To Beep or Not to Beep Is Not the Whole Question

Kerstin Fischer[1], Lars Christian Jensen[1], and Leon Bodenhagen[2]

[1] University of Southern Denmark
Institute for Design and Communication
Alsion 2, 6400 Sonderborg
[2] University of Southern Denmark
Maersk-McKinney-Moller Institute
Niels Bohrs Allé, 5230 Odense M
kerstin@sdu.dk,
larje09@student.sdu.dk,
lebo@mmmi.sdu.dk

Abstract. In this paper, we address social effects of different mechanisms by means of which a robot can signal a person that it wants to pass. In the situation investigated, the robot attempts to pass by a busy, naïve participant who is blocking the way for the robot. The robot is a relatively large service robot, the Care-o-bot. Since speech melody has been found to fulfill social functions in human interactions, we investigate whether there is a difference in perceived politeness of the robot if the robot uses a beep sequence with rising versus with falling intonation, in comparison with no acoustic signal at all. The results of the experimental study (n=49) shows that approaching the person with a beep makes people more comfortable than without any sound, and that rising intonation contours make people feel more at ease than falling contours, especially women, who rate the robot that uses rising intonation contours as friendlier and warmer. The exact form of robot output thus matters.

Keywords: Human-robot interaction, attention getting, acoustic signals, social spaces, intonation.

1 Introduction

As robots increasingly leave the labs and become more present in everyday situations, the need to ensure that they interact smoothly with naïve, unsuspecting users increases as well. For instance, if a service robot drives around in a care institution, it will encounter doctors, nurses, patients, as well as visitors, relatives, and service personnel. One of the possible encounters is when the robot drives along a corridor or narrow space and approaches a person who does not see it coming.

In similar situations between humans, when a person is unknowingly blocking the way for another, various techniques are used, and the practices employed may differ slightly according to the cultural background of the participants. A big role is played by eye contact, by means of which people indicate to each other that they perceive each other and also in which direction they are heading [7]. In case one person approaches another from behind, speech and body contact are used

M. Beetz et al. (Eds.): ICSR 2014, LNAI 8755, pp. 156–165, 2014.

to different degrees in different cultures, such as saying "excuse me," tapping the other on the shoulder or pushing him or her aside a little. A big concern in these kinds of encounters is politeness; safety is not much at issue as long as people are moving 'normally', but generally people invest some interactional effort into creating and maintaining social relationships in encounters like the one under consideration. Thus, they will not only try to reach their goal, to be able to pass, but they will also attend to politeness [1], for instance, by minimizing the imposition (e.g. "could you move just a tiny bit to the side?") or by asking instead of ordering (e.g. "may I possibly pass behind you?"). Intonation plays a role in such situations since especially rising intonation is associated with openness towards the other and thus potentially with politeness [15].

To sum up, in interactions between humans, encounters in which one participant needs the other's collaboration in order to pass through constitute interactional problems that are solved with interactional effort especially regarding the maintenance of politeness. The question addressed in this study is how a robot could initiate such an encounter, and what kinds of behaviors are suitable to a) draw a person's attention to its presence and b) to communicate its intention to pass through in socially acceptable ways.

2 Previous Work

Few studies have so far systematically addressed how robots can get a person's attention when the person is occupied; robots are still generally confined to laboratory settings, or, if they are brought 'into the wild', then these are mostly contexts in which people are in exploration mood and not involved in particular activities themselves, like, for instance, in museums or at science or technology exhibitions. Accordingly, there is as yet not much knowledge on how a robot can get a busy user's attention and possibly even ask for a favor. One such study is Hüttenrauch and Severinson-Eklundh [6] who had a robot request naïve participants for help. The authors argue that acoustic signals, such as beeps, are effective means to get people's attention. Their results also show that whether people are engaged in an activity themselves has a considerable impact on their willingness to attend to the robot.

Furthermore, several studies by Sidner and colleagues (e.g. [12,13]) address how a zoomorphic robot can initiate interactions with people. They distinguish between different situations depending on whether users are already perceived to be attending, perceived to be non-attending or only suspected to be present. The authors report that methods based on eye-gaze may not always be effective [12] and suggest to combine the robot's initiative with an acoustic signal (speech) or a gesture.

In a recent study, Fischer and colleagues [2] investigated attention getting by means of speech in comparison to attention getting by means of a beep. We found that busy people may react only reluctantly to the robot's gesture and may not respond to a robot beeping at all; in contrast, all participants reacted to the robot's use of speech by looking at the robot.

