Making Social Robots More Attractive: The Effects of Voice Pitch, Humor and Empathy

**Andreea Niculescu · Betsy van Dijk · Anton Nijholt ·
Haizhou Li · Swee Lan See**

Abstract In this paper we explore how simple auditory/ verbal features of the spoken language, such as voice characteristics (pitch) and language cues (empathy/humor expression) influence the quality of interaction with a social robot receptionist. For our experiment two robot characters were created: Olivia, the more extrovert, exuberant, and humorous robot with a higher voice pitch and Cynthia, the more introvert, calmer and more serious robot with a lower voice pitch. Our results showed that the voice pitch seemed to have a strong influence on the way users rated the overall interaction quality, as well as the robot's appeal and overall enjoyment. Further, the humor appeared to improve the users' perception of task enjoyment, robot personality and speaking style while the empathy showed effects on the way users evaluated the robot's receptive behavior and the interaction ease. With our study, we would like to stress in particular the importance of voice pitch in human robot interaction and to encourage further research on this topic.

A. Niculescu (✉) · B. van Dijk · A. Nijholt
Human Media Interaction, University of Twente,
7500 AE Enschede, The Netherlands
e-mail: niculescuai@utwente.com

B. van Dijk
e-mail: e.m.a.g.vandijk@utwente.com

A. Nijholt
e-mail: a.nijholt@utwente.com

H. Li · S.L. See
Institute for Infocomm Research, 1 Fusionopolis Way,
138632 Singapore, Singapore

H. Li
e-mail: hli@i2r.a-star.edu.sg

S.L. See
e-mail: slsee@i2r.a-star.edu.sg

1 Introduction

Since the technology advances in engineering and computer science of the last decade has brought the use of robots outside their traditional industrial 'playground' there has been a growing interest in designing socially competent robots for entertainment, educational purposes, health care assistance or as museum tour-guides. Especially, as receptionists and office companions social robots are becoming increasingly popular. For example, Wakamaru [1], a yellow midget robot, developed by Mitsubishi, can greet in four languages and is able to escort visitors to different destinations within a building. Wakamaru has been on the market since September 2005 and in the meantime has even found a 'job' at a temporary staffing agency in Japan as dispatch worker. Ms. Saya [1] is another humanoid robot that was 'employed' as a receptionist by Tokyo University of Science for a period of two years. Ms. Saya was able to do some basic conversation, guiding visitors to locations of interest and luring university graduates to sign up to courses. Other notable examples are Valerie [2], Tank [3] and Hala [4], all social robot receptionists developed within the Roboceptionist Project[1] at Carnegie Mellon University. These robots were placed at the entrance of the building giving directions, answering phone calls, looking up weather forecasts and giving information about local events, using English or even Arabic (Hala). Apart from their conversational capabilities, the robots represent different characters, each one having a different personal story revealed to visitors through daily interaction over a longer period of time. The purpose of these

---

[1]http://roboceptionist.org/project.htm.

robots was not just to inform, but also to be agreeable to the visitors, encouraging them to return in the future. This approach is adopted in real life by human receptionists who are responsible for giving customers a good first impression of the organization they represent.

What qualities are essential for being a good receptionist? Apart from being efficient, multitasking, flexible or computer literate, most of the on-line adverts for receptionist jobs emphasize excellent interpersonal communication and customer service skills as absolutely essential requirements. Thus, a receptionist should be able to communicate effectively with the visitors, to listen actively to their problems showing empathy and providing help, to be polite and friendly and to possess an appropriate sense of humor. Since the 'Computers Are Social Actors' (CASA) paradigm [5] proved that humans treat computers—and consequently robots—as social entities we believe that a good social robot receptionist would have to meet similar requirements. Therefore, in this paper we focus on two interpersonal skills relevant in interaction with a social robot receptionist: humor and empathy. We chose these two social abilities for our experiment on one side, because of their proven positive effects in human social interactions as promoting satisfying and healthy interpersonal relationships [6] and, on the other side, because they have not yet been studied in combination with a social robot in a receptionist scenario.

Since humor is related with more extrovert personalities we followed the Roboceptionist example and created two robot characters by manipulating the voice pitch: Olivia, the more extrovert, exuberant and humorous robot with a higher voice pitch and Cynthia, the introvert, calmer and more serious robot with a lower voice pitch. The manipulation allowed us to study the effects of an important element in the design of a robotic voice user interface: the voice pitch.

With this study we aim to determine how the manipulation of our auditory/verbal spoken language features influenced the ratings concerning the quality of interaction with the robot, how the voice pitch affected the user experience and whether empathy and humor can be considered prerequisites for a successful human-robot interaction. The paper is structured as follows: Sect. 2 offers an overview of relevant background work concerning the effects of voice pitch, humor, empathy and the socio-cultural experiment context. Section 3 presents the experiment design, setup and questionnaire design. Section 4 discusses the results and finally, Sect. 5 concludes the paper.

2 Related Work

2.1 Voice Pitch

In human social groups, sensitivity to voice and language cues has always played a critical role in the evolutionary history [7]. In face-to-face communication, the voice conveys the intended message and simultaneously contains highly relevant cues for social interactions. Such cues point to the speaker's gender, age, personality, emotional state or place of origin enabling socially intelligent individuals to decide who to like, who to trust and who to mate with.

A very important characteristic of the voice is the pitch referring to how high or low the voice is. The pitch is determined by the fundamental frequency. Average pitch values for male and female voices are respectively 120 Hz and 210 Hz [8].

The voice pitch has been found to have influence on the perceived attractiveness of a person. The study by Riding et al. [9] showed that women found men with high-pitched voices significantly less attractive than men with medium or low pitched voices; for female voices opposite results were found, that means high-pitched female voices were rated as significantly more attractive. Further, the study by Colin and Missing [10] showed that men found high-pitched female voices more attractive. According to Jones [11] women also show a preference for high-pitched female voices. Further, voices rated higher in attractiveness are associated with more favorable impressions of overall personality and even with a higher degree of similarity between participants and the rated voice [12].

A recent study by Helfrich and Weidenbecher [13] associating voice pitch with the retention of content in long term memory found that both high and low pitched voices led to better results than medium-pitched voices. This result was independent of whether the speech samples represented natural or manipulated voices.

Pitch, pitch range, volume, and speech rate are the four fundamental characteristics of the voice that indicate personality [14]. People who talk loudly, rapidly, in a high pitch and using a wide pitch range are most likely extrovert, while introvert people often speak slowly with a soft, deep and monotone voice [7]. These four voice characteristics appeared to be sufficient to model the personality of synthetic voices as shown in the studies of Nass and Brave [7] where test subjects identified the 'personality' behind the synthetic voice correctly and even used this knowledge to guide their feelings and behavior towards the product represented by the voice.

The voice pitch is also an indicator of a person's maturity [15]. For both female and male voices, a lower pitch was found to correspond to a higher maturity. A voice with a higher pitch, on the other hand, was perceived as more emotional and immature, indicating greater levels of emotional instability, weakness and psychological tension for both male and female voices [16]. Impression formation studies investigating the voice have found that more mature voices result in impressions of lower warmth and agreeableness [16] but higher dominance and assertiveness [17].

Surprisingly, research in the field of Human Robot Interaction (HRI) has hardly ever focused on the psychological effects of the voice and of the voice pitch on users' perceptions. One study by Walters et al. [18] investigated how people approached a robot when its voice was female, male, highly synthesized (robotic) or when no voice was involved. They found that approach distances were significantly higher when the voice was highly synthesized. Another study by Scheutz and Schermerhorn [19] investigated the effects of manipulating affect expression the robot's voice on task performance. The results showed that voice affect expression motivated people to perform better on joint tasks when the robot was present in the environment. Apart from these two studies there are no others known to the authors dealing with voice manipulation in HRI.

2.2 Humor

Humor is pervasive in human social relationships [20] and one of the most common ways to produce a positive affect in others. Research studies have shown that innocent humor increases likeability and interpersonal attraction [21, 22], boosts friendship and trust [23] and fosters social cohesion [24].

Humor can be spontaneous or be deliberately used not only in conversations with friends but also in more formal environments [25]. For example, the use of humor in teaching stimulates students' attention, promotes comprehension, retention of information and more positive feelings towards the content [26]. Also, in the workplace, humor seems to be a successful method to alleviate stress [27], reduce tensions, encourage creativity [28] and improve teamwork and socialization [29]. The presence of humor—which is apparently correlated with more expressive, self-confident and extrovert personalities [30]—was proven to increase trust in online negotiations [31]. Large corporations such as IBM, Kodak and AT&T hired humor experts to help improve teamwork, stimulating creativity and motivating employees [32].

Despite the positive effects of humor in work situations, scientifically proven by decades of research, the HCI field holds a rather negative view about the use of humor in interfaces. The general trend in interaction design is to develop interfaces that increase task performance on one side, and minimize task duration, learning time and error rate, on the other side [33]. Since the use of humor would distract the users from their tasks increasing the total completion time it would contradict HCI policies of maximizing efficiency in interaction.

However, an exception seems to be the artificial intelligence and natural language processing communities where computational humor is a well-established branch. Here, the research is split in two directions: the first one concerns the automatic generation and/or recognition of jokes and humor.

The progress in this direction is relatively slow due to the complexity of the humor phenomenon, which is highly reliant on the context. Some successful implementations of language based humor interfaces already exist—see [34] and [35]—but the development of fully automatic humorous machines that spontaneously produce jokes and react to humor might still take a while.

Studies belonging to the second research direction investigate the effects of pre-programmed humor in interaction with artificial entities. Morkes et al. [36] showed that participants who received humorous comments during the interaction with a computer, rated the system as more likeable, more competent and much more cooperative. They also smiled and laughed more, behaving much more socially during the experiment. Their experiment demonstrated that, contrary to the traditional HCI view, humor enhances the users' experience with the system without distracting them from their tasks. These findings also suggested positive effects of humor on the overall system usability.

The study of Huan and Szafir [37] investigated the effect of humor in an agent-based educational context. Their results showed that humor improved the instructor likeability significantly regardless of whether they were human or robotic. However, no effects were found on the content memorization.

Another study by Dybala et al. [38] performed with conversational agents showed that test users evaluated the humorous agent as more human-like, funny and likeable. In general, the humorous agent received much higher scores as compared with the non-humorous one.

Some other studies, such as the one by Babu et al. [39], suggest that the use of humor by a virtual receptionist agent may play a major role in engaging users in social conversations, since 50 % of the social conversations between users and the agent Marve contained jokes.

