Character Expression of a Conversational Robot for Adapting to User Personality

Kenta Yamamoto, Koji Inoue, Tatsuya Kawahara

Graduate School of Informatics, Kyoto University, Kyoto, Japan

ARTICLE HISTORY
Compiled January 5, 2024

ABSTRACT
User adaptation is important in conversational robots to enhance the user experience and engagement. One way of user adaptation is to change the dialogue behavior. A character (e.g., extrovert or agreeable) can be defined to achieve human-like behaviors and user adaptation, and the appropriate character differs depending on each user. In this study, we investigate user adaptation by character expression of conversational robots. Our character expression model adopts the Big Five traits for controlling four dialogue behaviors: amount of utterance, backchannel frequency, filler frequency, and switching pause length. We cluster the system character and user personality into four classes based on an analysis of a speed dating dialogue corpus, and we found the best combinations between the system character and user personality. We implement the character expression model into a laboratory guide and a chit-chat robot, and conducted subjective experiments, where the subjects talked with robots with four different characters and evaluated their impressions of each character. The results shows that the Role model character was preferred for the task-oriented dialogue of the laboratory guide, but a robot character complementary to the subject personality was preferred for the non-task-oriented dialogue of the speed dating and chit-chat.

KEYWORDS
Conversational robot; User adaptation; Character; Personality

1. Introduction

User adaptation of conversational robots is to make the systems generate behaviors appropriate to the user, which leads to increasing user satisfaction with human-robot interaction. Usually, classification of the user is conducted for user adaptation. A widely-used user classification is personality. Many studies on personality estimation [1–3] suggest that the user personality can be estimated through spoken dialogue. It was also shown that users with different personalities had different impressions of a robot [4,5]. Therefore, user adaptation based on the user personality is desired to achieve a satisfactory user experience. In this study, we analyze the effect of a character expression, where a robot expresses the character appropriate to the user personality.

It has been shown that the character expression of robots leads to increasing user engagement and naturalness in dialogue [6]. Here, "personality" is used as a psychological dimension for classifying users, and "character" is used as an impression that

CONTACT Kenta Yamamoto. Email: kentayamamoto@sanken.osaka-u.ac.jp

the robot gives to the user in this study. Character is the user's impression of the consistency of the robot's behavior. Therefore, by investigating the optimal character for the user, we can simultaneously investigate optimal actions from multi-modal behaviors. First, we define the user personality with four classes: Role model, Reserved, Self-centered and Introvert. Then, we define the robot character with four classes: Role model, Reserved, Introvert and Neutral. We investigate the validity of these classifications using a speed dating corpus. Then, we conduct two subjective experiments, where the robot is given a social role as a laboratory guide (task-oriented dialogue) and a chit-chat (non-task-oriented dialogue), to confirm whether the combination of the user personality and the system character affects the impression of the dialogue.

The rest of this paper is organized as follows. In Section 2, we describe the related work on character expression and user adaptation. In Section 3, we explain the methods of the proposed character adaptation. We present the evaluation of character adaptation on task-oriented dialogue in Section 4 and on non-task-oriented dialogue in Section 5.

2. Related work

2.1. Character expression model

In previous studies, many character expression methods have been proposed. A sentence generation system PERSONAGE [7] creates response sentences according to the Big Five parameter. The widely-used approach for defining a character uses sentences providing system's information like "I like to travel", which is called *persona* [8]. Then, it generates utterances with personality assigned to them by inputting dialogue act tags and personality-related features to a sequence-to-sequence model based on LSTM [9]. A chat corpus with speaker information for model training is also available [8]. Then it is possible to generate characterized utterances by embedding speaker information in the decoder of a sequence-to-sequence model [10]. A character expression method using role-playing dialogue data was also proposed [11]. Another work proposed to maintain dialogue consistency according to the persona using large-scale language models [12].

However, these methods require constructing a new dialogue system for character expression. In this study, we address a character expression method that uses spoken dialogue behaviors, enabling character expression using existing spoken dialogue systems.

2.2. User adaptation of conversational agents and robots

Studies about human robot interaction showed that personality affects the user's preferences and impressions of the system [4,5]. These studies use the behavior of robots and agents. In previous studies of dialogue systems, user adaptation methods controlled system utterances. It was shown that users are more trusting and satisfied with chat systems expressing the character similar to user personalities [13]. Some models that obtain user persona information from the dialogue history were proposed [14–17]. The system expresses personas according to the acquired user personas [18]. Other methods were proposed using the user's intent and proficiency with the system, or identifying the user's preferences [19,20]. The goal of these studies is to generate utterances on topics appropriate to the user. In this study, we propose a character expression method