To sum up, these studies suggest that an acoustic signal may be more effective than just the use of gaze and gesture to get a person's attention, yet that social signals such as speech may still be more effective than a beep sequence – this, in turn, suggests that socially relevant forms such as intonation contours may influence the social perception of beep sequences produced by a robot. Since a robot that uses speech implicitly suggests that it also understands speech, a beep sequence may still be preferable, and it is open whether the melody of the beep sequence may turn the acoustic signal into a social signal as well.

A second area of previous work concerns research on comfortable distance and direction of approach; one such study is Walters et al. [16], which addresses the role of the direction from which the robot should approach. The authors find that people do not like the robot to approach from the back and prefer it to come from either side. Furthermore, they find that people's degree of comfort when the robot approaches also depends on whether they are sitting or standing. Concerning distance, Yasumotu et al. [17] investigate people's preferences regarding the distance to the humanoid robot Asimo. Participants were asked to indicate when the robot should not come closer, which on the average was at about 78cm. However, the authors tested only trajectories towards the person from the front.

Mumm and Mutlu [9] investigate the effects of likeability of the robot (manipulated by means of polite versus rude introductory statements by the robot) and eye gaze (mutual versus averted) on the physical distance by means of which participants move around the robot. They find that in the mutual gaze condition, participants increased the distance to the robot; similarly, if participants disliked the robot, they also increased the physical distance. However, if participants liked the robot, gaze had no influence on physical distance. These results correspond to earlier findings by Takayama and Pantofaru [14] who report negative attitudes towards robots as well as the personality trait of neuroticism to be the best predictors of the distance people place themselves in when interacting with a robot. Furthermore, experience with robots and pet-ownership make people decrease the distance to a robot.

All of these studies show that distance is determined by a complex set of variables, and that the robot should be equipped with capabilities to negotiate how close it is allowed to come. However, none of these studies report on how close the robot should come when approaching a person from behind and when it actually needs to pass through. Furthermore, whether the distance is mitigated by acoustic signals is also open.

Finally, there is some relevant previous work on the kinds of sounds robots should produce. In particular, Read and Belpaeme [10] put a beep sequence into different interactional contexts, and people understood the same beep sequence as having very different meanings, depending entirely on the context in which it occurred. This finding suggests that a beep series itself should not have an effect on how people perceive a robot. However, if a robot approaches from behind, some kind of acoustic signal may be necessary to warn a user that a robot is approaching since other modes of communication are not available (e.g. gesture or eye contact, see, for instance, [5] and [12]). While speech seems to be the most effective strategy (see [2]), previous work on intonation suggests a considerable role of intonation contours, i.e. the melody of an utterance or beep sequence. For instance, Tench (1996: 105) summarizes: "a fall indicates the speaker's dominance in knowing and telling something, in telling someone what to do, and in expressing their own feelings; a rise indicates a speaker's deference to the addressee's knowledge, their right to decide, and their feelings" [15]. If these cues are consistently relevant in natural language exchanges between people, it can be expected that they will have an effect on how people will respond to robot beeps as well.

3 Hypotheses

The aim of this study is to determine the effects of different kinds of robot behaviors when the robot is approaching a person from behind. In particular, we test the effects of beep sounds versus no sound at all. Given the effects of speech melody on people's perception of another person's politeness, we hypothesized that a robot that uses a beep sequence with rising intonation contour will be perceived as friendlier and more polite than a robot that uses a beep sequence with falling intonation contours.

4 Methods

The experiment is an empirical study of HRI in three conditions, comparing social effects of robot behavior without sound and with two different beep sequences as stimuli.

4.1 Stimuli

The stimuli were generated on the basis of natural language utterances: First, the phrase "excuse me, please" was recorded with two different intonation contours, rising and falling. Since rising intonation contour generally signals openness and hearer involvement, in connection with the phrase "excuse me, please", it results in a very polite utterance that is best understood as a friendly request. Falling intonation contours are generally associated with statements and the assertion of facts. In connection with "excuse me, please," it serves rather to signal that a disturbance is going to take place. It is thus less polite.

In a second step, the natural language utterances of "excuse me, please" with the two different intonation contours were analyzed and the intonation contours used visualized using Praat.