2.3 Empathy

Empathy is described as the capacity to react emotionally when perceiving another person experiencing, or about to experience an emotion [40]. Thus, expressing empathy involves taking perspective, understanding non-verbal cues, sensitivity to other's affective state, communicating feelings of care and desire to help appropriately [41].

Research regarding empathetic agents is divided into two categories: one concerns agents simulating empathetic behavior towards users, as is the case in our experiment; the other concerns agents fostering empathetic feelings in users [42].

Many studies in the past have shown the overall positive effects achieved by agents and robots expressing empathy towards users. The study by Brave et al. [43] showed that modeling empathetic emotion in an agent was found to

increase the positive ratings concerning its likeability and trustworthiness. Additionally, the empathetic agent was also perceived as more caring and more supportive.

Other studies have also reported that expression of empathy in agents reduces frustration and stress [44], increases user's satisfaction [45, 46], comfort [47], engagement [45] and performance in task achievement [48]. Pereira et al. [42] experimented with an iCat robot displaying empathetic and neutral behaviors towards chess players. The results showed that the chess player who received empathic comments perceived the robot more as a friend than the player who did not receive such comments.

On the other side, the studies by Cramer et al. [49] and Ochs et al. [50] showed that empathy expression produced positive effects only in situations when it correctly matched the users' affective state. Hence, incongruous emotional responses can lead to negative user rating of the agent.

Factors such as interaction environment, the subject's culture, interaction style or task type might play a role in the way empathy is perceived and reflected on the agent's ratings. However, a full overview of the empathy effects on user's perceptions of robots is not available yet [51].

2.4 Socio-cultural Experiment Context

Although humor and empathy are universal behaviors, their specific content is expected to vary according to social situations and cultural influences. Therefore, it is important when designing an experiment involving the manipulation of such interpersonal skills to take into account the social-cultural context in which the study is performed.

In our case the experiment was carried out in Singapore, one of the most westernized Asian countries. However, despite its modern lifestyle Singapore's eclectic society, dominated by a Chinese majority—74 % Chinese, 13 % Malay, 9 % Indians and 4 % expatriates—remains deeply anchored in its collectivistic traditions. According to the studies conducted by Hofstede [52] Singapore scored relatively high on power-distance and low on individualism which aligns the society with the Confucian teachings [53]. Hofstede refers to high power-distance as to the extent to which members of a society accept that power in institutions is unequally distributed among individuals. Further he describes individualism, as opposed to collectivism referring to a society in which the ties between the individuals are loose. Both power-distance and collectivism have implications on people's social behavior, including the expression of humor and empathy.

In a high power-distance society hierarchical relationships are strongly valued while deference, respect and formality towards superiors are considered the norm. Subordinates or youngsters are expected to obey authority without questioning. As such, they are unlikely to initiate upward communication unless requested to do so [54]. Because of the importance of status differences and out of concern for being respectful South-East Asians are generally more cautious in using humor than Westerners, especially when interacting with a hierarchical superior. Additionally, not all humor styles are appreciated. Studies have shown that for example self-deprecating humor is not considered as funny, since Chinese do not see any fun in misfortune [55]. Also teasing or making jokes at someone else's expenses can be perceived as extremely inappropriate and even offensive. However, Singaporeans do smile, laugh and enjoy humor at the same rate as other Westerners. This was demonstrated in a study conducted by Nevo et al. [56]. The study also showed that Singaporean humor style tends to be more conservative, that is, the jokes with sexual content were less frequent as compared with jokes containing aggressive, social or neutral topics.

Cultures characterized by collectivism emphasize relationships between people to a greater degree than individualistic cultures. Group conformity and commitment are considered more important than personal interests while harmony, getting along with others and maintaining 'face' are seen as crucial. Thus, the Singaporean society is expected to promote an interdependent view of the self through values and norms that consider the self as being a socially interconnected entity stressing the importance of attending, adjusting, and being attuned to the needs and experiences of others [57]. As a consequence empathetic reactions, such as perspective taking, expression of concern and readiness to help are deeply rooted in people's culture as norms of polite and socially acceptable behavior within the community. Thus, behaving empathetically is not only welcomed, but also expected within a collectivist culture.

3 Experiment Design

3.1 Voice Pitch Manipulation

Due to resource limitations we used only one robot in our experiment: Olivia.[2] However, to study the effects of different variable manipulations within the same group of participants two different robot characters are obviously needed. Thus, we 'created' a second robot character by modifying the voice pitch. We called the new character Cynthia.

To generate the robot's voices the TTS engine from Loquendo was used. The engine allows pitch variations between 0 and 100. A normal inflected voice would have a pitch of 50. Highly animated voices would display pitch values greater than 50 while a pitch value of 0 would produce a flat, monotone voice. For our experiment we chose 'Allison',

---

[2]Technical details about Olivia can be found in [58].

Fig. 1 Olivia (*left*) and Cynthia (*right*)

a U.S. English accented voice. We set the pitch value for Cynthia's voice at 20 and for Olivia's voice at 70. All other voice parameters were tuned to the same values for both characters: timbre = 50, speech rate = 40, volume = 50. Additionally, to make the difference between the two robot characters even more explicit, we added two pigtails and a pink band to Cynthia's fabric hair (see Fig. 1). The differences in look and voice were explained as being related to different software modules used for Olivia and Cynthia. Such simple visual modifications, even if minimal, are required in order to keep the robot character consistent. Literature studies in the past have warned about the negative effects of multiple voice personalities associated with the same system [7]. Thus, an experiment design in which we would alternate voice and look in a 2 × 2 matrix should be avoided, since it could confuse the participants about the robot's personality inducing unwanted negative effects.

3.1.1 Hypotheses

Following the literature findings (see Sect. 2.1), our hypothesis was that a high pitch 'female' robot would be perceived as generally more attractive, but also more emotional and possibly more extrovert. Additionally, the high pitch robot would achieve higher ratings concerning personality appeal and degree of similarity with the test participants. Further, we hypothesized that an increased level of attractiveness and personal appeal would lead to more positive user feelings and a better interaction evaluation in terms of enjoyment and overall interaction quality. On the other hand, we would expect the low pitch robot to be perceived as less pleasant, but as having a more assertive and stronger personality.

3.2 Designing Humor: Let's Laugh a Bit!

Humor is a powerful tool, however not every type of humor is advisable. For example sarcastic and irreverent humor can

be perceived as hostile, especially when delivered by a female voice [7]. Ethnical, racial and sexist jokes have offensive connotations, intellectual jokes might not be always understood by all people, while self-deprecating jokes can affect the own image. Also dark, vulgar or toilet humor should be avoided, since it suggests an interlocutor with rather low class attributes [7]. The only type of humor recommendable and conform with the social-cultural context of our experiment is the innocent humor, also defined as inoffensive and light humor.

For our experiment, we chose two jokes of innocent humor type (see Table 1). The first joke is a punning riddle—a question-answer joke based on the pun—while the second joke is based on incongruity, on a surprise element, that is to say, the question gets an unexpected, out of place answer.

The decision to pre-program the jokes was taken in order to facilitate the effect of the jokes, as in the studies conducted by Morkes [36] and Babu [39]. Humor relies on sensing the right situation and the appropriate time. Thus, using a scenario helps to create the right environment for the joke. The jokes were selected from a local website for fun and humor[3] to ensure that the humor type was not unfamiliar to our participants. However, the jokes had no 'local' content, that is, they could be found on other English humor websites as well. We chose to avoid 'local' humor for two reasons: firstly because the jokes might be interpreted as offensive, that is, the robot-host might be perceived as making jokes at the expense of a local visitor-guest. Secondly, 'local' humor might not be understood by foreign participants. Both jokes were embedded in the scenario in such a way that they would appear as humorous answers.

3.2.1 Hypotheses

Conform to the findings presented in Sect. 2.2 we expect the humor to contribute to the robot's likeability, improving the users' perception of its social skills and behavior, including friendliness, helpfulness and trustworthiness. We also expect the robot to be perceived as more extrovert. Furthermore, we expect an increase of task enjoyment and more positive user feelings during the interaction which would eventually lead to better ratings concerning the overall enjoyment and interaction quality.

3.3 Designing Empathetic Reactions: I Can Feel What You Feel!

Empathic reactions emerge as a consequence of interpreting the user's internal affective state. In a game situation, such as the one described in [42] the achievement of an internal affective state is favored by the task itself: participants are required to play a chess game and subsequently are confronted

---

[3]http://sgforums.com/forums/2223.

Table 1 Jokes told by the robot during the experiment

Joke 1	Joke 2
Robot: …but you better don't go there *(i.e. to the sky-garden)* in the evening because there are a lot of mosquitoes attracted by the pond. But, if I have to choose between flies and mosquitoes definitively prefer mosquitoes. Do you know why? **Visitor**: No …why? **Robot**: Because mosquito can fly, but fly cannot mosquito! He, he, he *(laughing)*! **Visitor**: He, he, he… I also hate flies *(laughing)*!	**Visitor**: Do you know if they have chicken feet[a] in the canteen? **Robot**: Hm …I cannot tell you if they their shoes on! He, he, he *(laughing)* **Visitor**: Ah …he, he, he *(laughing)* you are funny! **Robot**: Of course they have the best chicken feet in whole Singapore, at stall no. 4!

[a]Chicken feet are a Chinese delicacy

with an empathetic agent while winning or losing the game. Unfortunately, the settings of a realistic receptionist scenario do not include game playing. Therefore, in our experiment we tried to come up with a scenario task that would artificially induce an affective state by asking the user to imagine herself in the following situation: she/he is a famous scientist traveling from China on business purposes; on the way from the airport the user experiences a quite unpleasant event: in the taxi she/he forgets a bag containing all his/her important documents, such as passport, wallet, mobile phone and secret work information sheets. Tired and upset the user runs to the lobby asking for help.

There is, of course, no guarantee that the scenario would induce the desired effect in the user. But, on the other side losing a wallet is a particularly uncomfortable situation to which participants could relate from own or others' experience. Also, using a scenario creates the same conditions for all test participants and translates them—at least mentally if not emotionally—in a situation where an empathetic response from the robot would not be perceived as inappropriate. Thus, the scenario should also overcome possible problems encountered in [49] and [50] concerning incongruous emotional responses leading to negative agent ratings.

