# Effects of an Agent Feature Comprehension on the Emotional Attachment of Users

Yoshito Ogawa[1] Kaho Harada[2] Hideaki Kikuchi[3]

[1] Graduate School of Human Sciences, Waseda University
[2] School of Human Sciences, Waseda University
[3] Faculty of Human Sciences, Waseda University

**Abstract:** In this paper, we consider effects of an agent's comprehension of user features on a user's sense of emotional attachment to the agent. Recently, some research studies in Human-Agent Interaction (HAI) have examined the strategies that make users continue to use spoken dialogue systems over a long period of time. In this paper, we propose a system that discriminates a user's activity from the speech prosody and accumulates the prosody and the actual activity as training data to use in subsequent discrimination. Our results indicate that our proposed system performs more stable discrimination than heretofore, and that higher discrimination accuracy inspires a sense of emotional attachment more effectively.

## 1 INTRODUCTION

Currently, facilitated by advances in key technologies such as speech recognition and image recognition, advanced systems for interaction with users are being implemented. These systems are not only handling tasks with fixed roles such as book searches and booking of tickets, but also the interaction itself or the entertaining of users via interaction, such as pet robots [4]. However, the problem with such systems is that at the outset, users actively dialogue with the system, but as time progresses, boredom sets in and users lose interest in having dialogues with the system, and so begin to use it less frequently. It is believed that factors such as novelty and humor can improve the sustainability of dialogue continuance desire in the Human-Agent Interaction (HAI) field and, as a result, research is actively being conducted in this direction.

We believe that there are factors that can improve the sustainability of dialogue continuance desire besides novelty and humor, and therefore focus on a user's sense of emotional attachment to a system. Munakata et al. [8] have already shown that a sense of emotional attachment to an interface makes a user want to use it for a long time even if there is no new development on the content side. In this paper, using the hypothesis that the sustainability of dialogue continuance desire with a spoken dialogue system is also improved by a user's sense of emotional attachment to a system, we propose and verify a technique to inspire a sense of emotional attachment to a spoken dialogue system.

In this paper, we define "sense of emotional attachment to a system" as the sense that a user is conscious that a new value is not inherent in the system and wants to continue to use the system rather than other systems. Emotional attachment to a family or a friend has been studied in the fields of developmental and social psychology. In those fields, it is often defined as an emotional bond between one person and another another [1]. Nakata et al. classified emotional attachment as a sense of belonging, and defined emotional attachment as a feeling of wanting to strengthen a connection between a target and oneself. They also argue that emotional attachment is able to target even things that do not posses a will, such as plants, non-biological objects, and locations [3]. In addition, Kino et al. showed that similarities exist between emotional attachment to a person and to a thing [6]. Saito et al. further clarified that a user feels "Personification Partnership," meaning that users tend to personify and characterize information equipment during their interactions with it [9]. Hashimoto et al. cited the following four factors for sensing emotional attachment to a thing, and argues that by possessing them users are induced to feel "Not abandoned" and "I want to continue to use it" [10]:

**(1)** Structure and function of the thing is excellent, therefore users are able to experience the height of performance and ease of use.

**(2)** Material and texture of the thing is superb, therefore users feel pleasant.

**(3)** The thing has features; therefore, users take pleasure in possessing it.

**(4)** Users are conscious that a new value is not inherent in the thing, such as experience of encounter and memories.

Of the above factors, factors (1) to (3) are values that are innate in the thing (primary value), factor (4) is value that a user adds to it (secondary value). The system to be used in our study was not owned by an individual user. It operated by user speech, and its internal processes were hidden from users, therefore, its users did not know its structure, composition, and exact function. Accordingly, factors (1) to (3) ware

not applicable in this study, and so we defined emotional attachment with focus on factor (4).