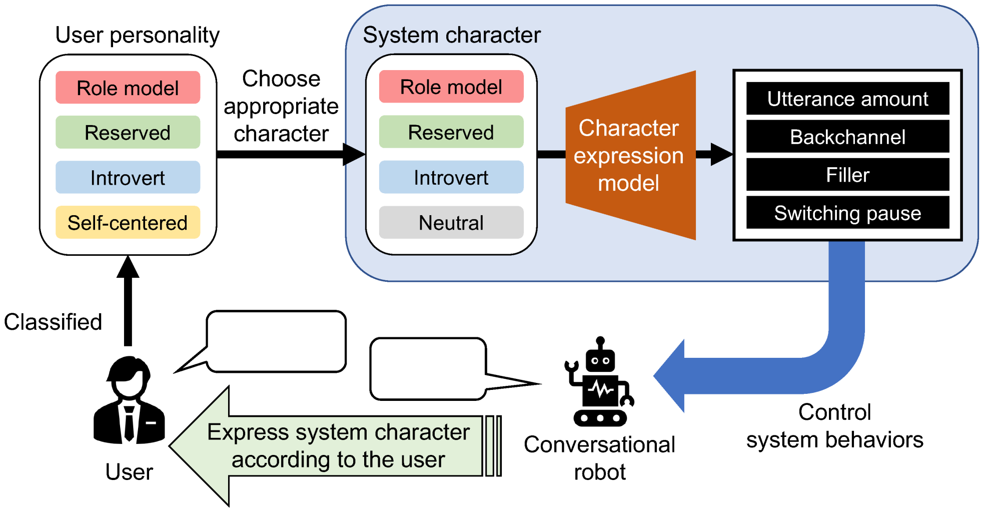

Figure 1. An overview of the character expression method

by controlling the spoken dialogue behaviors according to the user.

3. Character expression according to user's personality

An overview of the proposed character expression model is shown in Figure 1. The system first identifies the user personality among four classes. Then, the system selects the character which is appropriate for the user personality. The model assumes that it is valid for the system to change its character based on the user's personality. In the following sections, we will examine the validity of this assumption. This classification is derived in Section 3.1. This is described in Section 3.2 and examined in Section 4 and 5. The system calculates the control amount for spoken dialogue behaviors using a character expression model according to the selected character. This model is explained in Section 3.3.

3.1. Classification of robot character and user personality

We define the personality and character class based on the Big Five traits [21] as summarized in Table 1. The Big Five traits are widely used for personality studies in psychology. We extract typical classes from the Big Five parameters and used them as user personalities and system characters. One classification was done in a previous study [22] that analyzed the personality ratings of over 140,000 people and then found four template clusters: Role model, Reserved, Self-centered, and Average. We apply the same methodology.

We analyzed the user personality using the speed dating dialogue corpus, where male subjects talked with the android ERICA [23], which was remotely controlled by a female operator using the Wizard of Oz (WOZ) method. They talked about getting to know each other in their first meeting. After the dialogue, each subject evaluated the item "did you have a favorable impression of the counterpart?" on a 7-point scale. However, the ratings of the subject personality and ERICA's character were not collected in this corpus.

Table 1. Description of the Big Five traits

Traits	Typical properties		
Emotional instability (Em)	sensitive/nervous	vs.	resilient/confident
Extrovert (Ex)	outgoing/energetic	vs.	solitary/reserved
Openness (Op)	inventive/curious	vs.	consistent/cautious
Agreeableness (Ag)	friendly/compassionate	vs.	critical/rational
Conscientiousness (Co)	efficient/organized	vs.	extravagant/careless

Table 2. Mean of favorable scores on 7-point scale for combination between subject personality and the robot character (data size)

		Robot character (WOZ)			
		Role model	Reserved	Introvert	Self-centered
Subject personality	Role model†	4.90 (83)	4.55 (40)	**5.26** (34)	5.11 (28)
	Reserved*	4.97 (76)	5.05 (77)	**5.10** (42)	5.02 (43)
	Introvert*	4.75 (51)	**5.03** (60)	4.94 (50)	4.92 (26)
	Self-centered†	**5.51** (47)	5.15 (54)	5.17 (46)	4.86 (32)

$\dagger$ $p < .10$, $*$ $p < .05$

Thus, we conducted an experiment to annotate the subject personality and the robot (WOZ) character. In this experiment, 39 university students watched the video of the corpus and answered the questionnaires. The annotators answered TIPI-J [24] as their impressions of the subject and the robot. TIPI-J is a Japanese translation of TIPI [25] and consists of the 10 items about personality traits on a 7-point scale. In this questionnaire, there are a couple of question items about each Big Five trait. For example, items about extroversion are "do you see yourself as extrovert and enthusiastic?" and "do you see yourself as reserved and quiet?" We prepared 195 video clips sampled from the 65 dialogues in the corpus. Each annotator evaluated approximately 20 video clips and we finally collected 778 annotations.

We normalized the evaluation score using the mean and standard deviation of each annotator's rating and clustered the Big Five scores into four classes using K-means clustering. K-means++ was used for clustering and the number of iterations was set to 300. The clustering results of the subject personality and robot (WOZ) character are shown in Figure 2. We name each class referring to the previous study. However, subjects clustered in the "Average" tended to be introverted. Thus, we name this class "Introvert" instead of "Average."

3.2. Best combination of robot character and user personality

Then, we analyze the effect of the combination of the WOZ character and the user personality on the impression of the dialogue. The results of the relationship between the favorable scores and the personality classes are presented in Table 2. We conducted a one-way factorial analysis of variance (ANOVA) in each subject personality class. This analysis examined whether differences in the robot (WOZ) character affected users' favorable scores. The result shows that the subjects preferred different robot's characters depending on their personality. Highly agreeable subjects such as "Role model" and "Reserved" preferred "Introvert" robot (WOZ) with less extroversion and agreeableness. Conversely, subjects with low agreeableness such as "Self-centered" and "Introvert" preferred highly agreeable robots such as "Role model" and "Reserved." "Self-centered" robots were not preferred in any case. This result suggests that a

