# Usability of Anthropomorphic Sound Feedback for Service Robot Malfunction through Repeated Experiences

With Emphasis on Vacuum Cleaning Robots.

Kim, Bomi* Kwak, Sonya Sona** Kim, Myung-suk

** Dept. of Industrial Design, KAIST*
*Seoul, Korea, wizard-5@hanmail.net*
*** Design Evaluation Division I, Korean Intellectual Property Office*
*Daejeon, Korea, kwak@kipo.go.kr*
**** Dept. of Industrial Design, KAIST*
*Daejeon, Korea, sonakwak@kaist.ac.kr*

**Abstract**: The service robot has a complicated interface and therefore users may find it hard to understand the system and structure of the robot. Users' misunderstanding of the system of the robot can cause malfunctions of the robot, which can bring users serious problems and unexpected danger. In order to reduce the damage from malfunctions, research on the feedback of malfunction is required. Moreover, a service robot can produce a novelty effect, the tendency for performance to initially improve when new technology is instituted due to increased interest in the new technology. For this reason an investigation of the usability of feedback through repeated serial malfunction is needed. The aim of this study is to apply the anthropomorphic elements to the feedback of malfunction for service robots and evaluate the usability of the anthropomorphic sound feedback for a service robot through repeated malfunction. The experiment was carried out with two types of feedback (anthropomorphic vs. non- anthropomorphic) to evaluate the usability of the anthropomorphic feedback. Sixty-seven university students or graduate students at an engineering college were recruited as subjects for the experiment because demographically users of vacuum cleaning robots were mostly in their twenties and accustomed to new technology. In the experiment, subjects were required to deal with repeated malfunctions. The usability scale is comprised of effectiveness, efficiency, and satisfaction, which was evaluated through a questionnaire. Participants' behaviors were analyzed by protocol analysis, video analysis, and interviews. The results from the experiment are summarized into four parts: First, the more repeatedly a malfunction occurred, the more effectible the anthropomorphic feedback is than non-anthropomorphic feedback; second, the anthropomorphic feedback is more efficient than non-anthropomorphic feedback in similar cases; third, the anthropomorphic feedback is overall more satisfactory than non-anthropomorphic as the feedback of malfunction; fourth, anthropomorphic feedback helped relieve temper and reduce aggressive actions caused by stress. These findings will provide an essential basis for feedback design of service robot malfunction.

***Key words: repeated malfunction, sound feedback, anthropomorphic.***

## 1. Introduction

Interests in service robots are increasing, showing the highest growth rate among robotic products [1]. The service robot is defined as a robot which operates semi or fully autonomously to perform services useful to the well being of humans and equipment, excluding manufacturing operations [2]. The main function of a service robot is to interact and communicate with humans because it plays an important part in providing a specific service [3].

When interacting with the service robot, due to the complicated interface of the robot users may find it hard to understand the system and structure of the robot. Users' misunderstanding of the system can cause malfunctions of the robot. If users do not properly respond to the malfunctions, serious problems may occur by bringing users unexpected danger. In order to minimize the damage from the malfunctions, research on the feedback of malfunction is required.

Most of the latest feedback is composed of text, beeps, or light signals. It is hard to understand the meaning of feedback without a manual as the feedback provides only partial information [4]. Feedback of malfunction that is perceived easily can lead users to take some actions to solve the problems promptly and correctly. The actions can provide not only higher reliability with satisfaction toward the product, but also longer product life. These days, as a way to increase natural and emotional interactions between a user and a product, anthropomorphism is effectively used for the interaction design of the product, such as a lamp of a Mac power book [5].

In authors' previous study [6], we applied the anthropomorphic elements to the feedback design of service robot malfunction and investigated the usability of the feedback in the condition of a single experience of malfunction. The result of the previous study showed that anthropomorphic feedback was more effective, efficient, and satisfactory than non-anthropomorphic feedback. However, the previous study was limited to examine the effect of the anthropomorphic sound feedback in real, as a service robot can produce a novelty effect, the tendency for performance to initially improve when new technology is instituted due to increased interest in the new technology. To overcome this limitation, an investigation of the usability of anthropomorphic sound feedback through repeated serial malfunction is needed. The aim of this study is to evaluate the usability of the anthropomorphic feedback for a service robot through repeated malfunctions.