# 2 PROPOSED APPROACH

In this paper, we propose comprehension of user feature by a system as an approach to inspire a sense of emotional attachment in the user. In this approach, the spoken dialogue system learns about the user through interaction with the user, and gives feedback as system reactions. The feedback imbues in the user a secondary value, "This system understands my own feature," and inspires a sense of emotional attachment. In this study, we adopt the relationship between activity and prosody, which is believed to make a large individual difference in user feature comprehension, and we construct a model to discriminate between speech emitted in a state of high and low activity as the method to comprehend user's activity. In addition, we adopt the changing of system utterances according to user activity as system reaction. Further, to efficiently inspire a sense of emotional attachment, our system collects training data about users via practical interaction, which it then uses to gradually increase discrimination accuracy by constructing a discriminant model specialized to the user. We believe that by gradually increasing discrimination accuracy, a user feels the system that did not understand his feature gradually comprehends it—which inspires a sense of emotional attachment to the system.

# 3 SYSTEM DESIGN

In our study, we used a spoken dialogue system for book management by means of RFID tags in experiments. The system comprised communication robot PaPeRo (manufactured by NEC) and RFID reader (manufactured by Sobal Co., Ltd.). The borrowing and returning of books by each user is done by allocating an RFID tag to each book and user. Operations other than closing the RFID tag is performed using speech dialogue in this system. In our study, we conducted evaluation experiments in which we added a dialogue to discriminate user activity before book management tasks.

## 3.1 Flow of Activity Discrimination Dialogue

Fig. 1 illustrates the flow of dialogue that is used to discriminate user activity. Explanation for the numbers in Fig. 1 are given below.

**(1)** When the user brings the RFID tag close to the RFID reader, the system recognizes the user.

**(2)** After obtaining the user ID, the system greets the user with "Hello," to which the user replies with "Hello" too.

S: System, U: User

| | | |
|---|---|---|
| U : | ( Closing RFID tag to the RFID Reader ) | ・・・(1) |
| S1 : | Hello. | |
| U1 : | Hello. | ・・・(2) |
| S2 : | You look down, | |
| S3 : | aren't you? | ・・・(3) |
| U2 : | Yes. | ・・・(4) |
| S4 : | Are you all right? Cheer up! | ・・・(5) |

Figure 1: Flow of Dialogue

**(3)** System discriminates user activity from greeting speech, and, according to user activity, replies with "You are looking good!" or "You look down. "

**(4)** System questions the success or failure of discrimination in (3). After the user answers, system stores in the database the feature value of speech with the correct label that system created from the user answer.

**(5)** In order to make the dialogue flow naturally, according to success or failure of activity discrimination and correct activity, system makes utterances such as "I'm glad to be correct! ", "I'm sorry made a mistake. " or "Are you all right? Cheer up! ".

## 3.2 Disicrimination Method

In our study, we adopted the maximum likelihood estimation model based on Gaussian Mixture Model (GMM) as the activity discrimination model. In order not to disrupt the discrimination process even when training data was small, we selected a few features from the previous studies [5][11]. The features adopted are displayed in Table 1. We used the z-score of these features practically. In this system, there is a need to re-build the discriminant model when system discriminates user activity, because training data increase in every dialogue. The flow of activity discrimination in the system in as follows:

**(1)** Load training data (high activity group and low activity group)

**(2)** Learn GMMs from training data (GMM1 from high activity group, GMM2 from low activity group)

**(3)** Extract features from discrimination target speech

**(4)** Calculate likelihoods from GMM1, GMM2

**(5)** Discriminate user activity using Maximum likelihood estimation

This system held 40 pre-recorded speeches emitted in a state of high and low activity from eight speakers

Table 1: Feature Values

| Type | Feature Value |
|---|---|
| Fundamental Frequency (F0) | Mean Value |
| | Maximum Value |
| | Minimum Value |
| | Range |
| | Mean of Frame Difference |
| Power | Mean Value |
| | Maximum Value |
| | Minimum Value |
| | Range |
| | Mean of Frame Difference |
| | Value when F0 is Maximum Value |
| Time | Duration of Speech |

(men and women) in a database as general discriminant criterion before it comprehended user feature. Before a user used the system two times, the system discriminated at random because it could not calculate z-score. Following that, the system used pre-recorded speeches and the user's past speeches as training data. The system treats a user's past speech as 10 speeches to learn the discriminant criterion specific to the user quickly.

