Arts and Design Studies
ISSN 2224-6061 (Paper) ISSN 2225-059X (Online)
Vol.69, 2018
www.iiste.org

A Robot is a Smart Tool: Investigating Younger Users' Preferences for the Multimodal Interaction of Domestic Service Robot

Yiqi Xiao
School of Design, Hunan University, No.2 South Lushan Street, Changsha 410082, China

Abstract
The degree that domestic service robots are generally accepted mainly depends on the user experience and the surprise that the design brings to people. To make the design of robots to follow the trend of interactions of smart devices, researchers should have insights into young people's acceptance and opinions of emerging new interactions. The main content of this study is a user elicitation through which the users' suggestions for commanding a robot in specific contexts are gathered. Accordingly, it sheds light on the features of user preferences for human-robot interaction. This study claims that younger users regard service robots merely as intelligent tools, which is the direct cause of the above interaction preferences.
Keywords: Service robot, Interaction design, User preference

1. Introduction

As an integral part of household intelligent equipment, the domestic service robot plays an important role in the user experience of a smart home. The ways the domestic robots provide services for people through interactions have to meet the need, habits and preferences of users, however, these expectations and preferences differ from person to person. As the sensor technology develops, the multimodal interactions, for example, the voice command and gesture, are more frequently applied in the non-touch interaction scenarios. These interaction modes facilitate user's ability to issue single commands directly and quickly, and avoid the use of touch screen and buttons. Therefore, it is necessary to investigate the user-preferred inputs of multimodal interactions in order to better control the service robots.

2. Related work

2.1 Human-Robot Interaction (HRI) of service robot

Mainstream users more often use the smartphone and remote-control to silently interact with domestic service robots. Coskun et al. (2017) found that users preferred smartphone to control the smart appliances, and they viewed the voice input as an assistant to direct dialogue with computers. On the whole, the participants thought that the buttons and knobs are always indispensable, while the gestures are only suitable for a limited number of contexts. In addition, the participants did not want constant feedback from the appliances as important information will be ignored among the stream of notifications. In fact, this study did not make detailed analysis of each specific scenario, so there are reasons to believe that the users expect more natural interactions such as voice and gesture inputs when they are far from the device or controller and when they are attending to a primary task.

Nowadays voice command has become the standard mode of interaction for the smart appliances, by contrast, some technological difficulties like image segmentation, feature extraction and recognition stand in the way of developing the gesture control. Nevertheless, gesture is a powerful mode of interaction for the targeted user group of laymen and elderly or disabled people. Schiffer et al. (2011) presented a very simple gesture recognition method for deictic gestures, but in their paper the study of relationship between gestures and user's intentions is absent. Similarly, Chapa Sirithunge (2017) suggested an identification model which extracts and analyzes upper body skeletal information for a period of time to recognize the deictic calling gestures and friendly behaviors. A gesture recognition method based on flight-of-time camera was devised by Droeschel et al. (2011). This method enabled users to simply point towards the object's location or show the object to the robot, and do not require the person to be at a predefined position. The robot can also have joint attention that combines the detection of human with the understanding of gestures. These researches justify the feasibility of using gestures in the domain of HRI.

Implicit interaction is another attractive concept. Funk et al. (2018) presented the Interactive Intentional Programming Model, aiming to replace the sensors by a new system that can judge the scenes, interpret human intention and generate the desired outcome of users from preset preferences. However, in this paper, the main focus is intentional interaction.

2.2 User elicitation study

User elicitation is a noteworthy design strategy which accords with the philosophy of participatory design and co-design. The paradigm of user elicitation study consists of three steps in general: (1) showing the participants some interactive tasks and portraying their effects, (2) asking the participants to define appropriate inputs, regardless of any technological limitations, and (3) electing the most frequently suggested proposal as the design solution. This working process was pioneered by Wobbrock (2009) and has been widespread in the studies on user-defined interactions.

Peshkova (2016) adopted the user-elicited paradigm to gather users' suggestions for suitable voice and gesture commands. Finally, three mental models of user gestures were revealed. They are (1) imitative (using gestures to imitate the effects), (2) intelligent (an operator expects a certain level of intelligence of a device that enables the device to interpret a given command correctly) and (3) instrumented (operator gives the instructions using an imaginary tool). The study implies that the explicit attitudes and implicit preferences of users can be elicited from the user-defined inputs, especially gestures.

