Character adaptation of spoken dialogue systems based on user personalities

Kenta Yamamoto, Koji Inoue, and Tatsuya Kawahara

Abstract Character expression (e.g., extrovert or agreeable) is important for spoken dialogue systems to achieve human-like dialogue. The appropriate character is different depending on each dialogue task and the user. In this study, we propose a character expression method according to the user personality in task-oriented dialogues. A previous psychological study identified four representative character classes based on the large-scale ratings on the Big Five traits. We use these four-character classes for character adaptation to the user personalities. Specifically, we investigate how the combination of the user personality and the system character affects the impression of the dialogue. Our analysis of a human-robot dialogue corpus using the Wizard of Oz (WOZ) method shows that the combination of the subject personality and the robot character affects the favorable impressions toward the robot. Based on the analysis, we have designed and developed a character adaptation model that controls spoken dialogue behaviors: utterance amount, backchannel frequency, fillers frequency and switching pause length. In a subjective experiment, a robot talked with subjects as a laboratory guide in four different character conditions, and each subject evaluated the impression of each robot. The results shows that the extrovert character was preferred for items on the laboratory guide's skill, and that the appropriate character to the user personality was preferred for items on how easy to talk with the robot.

1 Introduction

User adaptation of spoken dialogue systems is to make systems generate behaviors appropriate to the user's attribute or situation, which leads to increasing user's satisfaction with dialogues. One of the system elements that can be adapted is its character. Character expression (e.g., extrovert or agreeable) is important for spo-

Graduate School of Informatics, Kyoto University, Japan
e-mail: `[yamamoto][inoue][kawahara]@sap.ist.i.kyoto-u.ac.jp`

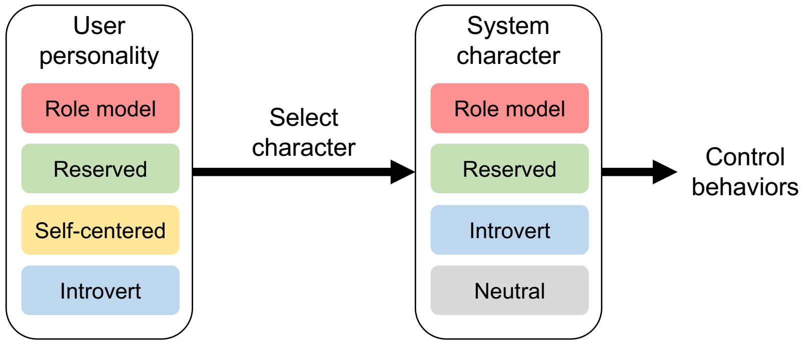

Fig. 1 Overview of character adaptation based on user personality

ken dialogue systems for human-like interaction [5, 21]. It is shown that the character expression of spoken dialogue systems leads to increasing user engagement and naturalness in dialogue [18, 10, 20]. Therefore, previous studies addressed character expression models for dialogue systems [16, 9, 8, 19].

On the other hand, classification of the user is necessary for user adaptation. A widely-used user classification is personality. As many studies on personality estimation have also been conducted [12, 1, 22], it is possible that the user personality is estimated through spoken dialogue. It was confirmed that users with different personalities had different impressions of a dialogue system [2, 25]. Therefore, user adaptation based on the user personality is required to achieve a satisfactory dialogue.

In this study, we propose character adaptation, where a spoken dialogue system expresses the character appropriate to the user personality. Note that, "personality" is used as a psychological dimension for classifying users, and "character" is used as the impression that the dialog system gives to the user in this paper. An overview of character adaptation is shown in Fig 1. First, a user is classified into four classes: Role model, Reserved, Self-centered and Introvert. According to the user personality, the system expresses one of four characters: Role model, Reserved, Introvert and Neutral. The character expression model [24] is used to control the behavior of the spoken dialogue system.

We conducted two experiments to evaluate the proposed method. In Section 3, we analyze the classifications of the system character and the user personality by a corpus-based analysis. In Section 4, we identify the appropriate system character for the user personality in a subjective experiment.

2 Related work

Character adaptation consists of user adaptation and a character expression model. We discuss how each method in this study differs from previous studies.