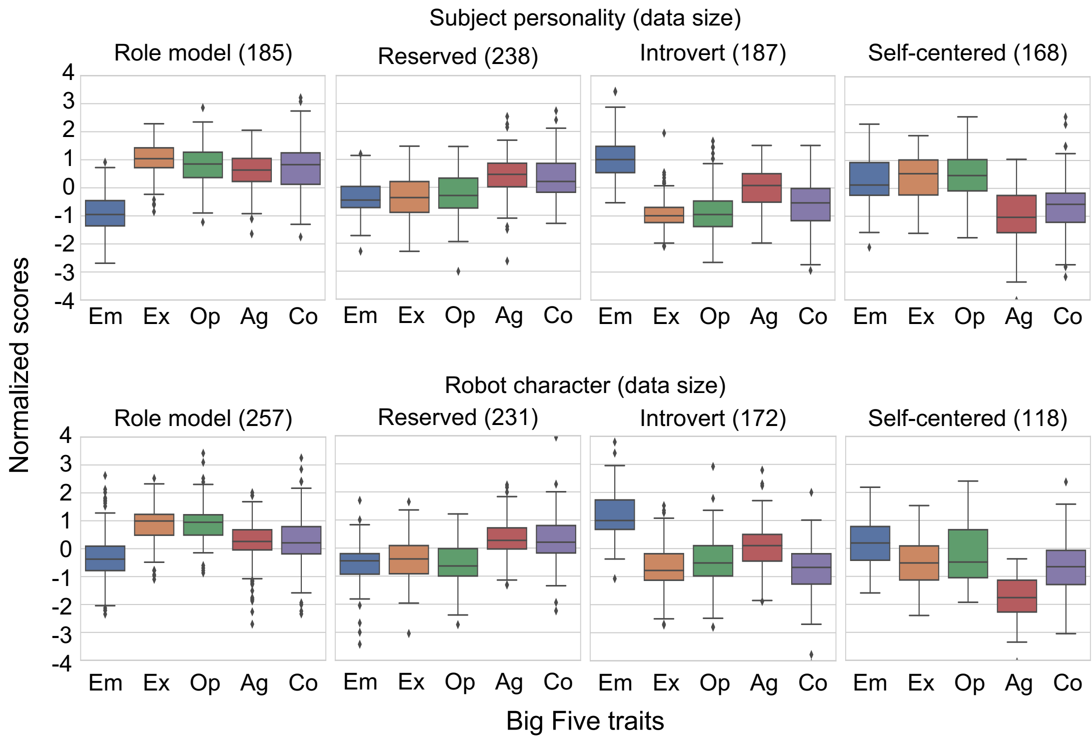

Figure 2. Classification results of Big Five traits on the subject personality and the robot (WOZ) character in corpus

Table 3. User personality and robot character classes

Class name	User personality	Robot character	Big Five traits Em	Ex	Op	Ag	Co
Role model	○	○	▼	△	△	△	△
Reserved	○	○	▼	▼	▼	△	△
Introvert	○	○	△	▼	▼	▼	▼
Self-centered	○		△	△	△	▼	▼
Neutral		○	-	-	-	-	-

△: High, ▼: Low, -: Middle

"Self-centered" character is not appropriate for the dialogue systems. Therefore, we introduce "Neutral" instead of "Self-centered" for the robot (WOZ) character. The results in Table 2 shows that it is necessary to express different characters according to the user personality in order to make a favorable impression on the user. According to this result, we defined the user personality classes and the robot character classes shown in Table 3.

3.3. Character expression of the robot

We have proposed a character expression model for spoken dialogue systems [26]. This model controls four spoken dialogue behaviors: utterance amount, backchannel frequency, filler frequency, and switching pause length. This model is based on a variational auto-encoder (VAE) and composed of a three-layer neural network. The inputs of this model are the Big Five traits ranging from 0 to 1, which we want the system to

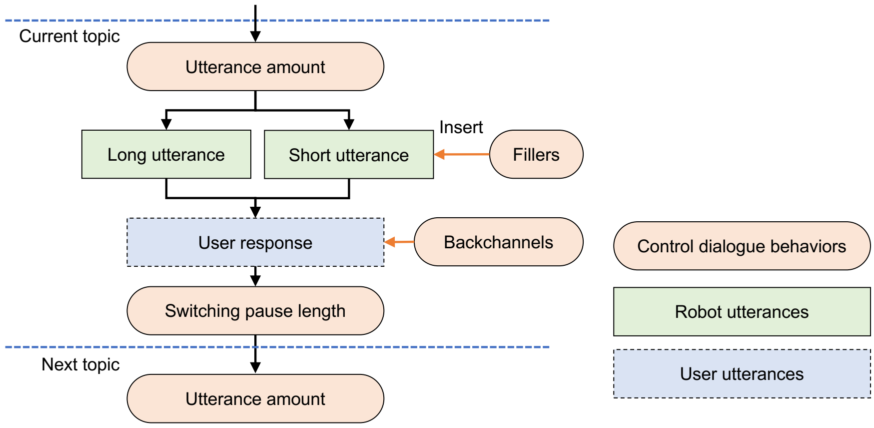

Figure 3. The proposed method of controlling dialogue behaviors for character expression

Figure 4. A snapshot of the dialogue in the experiment

express. The outputs are the control values of the four dialogue behaviors, also ranging from 0 to 1. We trained the model by using the annotated dialogue data explained in Section 3.1.

The control values correspond to the behavior settings as follows. We prepared two utterance patterns corresponding to the long and short utterance amount. According to the control value of utterance amount, the system selects one of the two-utterance patterns: long or short utterances. We used the backchannel generation module [27] to control the backchannel frequency. The model determines the generation of backchannels every 100 ms by using the prosodic features of the user utterance with a logistic regression model. The control value of the backchannel frequency corresponds to the threshold of the output probability of the backchannel generation module. The control value of the filler frequency corresponds to the threshold of its probability. Fillers are inserted stochastically at the beginnings of the system utterances. The switching pause length is the length of silence before the system takes a turn. The control value of the switching pause length is linearly mapped to the switching pause length from 700 to $3,000$ ms.