A widely popular model of emotions among computer scientists is the OCC model [59]. The model describes a hierarchy of 22 emotion types. In our scenario we used two of them: feeling 'sad' for the visitor's loss—which triggers the robot's prompt reaction to help recovering the bag—and feeling 'happy' once the bag is found. Hence, the robot response strategy is to mimic the user's affective state (parallel empathy) and to offer immediately help (reactive empathy) [60].

Ideally, the expression of empathy involves different communication channel, such as facial, vocal non-verbal, verbal etc. Unfortunately, the robot we used for our experiments had a synthetic fabric face, which made the display of facial expressions impossible. Thus, our empathy manipulation had to rely only on a verbal and vocal non-verbal cues. However, research has shown that it is possible to model empathy using only oral style of interaction: Prendinger and

Ishizuka [61] implemented empathetic behaviors in a life-like character companion attached to a web-based application with successful results. The companion aimed to help users to feel less stressed during a job interview by providing real-time, verbal-empathetic feedback based on the user's elicited emotions.

The TTS from Loquendo enables the robot to perform various vocal non-verbal cues, such as sighs, coughs, laughs and several interjections. To design the robot's empathetic reactions we used sighs (**1**) and interjections in lowering tone (**2**)—to express sadness—and in rising tone (**3**)—to express happiness or positive surprise. We also varied the intonation using pre-built TTS features to stress certain words (**4**) with the purpose of giving them additional affective value (happy, sad, surprised, unsure, etc.).

In terms of verbal empathetic communication the prompts were designed in concordance with the socio-cultural background of our experiments, that is we included empathetic reactions such as perspective taking (**5**), expression of concern (**6**) and joy (**7**), as well as readiness to help (**8**). As such, the robot expressed its 'feelings' of being sorry for the bag loss (**6**), took the visitor's perspective while commenting on the event (**5**) and showed its 'understanding' concerning the visitor's internal affective state (**5**). Further, the robot offered to call the taxi company assuring the visitor to do its best to find the bag (**8**). When the bag was found the robot expressed joy (**7**) and announced the good news with a happy tone. At the end of the conversation the robot once again reassured its satisfaction about finding the lost bag (**7**).

In the non-empathetic version, the robot used a neutral voice without voice pitch variation, did not take the visitor's perspective nor any initiative to help and did not make use of affect to express joy or sadness. Thus, the dialogue between the visitor and the robot was in this version much shorter—for comparison see Table 2.

When we designed the manipulation of empathy we took into account that the experiment was performed in the lab. Effects of empathy expression have also been studied by other researchers—see [42, 49, 50]—in similar interaction

Table 2 Empathetic vs. non-empathetic dialogue

	Empathethic speech	Non-empathetic speech
	Visitor: I've just arrived and I forgot my bag in the taxi... **Robot**: Oh..**(1)** Ah..**(2)** Your bag is gone **(2)**!	**Visitor**: I've lost my bag in the taxi on the way here . . . **Robot**: Okay
	Visitor: Yes! **Robot**: I am so sorry **(4)** to hear that **(6)**!	**Visitor**: Yeah . . . ah. . . **Robot**: I see your bag is gone!
	Visitor: Yeah . . . **Robot**: This is really terrible **(6)**! There is nothing worse than to lose your bag together with your whole documents **(5)**!	**Visitor**: Right. . .ah. . .I wish to get it back!
	Visitor: Indeed! **Robot**: You must be very worried **(5)**! Should I call the taxi company to help you find the bag **(8)?** **Visitor**: Yes, that would be really very good. . .	
	Robot: What taxi was that? **Visitor**: It was a blue one.	**Robot**: What taxi was that? **Visitor**: I think, it was a COMFORT taxi.
	Robot: Was it a blue color COMFORT taxi? **Visitor**: Yes, yes it was!	**Robot**: Was it a blue color COMFORT taxi? **Visitor**: Yeah . . .
	Robot: OK. Just give me a minute, please. I will try my best to find you bag **(8)** . . . *[calls]*	**Robot**: OK. Just give me a minute, please. *[calls]*
	Robot *(on call)*: You found the driver **(4)**? Oh **(3)**. . . Great **(4)**! He will bring the bag right now? Good **(4)**! Thank you so much **(4)**! Bye-bye!	**Robot** *(on call)*: [. . .] You found the driver? Okay. Thank you! Bye!
	Robot: Yes **(4)**! I have good news **(4)**! **Visitor**: Oh!	**Robot**: Okay, okay . . . **Visitor**: Yes?
	Robot: They found your bag **(4)**! The taxi driver will be here in 2 minutes with your bag!	**Robot**: They found your bag. The taxi driver will be here in 2 minutes with your bag.
	Visitor: Thank you so much! **Robot**: You are welcome! I am happy that we managed to find your bag **(7)**!	**Visitor**: Thank you a lot! **Robot**: You are welcome!

environments. In order to reduce the impact that an artificial environment could have on the empathy perception we set up the scenarios and the observation room in similar settings to those in which the robot would later perform—for more details see Sect. 3.4.

3.3.1 Hypotheses

Based on the results of past studies (see Sect. 2.3) we expect that an empathetic social robot would be generally higher rated in terms of likeability, trustworthiness, behavior and social skills. Consequently, we expect the empathetic robot to induce in users more positive feelings during the interaction. This could eventually lead to better evaluation results in terms of enjoyment and overall interaction quality.

3.4 Experimental set-up

The experiment was performed at the Institute for Infocomm Research (I^2R) Singapore. It contained 5 scenarios

divided into three parts: the first part focused on exploring the robot's empathetic reactions on users (scenario 1), the second part was concerned with voice pitch manipulations only (scenario 2 and 3), while the third part studied the effects of humor in combination with voice pitch manipulations (scenario 4 and 5)—for an overview see Table 3. All scenarios had the same experimental settings, evaluation methodology and test participants. The robot's tasks described in the scenarios were designed conform to real life settings[4] and included making phone calls, arranging appointments, giving directions and informing about building amenities, such as the staff canteen, the sky garden and the swimming pool.

In the first part of the experiment (scenario 1) we split the participants into two groups: one half interacted with an em-

---

[4]We conducted a short interview with 5 professional receptionists working in I^2R Singapore who reviewed and confirmed the accuracy of the scenarios as being similar to the situations they encountered in their daily work.

Table 3 Overview of experiment conditions, variables, scenario topics, number of participants and average duration (in minutes)

Scen.	Condition	Variable	Topic	Nr. part.	Duration (min)
1	± empathy	High pitch	Calling taxi	14/14	2.32/1.42
2	+ voice only	High pitch	Arranging appointments + giving directions	28	3.52
3	+ voice only	Low pitch	Arranging appointments + giving directions	28	3.50
4	− humor	Low pitch	Informing about building amenities	28	3.56
5	+ humor	High pitch	Informing about building amenities	28	4.02

pathetic robot character (group A) and the other half with a non-empathetic robot character (group B). The decision was taken in order to avoid repeated exposure to an emotional stimulus—in this case the bag loss—which would create habituation and consequently, effect diminution [62]. Since the groups were independent, that is, they were not interacting twice with the same robot, only one robot character was used. For this particular scenario we chose Olivia because of the more expressive voice cue variations of the high-pitch TTS. In the second part of the experiment the test participants interacted with Olivia (scenario 2) and Cynthia (scenario 3). The robots' task in these scenarios was to assist the participants getting appointments with I^2R staff members. The scenarios were identical, apart from the staff member's name. In this part we focused on voice pitch manipulations only. Finally, in the third part of the experiment Cynthia (scenario 4) and Olivia (scenario 5) informed the participants about buildings amenities, such as the sky-garden, the swimming pool and the staff canteen. The information presented in the scenarios was identical. The only difference was that scenario 5 contained two jokes. The scenario order was randomized to exclude any potential biases that might arise from being exposed to one particular robot character before the other.

The robot was placed in a small observation room designed for usability experiments. Attached to the robot was a screen where information about name, office location, phone number, staff's picture and other building related issues could be displayed.

To ensure homogeneity between the trials we substituted the speech recognition module by a human wizard. The prompts were prepared in advance and played back by the wizard during the experiment. Equipped with headphones and sitting behind a dark screen the wizard team could see and hear the participants interacting with the robot.

Before entering the room the participants received a short briefing on the experiment procedure and the scenario for the current trial. After each interaction with the robot the participants were given a questionnaire to fill in.

The experiment was recorded with two cameras placed to the side and behind the robot. A total of 28 persons participated: 25 % were I^2R staff members and 75 % were people recruited from outside I^2R. At the end of the test each participant received a 25 Singapore Dollars shopping voucher (approx. 20 US Dollars).

The experiment took one hour for each participant to be completed. The study was conducted over a period of three days.

3.5 Questionnaire Design

In order to test how the variable manipulations affect users' perception concerning the quality of interaction we needed a questionnaire that addresses this question. According to Hassenzahl et al. [63], the user evaluation of a system is influenced by its pragmatic and hedonic quality. Applied to conversational, spoken interactions with a robot, pragmatic quality would refer to functional aspects determining how well a certain communicative goal is achieved, while hedonic quality would relate to non-functional aspects indicating how much the user enjoyed the interaction. Looking closer at the human robot interaction we can identify the following factors that could shape the people's perceptions about the robot and eventually influence the interaction evaluation in the terms described by Hassenzahl (see Fig. 2):

- Robot appeal—how the robot presents itself to the outside world, i.e. how it looks, talks, behaves, expresses personality—(hedonic aspects)
- Task appeal—how enjoyable (hedonic aspect) and how effective, efficient and easy (pragmatic aspect) it is to interact with the robot
- Content appeal—how attractive is the content presented (hedonic aspect), how interesting (hedonic aspect) and good (pragmatic aspect) are the answers delivered