2.1 User adaptation

Some user adaptation methods of dialogue systems were addressed in previous studies. One method accumulates the user's dialogue history and makes the user's profile [11]. This method assumed that the same user continues to use the system such as a smart speaker. Other methods were proposed using the user's intent, and proficiency with the system or identifying the user's preferences [4, 23]. In these methods, continuous use of the system is necessary to create the model of the user.

We propose a method to adapt the behavior of the system using personality to deal with first-time interaction. We presume personality affects a user's preferences and impressions of the system [2, 25].

2.2 Character expression model

In this study, the system character is adapted according to the user personality. Previous studies indicated that different users have different preferences for the characters. Some methods were proposed to set up specific characters or personas of dialogue systems and control system utterances [9, 15]. These methods are not suitable for controlling the character according to the user. On the other hand, some methods are proposed to express system characters using the Big Five [16, 8, 19] which is a personality trait in psychology. However, the search space for user adaptation is huge if we directly use Big Five traits, so we propose a method to classify the Big Five traits into four template classes and also to express appropriate characters from the four classes.

3 Character adaptation

An overview of character adaptation in this study is shown in Fig 1. In this method, the dialogue system classifies users into four classes and expresses the best character from the four characters according to the user personality class. In Section 3.1, we explain the four classes of user personality and the system character. In Section 3.2, we analyze the effectiveness of the classification using a dialogue corpus.

3.1 Classification of system character and user personality

We define the personality and character class based on the Big Five traits [17] as summarized in Table 1. Big Five is widely used for personality in psychology. However, to extract four classes from the Big Five and use them as system characters and personalities. This classification was based on a previous study [6] that analyzed us-

Table 1 Description of 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 Character and personality classes

Class name	User personality	System Character	Big Five traits Em	Ex	Op	Ag	Co
Role model	○	○	Low	High	High	High	High
Reserved	○	○	Low	Low	Low	High	High
Self-centered	○		High	High	High	Low	Low
Introvert	○	○	High	Low	Low	Low	Low
Neutral (Baseline)		○	Middle	Middle	Middle	Middle	Middle

ing the personality ratings of over 140,000 people and then found four template clusters: Role model, Reserved, Self-centered and Average. The relationship between these classes and the original Big Five scales is summarized in Table 2. Note that we remove “Self-centered” from the system character, which is inappropriate for a dialogue system, and use “Neutral” instead.

3.2 *Analysis of classification using spoken dialogue corpus*

We analyzed the effect of the combination of the dialogue system character and the user personality on the impression of the dialogue. We used a human-robot interaction corpus with android ERICA. In this corpus, the subject talked with the android ERICA [13] in Japanese, which was remotely controlled by an operator using the Wizard of Oz (WOZ) method. However, the subjects did not know that ERICA was being controlled by the operator. The dialogue task was speed-dating, in which each subject and ERICA talked to get to know each other in their first meeting. After the dialogue, each subject evaluated an item of “Did you have a favorable impression of the robot?”, called a favorable score, on a 7-point scale. On the other hand, the ratings of the subject personality and ERICA’s character were not collected in this corpus. In this section, we refer to the subject in this corpus as “subject” and ERICA as “system”.

We conducted an annotation for the subject personality and the system character. In this experiment, 39 university students watched the video of the corpus and answered the questionnaires. In this section, we refer to them as “evaluators”. The evaluators answered TIPI-J [3] as their impressions of the subject and the system. TIPI-J is a Japanese translation of TIPI [7], in which the evaluators answered the 10 items on a 7-point scale. For example, a couple of question items about extroversion