2.3 Demographic factors of user preferences

User preferences are generally related to demographic factors. The personas that are more positive to form an emotional bond with robots are more likely to recognize the necessity of multimodal interaction with robots (Dos Santos, 2014). The existing studies about the influence factors of user preferences usually focused on age, gender and prior experience.

Kleanthous et al. (2016) found that elderly people prefer voice command to other interaction modes. Specifically, elderly people wish for more proactive service robots rather than passive robots, and hope that human can chat with a robot by calling each other's first name. Elderly people also like the robots that can be activated accordingly. Martínez-Miranda et al. (2018) employed the Wizard-of-Oz method to inspect how the children aged 6-11 guided a personalized robot via voice commands through a maze while collecting sweets and avoiding obstacles. It was found that children of different ages have significant differences in affective reactions and preferences towards robot's personality. So far, few studies have been concerned with differences in interaction preferences between young people and other age groups.

In respect of gender, Fink et al. (2011) argued that it does not affect user's evaluation (including intelligence, usefulness, ease of use, fun, attachment, etc.) of domestic service robots. Family members are apt to have a similar view of usability of their robot to each other.

Several studies indicated that the habituation effect changed user's evaluation of robots and resulted in different preferences. For example, an empirical study conducted by Koay et al. (2006) showed that users did not as much care about the approach distance of robot with neither a humanoid nor mechanoid appearance as before after using it for a period of time. In an ensuing research (2014), they found users who had previous experience of interacting with robots were more adaptive to robot's approach, and tended to give positive ratings to the tasks requiring more coordination with the robot. Buchner's study (2012) also corroborated the assumption that differences in estimating a service robot decrease over time. This means the experience in interacting with robots may have a considerable impact on user-preferred interactive input.

3. Methodology

3.1 Research questions

According to the literature reviews above, the main topic of this study is younger users' preferences for interacting with service robot through voice and gesture. Research questions include: (1) younger users' preferences for voice command and gestural interaction with robot under specific tasks and contexts, (2) differences in user-elicitation between the users who have experienced service robot and those who have not, and (3) understanding users' attitudes and expectations by observing how the user-preferred inputs are described and evaluated.

3.2 Selecting the interaction tasks

Six interactive tasks that are common to the use of various types of robots were chosen before inviting participants (Table 1). A desktop research was targeted at robotic vacuum cleaner, air cleaner robot and companion robot to obtain a list of available functions. The functions that are more appropriate for traditional interactions were excluded, for example, smartphone serves as a monitor of the detailed data about working condition, and remote control can be used to fine-tune the device or deal with multi-steps tasks. Consequently, the remaining tasks are frequently-used simple interactions, among which only one—*Delayed Start*—does not require gestural commands.

Table 1. The selected tasks of human-robot interaction

Start	The robot starts to work.
Stop	The robot stops working.
Specify location	Specifying a location where user wants the robot to go to.
Move away	Changing the moving path of robot to prevent it from hindering the user.
Return to charge station	The robot finds its way automatically to return and begin charging.
Delayed start	The robot will start to work at an appointed time.

3.3 Experimental Setup

The robot used in experiment is Ecovacs AA30. As an air cleaner robot, the height (560mm) of this robot is greater than vacuum cleaner and less than humanoid companion robot, thus the gestures that adapt to this medium height will match robots of various sizes to some extent. The experimental area is approximately 5m×3.5m. In the experimental environment, there were one chair, one robot, one camera mounted on a tripod, and some closed areas demarcated by colored tapes—which were prepared for the *Specify Position* task. LEDs and loudspeakers were attached on the top of robot.

The experiment is comprised of two sessions. 40 participants (16 men, 24 women, age range: 20-28) with diverse backgrounds joined in the first session. Among them, 23 were fresh users, while 17 had used a service robot for more than one year. The number of participants in the second session was 30. Since they were invited to evaluate the user-elicited gestures and voice commands, the prior experience of users was not considered in the process of recruiting participants for Session 2.