Third, based on the intonation contours used in the two utterances, two beep series were created. In particular, to synthesize the beep sequence, a tone generator which is part of the free software package Audacity was used. The intonation contours of the two utterances of "excuse me, please" were simplified in such a way that they were taken as two sequences of four beeps, corresponding to each syllable, which are five semi-tones apart respectively, yielding  in the rising contour condition  in the falling contour condition.

4.2 Robot

The robot used is the Care-O-bot 3 (see Fig 3), developed by Graf and colleagues [3]. The robot is approximately 1.4 meters tall and composed of a 4-wheeled omnidirectional base, manipulator with a 3-finger gripper attached (SDH2) and a tray with 1 DOF that can be either in front of the robot or at its side. During the experiments, the configuration of the robot was kept constant, with the tray being in front, the torso being in an upright configuration and the gripper being at the side of the robot. All motions of the robot were controlled remotely by a hidden human operator.

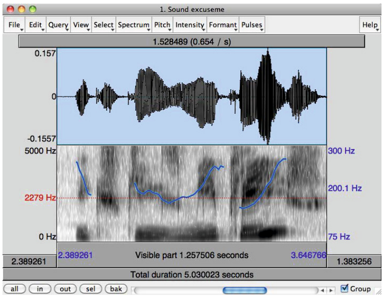

Fig. 1. The speech signal of 'excuse me, please' with rising intonation contour (blue line in the lower half of the window)

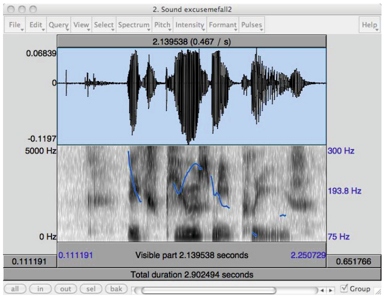

Fig. 2. The speech signal of 'excuse me, please' with falling intonation contour (blue line in the lower half of the window)

4.3 Procedure

Participants were greeted and asked to fill out a consent form in another room. Then, they were led into the experiment space and first asked to step in front of a white-board so that a picture of them with their participant number could be taken. While the participant was standing close to the white-board, the experimenter picked up a questionnaire and began asking the initial, demographic, questions of the questionnaire, positioning the participant implicitly with the back to the hallway through which the robot would have to pass. While she was asking the questions, the robot drove up behind the participant, who was facing the experimenter (see Fig 6). In conditions 2 and 3, the robot used one of two beep series when approaching (approximately 50cm before reaching the subject). In condition 1, the robot only attempted to drive through and relied on participants' peripheral vision or their perception of the robot's engine sounds to notice its presence.

Participants were videotaped for their behavioral responses if they had agreed to being recorded. Since some participants did not notice the robot when it was coming close, and some participants did not notice the beep, either, the experimenter, once the robot had passed, put the consent forms onto the robot's tray. This ensured that all participants had seen the robot before asked to fill out the questionnaire.

Then, they were handed the questionnaire the experimenter had begun to read to them, saying, "oh, actually you can fill this out yourself". After filling out the questionnaire, participants were asked to take part in a second study, at the end of which they were de-briefed.

4.4 Participants

We recruited 49 participants from the technical faculty of the University of Southern Denmark who were either undergraduates (49%), graduates (24.5%), faculty members (22.4%) or non-academics who had an affiliation with the university (4.1%). Even though most participants were students between the ages 20 and 40 (81.6%) and with an overrepresentation of males (71.4%), most participants had only little previous experiences with robots. 38.8% had worked or