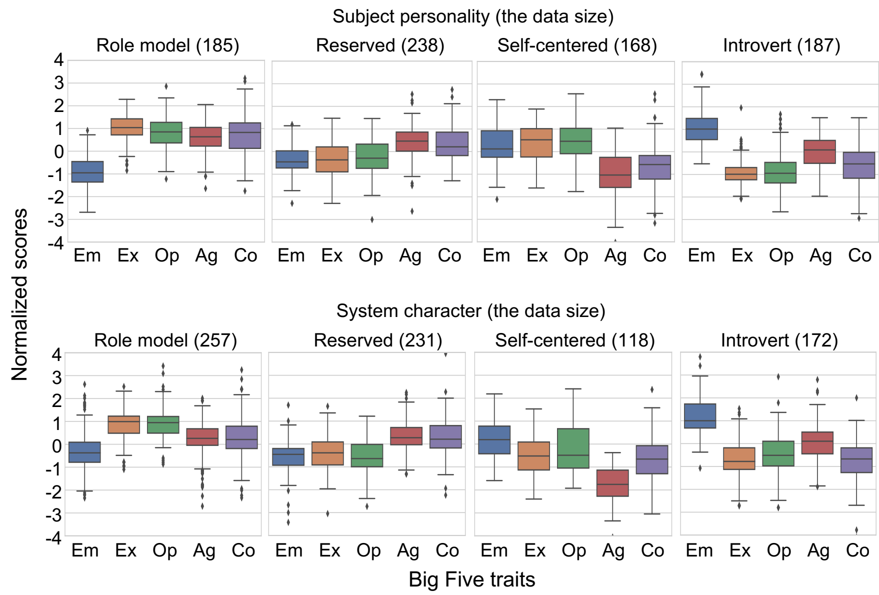

Fig. 2 Clustering results of subject personality (Upper) and system character (Lower)

are "Do you see yourself as extroverted 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 evaluator evaluated about 20 video clips and we finally collected 778 annotations, excluding some missing data.

We normalized the evaluation score using the mean and standard deviation of each evaluator's rating. We then classified the Big Five scores into four classes using K-means clustering. K-means++ was used for initialization and the number of iterations was set to 300. The clustering results of the subject personality and ERICA character are shown in Figures2.

The results of the relationship between the subjects' favorable scores and the personality classes are shown in Table 3. This table shows the mean of the favorable scores and the data size for each combination of the subject personality and the system character. We conducted a one-way factorial analysis of variance in each subject personality class. This analysis examines whether differences in the system character affect users' favorable scores. The result shows that the subjects preferred different system's characters depending on their personality. Therefore, we can conclude it is necessary to express different characters according to each user personality in order to make a favorable impression of the user.

Table 3 Mean of favorable scores on 7-point scale for combination between system character and subject personality (the data size)

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

†$p < .10$, * $p < .05$

4 Subjective experiment

We conducted a subjective experiment to evaluate that the impression of the dialogue is improved when the system expresses an appropriate character to the user personality. Although we used the WOZ dialogue in the previous section, in this section, we evaluate the effect of character adaptation with an autonomous spoken dialogue system. This dialogue system was designed for an explanation of the research topics as a laboratory guide. We used our character expression model [24], which controls the system's spoken dialogue behaviors according to the specified system character.

4.1 Laboratory guide system with character expression model

We proposed a character expression model for spoken dialogue systems [24]. This model controls four spoken dialogue behaviors: utterance amount, backchannel frequency, filler frequency, and switching pause length. We trained the model by using the annotated dialogue data explained in Section 3.2. The inputs of this model are the Big Five traits that we want the system to express and the outputs are the control values of the four dialogue behaviors. The control values of the four behavior are continuous values of $0-1$ that can be directly used in the ERICA's system, as explained later. The input value of each Big Five trait was the average value of each character class in Fig 2. However, we input 0.5 to the model in the Neutral condition. The control values for each condition are summarized in Table 4.

The control values corresponded to the behavior settings as follows. We prepared two utterance patterns corresponding to the long and short utterance amount as a scenario-based dialogue system. According to the control value of utterance amount, the system selects one of the two-utterance patterns: long or short utterances. We used the backchannels generation module [14] to control the backchannel frequency. The model determines generation of backchannels every 100 milliseconds by using prosodic features of the user utterance with a logistic regression model. The control value of backchannel frequency corresponded to the threshold of the output probability of the backchannel generation module. The control value of filler frequency corresponded to the threshold of its probability. Fillers are inserted

Table 4 The control values of dialogue behaviors in each character condition

Character condition	Control values of dialogue behaviors $(0-1)$			
	Utterance amount	Backchannel	Filler	Switching pause length
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