3.2 Procedure

3.2.1 Study 1

Session 1 is user elicitation study. Before making suggestions for reasonable commands, participants were required to assess the suitability of applying gestures and voices in every tasks by their own standards, using a 5-point Likert scale. If they did not really sympathize with the value of novel interactions, they would certainly give casual answers or improvise some proposals to satisfy the experimenters. Only when a participant gave a positive score were they verbally told and shown the effect of the task through a video. Afterwards, experimenters videotaped and transcribed the user-preferred commands and how participants interpreted them.

It was noted that the participants were also asked to conceive a wake-up word and suitable outputs of robot, thereby forming a complete dialogue process. There are six modes of outputs, or so-called feedbacks: sound, voice response, light, sound & light, voice & light and mechanical motion. Some more expressive outputs, e.g., co-verbal gesture (Brember, 2011), are merely available for humanoid robot, but not for most of domestic service robots.

In addition to users' subjective evaluation, researchers should first generalize user-defined interactions into several types so as to infer users' preferences by comparing the frequency of proposals in each type. Therefore, frequency of proposals and subjective ratings are the two measurements of this experiment. By the conclusion of user elicitation study, researchers had found three types of gestures, and the voice commands, wake-up words and feedbacks can be all classified into two types, see Table 2. The most frequent proposal in each type was the selected member for a evaluation study based on Wizard-of-Oz experiment.

Table 2. The different types of user-preferred interactions and the definitions of every type

Three types of user-preferred gestures	
Human-human interaction (HHI)	Using the co-verbal gestures or gesticulations that are performed to communicate with human beings or animals to command the robots.
Sign language (SI)	The emblems which are designed to represent specific tasks and indicate the meanings of tasks.
Borrowed gesture (BG)	Borrowing the gestures or motions used for interactions with other products to command the robots.
Two types of user-elicited voice commands and wake-up words	
Brief introduction (BI)	Giving an order to the robot by directly describing the task with short words or an unemotional speech.
Spontaneous spoken dialogue (SSD)	Expressing an intention to the robot with a tone of calling, asking or teasing a creature with which we can talk freely in daily life.
Two types of user-preferred system outputs (feedbacks)	
Efficient	The system outputs inform me of system states as effectively as possible.
Emotional	The system outputs increase a feeling of warmth, and make people relaxed and joyful.

3.2.2 Study 2

Session 2 is user evaluation of the selected proposals. In this session, the order of tasks for every interaction mode (gesture, voice command, wake-up word, and outputs/feedbacks) was counterbalanced. An experimenter secretly manipulated the robot with a remote control to respond to the instructions issued by a participant immediately, then the participant used a 7-point Likert scale to rate the interactive inputs in terms of the pragmatic and hedonic quality. Pragmatic quality is also called EQ (ergonomic quality). It refers to the perceived usability of the product, which addresses the underlying human need for security and control. Hedonic quality refers to the non-task-oriented quality such as innovativeness and novelty (Hassenzahl, 2001).

4. Results

4.1 Inter-rater reliability

Two coders categorized the proposals for every task according to the descriptions in Table 2. In the next stage, inter-rater reliability of the results was computed to ascertain whether the frequency of each type of command was reliable, in order to avoid subjective judgments of the coders. Statistical analysis showed that Cohen's kappa of four tasks with respect to gestural control were substantial ($0.6<k\leq0.8$), and one was perfect ($0.8<k\leq1$). In respect to voice command, three tasks were substantial and the other three were almost perfect. For wake-up word and feedback, the k values were substantial and perfect respectively.