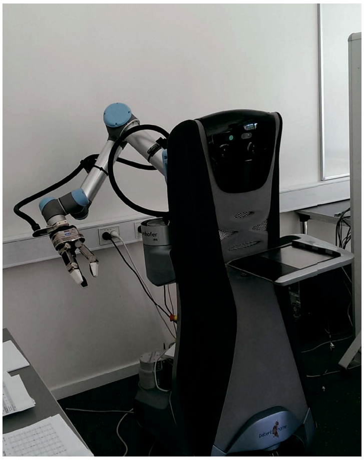

Fig. 3. The Care-O-Bot 3

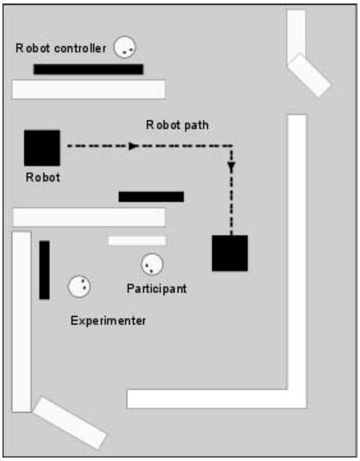

Fig. 4. Map of the experiment area

	Male	Female	Total
No Sound	11	4	15
Rising	11	6	17
Falling	12	4	16

Fig. 5. Participants per conditions

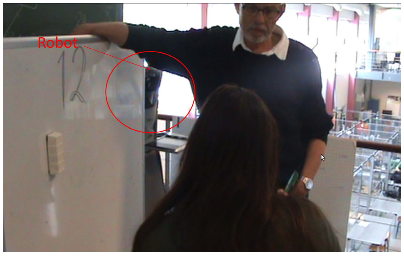

Fig. 6. The robot attempts to pass by

played with one, 36.6% had seen one or a few, 14.3% knew robots only from TV, while a minority (10.2%) worked with robots regularly.

4.5 Questionnaire

The first part of the questionnaire consisted of demographic questions and questions concerning people's prior experience with robots. This part was carried out as an interview until the robot had passed. The rest of the questionnaire concerned three aspects (based on the questionnaire developed in [4]): First, we asked for participants' perception of the robot's capabilities, politeness and other characteristics. Second, we asked to what degree people ascribe certain human characteristics to the robot. Third, we asked for participants' own feelings while they encountered the robot. People could mark their choices on a 7-point Likert scale.

4.6 Analysis

The statistical analysis of the questionnaire results (one- and two-way ANOVA) was carried out using the statistical software package SPSS. The behavioral data were analyzed according to when participants noticed the robot: when it was approaching, when it was close, or after it had passed.

5 Results

We first report the questionnaire results and then the results from the behavioral analysis.

5.1 Effect of Different Intonation Contours

There are no significant differences in the way the participants perceived the robot. That is, contrary to our expectations, the three conditions no sound, rising contour and falling contour, did not produce significantly different judgments of the robot's capabilities or politeness. In particular, participants were asked to rate the robot on a 7-point semantic differential scale on the following characteristics: Appeal, intelligence, competence, subordination/superiority, safety, approachability, confidence, friendliness and cooperativeness. None of the characteristics yielded any statistical differences between the conditions.

Similarly, participants do not generally ascribe different personality characteristics to the robot depending on the condition. Participants were asked to rate the robot on eight adjectives (cheerful, kind, likable, aggressive, assertive, bigheaded, harsh and rude) on a 7-point Likert scale ranging from 'describes poorly' (1) to 'describes very well' (7). Univariate ANOVA testing yielded one near-significant difference for the adjective *assertive*: no sound (M = 3.67, SD = 1.11), rising contour (M = 2.56, SD = 1.32) and falling contour (M = 3.20, SD = 1.32) [F(2,43) = 4.77, p = 0.059]. Tukey post-hoc pairwise comparison shows that this effect is due to a significant difference between the no sound and the rising intonation conditions.

However, participants' evaluation of their own states differs significantly between conditions: When asked to rate their own feelings on seven adjectives (angry, comfortable, cooperative, relaxed, uncomfortable, warm, afraid) on a 7-point Likert scale when the robot approached, statistically significant differences were observed for the degree with which they felt uncomfortable; no sound (M = 1.53, SD = 0.74), rising contour (M = 1.94, SD = 1.25) and falling contour (M = 2.88, SD = 2.00) [F(2,45) = 3.61, p = 0.035]. Post-hoc tests show that this effect is due to a significant difference between the no sound and the falling intonation conditions. Similarly, the reverse question concerning the degree with which they felt comfortable reaches near-significance; no sound (M = 4.53, SD = 1.89), rising contour (M = 5.29, SD = 1.61) and falling contour (M = 3.88, SD = 1.41) [F(2,45) = 3.10, p = 0.055]. In this case, Tukey post-hoc comparisons reveal that this effect is due to a significant difference between the rising and the falling intonation conditions.