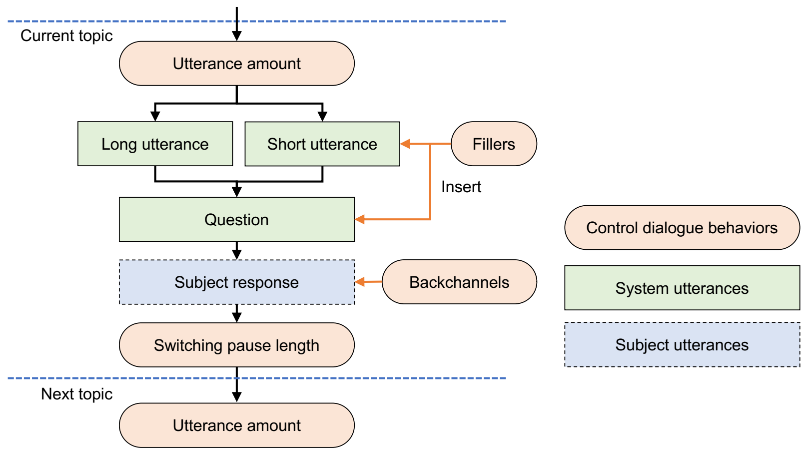

Fig. 3 Dialogue flow of the laboratory guide system

stochastically at the beginnings of the system utterances. The switching pause length is the length of silence until the system takes a turn. The control value of the switching pause length is linearly mapped to the switching pause length from 700 to $3,000$ milliseconds.

We implemented the character expression model in the laboratory guide system of android ERICA. This dialogue system is a scenario-based one that introduces research topics to a student who visits the laboratory. The system reads pre-defined sentences and sometimes asks the visitor to ask questions (Fig 3). After the visitor's response, the system proceeds to the next topic.

4.2 Experimental setting

In this experiment, 40 undergraduate and graduate students talked with the laboratory guide system with four different conditions: Neutral, Role model, Reserved, and Introvert. In each condition, the system introduced one of four different research topics: speech recognition, spoken dialogue system, acoustic signal processing, and music information processing. At the beginning of the experiment, each subject answered his/her Big Five personality traits using TIPI-J [3]. At the end of the ex-

Table 5 Questionnaire used in subjective evaluation

Items	
Q1	Is it easy for you to talk with the robot?
Q2	Do you think that the robot was good at explaining?
Q3	Do you think that the robot adapted to you?
Q4	Do you have a favorable impression with the robot?
Q5	Do you think that the robot spoke naturally?
Q6	Would you like the robot to explain other research topics?
Q7	Would you like the robot to talk about topics other than research?
Q8	Do you think the robot understand your personality?
Q9	Do you think the robot was a good laboratory guide?

Table 6 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)

periment, he/she answered the questionnaires about subject's impression with the dialogue as shown in Table 5.

4.3 Experimental results

At first, we analyzed the Big Five rating for each character condition in Table 6. The scores normalized for each subject are shown in Fig 4. The score in each character condition was consistent with the character class tendencies in Table 2. This result confirms that the subjects recognized the differences in the different system character conditions. The results of the clustering of the subject personality are shown in Fig. 5. Based on the combination of these four personality classes and the system character conditions, we analyze the impression evaluation results.

We conducted a two-way repeated measures analysis of variance (ANOVA) on the conditions. The subject's evaluation scores and the results of the ANOVA are shown 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 shown among the system characters. "Subject factor" means whether significant differences are shown among the subject personalities. "Interaction effect" means whether significant differences is shown among the combination between the system character and the subject personality. The result showed significant differences in the system factor of Q2, Q6, and Q9. The evaluation results for each system character in Table 7 show that "Role model" and "Reserved" are highly rated. Since Q2, Q6, and