The effect produced by the interplay of functional and non-functional interaction aspects is reflected on the users' feelings during the experiment—in Fig. 2 several users' feelings (orange) are highlighted in possible associations with
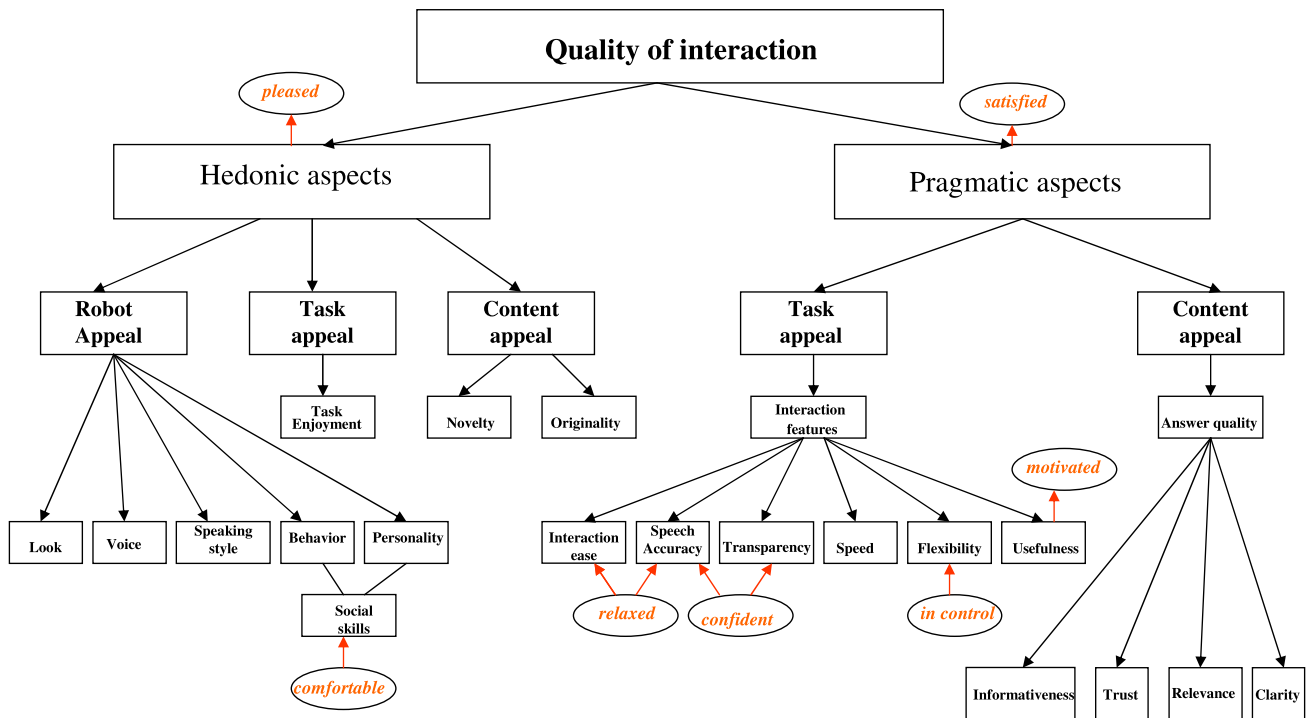
Fig. 2 Overview about hedonic and pragmatic aspects of quality of interaction, as well as their impact on users' feelings (*in orange*)

different hedonic and pragmatic interaction aspects. Therefore, the questionnaire included questions addressing this factor.

Finally, a fifth dimension was included in the questionnaire which directly linked the robot's personality and behavior with our variable manipulation concerning empathy and humor. We call this factor robot social skills.

Since there is no single questionnaire that includes all these factors, we designed our own survey using AttrakDiff [64], SASSI [65], ITU MOS [66]. We also incorporated in our survey items originating from the four Gricean conversational maxims which address basic principles for an effective communication [67]. The questionnaire was built, similar to AttrakDiff—from which most of the items were taken—using semantic differentials on 7 point bipolar scales.

3.5.1 Robot Appearance Appeal

The appeal of the robot's appearance addresses 4 different aspects comprised in the following subscales: look, voice, speaking style and behavior. Additionally, there is a fifth component—the robot's personality—which includes variables meant to assess different personality dimensions that could not be subsumed in one single subscale.

The robot's look subscale contains 5 items: presentable/unpresentable, professional/unprofessional, pleasant/unpleasant, inviting/rejecting and overall impression (AttrakDiff).

For the robot's voice appeal we used 5—out of 8—items from the ITU MOS questionnaire which measures the quality of TTS voices. These were: pronunciation, articulation, speaking rate, pleasantness and overall impression. Two remaining items, listening effort and comprehension problems were referred to on a further subscale addressing the answer quality, while a third one—acceptance—was reformulated into another item referring to the matching degree between the robot's voice and look.

For the third subscale—the speaking style—we used the AttrakDiff word pair stylish/tacky split to two different, more appropriate concepts for a conversational speech style: polite/impolite and elegant/rough. The subscale also contains a third item concerning the overall speaking style impression.

The robot's behavior scale includes mostly word pairs from AttrakDiff questionnaire, such as boring/funny, likeable/disagreeable. The word pair sympathetic/unsympathetic was split into two more precise concepts: friendly/unfriendly and empathetic/not empathetic. Additionally, we added another word pair especially relevant for a social robot receptionist: helpful/unhelpful. The scale also contains a statement about the overall behavior impression.

Regarding the robot's personality, we included 4 questions inspired from the work of Kahn and De Angeli [68] concerning personality traits dealing with potency, social and emotional competence: extrovert/introvert, rational/emotional, strong/weak and assertive/submissive. The

personality 'module' also included a statement about the overall personality impression.

All items related to the robot's appeal subscales held only hedonic values, meaning that they belong to the non-functional aspects of the interaction.

3.5.2 Task Appeal

The task appeal included functional as well as non-functional interaction aspects grouped into two subscales. The non-functional subscale refers to the task enjoyment and subsumes the word pairs: new/common, undemanding/challenging and simple/complicated (AttrakDiff). The functional aspects refer to interaction features emerging from performing the tasks. The interaction features—formulated as semantic differentials—are originated from the SASSI questionnaire and concern speech accuracy, interaction ease, transparency, speed, flexibility and usefulness.

3.5.3 Content Appeal

For the non-functional aspects of the content appeal we chose two statements concerning the content interest (exciting/lame) and the content presentation (original/conventional). Both word pairs were taken from the AttrakDiff questionnaire. For the functional aspects we constructed a subscale regarding the quality of the answers delivered. The subscale includes variable pairs inspired by the Gricean conversational maxims and concerns answers' informativeness, trustworthiness, relevance and clarity.

3.5.4 User Feelings

This subscale contained items addressing the user feeling during the interaction with the robot. The following word pairs were included: motivated/unmotivated (AttrakDiff); annoyed/pleased, tense/relaxed, confident/insecure, frustrated/satisfied, comfortable/awkward, in control/out of control (SASSI).

3.5.5 Robot Social Skills

This subscale was used in one of our previous studies [58] and contains 4 items referring to the robot's capabilities to socialize with humans and to express personality traits, emotions and humor.

3.5.6 Overall Judgments and Other Personal Details

The questionnaire also included global ratings concerning the robot's overall aesthetic appeal (look, voice, speaking style), overall degree of entertainment, overall enjoyment and overall interaction quality. Additionally, we asked the users about the similarity degree between the robot and themselves and about how much they liked the robot in each particular scenario context. At the beginning of the questionnaire the users filled in details concerning their gender, age, work background, knowledge/experience with social robots and speech recognition applications, expectations regarding the robot and personality type. The personality type was determined via an on-line Myers Briggs personality test.[5] This particular data was collected with the purpose of determining whether there were interaction effects between users' and robots' personality. At the questionnaire's end the participants had the possibility to fill in their own comments about what they liked and/or disliked concerning their experience with the robot and to give some suggestions for improvements, if they had any.

4 Results and Discussion

The majority of users (77.8 %) were under 30 years while 22.2 % were between 31–41 years. 63 % were male and 37 % were female.

The distribution of the ethnic groups in our data is not equal but corresponds roughly to the distribution of the Singaporean population (as presented in Sect. 2.4): 75 % were Chinese, 10.7 % Malay, 7.1 % Indians and another 7.1 % European/North Americans. Thus, our sample can be considered representative for the cultural context in which the robot would later be deployed.

Concerning the work background 35.7 % were working in IT, 17.9 % in social sciences, 14.2 % in finance and business and 10.7 % in health care. 17.9 % were students and one person was unemployed.

44 % of the participants had no knowledge of social robots while 45.2 % had seen and/or read about them. A small percentage (7.1 %) had interacted with a social robot while one person was involved in robotic design.

66.7 % of the participants had no experience with speech recognition systems. 33.3 % indicated having some occasional experience with speech recognition built-in applications for mobile phones, PC software, social robots or web interfaces. However, they used such applications relatively seldom (1–2 times); only one person reported a frequent use (more than 10 times). Half of the people who used speech recognition devices did it for real purposes, such as for airport enquiries, database searches, ordering a pizza or interacting with speech handicapped people. The other half did it for testing or hobby purposes. Four people reported positive experiences with such devices, one was neutral and three reported negative experiences.

---

[5]http://www.humanmetrics.com/cgi-win/jtypes2.asp.

46.1 % of the participants had high expectations, 19.2 % had standard or medium expectations while 34.7 % indicated low expectations. Some of the participants mentioned they would expect the robot to be able to recognize speech and gestures (21.7 %), and to respond well (25 %) and fast (2 people) to standard questions. Others expected the robot to speak clearly and to express a proper body language (2 people). While some participants expected the robot to be somewhat restricted to basic tasks and to have limited capabilities (17.9 %), others expected the robot to act as in real life (2 people) to have a friendly disposition (2 people), a nice voice tone (2 people) and even to be able to understand the feelings of users (2 people).

Concerning participants' personality profiles 44.4 % were extrovert while 55.6 % were introvert.

4.1 Scale Reliability

Firstly, we performed a reliability analysis on our scales to verify their internal consistency. All our scales showed relatively high Cronbach values, between .760–.903.

Table 4 presents the Cronbach values for robot's look ($\alpha = .856$), voice ($\alpha = .801$), speaking style ($\alpha = .782$) and behavior appeal ($\alpha = .903$). Only one variable was removed from the robot's look subscale—professional look—which shows a low correlation ($\alpha = .214$) with the other subscale items. The subscales addressing the task enjoyment and answer quality, as well as the user feelings and robot's social skills also show high internal consistency: $\alpha = .760$ for interaction features, $\alpha = .795$ for answer quality, $\alpha = .856$ for task enjoyment, $\alpha = .846$ for user feelings and $\alpha = .825$ for robot's social skills (see Tables 5 and 6).

4.2 Voice Pitch

To detect differences in ratings between the two robot characters, we performed a Wilcoxon signed rank test. The results show that our hypothesis concerning the attractiveness of the high pitch robot (Olivia) was correct: she was perceived as having significantly more appeal in terms of voice, ($p = .038$) and overall aesthetic appeal ($p = .017$). The test participants also found her to have a more appealing behavior ($p < .001$), better social skills ($p = .012$) and a more pleasant overall personality, ($p = .029$). Users also had more positive feelings while interacting with Olivia ($p = .027$). As expected, Olivia also appeared to be a more extrovert robot ($p = .031$) and users found her much more like themselves ($p = .009$). On the other hand, the low pitch robot (Cynthia) was perceived as being stronger ($p = .046$) but, contrary to our prediction not more assertive. The overall robot likeability in the high pitch condition was statistically significantly higher as compared with the low pitch condition ($p < .001$).