## 2. Related works

### 2.1 Cleaning Robot

A service robot can be divided into three categories comprising of a robot for servicing humans like personal safeguarding, for servicing equipment such as cleaning, maintenance, and for servicing other tasks like transport [2]. Most service robots are not used commonly because of high manufacturing costs and imperfection in autonomy. Among various types of service robots, the cleaning robot is the only one used commonly because it is provided through a market with a wide range of prices. Sung reported that users of cleaning robots and domestic robots have been expanding gradually [7]. Accordingly, this study focused on the cleaning robot which has been widely used and has growing market potential.

### 2.2 Anthropomorphic Sound Feedback

A cleaning robot which cleans the house automatically helps a user to carry out another task simultaneously during cleaning. Accordingly, when a user performs other tasks, sound interface is more suitable than visual interface to inform the state of a robot [8]. Especially, in the case of an emergency, sound interface is effective because information can be delivered during a short period of time [9]. Consequently in this study, the feedback of malfunction was designed as sound feedback.

People tend to perceive a service robot as a living thing because it has intelligence, movement, and an ability of communication. In order to increase lifelikeness of a service robot, anthropomorphic elements which aid natural communication between humans and robots [10, 11] can be effectively applied to the sound feedback design of a service robot.

## 3. Experimental Design

The experiment used a between-groups factorial design. The between-groups factors were types of feedback (anthropomorphic vs. non-anthropomorphic). Participants experienced only one type of feedback. The usability of sound feedback was evaluated through questionnaires, and participants' behaviors were analyzed by protocol analysis, video analysis, and interviews.

We were interested in examining which type of sound feedback is more effective, efficient, and satisfactory for repeated serial service robot malfunctions. The authors' previous research [6] found that anthropomorphic sound feedback is more effective, efficient, and satisfactory than non-anthropomorphic sound feedback in the condition of a single experience of malfunction. Likewise, we anticipated that anthropomorphic sound feedback would show higher usability than non-anthropomorphic sound feedback in repeated experiences of malfunctions. Moreover, we believed that the usability of sound feedback in repeated experiences is distinguished from that in a single experience. We were also interested in participants' attitudes towards the sound feedback of malfunction, as this would indicate participants' preferences for the feedback. Consistent with the usability of sound feedback, we believed that the participants displayed irritation and anger earlier when they experienced non-anthropomorphic sound feedback than when they experienced anthropomorphic sound feedback.

Thus, we formulate four hypotheses as follows:

**H1**: Anthropomorphic sound feedback will be more effective than non-anthropomorphic sound feedback when people experience repeated malfunctions.

**H2**: Anthropomorphic sound feedback will be more efficient than non-anthropomorphic sound feedback when people experience repeated malfunctions.

**H3**: Anthropomorphic sound feedback will be more satisfactory than non-anthropomorphic sound feedback when people experience repeated malfunctions.

**H4**: People who experience non-anthropomorphic sound feedback will display irritation and anger earlier than those who experience anthropomorphic sound feedback.

### 3.1 Participants

According to demographic study on cleaning robot users, it is shown that the twenties are the main users of the cleaning robot [7]. Most of the users were highly educated and were familiar with technology. Accordingly, university students or graduate students from an engineering college in their twenties (Male: 32, Female: 35) participated in the experiment.

### 3.2 Materials

#### 3.2.1 Robot Prototype

A cleaning robot prototype was developed because an existing robot wasn't under absolute control. The exterior covering of an existing cleaning robot was used for the realism. The mechanism was made of wireless speakers and Bioloid [12] which is an educational robot kit for building various robots. The movement and sound of the prototype was fully under control, using the Wizard of Oz technique [13].

#### 3.2.2 Malfunction

We extracted the list of malfunctions from the on-line community for cleaning robot users which is composed of 13060 members [14]. The following malfunction and the method to solve it were used in the experiment.

- Malfunction: The robot stops suddenly due to an incomprehensible problem
- The method to solve the malfunction

1. Connect the cleaning robot to the adapter.
2. Change the setting of the range of virtual unit 0-3m and turn off the virtual unit.
3. Send a text message to 016-421-6063 for a repair request.
4. Disconnect the robot from the adapter when you received a text message which indicates the completion of a repair.

For the experiment, the number of repeated malfunctions was determined by a pilot test. As participants' attitudes toward the sound feedback changed greatly at the eighth malfunction, in the main experiment malfunctions were repeated eight times.