# 4 EVALUATION

In order to demonstrate the effectiveness of our proposed approach, we conducted an evaluation experiment. In the experiment, we had subjects use our system, and evaluated their degree of emotional attachment to the system over a period of times.

## 4.1 Purpose

The purpose of this experiment was to verify the hypothesis that feature comprehension by the system inspires a sense of emotional attachment. We focused on the following 3 points:

**(1)** Discrimination Accuracy

**(2)** Change of Emotional Attachment

**(3)** Learning Discriminant Criterion Specific to the User

## 4.2 Procedure

### 4.2.1 Preparation

We had subjects answer a questionnaire to evaluate whether they liked PaPeRo, which is used in the system, and divided them into test group and control group based on its results. In the questionnaire, we implemented nine-grade evaluation with an improved version of Japanese Love-liking scale applied to an agent [7], and used the total score from 7 items as favor degrees. The following items comprised the questionnaire:

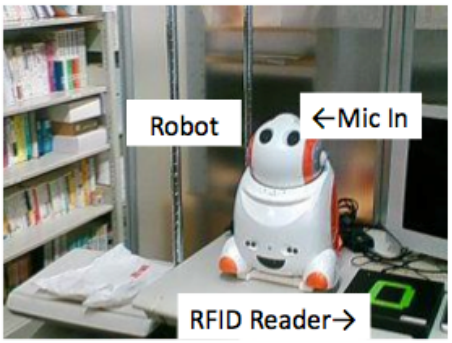


Figure 2: Experimental Environment

**(1)** Even with PaPeRo, I still did feel any change from the usual.

**(2)** I think that PaPeRo is a robot with the ability to adapt.

**(3)** I am able to leave responsible work to PaPeRo.

**(4)** PaPeRo is good compared with other robots.

**(5)** I trust PaPeRo's judgements.

**(6)** I think that PaPeRo is a being that is liked by many people.

**(7)** I prefer PaPeRo to other robots I know.

### 4.2.2 Using the System

Subjects used the system and assumed that they were borrowing books from and returning books to the laboratory. The experimental environment used is depicted in Fig. 2. The Test group used the system that comprehended user features, while the Control group used a system that discriminated using unchanged general criterion.

### 4.2.3 Evaluation of Impressions

In order to evaluate whether a sense of emotional attachment was inspired in the subjects, we had them answer a nine-grade questionnaire related to a sense of emotional attachment. The items comprising that questionnaire are listed below:

**(1)** I feel PaPeRo accepts me.

**(2)** If I lose PaPeRo, I have no problem with getting a replacement.

**(3)** For me, PaPeRo is not just a book management system.

**(4)** It is a problem for me that the performance of PaPeRo deteriorates.

**(5)** Even if there is a book management system that is better in terms of function and quality, I still want to use PaPeRo.