The interaction features subscale was better rated in the 'Olivia' condition ($p = .044$). The content, even though identical in both scenarios (apart from the staff member's name)—appeared as more exciting when it was formulated by the high-pitch robot ($p = .002$). In general, the interaction with Olivia was considered to be much more entertaining ($p = .002$). As predicted, the overall enjoyment experienced with Olivia during the interaction was higher ranked ($p = .006$), as was the overall interaction quality ($p = .001$).

No significant differences could be found for the originality of the content presentation, task enjoyment, answer quality or robot's look—the last one most probably due to

Table 4 Robot's appeal subscales: look, voice, speaking style and behavior

Look	Items	Item-total correlation (r)	Speaking style	Items	Item-total correlation (r)
Items: 4 $\alpha = .856$	1. Presentable	.717	Items: 3 $\alpha = .782$	1. Elegance	.680
	2. Pleasant	.741		2. Politeness	.611
	3. Inviting	.557		3. Ov. impression	.597
	4. Ov. impression	.797			
	*Professional (removed)	.214			
Voice	**Items**	**Item-total correlation (r)**	**Behavior**	**Items**	**Item-total correlation (r)**
Items: 6 $\alpha = .801$	1. Articulation	.544	Items: 6 $\alpha = .903$	1. Funny	.701
	2. Pronunciation	.568		2. Friendly	.782
	3. Match look	.488		3. Likeable	.843
	4. Speaking rate	.728		4. Empathetic	.668
	5. Pleasantness	.407		5. Helpful	.681
	6. Ov. impression	.642		6. Ov. impression	.784

Table 5 Answer quality, interaction features, tasks enjoyment and user feelings subscales

Answer quality	Items	Item-total correlation (r)	Tasks enjoyment	Items	Item-total correlation (r)
Items: 4 $\alpha = .795$	1. Trustworthiness	.471	Items: 3 $\alpha = .856$	1. New	.641
	2. Informativeness	.715		2. Challenging	.888
	3. Relevance	.617		3. Simple	.889
	4. Clarity	.635			
Interaction features	**Items**	**Item-total correlation (r)**	**User feelings**	**Items**	**Item-total correlation (r)**
Items: 6 $\alpha = .760$	1. Speech	.523	Items: 7 $\alpha = .846$	1. Motivated	.575
	2. Transparency	.474		2. Pleased	.687
	3. Ease	.463		3. Relaxed	.606
	4. Fastness	.496		4. Confident	.645
	5. Flexibility	.590		5. Comfortable	.662
	6. Usefulness	.528		6. Satisfied	.632
				7. In control	.451

Table 6 Robot's social skills subscale

Social skills	Items	Item-total correlation (r)
Items: 4 $\alpha = .825$	1. Express emotions	.621
	2. Express personality	.754
	3. Express humor	.561
	4. Socialize	.399

the minimal difference between Olivia and Cynthia's look (see Table 7).

4.3 Humor

According to a Wilcoxon post-hoc analysis with Bonferroni correction ($p = .012$) the ability to express humor was perceived as being the highest in scenario 5 ($p < .001$). Most of the participants (81 %) had a visible reaction (through smile and laughs) to the jokes. Thus, we can assume that our manipulation was successful.

Since the humor was manipulated together with the voice pitch, similar effects as in the previous experimental part were expected. Indeed, we found the same preferences concerning the robot's behavior ($p = .001$), voice appeal ($p = .001$), overall aesthetic appearance ($p = .002$), social skills ($p < .001$), interaction features ($p = .005$), overall entertainment ($p = .001$), enjoyment ($p = .028$) and interaction quality ($p = .042$).

Apart from the above results, there are also others that were not found in the previous experimental part. Thus, these results could be linked to the humor manipulation alone. For example, the speaking style of the humorous robot was considered to be much more appealing ($p = .005$). Also, the tasks to accomplish with the robot were considered more enjoyable ($p = .032$). The robot's personality appeared to be stronger ($p = .008$), but also more emotional ($p < .001$) and, as expected, more extrovert ($p = .011$), (see Table 8).

However, even though the humorous robot was considered as having a more appealing behavior (that is, more friendly, funny, likeable etc.) and better social skills it was not perceived as significantly more trustworthy, despite higher median values (5 > 4). This might be explained by the fact that the scenario task—giving direction and amenities information—was somewhat different from the one presented in the study conducted by Kurtzberg et al. [31], where users were mainly confronted with a highly trust sensitive topic, that is an on-line business negotiation. Thus, humor may have influence on users' perceptions of trust as long as trust plays a central role in the interaction process. In our scenario trust issues were only of secondary importance.

Also, the robot's likeability in the humor condition, as well the user feelings during the interaction did not show statistically significant differences—except for the feeling of being comfortable ($p = .018$). One possible explanation is that some of the participants might have preferred a different style of humor—in the comment section we received eight remarks concerning the humorous robot: four were positive and five were negative; the negative comments were requesting another type of humor or no humor at all. However, regardless of whether participants found the jokes funny, the vast majority understood the robot's intention to be humorous and felt entertained. This is shown by significantly higher values for overall entertainment (6>5) and enjoyment (5 > 4).

Table 7 Results high-pitch/low-pitch robot

Aspects/Subscale	Items	Low pitch Median	High pitch Median	Z-Values	p-Values
Robot's appeal/Voice	All	4	5	−2.080	.038
Robot's appeal/Behavior	All	4	5	−3.988	<.001
Robot's social skills	All	2.65	3	−2.525	.012
Task appeal/Inter. feat.	All	4.50	4.83	−2.014	.044
User feelings	All	3.87	4.57	−2.213	.027
Robot's appeal/Pers.	Ov. impression	4	5	−2.181	.029
Robot's appeal/Pers.	Extro./Intro.	4	4	−2.156	.031
Robot's appeal/Pers.	Strong/Weak	4	4	−2.006	.046
Content appeal	Content excitement	3	4	−3.123	.002
Ov. judgment	Aesthetic appeal	4	5	−2.392	.017
Ov. judgment	Entertainment degree	4	5	−3.035	.002
Ov. judgment	Enjoyment	4	5	−2.719	.006
Ov. judgment	Inter. quality	4	5	−3.274	.001
Ov. judgment	Robot likeability	4	5	−3.688	<.001
Ov. judgment	A lot like me	3	4	−2.600	.009

Table 8 Results humorous robot vs. non-humorous robot

Aspects/Subscale	Item	Humor Median	No humor Median	Z-Values	p-Values
Robot's appeal/behavior	All	5.33	4.58	−3.229	.001
Robot's appeal/Voice	All	5.50	5	−3.330	.001
Robot's social skills	All	5	3	−3.986	<.001
Task appeal/Inter. feat.	All	5.2	4.7	−2.362	.005
User feelings	Comfortable	5	4	−2.370	.018
Ov. judgment	Aesthetic appeal	5	4	−3.082	.002
Ov. judgment	Entertainment	6	5	−3.404	.001
Ov. judgment	Enjoyment	5	4	−3.348	.028
Ov. judgment	Inter. quality	5	5	−2.029	.042
Robot's appeal/Speak. style	All	5.33	5	−2.824	.005
Tasks appeal/Task enjoy.	All	5.50	4.66	−2.148	.032
Robot's appeal/Pers.	Strong/Weak	4.5	4	−2.653	.008
Robot's appeal/Pers.	Emot./Rat.	5	3	−3.724	<.001
Robot's appeal/Pers.	Extro./Intro.	5	4	−2.527	.011

On the other side, both user feelings and robot's likeability—significantly correlated along our whole data collection (r values between .825–.771)—showed a trend of higher median values in the humorous condition, that is, user feeling (4.85 > 4.29) and robot's likeability (5 > 4). Thus, we believe that repeating the experiment with a higher number of participants we might have a chance of proving our initial hypothesis.

Furthermore, no other significant differences were found concerning the robot's look and assertiveness, content appeal and similarity degree between users and robot.

4.3.1 Anecdotal Results

More than half of the people who participated in the study came back to complain that the famous stall no. 4, mentioned by the robot in scenario 4 and 5, did not sell chicken feet. The stall was, of course our invention and we did not expect anybody to check the veracity of the information provided in the experiment. However, our 'mistake' was not taking into account that eating is a grand passion in the Singaporean gourmet society. This result confirms once again

Table 9 Results empathetic/non-empathetic robot

Aspects/subscale	Item	Empathetic median	Non-empathetic median	U-Values	p-Values
Robot's appeal/Behavior	Empathetic	6	5	50.5	.039
Robot's appeal/Behavior	Ov. impression	6	5	50.0	.048
Robot's appeal/Pers.	Emot./Rat.	6	4	30.5	.002
Task appeal/Inter. feat.	Ease	6	5	50.0	.037
User feelings	Confident	5	4	48.0	.033

the CASA paradigm [5], proving that our participants took the robot seriously.

4.4 Empathy

According to a Mann-Whitney test no significant effects were found on our scale constructs between group A (exposed to the empathetic robot character) and group B (exposed to the non-empathetic robot character). However, on the level of each scale, we discovered significant differences between the two groups. Firstly, the robot in group A was perceived indeed as being more empathetic ($p = .039$) to visitors' problems and more emotional ($p = .002$). This finding was also confirmed by almost half of the participants (40 %) from group A who made several comments about the "extreme willingness to help", "empathetic", "very receptive" robot, etc. Similar comments were made by only 15 % of the participants in group B. Secondly, the overall behavior of the empathetic robot was better rated by group A than by group B ($p = .048$). Further, participants from group A felt more confident interacting with the robot ($p = .033$) and the interaction appeared to them to be easier ($p = .037$)—for an overview see Table 9.

However, despite higher median values in group A and an apparently successful manipulation our results did not achieve statistically significant values in terms of robot likeability, trustworthiness, social skills, overall enjoyment or interaction quality, as we initially expected. In general, the manipulation effects were minor as compared with those achieved in the other experimental parts. One possible explanation might be that the exposure to empathy expression might have needed a longer time until achieving 'visible' results. The scenarios duration for both groups was below 2.5 minutes, i.e. 2.32 minutes for group A and 1.42 minutes for group B, while most literature studies indicated scenario durations of at least 5 minutes exposure per trial.