4. Laboratory guide: Task oriented dialogue

We first examined the effect of character adaptation in a laboratory guide, which conducts a task-oriented dialogue. We implemented the laboratory guide system on the

Table 4. Control values of dialogue behaviors in each character condition

System character	Control values of dialogue behaviors (0∼1)			
	Utterance	Backchannel	Filler	Switching pause
Role model	0.8	0.8	0.1	0.2
Reserved	0.4	0.3	0.1	0.4
Introvert	0.1	0.1	0.7	0.7
Neutral	0.5	0.3	0.2	0.3

android ERICA. In this experiment, subjects interacted with the robot that expressed the four characters and evaluated the dialogue with each character.

4.1. Laboratory guide system

A laboratory guide is a scenario-based dialogue system that introduces research topics to a subject. The system reads pre-defined sentences and occasionally asks the subject to ask questions. After some interaction, the system proceeds to the next topic. The average number of turns in the experiment was 12, and the dialogue duration was about 8 minutes.

The control values of the four behaviors are continuous values in the range of 0∼1 that can be directly used in the spoken dialogue system. The input to the character expression model based on VAE was the mean value of each Big Five trait for each character class shown in Figure 2. However, we input 0.5 to the model in the Neutral condition. The control values obtained by the character expression model for each character class are summarized in Table 4. The system controls the behaviors according to the character in the manner described in Section 3.3.

4.2. Experimental setting

The experiment involved 36 students and graduate students aged 18 to 27. Among them, there were 24 males and 12 females. At the beginning of the experiment, each subject answered his/her Big Five personality traits using TIPI-J [24]. The subjects are instructed to listen to explanations about the research from the robot four times, and respond when asked questions. They talked with the laboratory guide system with four different conditions: “Role model”,“Reserved”, “Introvert”, and “Neutral.” The dialogue between a subject the robot in the experiment is shown in Fig.4. In each condition, the system introduced one of four different research topics: automatic speech recognition, spoken dialogue system, acoustic signal processing, and music information processing. The system characters and dialogue topics were randomly arranged. Each dialogue lasts approximately eight minutes. At the end of each dialogue, the subject answered the system character impression using TIPI-J and the questionnaire about the his/her impression with dialogue, as presented in Table 6. This questionnaire consists of four constructs: skill (Q1), engagement (Q2), adaptation (Q3), and naturalness (Q4). Each construct score was calculated as the average over multiple questions. “Skill” means how well the robot is suited for functioning as a laboratory guide. “Engagement” means how the subject is satisfied with the dialogue. “Adaptation” means how well the robot adapt to the subject. “Naturalness” is used to determine whether the dialogue has broken down.

Table 5. Subjective evaluation scores (7-point scales) on Big Five traits for each system character condition

Character condition	Big Five scores: Mean (SD)				
	Em	Ex	Op	Ag	Co
Role model	2.80 (0.87)	4.72 (0.95)	4.37 (1.06)	4.94 (1.27)	5.02 (1.09)
Reserved	2.79 (1.02)	4.31 (1.21)	4.21 (1.10)	4.97 (1.25)	4.75 (0.97)
Introvert	4.00 (1.19)	3.63 (1.29)	4.14 (1.15)	5.06 (0.93)	4.43 (1.30)
Neutral	3.26 (1.23)	4.13 (1.24)	4.15 (1.11)	5.22 (0.94)	4.79 (1.08)

4.3. Experimental results

First, we analyze the Big Five rating for each character condition in Table 5. The score in each character condition was consistent with the character class tendencies in Table 3. This result confirms that the subjects recognized the different system character conditions. The results of the clustering of the subject personality are shown in Figure 5. Based on the combination of these four personality classes and the system character conditions, we analyze the impression evaluation results. The Cronbach's alpha coefficients calculated from the evaluation results of the questions presented in Table 9 were 0.87 for Q1, 0.91 for Q2 and 0.74 for Q3. We did not calculate Cronbach's alpha for Q4, which has a single item. A high alpha value of each scale confirms, the validity of these evaluations.

We conducted a two-way repeated ANOVA on the conditions. The subjects' evaluation scores and the results of ANOVA are listed in Table 7. "System character" means the evaluation scores for each system character and "Subject personality" means the evaluation scores for each user personality. "System factor" means whether significant differences are observed among the system characters. "Subject factor" means whether significant differences are observed among the subject personalities. "Interaction effect" means whether the combinations between the system character and subject personality affect the evaluation scores. The Q4 results suggest that the character expression did not affect the naturalness. There were significant differences in the system factors of Q1 and Q2. Tukey's multiple tests for Q1 confirmed a significant difference between Role model and Introvert at a 5% level of significance. The evaluation results for each system character in Table 7 show that "Role model" is highly rated. Given that Q1 is constructed about whether the system is good as a laboratory guide, there is a suitable character for the laboratory guide. This result is reasonable for a task-oriented dialogue, since the completion of the task requires collaboration. Moreover, significant differences are observed in the subject factor of Q1, Q2, and Q3. Tukey's multiple tests for Q3 confirmed a significant difference at a 5% level of significance in two pairs: between Role model and Reserved, and between Reserved amd Self-centered. Generally, the Role model subjects were rated highest in four personality classes.