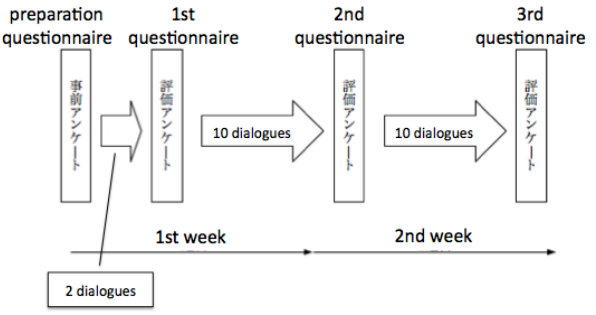


Figure 3: Flow of the Experiment

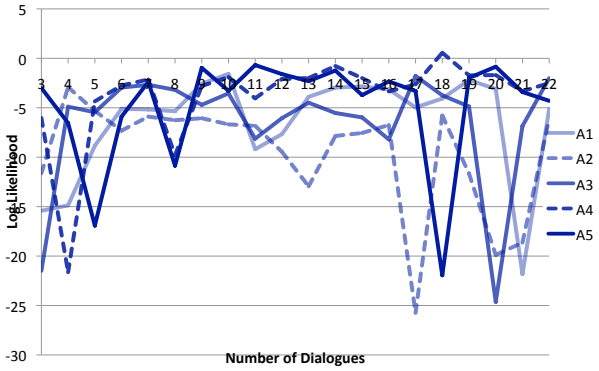


Figure 4: Log Likelihood

**(6)** My feelings for PaPeRo are one way.

**(7)** PaPeRo is able to discriminate whether I am cheerful or not.

Of the above items, items (1) to (6) are related to the sense of emotional attachment; item (7) measures whether subjects correctly recognize the discrimination accuracy of the system; items (2), (4) and (6) are inverse items. We prepared items (1) to (6) by referencing previous research related to emotional attachment to things [2] and scale of emotional attachment in interpersonal relationships. We considered the total score of items (1) to (6) ((2), (4), (6) to inverse) of the above as the sense of emotional attachment score.

The flow of the experiment conducted is shown in Fig. 3. To equalize the number of dialogues with the system and the experimental period as much as possible, we had subjects have a dialogue three days in a week; the first day twice, other days four times per day—a total of 22 times in two weeks—and answer a questionnaire related to their sense of emotional attachment after the 2nd, 12th, and 22nd dialogues. We call these questionnaires first questionnaire, second questionnaire, and third questionnaire below. Because the first two times it is used the system is not able to calculate z-scores and discriminate user activity at random, the subjects in both groups answered were under the same conditions for the first questionnaire. We observed changes in the sense of emotional attachment by using subsequent questionnaires.

## 4.3 Results and Considerations

Our subjects comprised 10 persons of researchers and students from our laboratory. According to the result of the questionnaire for measuring whether the subject likes PaPeRo, we had five subjects in the test group and the others in the control group. We labeled the subjects in the test group subject A1 to A5, and subjects in the control group subject B1 to B5. The results of our experiment are described and analyzed below with respects to each point described in section 4.1.

### 4.3.1 Discrimination Accuracy

The discrimination accuracy of each subject, and the mean of the test and control groups are shown in Table 2. Table 2 shows that in the test group, the discrimination accuracy of all subjects was 0.70, or more in the range of 13th to 22nd (in 13th-22nd) dialogues. For subject A5, in particular, the discrimination accuracy increased from 0.40 in the 3rd-12th dialogues to 0.70 in the 13th-22nd. In contrast, although there ware subjects with high discrimination accuracy (B2, B3), there was a subject with discrimination accuracy 0.50 (B1), and a subject whose discrimination accuracy in the 13th-22nd dialogues decreased compared to the 3rd-12th. From these results, we believe that subjects A5, B1, and B5 have features that are difficult to discriminate correctly by general discriminant criterion, and subjects B2 and B3 have features that are concordant with general discriminant criterion. Thus, we can say that our proposed approach functions effectively even for a user who has features that are difficult to discriminate correctly by general discriminant criterion from the discrimination accuracy of subject A5 increased to 0.70 in the 13th-22nd dialogues.

Furthermore, in order to analyze the effect of accumulation of training data on discrimination, we calculated the likelihoods of the test group. The log likelihood (The larger value of to GMM of high activity and low activity) in the 3rd-22nd dialogues in Fig. 4. From Fig. 4, it can be seen that although there are some small values up to the 5th dialogue, likelihood is stable at a comparatively large value in the 6th-16th dialogues. This suggests that the system learned rough features of each subject in five dialogues. However, there ware some subjects whose likelihood was reduced again after the 16th dialogue. This likely occurred because the speeches in those dialogues had features that were appearing for the first time. From these results, it can be inferred that learning rough features requires about five dialogues, whereas learning detailed features requires more dialogues.