#### 3.2.3 Feedback

Non-anthropomorphic sound feedback was designed using the electronic sound, Earcon, which was widely used as an alarm of a machine [15]. Onomatopoeia such as a groan was used as an anthropomorphic sound for natural mapping to malfunction.

### 3.3 Procedure

At the beginning, the experimenter introduced the scenario, the malfunction and the method to correct the cleaning robot. Then, participants heard either anthropomorphic or non-anthropomorphic sound feedback, and the session began. The cleaning robot malfunctioned eight times, and each occurred every one minute. The robot would malfunction in a specific area to maintain the distance between the participant and the robot. After the experiment, questionnaires were administered and interviews were conducted. The experiment took approximately 35 minutes in total.

### 3.4 Measures

According to ISO [16], usability is defined as:

> *The extent to which a product can be used by specified users to achieve specified goals with effectiveness, efficiency and satisfaction in a specified context of use*.

Effectiveness is accuracy and perfection. Efficiency is the resource related with accuracy and perfection. Satisfaction is the accessibility and comfort which a user feels. These three elements of usability were used as the criteria to evaluate the sound feedback of service robot malfunction in the experiment.

In this study, effectiveness was estimated by examining whether a participant succeeded in solving the malfunction or not. A participant's success was determined by whether a participant correctly followed the instructions to recover the robot completely from the malfunction by making less than three mistakes. Efficiency was evaluated by measuring the time elapsed between the start of the experiment and the participants' perception of the malfunction. Satisfaction was evaluated by questionnaires. The questionnaires were comprised of 47 different Likert-type items, which were combined into eight factors. The eight factors were 'novelty, pleasure, excellence, comfortableness, performance, aesthetics, familiarity, and satisfaction in usability,' which were drawn from Jung's research [17] and Kanda's research [18].

In order to understand participants' attitudes toward the feedback, we analyzed the text messages, which were sent by participants when they were coping with a malfunction, by protocol analysis and analyzed interviews and the video recording which captured participants' interaction with the robot. The Table 1 describes the definition of each category of the coding scheme with the corresponding examples. The categories of the coding scheme were divided into three by participants' emotional state. Each category indicates the degree of negativity of the participant's emotional state. Category C describes the most negative emotional state of the participant. We set criteria of negativity to determine the degree of negative emotional state of a participant. When the text message included one of the expressions listed in the criteria of negativity, it was coded as category B, and if the text message involved more than two expressions, it was coded as category C.

Table 1. The Coding Scheme

| Category | Definition | Criteria of Negativity and Examples |
|---|---|---|
| A | Neutral | 1. Insertion of the word, "again"<br>Ex) A break down again.<br>2. Some characters or repetition of the same character without meaning<br>Ex) sl&k#%j, kkkkkkkkkkkkkkkkkkkkk<br>3. Onomatopoeia or Emoticon<br>Ex) buzz, TT<br>4. Negative contents<br>5. Repeated expressions<br>ex) "Repair, Repair"<br>6. Message became long.<br>ex) "Repairrrrrrrrrrrrrrrrrrrr" |
| B | Annoyance | |
| C | Anger | |

## 4. Results

### 4.1 Effectiveness of the Feedback

In order to investigate the effectiveness of the anthropomorphic sound feedback, we counted the number of participants' failures in resolving each malfunction (as shown in [Fig. 4-12]). As the malfunction repeated, the number of failures with non-anthropomorphic sound feedback increased while the number of failures with anthropomorphic sound feedback decreased. That is to say, the more repeatedly the malfunction took, the more effectible the anthropomorphic feedback is than non-anthropomorphic feedback. In addition, we conducted a Chi-Square test to compare the number of failures with anthropomorphic sound feedback and the number of failures with non-anthropomorphic sound feedback in each malfunction. The number of failures was significantly different in the 6th and 8th malfunctions ($x^2$=4.561, $df$=1, $p$=0.049/$x^2$=4.561, $df$=1, $p$=0.049). These

results indicate that anthropomorphic sound feedback is more effective than non-anthropomorphic feedback in the repeated serial experience of malfunctions, supporting H1.