Another explanation could be that, even though the scenario intended to transpose the users in a particular mental situation, it did not achieve any change in their affective state.

On the other side, since the median values were higher in group A for all categories mentioned in our hypothesis the effects of empathy might have been significantly higher if the experiments were to have been performed with a large number of participants.

4.5 Effects Within Groups

Next, we checked the effects of users' gender, age, expectations, personality, work background, level of experience and behavior during the experiment on the evaluation results. For this purpose we used a Mann-Whitney test.

4.5.1 Gender Effects

Concerning the voice pitch manipulation male participants rated the overall personality appeal ($p = .013$) of the high pitch robot (Olivia) much better than the female participants did. They also indicated significantly higher values for the user feelings ($p = .029$) and overall enjoyment ($p = .013$). Additionally, they found the interaction more transparent ($p = .036$). On the other hand, female participants assessed Olivia's personality as being more extrovert than male participants did ($p = .006$).

Further, male participants found the low pitch robot's (Cynthia) look significantly more professional ($p < .001$), her voice matched the face better ($p = .039$) and her speech was faster ($p = .025$). The content presented by Cynthia appeared to them significantly more interesting ($p = .008$) and her overall behavior more appealing ($p = .045$).

In the humor manipulation condition female participants found the answers of the humorous robot more informative ($p = .019$) and more clear ($p = .012$) than male participants. Additionally, females also found the humorous robot as having better abilities to socialize ($p = .027$).

Finally, in the empathy scenario male participants felt significantly more comfortable with the empathetic robot than female participants did ($p = .012$). They also scored the voice of the non-empathetic robot significantly better ($p = .019$) and indicated more positive feelings during the interaction ($p = .028$). On the other hand, female participants found the non-empathetic robot more useful ($p = .019$).

Overall, we observed a tendency by men to give higher ratings for both robots—see in Table 10 male (upper part) vs. female participants (lower part). This tendency might be

Table 10 Results within groups: male (*upper part*) vs. female (*lower part*) participants

Aspects/ subscale	Item	Condition/ variable	Female median	Male median	U-Values	p-Values
User feelings	Comfortable	Empathetic/ high pitch	3	5	4	.012
Robot's Voice	All	Non-empath./ high pitch	4.1	5.91	4.5	.019
User feelings	All	Non-empath./ high pitch	3	4.57	3	.028
Robot's appeal/ Pers.	Ov. impression	Voice only/ high pitch	4	5	31	.013
User feelings	All	Voice only/ high pitch	4	5	36	.029
Ov. judgment	Enjoyment	Voice only/ high pitch	4	5	31.5	.013
Task appeal/ Inter.feat	Transparency	Voice only/ high pitch	4	6	27	.036
Robot's appeal/ Look	Professional look	Voice only/ high pitch	3	5	6.5	<.001
Robot's appeal/ Voice	Match look	Voice only/ low pitch	2	3	23.5	.039
Task appeal/ Inter.feat.	Speech rate	Voice only/ low pitch	2	4	35	.025
Content appeal	Content pres.	Voice only/ low pitch	2	4	28	.008
Robot's appeal/ Behavior	Ov. impression	Voice only/ low pitch	3	4	39	.045
Robot's appeal/ Pers.	Extro./Intro.	Voice only/ high pitch	5	4	27	.006
Content appeal/ Ans. qual.	Informativeness	Humor/ high pitch	7	6	38	.019
Content appeal/ Ans. qual.	Clarity	Humor/ high pitch	6.50	5.50	33.5	.012
Robot social skills	Able to socialize	Humor/ high pitch	6	4	38.5	.027
Task appeal/ Inter.feat.	Useful	Non-empath/ high pitch	7	5	4.5	.019

explained by general stereotypes which predicts men's preference for female robots, also found by other studies [69].

4.5.2 Work Background and Experience with ASR Devices

Users with non-IT work background found the empathetic robot as to be more extrovert ($p = .008$) and as having better social skills ($p = .047$) compared with IT background users—see Table 11.

Users with no experience using speech recognition devices found the empathetic robot more emotional ($p = .018$) and showed more positive feelings during the interaction ($p = .028$), as compared with experienced users. Additionally, this user category rated both robot characters significantly better concerning their abilities to socialize, as found in the studies of Reeves and Nass [70]. This result was con-

stant during the entire experiment: scenario 1—empathetic robot/$p = .034$, non-empathetic robot/$p = .050$—, scenario 2/$p = 0.12$, scenario 3/$p = .002$, scenario 4/$p = .008$, scenario 5/$p = .010$. This finding suggests less experienced and less technical users tend to be more open in perceiving the robot as sociable entity—see Table 12.

4.5.3 Expectations

Users with higher expectations towards the robot found the high pitch robot (Olivia) to be more flexible ($p = .036$). On the other hand, users with low expectations found the speaking style ($p = .023$) and answer quality ($p = .003$) of the low pitch robot (Cynthia) better—see Table 13 for an overview. This outcome might be explained by the fact that the majority of the users with high expectations (71 %) kept

Table 11 Results within groups: participants with non-IT vs. IT work background

Aspects/ subscale	Item	Condition/ variable	Non-IT median	IT median	U-Values	p-Values
Robot's appeal/ Pers.	Extro./ Intro.	Empathetic/ high pitch	5	3.5	4.5	.008
Robot social skills	All	Humor/ high pitch	5.2	4	36.5	.047

Table 12 Results within groups: inexperienced vs. experienced ASR users

Aspects/ subscale	Item	Condition/ variable	ASR inexp. median	ASR exp. median	U-Values	p-Values
Robot's appeal/ Pers.	Emot./Rat.	empathetic/ high-pitch	5.50	3.00	6	.018
User feelings	All	Empathetic/ high-pitch	4.64	3.64	7	.028
Robot's social skills	Able to socialize	Empathetic/ high-pitch	4.50	3	8	.034
Robot's social skills	Able to socialize	Non-empath./ high-pitch	4.50	2	4	.050
Robot's social skills	Able to socialize	Voice only/ high-pitch	5	3	33	.012
Robot's social skills	Able to socialize	Voice only/ low-pitch	5	3	28	.002
Robot's social skills	Able to socialize	Non-humor/ low-pitch	5	3	28	.008
Robot's social skills	Able to socialize	Humor/ high-pitch	5	3	30	.010

Table 13 Results within groups: participants with high vs. low expectations

Aspects/ subscale	Item	Condition/ variable	High expect. median	Low expect. median	U-Values	p-Values
Task appeal/. Inter. feat	Flexible	Voice only/ high pitch	5	4	43	.036
Robot's appeal/ Speaking style	All	Voice only/ low pitch	4.16	5	40	.023
Content appeal/ Ans. qual.	All	Voice only/ low pitch	4.75	5.50	40	.003

their questions strictly limited to the scenario settings, while almost half of the users with low expectations (48 %) asked some additional questions—most probably trying to explore the robot's capabilities. Most 'unexpected' questions caused some delay, since the answers could not be prepared in advance and the wizard team had to improvise on the spot. The delay might have caused the impression of inflexible answer behavior. On the other hand, the answers improved the information richness and therefore, the answer quality could have been perceived as being better. As for the speaking style we have no explanation, except maybe for the fact that the test participants who interacted longer with the robot could have been more aware of its polite and elegant style of speaking.

4.5.4 Personality Profile

Finally, introvert users found the interaction with the low pitch robot (Cynthia) much easier ($p = .012$) and her behavior significantly more empathetic ($p < .001$) than extrovert users. Since Cynthia was perceived as being a more introvert robot, our result confirms the findings of Nass and Reeves [70] concerning the attraction effects between peo-

Table 14 Results within groups: participants with extrovert vs. introvert personality

Aspects/ subscale	Item	Condition/ variable	Extrovert median	Introvert median	U-Values	p-Values
Task appeal/ Inter. feat.	Ease	Voice only/ low pitch	5	6	39	.012
Robot's appeal/ Behavior	Empathetic	Voice only/ low pitch	5	3	15	<.001

Table 15 Inter-annotators reliability for the human-robot corpus

Annotation	Krippendorf's alpha
Dialogue acts	.8729
Body	.7301
Facial expression	.7769
Gaze direction	.8216
Gestures	.7468

ple and computer voices with similar personalities (see Table 14).

4.5.5 User Behavior

The user behavior was annotated with the help of three annotators in terms of dialogues acts, body postures, facial expressions (smiles, laughs), gaze direction and gestures. The percentage of agreement was calculated on 10 % of the data and showed reliable values, between .7301–.8729 (see Table 15).

The behavior analysis showed that participants who looked predominantly to the screen (and not to the robot) gave the low pitch robot (Cynthia) significantly lower rating for the voice appeal ($p = .007$). Additionally, they rated the non-humorous robot significantly lower in social skills ($p = .041$)—in particular in the ability to express emotion ($p = .010$) and similarity degree with themselves ($p = .011$).

Test participants who smiled more often found the high pitch robot (Olivia) to have much better social skills ($p = .036$).

Also, users who leaned their upper body forwards rated the overall interaction quality with the low pitch robot (Cynthia) better ($p = .022$). Additionally, they felt more comfortable ($p = .018$) gave the non-humorous robot better ratings concerning overall enjoyment ($p = .004$), entertainment ($p = .028$) and interaction quality ($p = .019$)—see Table 16 for an overview.

Leaning forwards, smiling and looking towards an interlocutor are signs of interest and engagement in conversation. Being more engaged in conversation was shown in [71] to have positive influences on evaluation ratings. However, more research is needed to prove this hypothesis.

4.5.6 No Effects

No particular effects were found for different age groups. Additionally, no strong recurrent patterns could be found in our data for building user groups based on gestures and dialogue acts or other speech patterns.

5 Conclusions

The main focus of this study was to analyze the effects of voice pitch, humor and empathy in interaction with a social robot receptionist with the ultimate goal to determine how the variable manipulations influence the ratings concerning the quality of interaction with a robot.

The manipulation of voice pitch showed strong effects on how users perceived the robots and the entire interaction: Olivia, the higher pitched robot, was better rated in terms of overall appearance, voice appeal, behavior and personality; users had more positive feelings while interacting with Olivia and gave better ratings concerning interaction features, overall enjoyment, entertainment and interaction quality.