Table 2: Discrimination Accuracy

| Subjects | Dialogue Count | | |
|---|---|---|---|
| | 3rd-12th (First half) | 13th-22th (Second half) | 3rd-22th (Whole) |
| A1 | 0.70 | 0.80 | 0.75 |
| A2 | 0.70 | 0.70 | 0.70 |
| A3 | 0.70 | 0.70 | 0.70 |
| A4 | 0.80 | 0.90 | 0.85 |
| A5 | 0.40 | 0.70 | 0.55 |
| Mean of Test Group | 0.66 | 0.76 | 0.71 |
| B1 | 0.50 | 0.50 | 0.50 |
| B2 | 0.80 | 1.00 | 0.90 |
| B3 | 0.90 | 1.00 | 0.95 |
| B4 | 0.50 | 0.70 | 0.60 |
| B5 | 0.80 | 0.60 | 0.70 |
| Mean of Control Group | 0.70 | 0.76 | 0.73 |

#### 4.3.2 Change in Sense of Emotional Attachment

In order to analyze changes in the sense of emotional attachment of the subjects, we calculated the differences from the first questionnaire to the second questionnaire and from the first questionnaire to the third questionnaire. The changes in the scores across the questionnaires are illustrated in Fig. 5. Fig. 5 shows that score changes in the test group are in a comparatively narrow scope, whereas they are more dispersed for the control group. With the exception of subjects B2 and B3 whom the system was able to discriminate virtually correctly, there is a tendency for the score changes to be larger in the test group than in the control group in the second and third questionnaires. This shows that our proposed approach inspires a sense of emotional attachment that is more stable than the method using only general discriminant criterion. In addition, this tendency is similar to the discrimination accuracy shown in Table 2, and there is also a tendency for the sense of emotional attachment to increase with discrimination accuracy. We also analyzed how subjects recognized the discrimination accuracy correctly. The scores for item (7) in the second and third questionnaires and the discrimination accuracy are shown in Figs. 6 and 7. Although the sample size is small, there are a strong positive correlation between the scores and the discrimination accuracy. (Correlation Coefficient: 0.88 in the second questionnaire, 0.83 in the third questionnaire.) This shows that subjects recognized the discrimination accuracy correctly. In summary, we have shown that there is a tendency for a user to feel that a system understands his own feature more correctly and is effectively imbued with a sense of emotional attachment by using the system with higher discrimination accuracy.

Furthermore, for the five subjects whose discrimination accuracy was not reduced the sense of emotional attachment was reduced from the second questionnaire to the third (subjects A2, A3, A4, B2, B3).

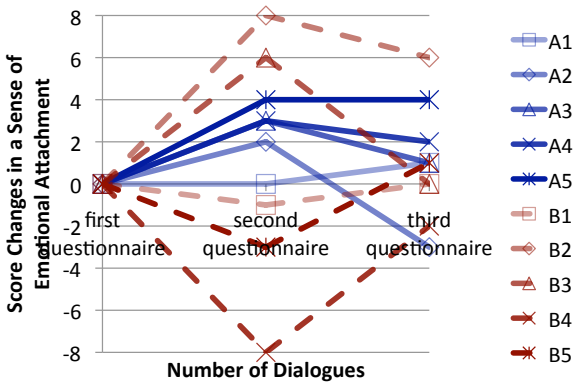


Figure 5: Score Changes in a Sense of Emotional Attachment

From the fact that these subjects have a comparatively high discrimination accuracy in the second questionnaire in common, we deduce that they had moved past the stage of recognizing the system understanding their own feature gradually over time and been satiated with no changes dialogue with the system. Fig. 4 shows that the system is not able to learn a users features in detail from only 20 dialogues. In spite of that, subjects had been satiated with the system dialogues. This indicates the following: If the system learns user features and shows the result of learning to the user, the user becomes satiated before the system learns his feature in detail. In order to avoid that, we considered the adjusting the timing of learning features and showing the result of learning (e.g., the system learns from all dialogues and takes some actions only when discrimination of user's low activity speeches) is effective.