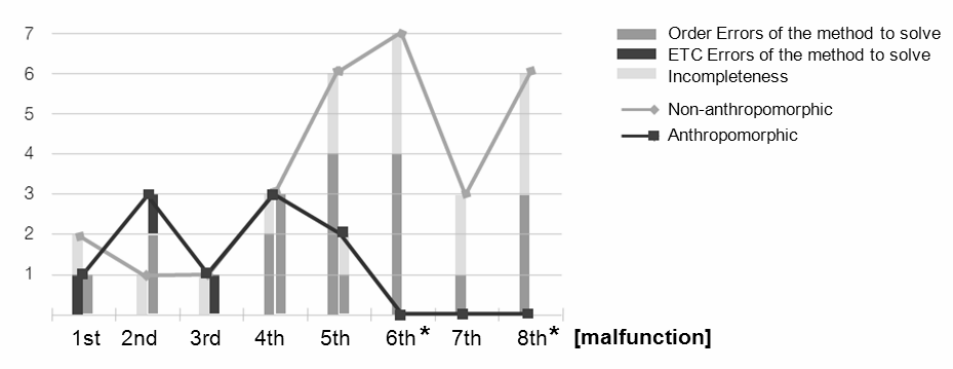


Figure.1 Effectiveness of feedback
(*denotes a case where $p<0.05$)

### 4.2 Efficiency of the Feedback

In order to examine the efficiency of the feedback, we measured the time elapsed between the start of the experiment and the participants' perception of the malfunction. In the case of non-anthropomorphic sound feedback, the time elapsed between the start of the experiment and the participants' perception of the malfunction became longer as the repeated number of malfunction increased, while it barely changed in the case of anthropomorphic feedback. Specifically, the time elapsed in the condition of anthropomorphic sound feedback in the eighth malfunction was significantly different from the condition of non-anthropomorphic sound feedback. ($U$=334.5, $p$=0.040). This result indicates that anthropomorphic sound feedback is more efficient than non-anthropomorphic feedback in the repeated experience of malfunctions which supports H2.

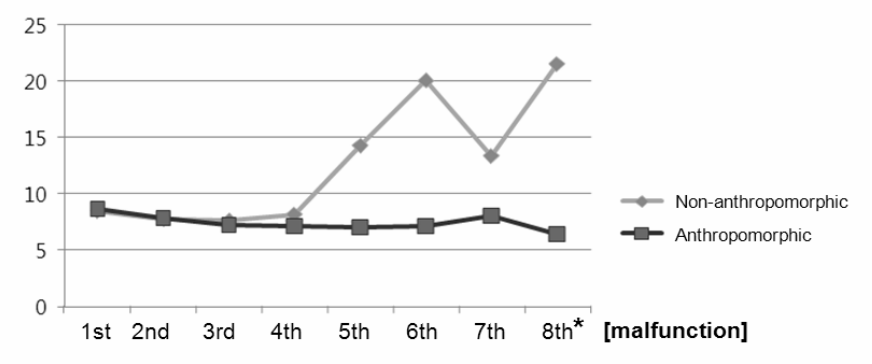


Figure.2 Efficiency of feedback
(*denotes a case where $p<0.05$)

### 4.3 Satisfaction of the Feedback

As predicted by H3, anthropomorphic feedback showed higher satisfaction than non-anthropomorphic feedback on all the factors of satisfaction except on performance and comfortableness factors. Especially, larger differences by feedback type were observed regarding novelty, excellence and pleasure factors.