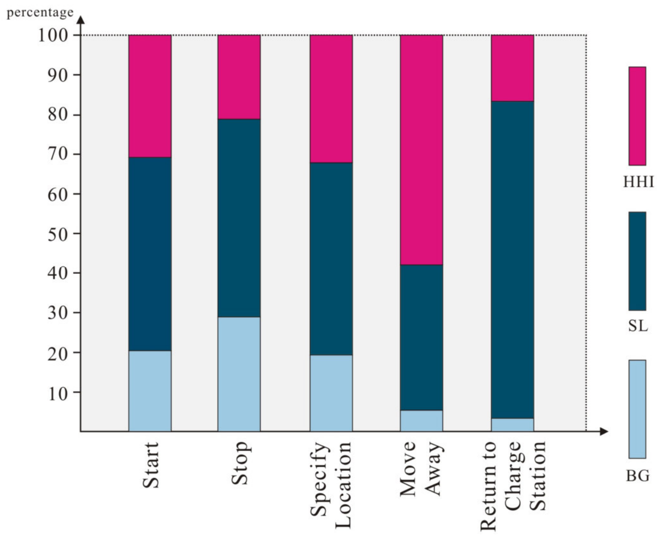

Figure 1. Proportions of The Three Types of Gestural Commands

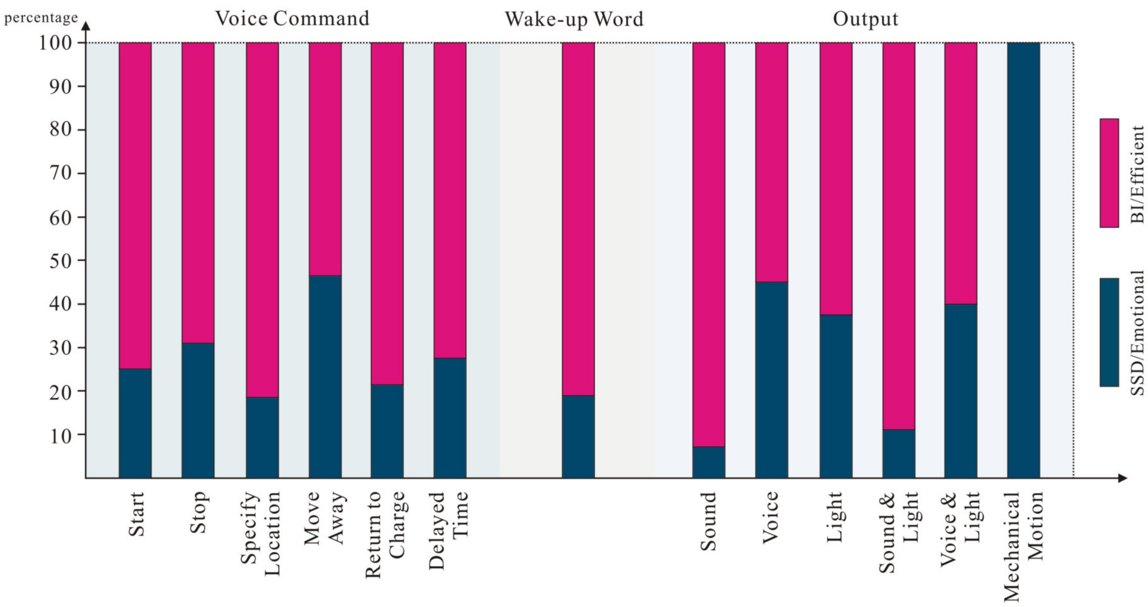

Figure 2. Proportions of The Two Types of Voice Commands and Wake-up Words and Two Modes of Outputs

4.2 Results of Study 1

4.2.1 Differences of the interaction tasks

In respect of gestural command, sign language makes up a relatively significant share (51.7%), followed by HHI (32.3%) and borrowed gesture (16%). As for the voice command, brief instruction occupies a bigger proportion (71.8%) than SSD (28.2%). In brief, some younger users are inclined to innovate communi- cative and imitative gestures in an attempt to activate the robot, and used to regard it as an obedient life form like a pet, a child or a servant. So they were less frequently motivated to borrow gestures directly from touchpad and manipulation of tools. In addition, younger users were usually not intended to initiate a dialogue with robot, instead they liked to issue an order by repeating the keyword of their intention.

Considering that the data was not normally distributed, Chi-square tests were conducted. Statistical analysis revealed significant differences in frequency of gestural types between tasks (χ^2=28.387, df=8, p<0.001). Pair-wise comparison among three types confirmed the significant effect of task in the frequency of gestures (SI-HHI: χ^2=14.228, df=4, p<0.01; HHI-BG: χ^2=10.114, df=4, p<0.05; SI-BG: χ^2=10.114, df=4, p<0.05). This led to the overall proportion as shown in Figure 1. In every task, the number of BI was larger than that of SSD (χ^2=6.805, df=5, p=0.236), see Figure 2. The distribution of two types of feedbacks was significantly different between the output modes (χ^2=11.454, df=5, p<0.05).