#### 4.3.3 Learning Activity Discrimination Model Specialized for Each User

We analyzed how the activity discrimination models were specialized for each subject after learning 22 dialogues in the experiment. For example, for subjects A2 and A5 the F0 mean model shows that the difference is significant with each other in Fig. 8 and 9. Although there is a tendency that F0 mean is higher in a state of high activity than low activity in both subjects, in subject A2 there are three peaks of the low activity state on position lower than the peak of the high activity state, and in subject A5, there are two peaks of the low activity state on position holding two peaks of the high activity state between. These figures show that activity discrimination models specialized for each subject were learned in our proposed approach.

## 5 CONCLUSION

In this paper, we focused on a sense of emotional attachment as a factor that can improve the sustainabil-

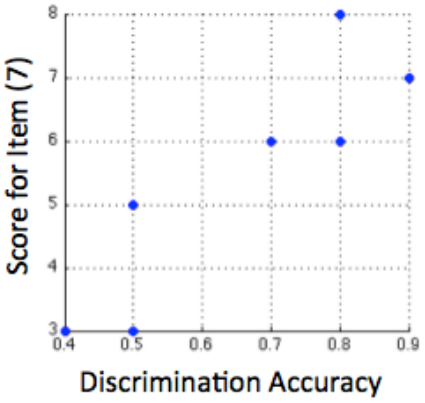


Figure 6: Score for Item (7) in the second Questionnaire and Discrimination Accuracy in 3rd-12th Dialogues

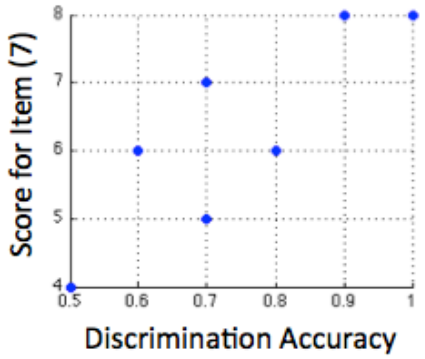


Figure 7: Score for Item (7) in the third Questionnaire and Discrimination Accuracy in the 13th-22th Dialogues

ity of dialogue continuance desire, and propose an approach that inspires a sense of emotional attachment by comprehending the features of each user. Specifically, the approach uses the relationship between activity and prosody as user features, uses maximum likelihood estimation using GMM as a discrimination method, and that discriminates user activity from prosody, gets the actual activity from user response, accumulates the prosody and the actual activity as training data to use in subsequent discrimination, increases discrimination accuracy gradually, and makes a user feel "this system understands my own feature" and inspires a sense of emotional attachment via these functions. The results of the evaluation experiment prove that discrimination accuracy is improved more steadily in our proposed approach than in the method using only general discriminant criterion, there is a tendency for the user to feel that the system understands his own feature more correctly and inspires a sense of emotional attachment more effectively by using a system with higher discrimination accuracy.