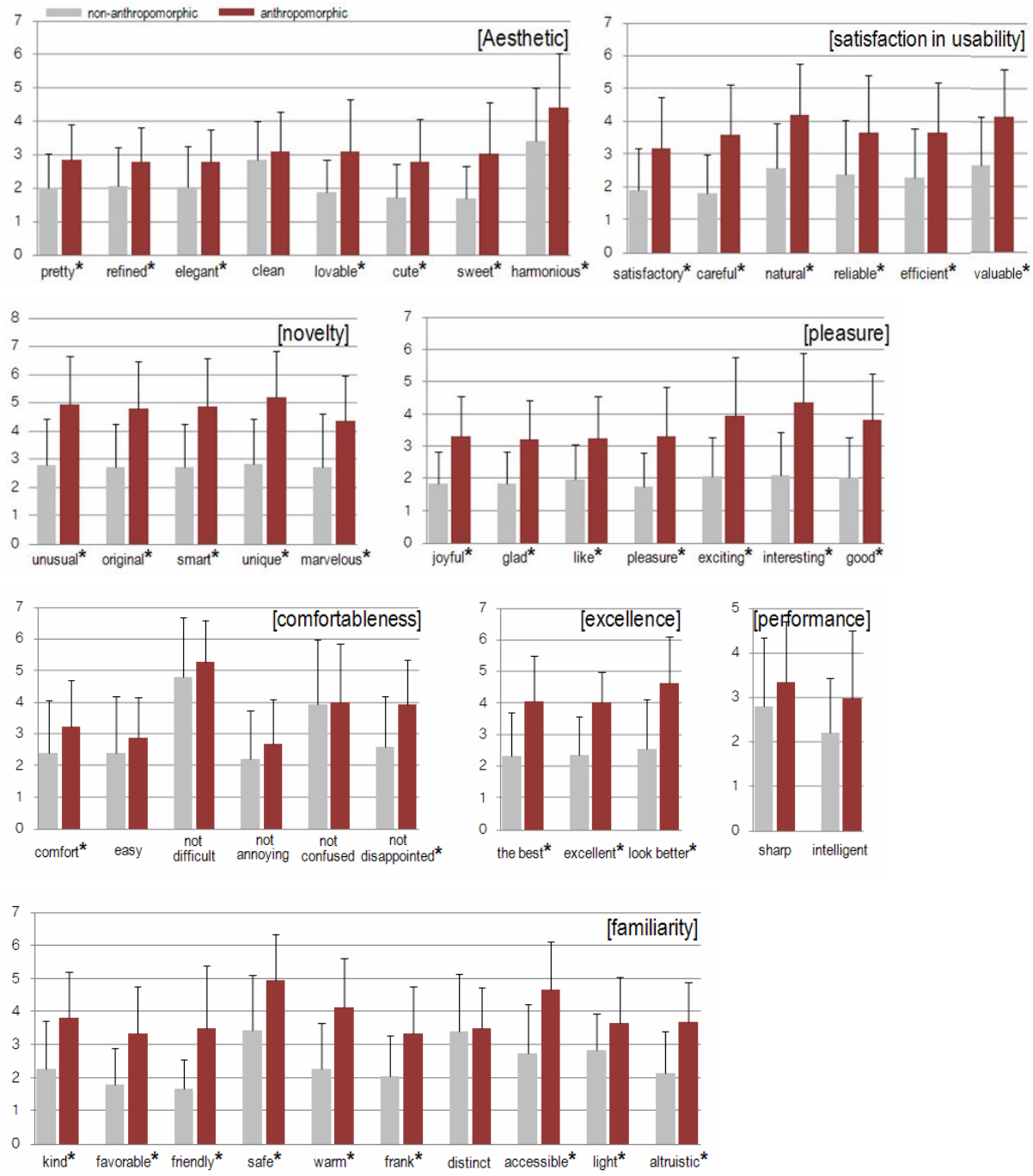


Figure.3 Satisfaction of the Feedback
(*denotes a case where $p<0.05$)

In addition, we compared the satisfaction on the feedback in the repeated experience of malfunctions with that in the single experience of malfunction which was examined in authors' previous study [6]. Except the novelty factor, the satisfaction of the feedback in the repeated experience of malfunctions is slightly smaller than that in the single experience of malfunction. This demonstrates the novelty effect of the feedback for service robot malfunctions. Also, the differences by feedback type in the repeated experience of malfunctions are larger than that in the single experience. This indicates that the effect of the anthropomorphic feedback on satisfaction increases as the malfunction repeats.

### 4.4 Participants' Attitudes towards the Feedback

In order to understand participants' attitudes toward the feedback, the multi-text messages, which were sent by participants, were analyzed by protocol analysis. In this study, we used the proportional reduction in loss (PRL) approach [19] to measure the interceders reliability, which is a measure of agreement between multiple coders about how they apply codes to the data [20]. The adequate level of agreement was defined as over .070 [19] because this study examined participants' emotional changes toward the feedback of malfunction.

Two coders participated in coding the data. Among 67 participants' samples, 26 samples were excluded because 26 participants sent prerecorded text messages. All samples except one exceeded the adequate PRL level ($M$=0.983). The 41 samples were analyzed by the coding scheme described in section 3.4 in order to examine how participants' emotion changed over repeated malfunctions. Figure.4 shows the results of protocol analysis.

As predicted by H4, most participants who experienced non-anthropomorphic feedback got irritated and angry earlier than those who experienced anthropomorphic feedback. For instance, participant no.52 displayed anger continually when experiencing non-anthropomorphic feedback. On the other hand, participant no.32 had no significant emotional change when experiencing anthropomorphic feedback.