4.2.2 Differences between user groups

Fisher's exact test was adopted to measure the effect of prior experience. The results showed that only the distribution of gestural types of *Start* might become significantly different between user groups (χ^2=6.010, df=2, p=0.05) among five tasks. The rest of test results were: *Stop* (χ^2=3.750, df=2, p=0.160), *Specify Position* (χ^2=2.431, df=2, p=0.332), *Move away* (χ^2=0.815, df=2, p=0.781), and *Return to Charge Station* (χ^2=1.586, df=2, p=0.614). Similarly, none of the rates of BI and SSD of all the six tasks was significantly different between two user groups. The comparison analysis of two types of wake-up word and feedback also showed insignificant effect of user groups in the frequency of proposals (Wake-up word: χ^2=1.130, df=1, p=0.637; Feedback: χ^2=8.335, df=5, p=0.107). These results indicate that the differences in user preference is not induced by prior experience of users.

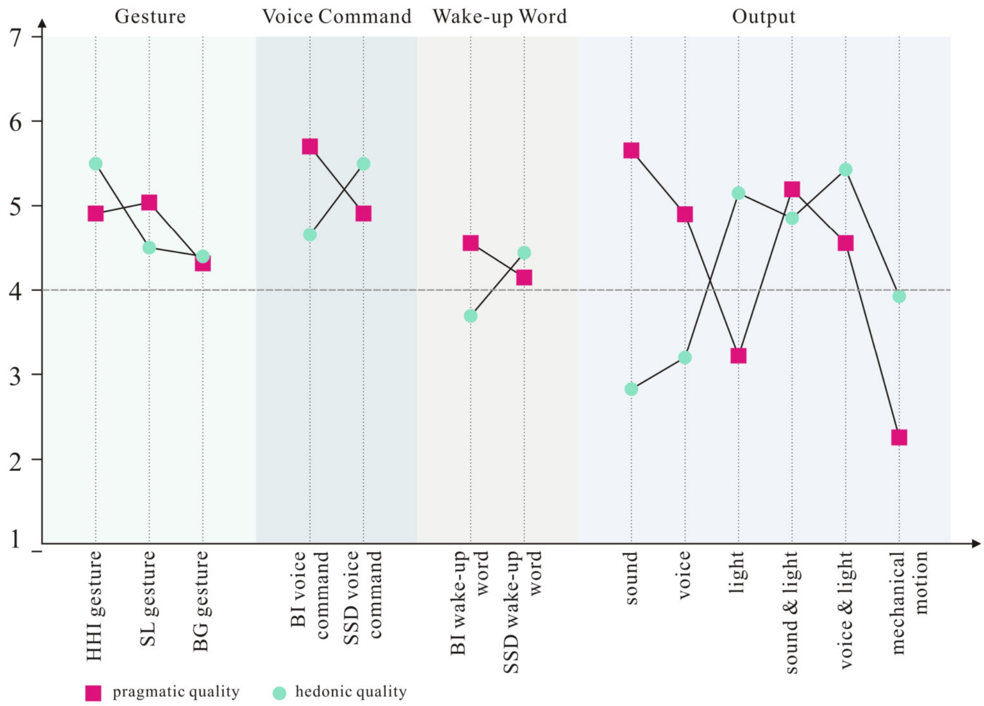

Figure 3. The Average Ratings of Pragmatic and Hedonic Quality of User-preferred Interactions

4.3 Results of Study 2

4.3.1 Ratings of pragmatic quality

The results of a Friedman test showed that the pragmatic quality of gesture did not depend on its types (χ^2=4.673, df=2, p=0.097). By comparison, user-perceived pragmatic quality of voice command varied from different types of verbal expression (Wilcoxon signed-rank test: Z=-2.384, p<0.05). Moreover, the pragmatic quality of SSD was not significantly lower than BI (Z=-1.765, p=0.078). Concerning the output mode, the rating of sound was remarkably higher than the others, by contrast light and mechanical motion are relatively lower rated, which magnified the overall differences (Figure 3) between modes (χ^2=89.030, df=5, p<0.001).