A challenge of this study was to adequately manipulate humor and empathy in such way that would achieve maximum effects on users' perceptions. Humor and empathy are interpersonal skills desirable in public services, such as in the receptionist domain. For a robot receptionist being humorous and empathetic may increase the robot's perceived cuteness and may improve the evaluation results. On the other side, these skills are highly context dependent requiring appropriate timing and situation sensing. Apart from this, the manipulation of humor and empathy in a controlled experiment involving receptionist tasks pose some additional difficulties: while manipulating the humor the receptionist scenario must also take into account that people's reactions depend not only on their own culture, but also on their individual sense of humor. Thus, the manipulation—even though appropriated to the situation and the cultural context—might not trigger the desired effects.

The manipulation of empathy requires the users to be mentally transposed in a situation that would match the robot's expression of empathetic feelings. Such manipulation is easily achieved in scenarios where users are active

Table 16 Results within groups: user behavior concerning gaze, smile and upper body movements

Aspects/ Subscale	Item	Condition	Look screen median	Look robot median	U-Values	p-Values
Robot's appeal/ Voice	All	Voice only/ low pitch	3	4.1	19	.007
Robot's social skills	All	No humor/ low pitch	2	3.5	35.5	.041
Robot's social skills	Emotions	No humor/ low pitch	2	3	22	.010
Ov. judgment	A lot like me	No humor/ low pitch	2	4	27.5	.011

Aspects/ Subscale	Item	Condition	Smile Median	No smile Median	U-Values	p-Values
Robot's social skills	All	humor	5.3	4.5	40.5	.036

Aspects/ Subscale	Item	Condition	Upperbody forwards median	Upperbody no move median	U-Values	p-Values
Ov. judgment	Inter. quality	Voice only/ low pitch	4	3	29	.022
Ov. judgment	Inter. quality	No humor/ low pitch	5	3	28.5	.019
Ov. judgment	Enjoyment	No humor/ low pitch	5	3	23.5	.004
Ov. judgment	Entertainment	No humor/ low pitch	5	3	30.5	.028
User feelings	Comfortable	No humor/ low pitch	5	4	26.5	.018

game players; here the robot's empathetic reactions when winning or losing would naturally match the users' emotional state. Unfortunately, the tasks of a receptionist do not include game scenarios. Further, it remains difficult, even if explicitly asked in a questionnaire, to determine the extent to which the users indeed assumed the emotional state described in the scenario.

Regarding the humor our results showed that the manipulation was successful: the ability to express humor was the highest in the scenario containing jokes. The effects of humor were studied in combination with the voice pitch. Combining the voice pitch with humor manipulations had the advantage of creating a more homogeneous personality profile and allowed us to collect more data on voice pitch effects. On the other side, the results obtained cannot be entirely related to the effects of humor, since the voice pitch had also an influence on users' perception: similar ratings to those found in the voice pitch scenarios concerned the overall appearance, voice appeal, social skills, interaction features, overall enjoyment, entertainment and interaction quality. However, apart from these results we also found others that could be linked directly to humor manipulations, that is,

they were not found in the voice pitch scenario. These results were related to the robot's speaking style appeal, task enjoyment and personality appeal (i.e. strong, emotional). No statistically significantly ratings were found for user feelings and robot's likeability in the humorous condition. Nevertheless, this should not be seen as contradictory, since there was a constant trend for higher median values for these two categories. It is likely that repeating the experiment with a higher number of participants would achieve significant results. In the future we plan to conduct an experiment focusing on humor effects alone, that is, without voice pitch manipulations.

Concerning the empathy our results showed that the manipulation seemed to be successful at least at the basic perceptive level: the robot was perceived as being significantly more empathetic and more emotional. However, the empathy manipulation, even though successful, achieved only minor effects on the evaluation results: the interaction was rated as easier, the overall robot's behavior appeared more appealing and the users felt more confident. The lack of more significant results might be partly explained by a too short exposure to the empathetic stimulus and partly because of a too low stimulus intensity, that is, the scenario setting

might not have achieved to transpose the user in the required affective state. On the other side, there was again a trend of higher median values for robot's likeability, trustworthiness, social skills, overall enjoyment and interaction quality, as we initially hypothesized. Thus, conducting the experiment with more participants may achieve more significant results. In the future we intend to reconsider the scenario settings, incorporate additional facial reactions to increase empathy effect, enlarge the exposure to the empathetic stimulus and the number of participants.

Regarding the effects within groups, we found that male participants in general gave better ratings to both robots, as compared with females, confirming stereotypes found by other studies which predict men's preference for female robots. More experienced users with IT background and/or knowledge in speech recognition devices rated the social skills of both robots significantly lower as compared with non-experienced users. This finding suggests that more experienced users tend to be less open in perceiving the robots as sociable entities. Participants with high expectations found the high pitch robot more flexible while participants with low expectations appreciated the speaking style and the answer quality of the low pitched robot better. Also, introvert participants found the low pitched robot much easier to interact with and more empathic to their needs.

The manipulation of humor and empathy showed positive effects on the way users perceived the interaction with the robot and the robot self. However, the major results of the study are related to the manipulations of the voice pitch: our results demonstrated the high impact of the voice pitch on robot's attractiveness, which further influenced the evaluation results of the entire interaction. We believe that choosing the right voice pitch should be a priority in social robot design. Finally, with this study, we would like to stress the enormous importance of the voice in human robot interaction and to encourage further research on this topic.

References

1. Bar-Cohen Y, Hanson D (2009) The coming robot revolution: expectations and fears about emerging intelligent, humanlike machine. Springer, New York
2. Gockley R, Bruce A, Forlizzi J, Michalowski M, Mundell A, Rosenthal S, Sellner B, Simmons R, Snipes K, Schultz AC, Wang J (2005) Designing robots for long-term social interaction. In: IEEE/RSJ int conference on intelligent robots and systems, pp 2199–2204
3. Lee MK, Makatchev M (2009) How do people talk with a robot? An analysis of human-robot dialogues in the real world. In: Proc of CHI, Boston, pp 3769–3774
4. Makatchev M, Fanaswala IA, Abdulsalam AA, Browning B, Ghazzawi WM, Sakr M, Simmons R (2010) Dialogue patterns of an Arabic robot receptionist. In: Proc of the 5th ACM/IEEE international conference on human-robot interaction, March 2010. ACM/IEEE, New York, pp 167–168
5. Nass C, Steuer J, Tauber ER (1994) Computers are social actors. In: Proc of the SIGCHI conference on human factors in computing systems: celebrating interdependence. ACM, New York, pp 72–78
6. Hampes WP (1994) The relation between humor styles and empathy. Eur J Soc Psychol 6(3):34–45
7. Nass C, Brave S (2005) Wired for speech. How voice activates and advances the human-computer relationship. MIT Press, Cambridge
8. Traunmueller H, Eriksson A (1994) The frequency range of the voice fundamental in the speech of male and female adults. Manuscript
9. Riding D, Lonsdale D, Brown B (2006) The effects of average fundamental frequency and variance of fundamental frequency on male vocal attractiveness to women. J Nonverbal Behav 30:55–61
10. Collins SA, Missing C (2003) Vocal and visual attractiveness are related in women. Anim Behav 65:997–1004
11. Jones BC, Feinberg DR, DeBruine LM, Little AC, Vukovi J (2008) Integrating cues of social interest and voice pitch in men's preferences for women's voices. Biol Lett 4:192–194
12. Zuckerman M, Miyake K (1993) The attractive voice: what makes it so? J Nonverbal Behav 17(2):119–135
13. Helfrich H, Weidenbecher P (2011) Impact of voice pitch on text memory. Swiss J Psychol 70(2):85–93
14. Apple W, Streeter LA, Krauss RM (1979) Effects of pitch and speech rate on personal attributions. J Pers Soc Psychol 37:715–727
15. Montepare JM, Zebrowitz-McArthur L (1987) Perceptions of adults with childlike voices in two cultures. J Exp Soc Psychol 23:331–349
16. Scherer KR (1979) Personality markers in speech. In: Scherer KR, Giles H (eds) Social markers in speech. Cambridge University Press, New York, pp 147–209
17. Zuckerman M, Miyake K, Elkin CS (1995) Effects of attractiveness and maturity of face and voice on interpersonal impressions. J Res Pers 29:253–272
18. Walters ML, Syrdal DS, Koay KL, Dautenhahn K, te Boekhorst R (2008) Human approach distances to a mechanical-looking robot with different robot voice styles. In: Proc of 17th IEEE international symposium on robot and human interactive communication (RO-MAN), München, Germany, pp 707–712
19. Scheutz M, Schermerhorn P (2009) Affective goal and task selection for social robots. In: Vallverd J, Casacuberta D (eds) The handbook of research on synthetic emotions and sociable robotics. IGI Global, Hershey, pp 74–87
20. Martin RA (2007) The psychology of humor: an integrative approach. Elsevier Academic, New York
21. Wilson CP (1979) Jokes: form, content, use and function. Academic Press, London
22. Cann A, Calhoun LG, Banks JS (1979) On the role of humor appreciation in interpersonal attraction: it's no joking matter. J Humor Res 10(1):77–89
23. Hampes WP (1999) The relationship between humor and trust. J Humor Res 12(3):253–259
24. Middleton R, Moland J (1959) Humor in negro and white subcultures: a study of jokes among university students. Am Sociol Rev 24:61–69
25. Nijholt A (2007) Conversational agents and the construction of humorous acts. In: Nishida T (ed) Conversational informatics: an engineering approach. Wiley, Chicester, pp 21–47