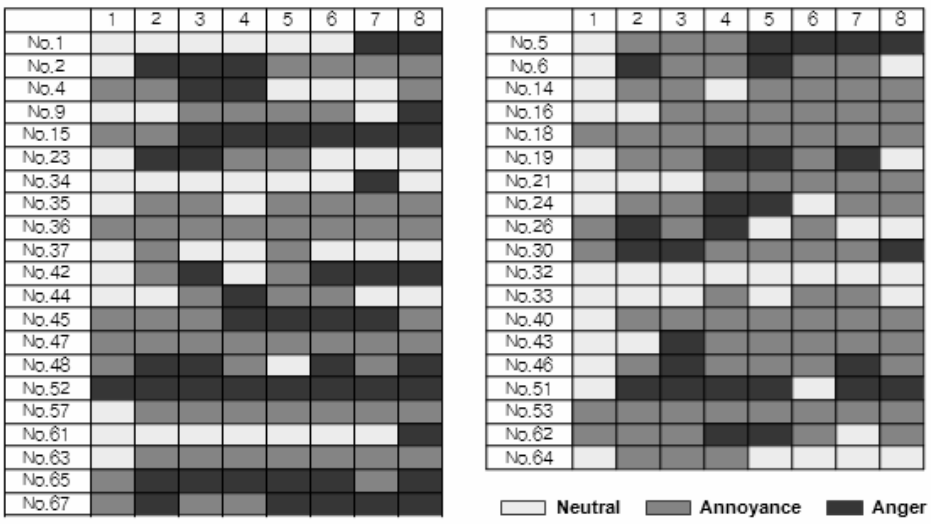


Figure.4 Results of the protocol analysis

We also analyzed the video recording of participants' interaction with the robot. Participants who experienced non-anthropomorphic feedback displayed more violent behaviors than those who experienced anthropomorphic feedback. Participant no.1 and no.45 slapped the cleaning robot with non-anthropomorphic feedback, and participant no.1 and no. 17 cursed the robot with non-anthropomorphic feedback. Even though some participants who experienced anthropomorphic sound feedback displayed violent behaviors, they showed relatively mild level of violence. Participant no.3 sighed toward the robot with anthropomorphic sound feedback.

Additional findings were drawn from the interviews. Participants who experienced anthropomorphic feedback became intimate with the robot as malfunction repeated. They felt as if the cleaning robot was relying on them and perceived the robot as a pet. This indicates that anthropomorphic feedback increases the lifelikeness of the robot which induces intimate relationships between the user and the robot, which supports H4.

Moreover, repeated experience of the non-anthropomorphic sound feedback increased lifelikeness of the robot. In the beginning of the experiment, participants who treated the cleaning robot as a living thing were mostly those who experienced anthropomorphic feedback. However, as malfunction repeated, participants who

experienced non-anthropomorphic feedback also perceived the robot as a living thing. This indicates that the autonomy of the robot increases lifelikeness of the robot when errors are repeated.

## 5. Conclusions

The aim of this study was to apply anthropomorphic elements to the feedback design of service robot malfunction and to investigate the usability of anthropomorphic sound feedback through repeated malfunction. We used a 2 (type of sound feedback: anthropomorphic vs. non-anthropomorphic) between-participants experiment design. Participants experienced either anthropomorphic sound feedback or non-anthropomorphic sound feedback through repeated malfunction. University students or graduate students from an engineering college participated in the experiment, and a cleaning robot prototype was developed for absolute control. The usability of sound feedback was evaluated through questionnaires, and participants' behaviors were analyzed by protocol analysis, video analysis, and interviews.

All of the hypotheses were supported by the data. As predicted by H1, anthropomorphic sound feedback was more effective than non-anthropomorphic feedback in the repeated serial experience of malfunctions. As the malfunction repeated, the number of failures with non-anthropomorphic sound feedback increased while the number of failures with anthropomorphic sound feedback decreased.

As predicted by H2, anthropomorphic sound feedback was more efficient than non-anthropomorphic feedback in the repeated experience of malfunction. As the malfunction repeated, the time elapsed between the start of the experiment and the participants' perception of the malfunction became longer in the case of non-anthropomorphic feedback while it hardly changed in the case of anthropomorphic feedback.