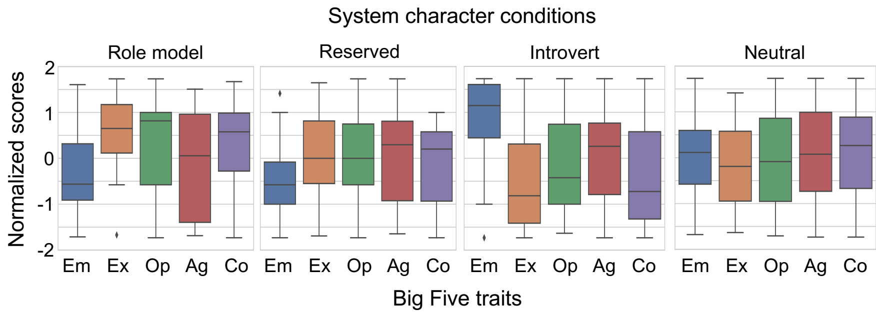

Fig. 4 Normalized Big Five scores for each system character condition in the subjective experiment

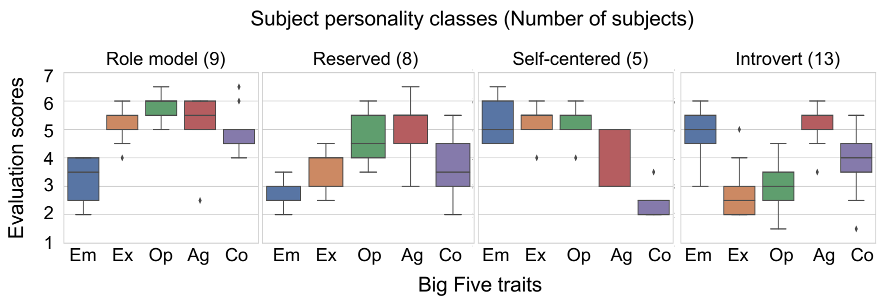

Fig. 5 Clustering of Big Five scores of the subjects

Table 7 Mean evaluation scores (7-point scales) and the results of ANOVA in each questionnaire item

Items	System character				Subject personality				ANOVA		
									System factor	Subject factor	Interaction effects
	Role	Res	Int	Ne	Role	Res	Self	Int			
Q1	4.42	4.58	4.19	4.39	3.97	4.53	4.65	4.50	0.41	1.35	2.08*
Q2	5.28	4.97	4.25	4.58	4.36	5.03	4.80	4.87	4.26**	1.75	2.96**
Q3	3.72	3.83	4.06	4.19	3.47	4.42	4.50	3.75	0.78	3.85**	1.67†
Q4	4.47	4.81	4.28	4.61	4.08	4.67	4.70	4.71	1.06	2.59†	1.71†
Q5	4.19	4.17	3.81	4.06	3.72	4.28	4.45	3.98	0.65	2.59	1.03
Q6	4.69	4.58	3.72	4.31	3.81	4.67	4.45	4.40	3.63**	2.59†	1.71†
Q7	4.08	4.06	3.81	3.89	3.39	4.31	4.50	3.90	0.31	3.52*	1.22
Q8	3.36	3.19	3.36	3.28	2.72	3.56	3.65	3.38	0.13	3.37*	0.56
Q9	4.91	4.69	3.83	4.25	4.08	4.83	4.70	4.27	3.78**	1.96	1.53

Role: Role model, Res: Reserved, Int: Introvert, Ne:Neutral
†$p < .10$, * $p < .05$, ** $p < .01$

Q9 are questions about whether the system is good at the laboratory guide, there is a suitable character for the laboratory guide. Moreover, significant differences are shown in the subject factor of Q3, Q7, and Q8. This means that subjects differ the evaluation scores depending on their personalities. For example, the reserved and self-centered subjects evaluated the high scores about questions. These subjects are