Significant differences are also observed in the interaction effects of Q1 and Q2. This means that the combination of the system character and the subject personality influenced on the evaluation scores. The subjects' evaluation results for each combination of system character and subject personality are shown in Figure 6. For example, Role model subjects prefer the Reserved system, Reserved subjects prefer the Role model system, Introvert subjects prefer the Neutral system, and Self-centered subjects prefer the Neutral system. These results show that the high agreeable subjects prefer a high agreeable character and low agreeable subjects like a neutral character. These results are not consistent with the observation in Section 3.1, but reasonable for the task-oriented dialogue.

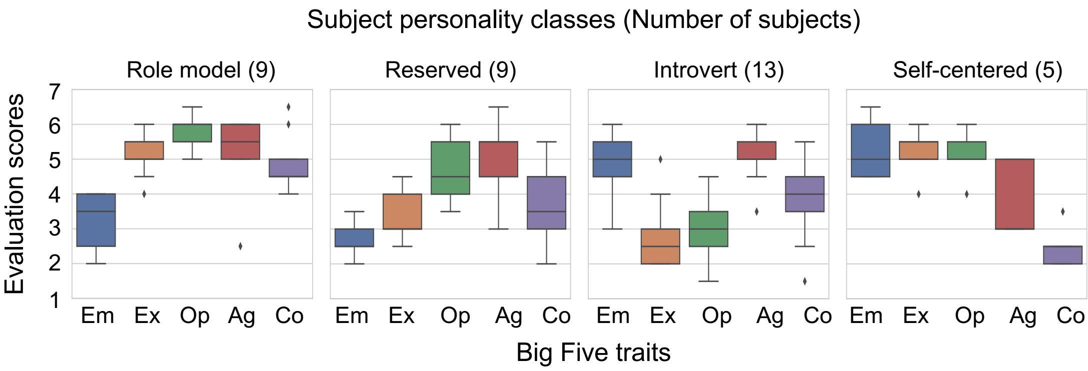

Figure 5. Clustering of Big Five scores of the subjects in the laboratory guide experiment

Table 6. Questionnaire used for the laboratory guide system

	Items
Q1 (Skill)	Do you think that the robot was good at explaining? Would you like the robot to explain other research topics? Would you like the robot to talk about topics other than research? Do you think the robot was a good laboratory guide?
Q2 (Engagement)	Is it easy for you to talk with the robot? Do you have a favorable impression of the robot?
Q3 (Adaptation)	Do you think that the robot adapted to you? Do you think the robot understand your personality?
Q4 (Naturalness)	Do you think that the robot spoke naturally?

Table 7. Mean evaluation scores (7-point scales) and the results of ANOVA in each questionnaire item (Role: Role model, Res: Reserved, Int: Introvert, Self: Self-centered, Ne:Neutral)

Items	System character				Subject personality				ANOVA System	ANOVA Subject	ANOVA Interaction
	Role	Res	Int	Ne	Role	Res	Int	Self	$F(\eta^2)$	$F(\eta^2)$	$F(\eta^2)$
Q1	**4.74**	4.58	3.90	4.26	**4.61**	3.91	4.36	4.71	4.57^{**}(0.08)	2.92^{*} (0.05)	2.39^{*} (0.12)
Q2	4.44	4.69	4.23	4.50	**4.68**	4.02	4.61	4.60	3.30^{*} (0.06)	3.32^{*} (0.06)	2.71^{*} (0.14)
Q3	3.54	3.51	3.71	3.74	**4.08**	3.56	3.57	3.99	0.71 (0.01)	7.27^{**}(0.13)	1.56 (0.08)
Q4	4.19	4.17	3.81	4.06	4.45	4.72	3.98	4.28	$2.31^{\dagger}$ (0.04)	$2.54^{\dagger}$ (0.05)	1.65 (0.09)

$^{\dagger}p < .10, ^{*}p < .05, ^{**}p < .01$

5. Chit-chat: Non-task-oriented dialogue

We also conducted a subjective experiment using a chit-chat as non-task-oriented dialogue. We implemented a chit-chat system on ERICA. In this experiment, subjects talked with the robot that expressed the four characters and talked about four topics, then they evaluated the dialogue with each character.

5.1. System architecture

The chit-chat system is a combination of an example-based system and a neural generation-based system. The example-based system searches for the example most similar to the user's utterance and outputs a response. We prepared 2000 examples for one topic using a crowd-sourcing service. The system converts the user utterance into a vector as a query. The system response is the example which has the highest degree

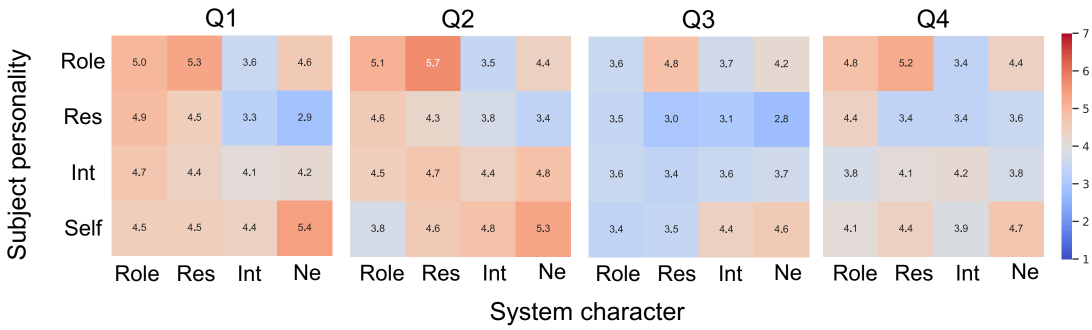

Figure 6. Mean evaluation scores (7-point scales) in each combination of system character and subject personality in laboratory guide. Role: Role model, Res: Reserved, Int: Introvert, Self:Self-centered, Ne: Neutral

