention region while disseminating the news. As shown in Table 1, during the interactions Talking-Ally synchronizes its body-gestures with vocal interaction. Table 1 lists the robot's bodily interaction, turn-initials, and entrust behaviors



Robot's speech interaction	Human behavior (attention)	Robot bodily interaction	Robot's turn-initials/entrust behaviors
News source(1)	Away from robot	-	-
"e-ttone"	-	Look around	Turn-initial
-	Robot	-	-
News source(2)	-	Initial-position	-
-	TV	-	-
"Kotti mite ne"(get attention)	-	Bending forward	Entrust behavior
-	Robot	-	-

Example for robot's utterance generation

Fig. 4. Overall process of Talking-Ally and example of utterance generation

that are used according to the addressees' attention regions, and randomly selects the relevant turn-initial and entrust behaviors within the reach of attention region. Variation of 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 (generation mechanism) to enhance the degree of communicative persuasion of Talking-Ally.

Our motivation is to explore 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). 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 toward persuasive communication in everyday life.

6 Experimental Design

All of the participants participated in four sessions (A, B, C, and D), and each of the sessions took approximately three minutes to complete. At the end of each session, the participants had to answer questioners using a rating scale of $(1-5)$. 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

Table 1. Talking-Ally utilizing eleven type of turn-initial (TI) and six types of entrust-behaviors (EB) while synchronizing its six kinds of bodily interactions (BHV) according the addressees' attention regions (AG1, AG2, AG3, AG4, AG5, and AG6).

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

(−), 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 the robot did not consider these channels in the condition, and the (+) sign indicates that the robot considered these channels in the interactions.

7 Results

Evaluate the performance of utterance generation mechanism is considered in two ways: (1) dynamic interactive history of robot and addressees, and (2) a questionnaire-based subjective evaluation. We collected the data of the attention region of the addressees, the response time of the human and robot, the robot's behaviors (bodily interactions), relevant turn-initials, and entrust-behaviors in every interaction. The direction of the analysis was as follows: (1) explore persuasive power of robot communication (mechanism of utterance generation), and (2) subjective evaluation of the robot's communication. The direction of the analysis is totally based on the psychological evaluation (subjective rating to the questioners) and performance of robot utterance generation in real-time interactions.

7.1 Performance of Persuasive Communication

It might be important to explore how eye-gaze behaviors and the utterance generation mechanism play their roles to enhance the persuasiveness of robotic communication. Within the four conditions of our experiment, B and D are the best combination to evaluate the power of persuasiveness, 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 not plainly consider the state of the hearership. But in condition D, the robot traced the addressee's attention (state of the hearership) in generating the utterances (mixing with turn-initial or entrust-behaviors) while synchronizing its bodily interaction (whole process of addressivity).

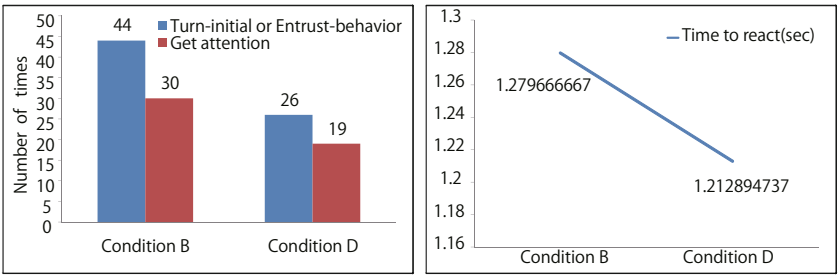


Fig. 5. The figure shows one of the addressee's responses (attention) within the selected segment for condition B and condition D. The right-hand of the figure shows the average responsive time within the segment.

During the experiment, the system collected 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 5(left-side)). The robot used a turn-initial or entrust behavior that was quite higher than the number of times in condition B, and it also proportionally increased its obtaining the attention of the addressees at 68%; however, in condition D, the usage of filler or entrust-behavior was reduced and also started to increase the acquisition of the attention of addressees to 73%.

The persuasive power of Talking-Ally communication can explore how the addressees' response time changes according to different conditions and also the clarity of the communication and degree of influence of the communication, etc. [11]. Figure 5(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 right side of the 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 hearership and addressivity are integrated, the addressee's response time begins to decrease.

7.2 Effectiveness of the Communication

Conducting a psychological experiment is useful to evaluate the participant's perception about the robot's communication. As such, subjective evaluations of

Table 2. After the end of all sessions (A, B, C, and D), the participants had to answer the following questions. A pair-wise comparison (using paired t-tests) was employed to explore the significant differences (based on rating scores) in each question.