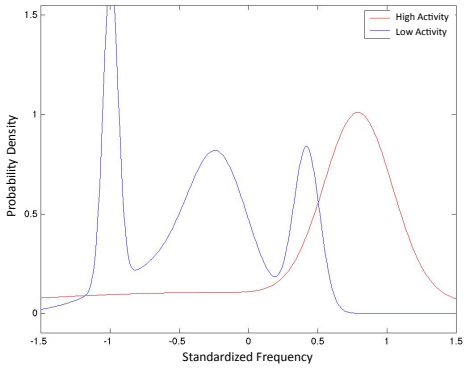


Figure 8: GMM of F0 Mean (Subject A2)

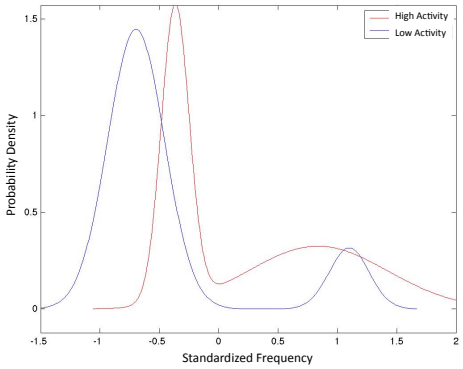


Figure 9: GMM of F0 Mean (Subject A5)

## 5.1 Future Works

In this study, we proposed an approach for inspiring a sense of emotional attachment more effectively in users. The effects of the sense of emotional attachment on dialogue continuance desire and the size of the sense of emotional attachment required to sustain dialogue continuance desire need further study. Detailed verification is also required for the reduction of the sense of emotional attachment that is observed in the evaluation experiment and which seems to be cause satiation. In addition, we plan to research the adjustment number and timing of learning features and how to show the result of learning.

Moreover, in this study we extracted 12-dimensional feature values and used all of them for learning GMM. We clarified that each user has different features that are effective for discrimination. We believe that by extracting many features and selecting only the valid features to learn the activity discrimination model for each user from those features will improve accuracy of user features comprehension.

## References


[1] 遠藤利彦, 数井みゆき: ”愛着理論の基本的枠組み”アタッチメント-生涯にわたる絆, pp.1-31, ミネルヴァ書房, 東京 (2005).

[2] 後藤真一, 椎塚久雄: モノにまつわる体験とモノに抱く愛着感との関連, 工学院大学研究報告 (110), pp.97-103 (2011).

[3] 橋本英治, 寺内文雄, 久保光徳, 青木弘行, 鈴木邁: モノに対する愛着の体系化, デザイン学研究 研究発表大会概要集, Vol.45, pp.28-29 (1998).

[4] 石黒浩, 日浦亮太: コミュニケーション支援ロボットビジネス, 日本ロボット学会誌, Vol. 20, No. 7, pp.672-675 (2002).

[5] 門谷信愛希, 阿曽弘具, 鈴木基之, 牧野正三: 音声に含まれる感情の判別に関する検討, 情報処理学会研究報告. SLP, 音声言語情報処理 2000(119), pp.43-48, (2000).

[6] 木野和代, 岩城達也, 石原茂和, 出来原裕順: モノへの愛着の分析-対人関係とのアナロジによる測定, 感性工学会研究論文集, Vol. 6, No. 2, pp.33-38 (2006).

[7] 小松孝徳, 山田誠二: 適応ギャップがユーザのエージェントに対する印象変化に与える影響, 人工知能学会論文誌, Vol.24, No.2, pp.232-240 (2009).

[8] 棟方渚, 小松孝徳, 松原仁: ユーザの愛着を促すインタフェース”あるくま”, 情報処理学会研究報告. EC, エンタテインメントコンピューティング, pp.15-20 (2008).

[9] 中田亨, 佐藤知正, 森武俊, 溝口博: ロボットの対人行動による親和感の演出, 日本ロボット学会誌, Vol. 15, No. 7, pp.1068-1074 (1997).

[10] 齋藤力也, 寺内文雄, 久保光徳, 青木弘行: モノに抱く愛着感と生活観との関係解明, デザイン学研究 研究発表大会概要集, Vol. 54, pp.38-39 (2007).

[11] 重永實: 感情の判別分析からみた感情音声の特性, 電子情報通信学会論文誌. A, 基礎・境界, Vol. J83-A, No. 6, pp.726-735, (2000).