Table 8. Subjective evaluation scores (7-point scales) on Big Five traits for each system character condition

Character condition	Big Five scores: Mean (SD)				
	Em	Ex	Op	Ag	Co
Role model	3.00 (1.01)	4.71 (1.48)	4.23 (1.10)	4.67 (1.42)	4.11 (1.40)
Reserved	2.98 (1.11)	4.73 (1.32)	4.25 (1.11)	5.13 (0.99)	4.23 (1.08)
Introvert	3.71 (1.55)	3.64 (1.59)	3.70 (1.00)	5.13 (0.98)	4.16 (1.11)
Neutral	2.94 (1.10)	4.43 (1.57)	4.20 (1.10)	4.80 (1.16)	4.11 (1.44)

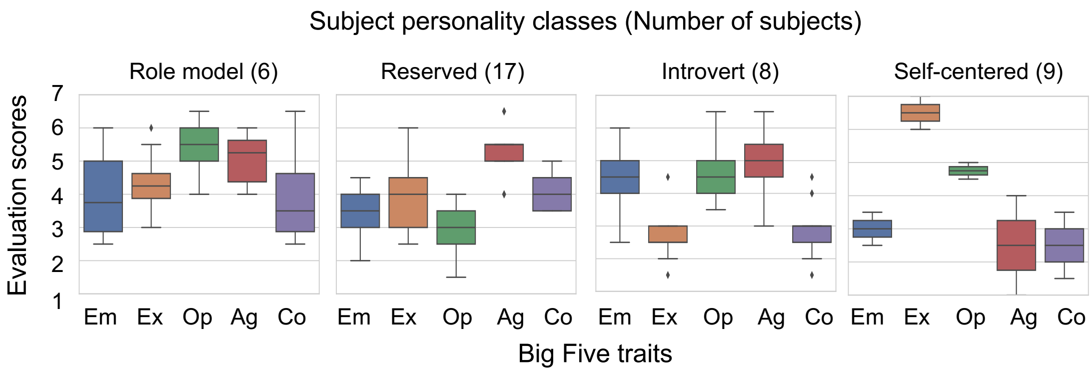

Figure 7. Clustering of Big Five scores of the subjects in the chit-chat experiment

of cosine similarity between the user utterance vector (key) and the response vector in the database (query). The query is the average vector that was computed from the content words (nouns, verbs, and adjectives) using Word2Vec[1]. The key is converted from the prepared user utterance in the database using the same manner. However, if the highest degree of similarity is lower than a threshold, the system determines that there is no matching example, and utters the generated response using the other generation model.

We prepared four items for each example in the database: expected user utterance, system response, episode, and question. The response is a reaction to the user utterance. For example, when a user asked "Do you like hot springs?", the system responds "Yes, I like hot springs." An episode is a system's self-disclosure, which follows the above response. For example, "I go to the hot spring every year." The system can ask a question to the user to continue the same topic. Therefore, the question should be

[1] hottoSNS-w2v: https://github.com/hottolink/hottoSNS-w2v

related to the system episode, such as “Do you often go to hot springs?” The speech amount is controlled by whether or not the system utters an episode.

For the neural generation model, we used the Transformer model, which was pre-trained and has been fine-tuned on the JPersonaChat dataset[2]. Four utterances of the user and the system are input into the model as the dialogue history. The beam width was set to 20. The following three filters are applied to the output of the model. First, we prepared dirty word lists for removing inappropriate utterances, referring to the corpus collected from internet forum[3]. The model response candidates that contain dirty words are not output. Second, we removed questions from the generated sentences to prevent transitioning to different topics by the system utterances. Third, if the speech recognition result of the user utterance does not contain any meaningful word, then the system asks a question such as “What did you say?” The generated response is divided into a piece of sentences. Each sentence is determined stochastically whether to be included in the system’s utterance. Specifically, if the value sampled from a uniform distribution between 0 to 1 is less than the threshold value, the sentence is used. We use the utterance amount for this threshold. The other behaviors are controlled as described in Section 3.3.

5.2. Experimental setting

The experiment involved 40 students and graduate students aged 18 to 27. Among them, there were 22 males and 18 females. These participants do not overlap with the participants of the experiment in Section 4. The appearance of the robot and the subjects in the experiment is the same as in Fig.4. They talked with the chit-chat system on four different character conditions as in Section 4.2. At the beginning of the experiment, each subject answered his/her Big Five personality traits using TIPI-J [24]. We prepared four topics: travel, movie, hometown, and student life. The subjects were instructed to freely talk about any topic related to the specified topics for each dialogue. The robot characters for the dialogue topics were randomly arranged so that there was no bias in the combination. The subject talked with the robot about 8 minutes in each condition. At the end of the each dialogue, the subject answered the questionnaire about three constructs: engagement (Q1), adaptation (Q2), and naturalness (Q3), as presented in Table 9. Each construct score was calculated as the average over multiple questions. “Engagement” means how the subject is satisfied with the dialogue. “Adaptation” means how well the robot adapts to the subjects. “Naturalness” is used to determine whether the dialogue has broken down.

5.3. Experimental results of the chit-chat system

The Big Five rating scores in each character condition are presented in Table 8. This results show that the system characters are accurately perceived by the subjects. The clustering results of the subject personality are shown in Figure 7. Based on the combination of these four personality classes and the system character conditions, we analyze the impression evaluation results. The Cronbach’s alpha coefficients calculated from the evaluation results of the questions presented in Table 9 were 0.87 for Q1 and 0.86 for Q2. We did not calculate Cronbach’s alpha for Q3, which has a single item. A high alpha value of each scale confirms, the validity of these evaluations.

