

Dialogue Behavior Control Model for Expressing a Character of Humanoid Robots

Kenta Yamamoto, Koji Inoue, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara

School of Informatics, Kyoto University, Sakyo-ku, Kyoto 606-8501, Japan

E-mail: {yamamoto, inoue, shizuka, takanasi, kawahara}@sap.ist.i.kyoto-u.ac.jp

Abstract—This paper addresses character expression for humanoid robots that play a social role via spoken dialogue so that the character matches to the given social role such as a lab guide or a counselor. While conventional methods of character expression mostly focused on changing the style of utterance texts, this study focuses on dialogue behavior features that may affect the impression of spoken dialogue. Specifically, we use five dialogue behavior features: utterance amount, backchannel frequency, backchannel variety, filler frequency, and switching pause length (the time until the system responds). We adopt three character traits of extroversion, emotional instability, and politeness for character expression. We then investigate the relationship between the dialogue behavior features and the character traits by conducting subjective evaluations. A statistical analysis of the subjective evaluations shows that the dialogue behavior features except for the backchannel variety are related to either of the character traits. By using the subjective evaluation scores on the relevant traits, we can train models to control the dialogue behavior features of a robot according to the desired character. Another experimental evaluation demonstrates the feasibility of character expression with regard to the traits of extroversion and politeness.

I. INTRODUCTION

Humanoid robots, which naturally interact with people, have been studied and developed [1], [2], [3]. Humanoid robots have an appearance and behavior similar to those of human beings, and thus users are expected to feel a character of the robot in the dialogue. A humanoid robot is usually given a social role such as a lab guide or a counselor depending on the dialogue task. Expressing characters matching to such social roles would not only give a good impression to users but also have a good effect on the performance of the task. For example, it is easy to talk to a calm counselor.

There are several studies on character expressions for dialogue systems. PERSONAGE is a system that generates a response sentence that matches the designated character [4]. The response generation is based on the relationship between the Big Five parameters and the corresponding features of sentences. There are also some methods of expressing characters for spoken dialogue systems by linguistic patterns [5], [6]. In these studies, characters are represented by changing the style of sentences. On the other hand, in spoken dialogue, factors such as the way speaker talks also have an impact on the impression. There is a spoken dialogue system that adjusts its speaking rate to those of the user [7]. It is expected that people who talk frequently are extroverted, and those who use a lot of fillers are likely to give an impression that they are restless.

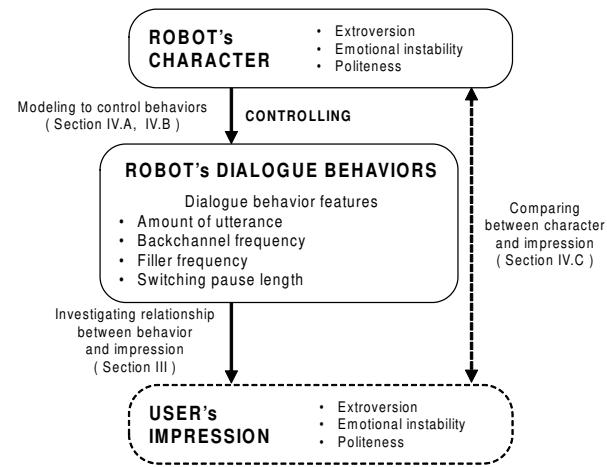


Fig. 1. Outline of this study

In this study, we present a model that expresses characters by controlling behaviors of the robot in dialogue. Fig. 1 depicts the outline of this study. These behaviors can be controlled by adjusting feature parameters of the robot's utterances. To investigate the feasibility of expressing characters, we conduct an experiment and analyze the effect of controlling the dialogue behavior feature parameters on the impression of three character traits: extroversion, emotional instability, and politeness. Based on the analysis, we construct a model to control the dialogue behavior features based on a given character.

II. CHARACTER EXPRESSION

We define characters in consideration of the personality traits of psychology and the usage of robots in society. Then, we choose effective dialogue behavior features for the character expression.

A. Definition of characters

In psychology, several scales expressing human personality have been proposed. Among them, extroversion and emotional instability are most widely used for many character classifications. For example, Eysenck [8] expressed a personality in two dimensions using the traits of extroversion and neuroticism (emotional instability). The Big Five scale expresses a per-

sonality using five traits of extroversion, emotional instability, integrity, agreeableness, and openness [9], [10].

Personality traits in psychology are also used as characters in dialogue systems. Extroversion and emotional instability have also been used in character expression of agents in previous studies [11]. For example, an extrovert character may be preferred for a lab guide, and low emotional instability is required for a counselor. Politeness was also considered to control the impression of agents [12], [13]. Politeness is required to spoken dialogue systems to serve many formal tasks such as a receptionist.