H3 was supported by questionnaire data. Anthropomorphic feedback was more satisfactory than non-anthropomorphic feedback on all the factors of satisfaction except the performance and the comfortableness factors. Moreover, the satisfaction of the feedback in the repeated experience of malfunctions was slightly smaller than that in the single experience of malfunction. This result explains the novelty effect of the service robot. Also, the differences by feedback type in the repeated experience of malfunctions are larger than that in the single experience of malfunctions.

H4 was supported by data from protocol analysis, video analysis, and interviews. Using protocol analysis, we found that most participants who experienced non-anthropomorphic feedback got irritated and angry earlier than those who experienced anthropomorphic feedback. Through video analysis, we discovered that participants who experienced non-anthropomorphic feedback showed more violent behaviors than those who experienced anthropomorphic feedback. Moreover, through interviews, we found that anthropomorphism increases lifelikeness of the service robot by inducing the intimate relationship between the user and the robot, and the autonomy of the robot increases lifelikeness of the robot with non-anthropomorphic feedback when feedback repeated. The results of the experiment provide an essential basis for designing the feedback of service robot malfunction.

## 6. Reference

[1] Roy.N., Baltus.G., Fox.D., Gemperle.F., Goetz.J., Hirsch.T. & Margaritis.D., Montemerlo.M., Pineau.J., Schulte.J. & Thrun.S. (2000) Towards Personal Service Robots for the Elderly. *Workshop on Interactive Robots and Entertainment 2000.*

[2] IFR., n.d., Service robots – classification and definition. Available at<,http://www.ifr.org/modules.php?name

=News&file=article&sid=14>

[3] Fong.T., Nourbakhsh.I. & Dautenhahn.K. (2003) A Survey of socially Interactive Robots. *Robotics and Autonomous Systems*, vol. 42, no.3-4, pp.143-166.

[4] Norman.D. (2007) *The Design of Future things*. Basic Books.

[5] Institute for Information Design Japan. (2002) *Information Source Book*. Graphic-sha Publishing co.,Ltd.

[6] Kim, B.M. Kwak, S.S. and Kim, M.S. (2009) Design guideline of anthropomorphic sound feedback for service robot malfunction – with emphasis on vacuum cleaning robot. *Submitted to the 18th IEEE International Symposium on Robot and Human Interactive Communication(RO-MAN).*

[7] Sung.J.Y., Grinter.R.E., Christensen.H.I. & Guo.L. (2008) Housewives or Technophiles?: Understanding Domestic Robot Owners. *Proceedings of the 3rd ACM/IEEE international conference on Human Robot interaction.*

[8] Nass.C., & Brave.S. (2005) *Wired for Speech : How voice activates and advances the human-computer relationship*, The MIT Press.

[9] Rauterberg.M., & Styger.E. (1994) Positive Effects of Sound Feedback during the Operation of a Plant Simulator. *Human Computer Interaction.*

[10] Lee. D.K. (2007) Application of biological signal as an idle feedback for emotion expression of robot. M.S. thesis, Dept. of Industrial Design, KAIST, Daejeon, Korea, 2007.

[11] Yamaoka.F., Kanda.T., Ishiguro.H., & Hagita.N. (2005) "Lifelike" behavior of communication robots based on developmental psychology findings. *2005 5th IEEE International Conference on Humanoid Robots.*

[12] Educational Robot Kit "Bioloid". Available at < http://www.robotis.com>

[13] Wizard of Oz. (retrieved in November, 2006). available from wizard.htm

[14] http://cafe.naver.com/robotclear.cafe

[15] Kim. H.J. (2004) A study on the development of earcon prototype for effective navigation in information architecture: with emphasis on the development of earcon prototype for mobile phones. KAIST, Daejeon, Korea, 2004

[16] (1998) ISO 9241-11: Guidance on Usability. Available at< http://www.usabilitynet.org/tools/r_international.htm>

[17] Jung. S.H. (2007) The effects of the usability of products on user's emotions. KAIST, Daejeon, Korea, 2007.

[18] Kanda.T., Ishiguro.I., & Ishida.T. (2001) Psychological analysis on human-robot interaction. *Proceedings of the 2001 IEEE International Conference on Robotics & Automation,* Seoul, Korea.

[19] Rust.R.T. & Cooil.B. (1994) Reliability measures for qualitative data: theory and implications. *Journal of Marketing Research*, vol.31, no.1, pp.1-14.

[20] Kurasaki, K. S. (2000) Intercoder Reliability for Validating Conclusions Drawn from Open-Ended Interview Data. *Filed Methods*, vol. 12, no. 3, pp.179-194.