[2]Japanese dialogue transformer: https://github.com/nttcslab/japanese-dialog-transformers

[3]Open 2channel dialogue corpus: https://github.com/1never/open2ch-dialogue-corpus

Table 9. Questionnaire used for the chit-chat system

	Items
Q1 (Engagement)	Do you enjoy the dialogue?
	Can you get along with the robot?
	Would you like to talk with the robot again?
	Is it easy for you to talk with the robot?
	Do you have a favorable impression of the robot?
Q2 (Adaptation)	Do you think that the robot is adapted to you?
	Do you think that the robot understands your personality?
Q3 (Naturalness)	Do you think that the robot spoke naturally?
	Do you think the robot's utterances were appropriate to the topic?

Table 10. Mean evaluation scores (7-point scales) and results of ANOVA in each questionnaire item (Role: Role model, Res: Reserved, Int: Introvert, Self: Self-centered, Ne:Neutral)

	System character				Subject personality				ANOVA System	ANOVA Subject	ANOVA Interaction
	Role	Res	Int	Ne	Role	Res	Int	Self	$F(\eta^2)$	$F(\eta^2)$	$F(\eta^2)$
Q1	3.46	3.97	3.58	3.57	3.81	3.51	3.80	3.64	1.76 (0.03)	1.24 (0.02)	1.21 (0.07)
Q2	3.33	**3.83**	3.13	3.46	3.58	3.29	3.52	3.53	2.62* (0.05)	1.01 (0.02)	2.00*(0.10)
Q3	3.80	4.19	3.91	4.03	3.81	4.47	3.71	3.56	1.22 (0.02)	2.72**(0.05)	1.09 (0.06)

$^{\dagger}p < .10$, $^{*}p < .05$, $^{**}p < .01$

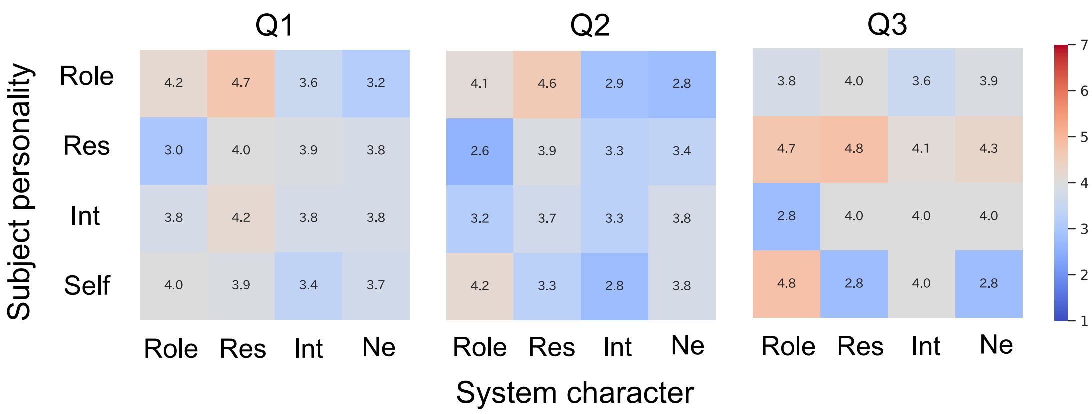

Figure 8. Mean evaluation scores (7-point scales) in each combination of system character and subject personality in chit-chat. Role: Role model, Res: Reserved, Int: Introvert, Self:Self-centered, Ne: Neutral

We conducted two-way repeated ANOVA. The evaluation scores and the results of ANOVA are listed in Table 10. The meaning of each column in the table is the same as in Section 4.3. There were showed significant differences in the system factor of Q2. Tukey's multiple tests for Q2 confirmed a significant difference between Reserved and Introvert at a 10% level of significance. The subjects felt that the Reserved system adapted to them. There were also significant differences in the subject factor of Q3. Reserved subjects felt that the systems spoke naturally. The result also showed significant differences in the interaction of Q2 and Q3. The subjects' evaluation results for each combination of system character and subject personality are shown in Figure 8. Introvert subjects preferred the Reserved system and Self-centered subjects preferred the Role model system. These are consistent with the observation in Section 3.1. However, the Introvert system was not preferred by the Role model subjects and Reserved subjects. One reason of these results is that the system did not accurately express the Introvert characters.

6. Conclusion

In this study, we investigated the effectiveness of character expression for a conversational robot appropriate to the user personality. First, we analyzed the speed dating dialogue corpus and observed that the favorable impression depended on the combination of system character and user personality. Then, we designed the character expression model using four spoken dialogue behaviors: utterance amount, backchannel frequency, filler frequency, switching pause length. This model was implemented into a laboratory guide and a chit-chat system on the android ERICA. We conducted subjective experiments to confirm the effect of character adaptation. As a result, "Role model" characters are found to be appropriate as a laboratory guide and similar tendency to speed dating was confirmed in the chit-chat dialogue.

In future works, we will investigate the effect of character adaptation in other dialogue tasks. Moreover, we will construct a real-time user adaptation model. This model will recognize user personality in the dialogue and adopt the best character for the user personality. In this study, we investigated the impacts of user adaptation using a female android. However, the appearance of the robot will affect the dialogue and its effect need to be explored by using many kinds of robots in the future. Additionally, it will also be important to analyze differences in users' impression of the robot based on their gender and other attributes.