4.3.1 Ratings of hedonic quality

In contrast to pragmatic quality, differences in the hedonic quality between gestural types were significant (χ^2=13.386, df=2, p=0.001), in part due to the HHI which was considered as a more appealing manner (SL-HHI: Z=-3.244, p=0.001; BG-HHI: Z=-3.125, p<0.01) of gestural interaction. Naturally, the voice commands (Z=-3.392, p=0.001) and wake-up words (Z=-2.582, p<0.01) using SSD were higher rated in terms of hedonic quality. Also, participants reported that lights were particularly helpful to enhance the fun, thus the ratings of sound, voice and mechanical motion were significantly lower (χ^2=76.632, df=5, p<0.001) (Figure 3) in comparison with light.

5. Findings

5.1 User preferences of interaction styles

The results of user elicitation study echo the view of Schiffhauer et al. (2016) that user's preferences and pleasantness of gestures are context-dependent. The semantics of *Start*, *Stop*, *Specify Location* and *Return to Charge Station* are so abstract that participants will consciously map these concepts onto other more familiar and concrete analogues before they express these analogous concepts with gestures to represent the corresponding tasks. The literal sense of *Move away* is specific. HHI gestures are more natural for this task because its literal sense offers potential for personifying service robot in the view of participants. Therefore, it is more likely for participants to prefer the interactions with emotional quality for this task. *Return to Charge Station* is a task only used in powered robots. Even though young participants are very familiar with electronic devices, they failed to copy many gestures from their prior knowledge of interacting with other products.

Regarding the feedback, difference between the proportions of two voice types is not obvious (see Figure 2), which indicates that younger users agree with the way of conveying emotional quality with voice output while they think the sound is not suitable for being designed with the aim of emotional interaction. Overall speaking, It is still necessary to enjoy interesting communications with robots in the eyes of some young people.

5.2 Effects of prior experience on user preferences

The study found that prior experience of using a robot has very limited effects on the user-elicited gestures and voice commands both. Fresh users do not anticipate the potential of robot in rich interaction, and experienced users do not deny the necessity for a robot to have the emotional attributes even if they have been accustomed to its presence. According to those studies on the long-term acceptance of service robots, user's evaluation are not as critical as it was when they use for the first time. To sum up, the young people's idea that a service robot should play the role of worker more than life partner could be the reason why the two user groups have similar preferences.

5.3 Users' attitudes towards the interactions with a service robot

Younger user's attitudes towards service robot do not change very much with the recent advances in interaction design. Regarding the voice command, as mentioned earlier, participants more frequently use instructions rather than lively spoken language to give a command or attract the robot's attention. For some tasks, participants prefer HHI gestures or casual conversations, but this does not mean they will keep these preferences constantly and say yes to equal relationship between human and robot. Forlizzi (2006) debriefed family users and found 7/14 families had built social relationships with a robotic cleaner called Roomba. Family members named this robot, let it work with other cleaners and emotionally interacted with it. The user study in this paper shows that young people will be less likely to behave like that. It is clear that personal bonding is important for elderly people who want to be talked softly with. However, younger users choose robots to ease the burden of housework, so they need a robot to be sensitive to the directives and be professional at its job.

Ratings of pragmatic quality and those of hedonic quality are usually contradictory. Participants prefer the designs of higher usability, that is, the more familiar, simpler and more efficient interactions. For example, many participants pointed out the poor practicality of lights as outputs, but they also need them to improve the atmosphere in home environment. The HHI gestures are attractive, but more complex. These attitudes are rooted in the viewpoint of taking service robots as an advanced tool which can interpret human intention.

6. Conclusions

This study investigated younger users' preferences for interacting with the domestic service robot. It may also have implications for design of service robot in the future and for user requirement analysis. The limitation of study is lack of cross-cultural comparison and of discussing the effects of strata of society, so further explorations are needed to answer these questions.

References

Bremner, P. et al. (2011). The Effects of Robot-Performed Co-verbal Gesture on Listener Behavior. IEEE-RAS International Conference on Humanoid Robots, Bled, Slovenia, October 26-28, 458-465.