5.2 Interpersonal Differences

Univariate ANOVA shows very different ratings depending on participants' gender: The robot was in general rated more intelligent by women (M = 4.38, SD = 0.77) than by men (M = 3.38, SD = 1.45) [F(1,43) = 5.58, p = 0.023]. We find similar differences for competency; men (M = 3.44, SD = 1.48) rated the robot less competent than women (M = 4.67, SD = 0.65) [F(1,42) = 7.63, p = 0.008]. Moreover, regarding subordination/superiority, men (M = 2.44, SD = 1.44) and women (M = 3.54, SD = 1.13) differ significantly [F(1,43) = 6.09, p = 0.018] such that women understand the robot to be more superior. In addition, regarding confidence, men (M = 3.56, SD = 1.56) ascribe less confidence to the robot than women (M = 4.54, SD = 1.13) [F(1,43) = 4.16, p = 0.048]. Furthermore, a comparison between men and women reveals that men (M = 2.4, SD = 1.7) find the situation more uncomfortable than women (M = 1.43, SD = 0.76) [F(1,46) = 4,52, p = 0.039]. Finally, women found the robot more aggressive (M = 2.36, SD = 1.74) than men (M = 1.48, SD = 0.87) [F(1,45) = 5.30, p = 0.026].

In addition to the gender differences, also differences for participants of different age were found. In particular, participants who were older were also more likely to be more afraid of the robot; participants below the age of 20 rated the robot as more friendly (M = 1.00, SD = 0.00) than participants between 20 and 40 (M = 1.41, SD = 0.82) and than participants between 40 and 60 (M = 2.67, SD = 2.34) [F(2,45) = 3.90, p = 0.027].

5.3 Interactions between Condition and Gender

We furthermore observe two significant interactions between gender and condition for likeability, such that female participants rate the robot significantly more likeable if the robot uses a beep with rising intonation contour (F(2,44) = 3.711, p =.033; see Fig 7). Similarly, women felt significantly less warmth if the robot does not produce any sound (F(2,46) = 5.698, p =.007; see Fig 8).

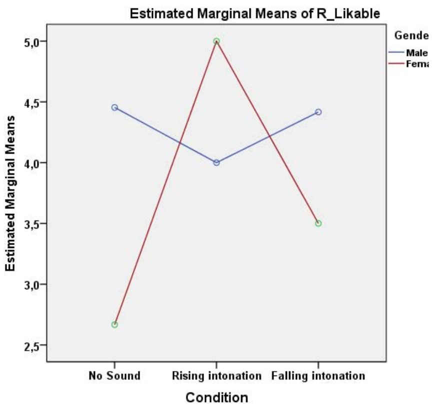

Fig. 7. Likeability by condition and gender

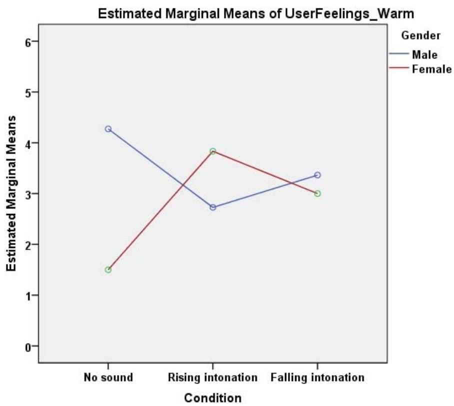

Fig. 8. Warmth by condition and gender

5.4 Behavioral Analysis

We analyzed the videos on the basis of obvious signs of attention to the robot, such as turning the head or stepping out of the way. The analysis[1] shows that 8 of the participants noticed the robot already when it was approaching, 15 noticed it when it was right behind them, and 14 only responded to it when it had passed. Whether the robot beeped or not did not influence people's attention when the robot was approaching such that even fewer participants noticed the robot approaching when it was beeping than when it played no sound. However, once the robot was close, more participants in the rising and falling intonation conditions noticed the robot than in the silent condition. This difference is marginally significant (Chi-square (6,N=40) = 12.197; p = .058).

6 Discussion

While beeps, irrespective of their intonation contour, do not seem to work well as attention getters (which is in accordance with previous findings [2]), our results suggest that people feel less uncomfortable around the robot if it approaches them using a beep than without producing any acoustic signal. Moreover, the results show that the melody of the beep sequence plays a role for participants' level of comfort: rising contours make people feel more at ease than falling contours. This is according to our predictions, which were based on natural language interactions between humans. However, the effects observed do not extend to the characteristics ascribed to the robot.