Question	Contains	Results		
		Combination	t-value	p-value
Q1	I wanted to listen to the robot's speech more.	(A,B)	$t(13) = 2.2804$	$^{*}p = 0.04 < 0.05$
Q2	Robot recognized you as a hearer.	(A,B)	$t(13) = 2.3470$	$^{*}p = 0.035 < 0.05$
		(A,D)	$t(13) = 2.7378$	$^{*}p = 0.017 < 0.05$
Q3	I felt the life-likeness to the robot.	(A,B)	$t(13) = 2.9245$	$^{*}p = 0.012 < 0.05$
		(A,D)	$t(13) = 5.0902$	$^{**}p = 0.0002 < 0.01$
Q4	I felt a kind of intention to convey speech from the robot's utterances.	(A,B)	$t(13) = 4.8374$	$^{**}p = 0.0003 < 0.01$
		(A,C)	$t(13) = 2.2234$	$^{*}p = 0.045 < 0.05$
		(A,D)	$t(13) = 3.8894$	$^{**}p = 0.002 < 0.01$
Q5	I felt robot was speaking and moving autonomy.	(A,B)	$t(13) = 2.2234$	$^{*}p = 0.045 < 0.05$
		(A,D)	$t(13) = 2.6874$	$^{*}p = 0.019 < 0.05$

the questionnaire were administered to evaluate the effectiveness of the robot's communication: clearness of the utterance, neutrality of the speech, sense of autonomy for behavior, and utterance generation, etc. The analysis considered pair-wised (interactive condition of the robot) comparisons of each of the questions through a paired t-test by considering the subjective scores of each question. Table 2 shows the available significant differences of the pairs for each of the questions. In general, the (A,B) and (A, D) combinations revealed a significant difference, but the pair-wise differences of (A,D) had a higher rating score difference ($A < D$) than the combination of (A,B).

8 Results and Discussion

The section on the persuasive power of the communication 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) in persuading the addressees. Additionally, the analysis of dynamic interaction showed that both 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 persuasiveness of the robot's communication. The performance of our proposed approach is targeted toward persuasive communication in everyday life for Talking-Ally.

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. Chidambaram, V., Chiang, Y.-H., Mutlu, B.: Designing persuasive robots: how robots might persuade people using vocal and nonverbal cues. In: HRI, pp. 293–300 (2012)
3. D'Mello, S., McCauley, L., Markham, J.: A mechanism for human - robot interaction through informal voice commands. In: IEEE International Workshop on Robot and Human Interactive Communication, ROMAN 2005, pp. 184–189 (August 2005)
4. Fogg, B.: Persuasive computers: perspectives and research directions. In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, CHI 1998, pp. 225–232 (1998)
5. Ford, C.E., Thompson, S.A.: Interactional units in conversation: Syntactic, intonational, and pragmatic resources for the management of turns (1996)
6. Gibbs, R.W.: Embodiment and Cognitive Science. Cambridge University Press (2005)
7. Goodwin, C.: Embodied hearers and speakers constructing talk and action in interaction, vol. 16(1), pp. 51–64. Wadsworth, Belmont (2009)
8. Ham, J., Bokhorst, R., Cuijpers, R.H., van der Pol, D., Cabibihan, J.-J.: Making robots persuasive: The influence of combining persuasive strategies (gazing and gestures) by a storytelling robot on its persuasive power. In: ICSR, pp. 71–83 (2011)
9. Ham, J., Midden, C.: Ambient persuasive technology needs little cognitive effort: The differential effects of cognitive load on lighting feedback versus factual feedback. In: Ploug, T., Hasle, P., Oinas-Kukkonen, H. (eds.) PERSUASIVE 2010. LNCS, vol. 6137, pp. 132–142. Springer, Heidelberg (2010)
10. Looije, R., Neerincx, M.A., Cnossen, F.: Persuasive robotic assistant for health self-management of older adults: Design and evaluation of social behaviors. International Journal of Human-Computer Studies 68(6), 386–397 (2010)
11. Lutz, R., Swasy, J.L.: Integrating Cognitive Structure And Cognitive Response Approaches To Monitoring Communications Effects. Association for Consumer Research (1977)
12. Mehrabian, A., Williams, M.: Nonverbal concomitants of perceived and intended persuasiveness. Journal of Personality and Social Psychology 13(1), 37–58 (1969)
13. Okada, M.: Why doesn't the computer speak with hesitation. Kyoritsu Shuppan, Tokyo (1995)
14. 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)
15. Roubroeks, M.A.J., Ham, J., Midden, C.J.H.: When artificial social agents try to persuade people: The role of social agency on the occurrence of psychological reactance. I. J. Social Robotics 3(2), 155–165 (2011)
16. Siegel, M., Breazeal, C., Norton, M.I.: Persuasive robotics: The influence of robot gender on human behavior. In: IROS, pp. 2563–2568 (2009)