Acknowledgement

This work was supported by Moonshot R&D (JPMJPS2011) and Grant-in-Aid for JSPS Fellow (JP20J22284). We are grateful to Prof. Takatsune Kumada and Dr. Ritsuko Iwai for helpful discussions.

References

[1] Kampman O, J Barezi E, Bertero D, et al. Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In: ACL; 2018. p. 606–611.
[2] An G, Levitan SI, Hirschberg J, et al. Deep personality recognition for deception detection. In: INTERSPEECH; 2018. p. 421–425.
[3] Suman C, Saha S, Gupta A, et al. A multi-modal personality prediction system. Knowledge-Based Systems. 2022;236:107715.
[4] Andrist S, Mutlu B, Tapus A. Look like me: Matching robot personality via gaze to increase motivation. In: International Conference of Human-Computer Interaction (CHI); 2015. p. 3603–3612.
[5] Zhang Q, Esterwood C, Yang XJ, et al. An automated vehicle (AV) like me? the impact of personality similarities and differences between humans and avs. In: The Artificial Intelligence for Human-Robot Interaction (AI-HRI); 2019.
[6] Serban IV, Lowe R, Henderson P, et al. A survey of available corpora for building data-driven dialogue systems. Dialogue and Discourse. 2018;9(1):1–49.
[7] Mairesse F, Walker MA. Controlling user perceptions of linguistic style: Trainable generation of personality traits. Computational Linguistics. 2011;37(3):455–488.
[8] Zhang S, Dinan E, Urbanek J, et al. Personalizing dialogue agents: I have a dog, do you have pets too? In: Annual Meeting of the Association for Computational Linguistics (ACL); 2018. p. 2204–2213.
[9] Oraby S, Reed L, Tandon S, et al. Controlling personality-based stylistic variation with

neural natural language generators. In: Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial); 2018. p. 180–190.

[10] Li J, Galley M, Brockett C, et al. A persona-based neural conversation model. In: Annual Meeting of the Association for Computational Linguistics (ACL); 2016. p. 994–1003.

[11] Higashinaka R, Mizukami M, Kawabata H, et al. Role play-based question-answering by real users for building chatbots with consistent personalities. In: Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial); 2018. p. 264–272.

[12] Song H, Wang Y, Zhang K, et al. BoB: BERT over BERT for training persona-based dialogue models from limited personalized data. In: Annual Meeting of the Association for Computational Linguistics and International Joint Conference on Natural Language Processing (ACL-IJCNLP); 2021. p. 167–177.

[13] Fernau D, Hillmann S, Feldhus N, et al. Towards personality-aware chatbots. In: Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial); 2022. p. 135–145.

[14] Ma Z, Dou Z, Zhu Y, et al. One chatbot per person: Creating personalized chatbots based on implicit user profiles. In: International ACM SIGIR Conference on Research and Development in Information Retrieval. Association for Computing Machinery; 2021. p. 555–564.

[15] Qian H, Dou Z, Zhu Y, et al. Learning implicit user profile for personalized retrieval-based chatbot. In: Conference on Information and Knowledge Management (CIKM); 2021. p. 1467–1477.

[16] Zhong H, Dou Z, Zhu Y, et al. Less is more: Learning to refine dialogue history for personalized dialogue generation. In: Annual Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (NAACL-HLT); 2022. p. 5808–5820.

[17] John NE, Rossi A, Rossi S. Personalized human-robot interaction with a robot bartender. In: ACM Conference on User Modeling, Adaptation and Personalization (UMAP); 2022. p. 155–159.

[18] Qian Q, Huang M, Zhao H, et al. Assigning personality/profile to a chatting machine for coherent conversation generation. In: International Joint Conference on Artificial Intelligence (IJCAI); 2018. p. 4279–4285.

[19] Cai W, Chen L. Predicting user intents and satisfaction with dialogue-based conversational recommendations. In: ACM Conference on User Modeling, Adaptation and Personalization (UMAP); 2020. p. 33–42.

[20] Uchida T, Minato T, Nakamura Y, et al. Female-type android ’ s drive to quickly understand a user ’ s concept of preferences stimulates dialogue satisfaction: Dialogue strategies for modeling user ’ s concept of preferences. International Journal of Social Robotics. 2021; 13:1499–1516.

[21] McCrae R R, John O P. An introduction to the five-factor model and its applications. Journal of Personality. 1992;60(2):175–215.

[22] Gerlach M, Farb B, Revelle W, et al. A robust data-driven apprach identifies four personality types across four large data sets. Nature Human Behaviour. 2018;2:753–742.

[23] Kawahara T. Spoken dialogue system for a human-like conversational robot ERICA. In: International Workshop on Spoken Dialogue Systems Technology (IWSDS); 2018.

[24] Oshio A, Abe S, Cutrone P. Development, reliability, and validity of the japanese version of ten item personality inventory (TIPI-J). The Japanese Journal of Personality. 2012; 21(1):40–52.

[25] Gosling SD, Rentfrow PJ, Swann WB. A very brief measure of the big-five personality domains. Journal of Research in Personality. 2003;37(6):504–528.

[26] Yamamoto K, Inoue K, Kawahara T. Semi-supervised learning for character expression of spoken dialogue systems. In: INTERSPEECH; 2020. p. 4188–4192.

[27] Lala D, Milhorat P, Inoue K, et al. Attentive listening system with backchanneling, response generation and flexible turn-taking. In: Annual Meeting of the Special Interest Group on Discourse and Dialogue (SIGdial); 2017.