Buchner, R., Wurhofer, D., Weiss, A., Tscheligi, M. (2012). User Experience of Industrial Robots over Time. In: Proceedings of HRI'12, Boston, Massachusetts, USA, 115-116.

Chapa Sirithunge, H. P., Arjuna S. Srimal, P. H. D. and Jayasekara, B. P. (2017). Identification of friendly and deictic gestures in a sequence using upper body skeletal information. In: Proceedings of the 2017 IEEE Region 10 Conference, Malaysia, November 5-8, 3051-3056.

Coskun, A., Kaner, G., Bostan, I. (2017). Is Smart Home a Necessity or a Fantasy for the Mainstream User? A Study on Users Expectations of Smart Household Appliances. International Journal of Design, 12(1), 7-20.

Dos Santos, T. F., De Castro, D. G., Masiero, A. A., Junior, P. T. A. (2014). Behavioral Persona for Human-robot Interaction: A Study Based on Pet Robot. In: proceedings of HCI International 2014, Greece, June 22-27, 687-696.

Droeschel, D., Stückler, J., Holz, D. and Sven, B. (2011). Towards Joint Attention for a Domestic Service Robot: Person Awareness and Gesture Recognition using Time-of-Flight Cameras. In: Proceedings of IEEE International Conference on Robotics and Automation, 251(3): 1205-1210.

Fink, J., Bauwens, V., Mubin, O., Kaplan, F. and Dillenbourg, P. (2011). People's Perception of Domestic Service Robots: Same Household, Same Opinion? In: Mutlu et al. (eds.): ICSR 2011, LNAI 7072, 204-213.

Forlizzi, J., DiSalvo, C. (2006). Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. In: Proceedings of HRI'06, Salt City, USA, 258-265.

Funk, M., Chen, L. L., Yang, S. W., Chen, Y. K. (2018). Addressing the Need to Capture Scenarios, Intentions and Preferences: Interactive Intentional Programming in the Smart Home. International Journal of Design, 12 (1), 53-66.

Hassenzahl, M. (2001). The Effect of Perceived Hedonic Quality on Product Appealingness. International Journal of Human-Computer Interaction, 13(4), 481-499.

Kleanthous, S. et al. (2016). Analysis of Elderly Users' Preferences and Expectations on Service Robot's Personality, Appearance and Interaction. In: J. Zhou and G. Salvendy (eds.): ITAP 2016, Part II, LNCS 9755, 35-44.

Koay, K. L., Syrdal, D. S., Walters, M. L. and Dautenhahn, K. (2007). Living with Robots: Investigating the Habituation Effect in Participants' Preferences. IEEE International Symposium on Robot & Human Interactive Communication, 564-569.

Koay, K. L., Syrdal, D. S., Ashgari-Oskoei, M., Walters, M. L. and Dautenhahn, K. (2014). Social Roles and Baseline Proxemic Preferences for a Domestic Service Robot. International Journal of Social Robotics, 6: 469-488.

Martínez-Miranda, J., Pérez-Espinosa, H., Espinosa-Curiel, I., Avila-George, H., & Rodríguez- Jacobo, J. (2018). Age-based Differences in Preferences and Affective Reactions Towards a Robot's Personality During Interaction. Computers in Human Behavior, 84, 245-257.

Peshkova, E., Hitz, M., Ahlström, D. (2016). Exploring User-Defined Gestures and Voice Commands to Control an Unmanned Aerial Vehicle. International Conference on Intelligent Technologies for Interactive Entertainment, 47-62.

Schiffer, S., Baumgartner, T., Lakemeyer, G. (2011). A Modular Approach to Gesture Recognition for Interaction with a Domestic Service Robot. ICIRA 2011: Intelligent Robotics and Applications, 348-357.

Schiffhauer, B., Bernotat, J., Eyssel, F., Bröhl, R. and Adriaans, J. (2016). Let the User Decide! User Preferences Regarding Functions, Apps and Interfaces of a Smart Home and a Service Robot. In: A. Agath et al. (eds.): ICSR 2016, LNAI 9979, 971-981.

Wobbrock, J. O., Morris, M. R. and Wilson, A. D. (2009). User-defined gestures for surface computing. Proceedings of CHI'09, pp. 1083-1092.