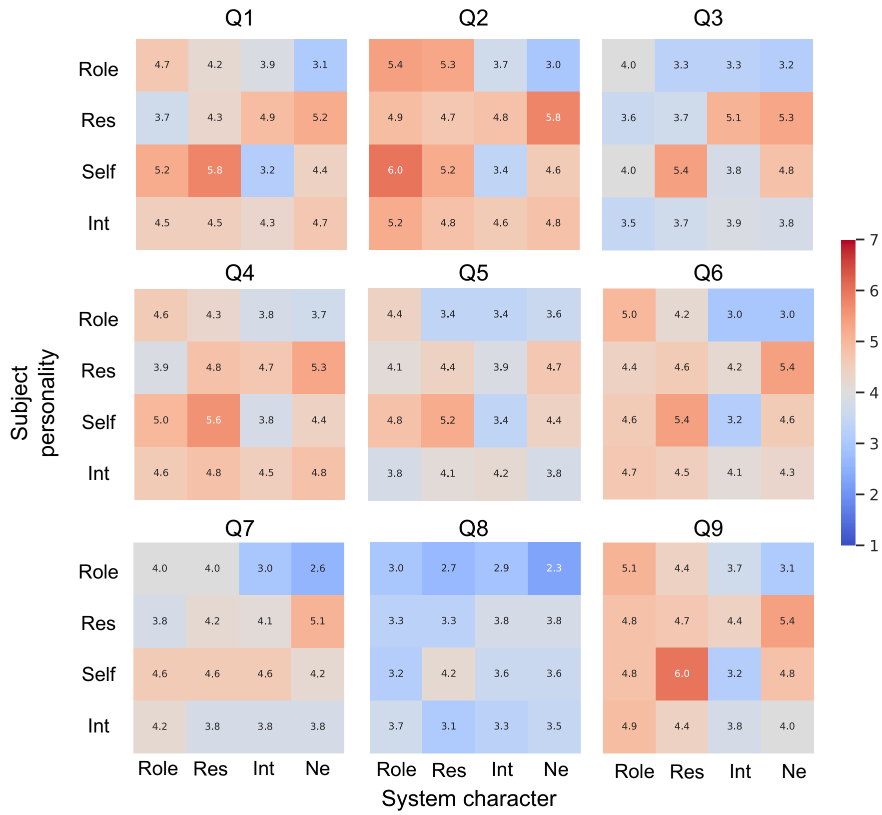

Fig. 6 Mean evaluation scores (7-point scales) in each combination of system character and subject personality. Role: Role model, Res: Reserved, Self:Self-centered, Int: Introvert, Ne: Neutral

considered to have a favorable impression of interacting with the robot. Significant differences are also found in the interaction effects of Q1 and Q2. This means that the subjects differ in their preferences for the system character they want to interact with depending on their personalities. We explain specific relationships below.

The subject's evaluation results for each combination of the system character and the subject personality are shown in Fig 6. The results show which characters are easier to talk with depending on the subject personality. For example, the role model subjects liked the role model systems, and the self-centered subjects liked reserved systems. Note that Q5 has no significant differences in any factors, which means that there is no relation between dialogue naturalness and system character.

In summary, it was seen that the character desired for the task differs from the character that the subject feels comfortable talking to. In other words, to achieve user-adapted dialogue, it is necessary to switch characters for each user.

5 Conclusion

In this study, we have shown that satisfaction with dialogue improves when the system expresses characters appropriate to the user personality. At first, we analyzed the speed-dating dialogue corpus and confirmed that the favorable impression depended on the combination of the system character and the user personality. Second, we conducted a subjective experiment and confirmed that there is a difference between the character desired as a laboratory guide and the character that subjects want to talk to as a dialogue partner. For example, “Role model” and “Reserved” characters are appropriate as laboratory guides. Moreover, the characters that user feels easy to talk to differ depending on their personalities. The results obtained in this study support that switching system characters according to the user personality is effective for user adaptation.

In future works, we will construct a real-time character adaptation system using a personality recognition model to confirm the effect of character adaptation in real dialogue scenarios. In addition, we will also evaluate the similar effect in non-task-oriented dialogues such as chatting.

Acknowledgement

This work was supported by Moonshot R&D (JPMJPS2011), Grant-in-Aid for Scientific Research on Innovative Areas “Communicative intelligent systems towards a human-machine symbiotic society” (JP19H05691) and Grant-in-Aid for JSPS Fellow (JP20J22284). We are grateful to Prof. Takatsune Kumada and Dr. Ritsuko Iwai for helpful discussions.