26. Ziv A (1984) Personality and sense of humor. Springer, Berlin
27. Roy DF (1960) Banana time: job satisfaction and informal interaction. Human Organ 18:158–168
28. Bolman LG, Deal TE (1992) What makes a team work? Organ Dyn 21(2):34–44
29. Clouse RW, Spurgeon KL (1995) Corporate analysis of humor. J Hum Behav 32(3–4):1–24
30. Ruch W (1994) Extraversion, alcohol, and enjoyment. Pers Individ Differ 16:89–102
31. Kurtzberg TR, Naquin ChE, Belkin LY (2009) Humor as a relationship-building tool in online negotiations. Int J Confl Manage 20(4):377–397
32. Gibson DE (1994) Humor consulting: laughs for power and profit in organizations. Humor 7(4):403–428
33. Shneiderman B (1998) Designing the user interface: strategies for effective human-computer interaction. Addison-Wesley, Menlo Park
34. Loehr D (1996) An integration of a pun generator with a natural language robot. In: Proc of the international workshop on computational humor. Twente, The Netherlands, pp 161–172
35. Tinholt HW, Nijholt A (2007) Computational humour: utilizing cross-reference ambiguity for conversational jokes. In: Masulli F, Mitra S, Pasi G (eds) 7th international workshop on fuzzy logic and applications (WILF 2007), Berlin, July. Lecture notes in artificial intelligence, vol 4578. Springer, Berlin, pp 477–483
36. Morkes J, Kernal HK, Nass C (1999) Effects of humor in task-oriented human-computer interaction and computer-mediated communication: a direct test of SRCT theory. Hum-Comput Interact 14:395–435
37. Huan CM, Szafir D (2001) No joke: examining the use of humor in Computer-mediated learning. http://pages.cs.wisc.edu/~dszafir/resources/HumorinComputer-MediatedLearning.pdf/. Unpublished material
38. Dybala P, Ptaszynski M, Rzepka R, Araki K (2009) Humoroids: conversational agents that induce positive emotions with humor. In: AAMAS (2), pp 1171–1172
39. Babu S, Schmugge S, Barnes T, Hodges LF (2006) What would you like to talk about? An evaluation of social conversations with a virtual receptionist. In: IVA, pp 169–180
40. Wispe L (1987) History of the concept of empathy. Cambridge University Press, New York
41. Goldstein AP, Michaels GY (1985) Empathy: development, training, and consequences. Erlbaum, Hillsdale
42. Pereira A, Leite I, Mascarenhas S, Martinho S, Paiva A (2010) Using empathy to improve human-robot relationships. In: Lamers MH, Verbeek FJ (eds) Proc of the 3rd international conference on human-robot personal relationships, vol 59. Springer, Berlin, pp 130–138
43. Brave S, Nass C, Hutchinson K (2005) Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. Int J Hum-Comput Stud 62(2):161–178
44. Hone K (2006) Empathic agents to reduce user frustration: the effects of varying agent characteristics. Interact Comput 2(2):227–245
45. Klein J, Moon Y, Picard RW (2002) This computer responds to user frustration: theory, design, and results. Interact Comput 14(2):119–140
46. Prendinger H, Mori J, Ishizuka M (2005) Using human physiology to evaluate subtle expressivity of a virtual quizmaster in a mathematical game. Int J Hum-Comput Stud 62(2):231–245
47. Bickmore TW, Schulman D (2007) Practical approaches to comforting users with relational agents. In: CHI extended abstracts'07, pp 2291–2296
48. Partala T, Surakka V (2004) The effects of affective interventions in human-computer interaction. In: Interacting with computers, pp 295–309
49. Cramer HSM, Goddijn J, Wielinga BJ, Evers V (2010) Effects of (in)accurate empathy and situational valence on attitudes towards robots. In: Proc of HRI'10, pp 141–142
50. Ochs M, Pelachaud C, Sadek D (2008) An empathic virtual dialog agent to improve human-machine interaction. In: Proc of AAMAS (1)'08, pp 89–96
51. Evers V, Winterboer A, Pavlin G, Groen FAC (2010) The evaluation of empathy, autonomy and touch to inform the design of an environmental monitoring robot. In: Proc of ICSR'10, pp 285–294
52. Hofstede G (1991) Cultures and organizations: software of the mind. McGraw-Hill, London
53. Craig J (1994) Culture shock! Singapore. Kuperard, London
54. Liu Q (2012) Core culture values and beliefs of Singapore. http://www.ncl.ac.uk/ecls/assets/documents/pdf/countryfiles/CCC-Singapore.pdf. Accessed February 7
55. Decker WH, Yao H, Calo TJ (2011) Humor, gender, and perceived leader effectiveness in China. SAM Adv Manag J 76(1):43–53
56. Nevo O, Nevo B, Yin LJS (2001) Singaporean humor: a cross-cultural, cross-gender comparison. J Gen Psychol 128(2):143–156
57. Cheon BK, Mathur VA, Chiao JY (2010) Empathy as cultural process: insights from the cultural neuroscience of empathy. World Cult Psychiatry Res Rev 5:32–42
58. Niculescu AI, van Dijk EMAG, Nijholt A, Limbu DK, See SL, Wong AHY (2010) Socializing with Olivia, the youngest robot receptionist outside the lab. In: Ge SS, Li H, Cabibihan J-J, Tan YK (eds) Proc of the 2nd international conference on social robotics (ICSR 2010), Berlin. Lecture notes in artificial intelligence, vol 6414. Springer, Berlin, pp 50–62
59. Ortony A, Clore G, Collins A (1988) The cognitive structure of emotions. Cambridge University Press, Cambridge
60. Bee N, Andre E, Vogt T, Gebhard P (2009) First ideas on the use of affective cues in an empathic computer-based companion. In: Proc of 8th int conf on autonomous agents and multiagent systems (AAMAS 2009), Budapest, Hungary, pp 5009–5014
61. Prendinger H, Ishizuka M (2005) The empathic companion: a character-based interface that addresses users' affective states. In: Applied artificial intelligence, pp 267–285
62. Oehman A, Hamm A, Hugdahl K (2000) Cognition and the autonomic nervous system: orienting, anticipation, and conditioning. In: Cacioppo JT, Tassinary LG, Bernston GG (eds) Handbook of psychophysiology. Cambridge University Press, Cambridge, pp 533–575
63. Hassenzahl M, Platz A, Burmester M, Lehner K (2000) Hedonic and ergonomic quality aspects determine software's appeal. In: Proc CHI 2000, Den Haag, The Netherlands. Guilford, New York, pp 201–208
64. Hassenzahl M, Burmester M, Koller F (2003) AttrakDiff: Ein Fragebogen zur Messung wahrgenommener hedonischer und pragmatischer Qualität. In: Mensch und computer '03, pp 187–196
65. Hone KS, Graham R (2000) Towards a tool for the subjective assessment of speech system interfaces (SASSI). Nat Lang Eng 6(3–4):287–303
66. ITU-T Recommendation P.85 (1994) Telephone transmission quality subjective opinion tests. A method for subjective performance assessment of the quality of speech voice out-put devices
67. Grice HP (1975) Logic and conversation. In: Cole P, Morgan JL (eds) Syntax and semantics. Vol. 3. Speech acts. Academic Press, San Diego, pp 41–58
68. Khan R, De Angeli A (2009) The attractiveness stereotype in the evaluation of embodied conversational agents. Interact 1:85–97
69. Siegel M, Breazeal C, Norton MI (2009) Persuasive robotics: the influence of robot gender on human behavior. In: Proc of IROS, Dresden, Germany, pp 2563–2568

70. Reeves B, Nass C (1996) The media equation: how people treat computers, television, and new media like real people and places. Cambridge University Press, New York
71. Niculescu AI, van Dijk EMAG, Nijholt A, See SL, Li H (2010) How humans behave and evaluate a social robot in real-environment settings. In: Brinkman WP, Neerincx M (eds) Proc of the 28th European conference on cognitive ergonomics (ECCE), Delft. Mediamatica, Delft, pp 351–352

Andreea Niculescu graduated in Computer Science from Ruhr-University Bochum, Germany in 2006. From 2006 until 2011 she was a full-time research assistant in the Human Media Interaction group at the University of Twente. She received her Ph.D. degree in Human Computer Interaction from the University of Twente in 2011. Her main interests are multimodal and speech-based interactions with conversational systems, such as dialogue systems and social robots, user-centered design and usability studies. In 2011 she received the Best Paper Award at the IEEE International Conference on User Science and Engineering (i-USEr), Subang Jaya, Malaysia. Currently, she is working as usability expert at the NT-Ware Systemprogrammierungs GmbH Germany which is part of the Canon Corporation.

Betsy van Dijk is assistant professor in the Human Media Interaction group of the University of Twente. She studied mathematics at the University of Nijmegen and has a Ph.D. on teaching methodology in computer science from the University of Twente. Currently, her research is in the field of human-computer interaction where the main focus is on user-centered interface and interaction design, natural interaction and experience design and evaluation. She is involved in several national and international research projects on multi-modal and multi-party interaction, ambient intelligence, socially engaging virtual agents and robots and tangible interfaces for children.

Anton Nijholt is professor of Computer Science in the Human Media Interaction group of the University of Twente. He studied mathematics and computer science at the Technical University of Delft and has a Ph.D. on theoretical computer science from the Vrije Universiteit Amsterdam. He held positions at various universities in and outside the Netherlands. His main research interests are multimodal interaction, affective computing, brain-computer interfacing and entertainment computing. In previous years he has been involved in some large-scale European and Dutch national research projects addressing these topics.

Haizhou Li received the B.Sc., M.Sc., and Ph.D. degrees in electrical & electronic engineering from the South China University of Technology (SCUT), in 1984, 1987, and 1990, respectively. Dr. Li's research interests include speech processing, natural language processing and social robotics. He is currently the Principal Scientist and Department Head of Human Language Technology, the Program Manager of Robotics at the Institute for Infocomm Research (I2R), Singapore. He is also a Conjoint Professor at University of New South Wales, Australia. Dr. Li was named one of the two Nokia Visiting Professors 2009 by the Nokia Foundation in recognition of his contributions to Speaker and Language Recognition research. He is now an Associate Editor for Springer International Journal of Social Robotics, IEEE Transactions on Audio, Speech and Language Processing, ACM Transactions on Speech and Language Processing, Computer Speech and Language (Elsevier). He was elected as a Board Member of International Speech Communication Association (ISCA, 2009-2013). He served as the General Chair of the 2nd International Conference on Social Robotics (ICSR 2010), and the 50th Annual Meeting of Association for Computational Linguistics (ACL 2012).

Swee Lan See received her B. Computing (Computer Science), B. Computing (Honours) and M. Computing (by Research) degrees from Monash University, Australia in 1995 and 2000 respectively. She was accredited the Certified Professional Marketer (CPM, Asia-Pacific) by APMF in 2000, Certified Behavioural & Career Consultant (CBCC) by IML, Inc. in 2007, and Certified International Project Manager (CIPM) by AAPM in 2009. She was listed in the Marquis Who's Who in Science and Engineering 2011–2012 (11th Edition). Her research interests include Human Computer Interactions, Social Robotics, Psychology, Speech & Natural Language Processing, Decision Support Systems, Entrepreneurship, and Education. She is currently a Manager in the Human Language Technology Department, Institute for Infocomm Research (I^2R), A*STAR, and the Chairman of the IEEE Singapore Section Women in Engineering Affinity Group (IEEE SG WIE AG).