We observed consistent gender differences on participants' ratings of the robot's competence, characteristics and on their relationship with the robot. While gender effects have been observed in previous work (see, for instance, [8], [9], [14] and [11]), we were nevertheless rather surprised concerning the extent to which men and women were found to differ in the current study. Moreover, the effects of the different intonation contours seem to be different for men and women; in particular, female participants responded significantly more positively to the rising intonation contours than male participants.

7 Conclusion and Design Implications

We can conclude that the exact form of robot output matters, and be it only the melody of the robot's beep sequence. Especially the female participants took the robot's intonation contours into account regarding robot likeability and warmth towards the robot. Thus, while much work in HRI concerns different modalities in which interaction takes place, the actual form of such interactions may play a role regarding the social acceptability of a robot. The results of this study suggest that social aspects of acoustic human-robot interaction may be relevant and need to be attended to in robot behavior design, especially since societal development suggests that most of the potential users of service robots are likely to be women.

[1] We have video data on only 40 participants as not all consented to be recorded.

Acknowledgements. We are grateful to Franziska Kirstein, Maria V. aus der Wieschen, Maria Aarestrup, Nathalie Schümchen, Janni Jensen and Julia Ruser for their help during the experiments. This work was partially supported by the Danish project Patient@home which is funded as a strategic platform for innovation and research by the Danish Innovation Fond.

References

1. Brown, P., Levinson, S.: Politeness: Some Universals in Language Use. Cambridge University Press (1987)
2. Fischer, K., Soto, B., Pontafaru, C., Takayama, L.: The Effects of Social Framing on People's Responses to Robots' Requests for Help. In: Ro-man 2014, Edinburgh (2014)
3. Graf, B., Reiser, U., Hägele, M., Mauz, K., Klein, P.: Robotic home assistant Care-O-bot3 - product vision and innovation platform. In: IEEE Workshop on Advanced Robotics and its Social Impacts (ARSO), pp. 139–144 (2009)
4. Groom, V., Takayama, L., Ochi, P., Nass, C.: I am my robot: the impact of robot-building and robot form on operators. In: HRI 2009 (2009)
5. Hoque, M.M., Das, D., Onuki, T., Kobyashi, Y., Kuno, Y.: Model for Controlling a Target Human's Attention in Multi-Party Settings. In: Ro-man 2012, Paris, France, pp. 476–483 (2012)
6. Hüttenrauch, H., Severinson-Eklundh, K.: To Help or Not to Help a Service Robot. In: Ro-man 2003, Millbrae, CA, pp. 379–384 (2003)
7. Kendon, A.: Conducting Interaction: Patterns of Behavior in Focused Encounters. Cambridge University Press (1990)
8. Kriz, S., Anderson, G., Trafton, J.G.: Robot-directed speech: using language to assess first-time users' conceptualizations of a robot. In: HRI 2010, New York, NY, pp. 267–274 (2010)
9. Mumm, J., Mutlu, B.: Human-Robot Proxemics: Physical and Psychological Distancing in Human-Robot Interaction. In: HRI 2011, Lausanne, Switzerland (2011)
10. Read, R., Belpaeme, T.: Situational Context Directs How People Affectively Interpret Robotic Non-Linguistic Utterances. In: HRI 2014, Bielefeld, Germany (2014)
11. Schermerhorn, P., Scheutz, M., Crowell, C.R.: Robot social presence and gender: Do females view robots differently than males? In: HRI 2008, pp. 263–270 (2008)
12. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in Engagement for Humans and Robots. Artificial Intelligence 166(1-2), 140–164 (2005)
13. Sidner, C., Lee, C.: Attentional Gestures in Dialogues between People and Robots. In: Nishida, T. (ed.) Engineering Approaches to Conversational Informatics. Wiley and Sons (2007)
14. Takayama, L., Pantofaru, C.: Influences on proxemic behaviors in human-robot interaction. In: IROS 2009, pp. 5495–5502 (2009)
15. Tench, P.: The Intonation Systems of English. Cassell, London (1996)
16. Walters, M.L., Dautenhahn, K., Woods, S.N., Koay, K.L.: Robotic Etiquette: Results from User Studies Involving a Fetch and Carry Task. In: HRI 2007, pp. 317–324 (2007)
17. Yasumoto, M., Kamide, H., Mae, Y., Kawabe, K., Sigemi, S., Hirose, M., Arai, T.: Personal space of humans in relation with humanoid robots depending on the presentation method. In: IEEE/SICE International Symposium on System Integration, pp. 797–801 (2011)