References

1. An, G., Levitan, S.I., Hirschberg, J., Levitan, R.: Deep personality recognition for deception detection. In: INTERSPEECH, pp. 421–425 (2018)
2. Andrist, S., Mutlu, B., Tapus, A.: Look like me: Matching robot personality via gaze to increase motivation. In: CHI, pp. 3603–3612 (2015)
3. Atsushi Oshio, S.A., Cutrone, P.: Development, reliability, and validity of the japanese version of ten item personality inventory (TIPI-J). The Japanese Journal of Personality **21**(1), 40–52 (2012)
4. Cai, W., Chen, L.: Predicting user intents and satisfaction with dialogue-based conversational recommendations. In: UMAP, pp. 33–42 (2020)
5. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems **42**, 143–166 (2003)
6. Gerlach, M., Farb, B., Revelle, W., Amaral, L.A.N.: A robust data-driven apprach identifies four personality types across four large data sets. Nature Human Behaviour **2**, 753–742 (2018)
7. Gosling, S.D., Rentfrow, P.J., Swann, W.B.: A very brief measure of the big-five personality domains. Journal of Research in Personality **37**(6), 504–528 (2003)

8. Gu, J.C., Ling, Z.H., Zhu, X., Liu, Q.: Dually interactive matching network for personalized response selection in retrieval-based chatbots. In: EMNLP-IJCNLP, pp. 1845–1854 (2019)
9. Higashinaka, R., Mizukami, M., Kawabata, H., Yamaguchi, E., Adachi, N., Tomita, J.: Role play-based question-answering by real users for building chatbots with consistent personalities. In: SIGDIAL, pp. 264–272 (2018)
10. Isbister, K., Nass, C.: Consistency of personality in interactive characters: verbal cues, non-verbal cues, and user characteristics. Human-Computer Studies **53**(2), 251–267 (2000)
11. John, N.E., Rossi, A., Rossi, S.: Personalized human-robot interaction with a robot bartender. In: UMAP, pp. 155–159 (2022)
12. Kampman, O., J. Barezi, E., Bertero, D., Fung, P.: Investigating audio, video, and text fusion methods for end-to-end automatic personality prediction. In: ACL, pp. 606–611 (2018)
13. Kawahara, T.: Spoken dialogue system for a human-like conversational robot ERICA. In: IWSDS (2018)
14. Lala, D., Milhorat, P., Inoue, K., Ishida, M., Takanashi, K., Kawahara, T.: Attentive listening system with backchanneling, response generation and flexible turn-taking. In: SIGDIAL (2017)
15. Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., Dolan, B.: A persona-based neural conversation model. In: ACL, pp. 994–1003 (2016)
16. Mairesse, F., Walker, M.A.: Controlling user perceptions of linguistic style: Trainable generation of personality traits. Computational Linguistics **37**(3), 455–488 (2011)
17. McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality **60**(2), 175–215 (1992)
18. Nass, C., Moon, Y., B.J.Fogg, Reeves, B., Dryer, D.: Can computer personalities be human personalities? Human-Computer studies **43**, 223–239 (1955)
19. Saha, S., Das, S., Srihari, R.: Stylistic response generation by controlling personality traits and intent. In: Workshop on NLP for Conversational AI, pp. 197–211 (2022)
20. Salem, M., Ziadee, M., Sakr, M.: Effects of politeness and interaction context on perception and experience of HRI. In: ICSR, pp. 531–541 (2013)
21. Serban, I.V., Lowe, R., Henderson, P., Charlin, L., Pineau, J.: A survey of available corpora for building data-driven dialogue systems. Dialogue and Discourse **9**(1), 1–49 (2018)
22. Suman, C., Saha, S., Gupta, A., Pandey, S.K., Bhattacharyya, P.: A multi-modal personality prediction system. Knowledge-Based Systems **236**, 107,715 (2022)
23. Uchida, T., Minato, T., Nakamura, Y., Yoshikawa, Y., Ishiguro, H.: 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)
24. Yamamoto, K., Inoue, K., Kawahara, T.: Semi-supervised learning for character expression of spoken dialogue systems. In: INTERSPEECH, pp. 4188–4192 (2020)
25. Zhang, Q., Esterwood, C., Yang, X.J., Jr., L.P.R.: An automated vehicle (AV) like me? the impact of personality similarities and differences between humans and avs. In: AI-HRI (2019)