In this study, characters are defined using three traits of extroversion, emotional instability, and politeness. These three traits can be easily perceived [10] by many users and they are expected to contribute to task achievement in dialogue. Note that the correlation among these traits was partly observed [10]. We will be able to choose an appropriate character in these three traits according to the social role of the robot.

B. Dialogue behavior features for controlling behavior

We examine dialogue behavior features that can affect the impression of the speaker in dialogue. The amount of utterance is considered to affect the impression in dialogue, for example, a person that speaks a lot seems more extroverted. Backchannels affect the impression of characters because they have various roles in dialogue such as expressing empathy and showing understanding. The frequency and type of backchannels have some effect on the impression of extroversion and emotional instability [11]. Fillers are also thought to have an influence on the impression of characters. Using a lot of fillers looks emotionally unstable. The switching pause, which is the time until the system responds in spoken dialogue, is an effective cue for users to recognize the system's character. It was investigated that the length of switching pause has an impact on impressions [14]. Based on the above, we adopt the amount of utterance, the frequency of backchannels, the variety of backchannels, the frequency of fillers, and the length of switching pause as dialogue behavior features.

III. ANALYSIS ON THE EFFECT OF DIALOGUE BEHAVIOR FEATURES ON IMPRESSIONS OF CHARACTERS

In order to construct a behavior control model, the effect of the dialogue behavior features on the character impression is investigated. In the experiments, the following hypotheses are made.

- 1) The extroversion is associated with the utterance amount and the backchannels frequency. A robot that talks a lot and generates many backchannels is perceived to be an extrovert.
- 2) The emotional instability is associated with the backchannel variety and the filler frequency. A robot that generates a same type of backchannels and many fillers is perceived to be emotionally unstable.
- 3) The politeness is associated with the switching pause length. A robot that waits for a long time before talking is perceived to be polite.

A. Speech samples

Audio samples of 20 conditions are prepared for experiments. The utterances of the robot are generated using the text-to-speech software in these scenarios. The utterances of the user are made by the experimenter. These utterances are spoken in Japanese. Two scenarios are prepared with reference to a human-robot dialogue corpus recorded in a WOZ setting [15], [16]. The contents of the dialogue are designed to be natural with any character and even when the dialogue behavior features (excluding latency) are adjusted.

A baseline dialogue of about one minute was created for each of the two scenarios. Backchannels and fillers appear moderately in the reference dialogue. The switching pause length is set to 0.5 seconds. In comparison dialogue used for each experimental condition, only one corresponding feature is adjusted from the baseline dialogue, and the remaining features are kept same as the baseline dialogue. In order to obtain clear results in the following experiments, we prepared the low (small, short) condition and the high (large, long) condition.

Prepared patterns of the dialogue behavior features are shown in Table I. When the backchannel frequency is high, backchannels are inserted at all clause boundaries in the user utterances. When the backchannel frequency is low, all backchannels are removed. With regard to the backchannel variety, the backchannels are changed to many kinds of backchannels in a large condition, and all the backchannels are replaced with "yes" in a small condition. In the high filler frequency condition, fillers are inserted at all clause boundaries and at the sentence beginning of the utterances. In the low condition, all fillers are removed. When the switching pause is long, the turn-taking time is set to 3 seconds. When the switching pause is short, the start of the system utterance overlaps the end of the user utterance by 0.5 seconds. With regard to the utterance amount of the system, we extended the original two scenarios. Specifically, each scenario was concatenated with its following part of the dialogue in the corpus so that the length of utterances becomes twice or more. This is needed to make it easy for the subjects to recognize the difference of the utterance amount. One of the extended scenarios was system-dominant (Large condition), and the other was user-dominant (Small condition).

B. Experimental procedure

Forty-six university students (28 men and 18 women, 18 - 23 years old) participated in the experiment. Each participant listened to the speech samples and responded to the questionnaire about an impression of the robot. We presented the baseline dialogue at first and the dialogue of different utterance amounts at last. The remaining conditions were randomly arranged between these for each experiment. We presented 20 samples in total to each participant.

Participants answered questionnaires on a 7-point scale, from 1 (not at all) to 7 (completely), whether the item is true to the system. For a questionnaire of extroversion and emotional instability, we used short versions of the Big Five scales [17], [10], which is widely used in personality psychology

TABLE I
CONTROL OF DIALOGUE BEHAVIOR FEATURES

| Dialogue behavior features | Conditions | Details |
|----------------------------|------------|---|
| Utterance amount | Large | Robot : 49.2 seconds, User : 25.3 seconds |
| | Small | Robot : 25.5 seconds, User : 38.8 seconds |
| Backchannel frequency | High | At all clause boundaries during user's utterance |
| | Low | Delete all robot's backchannels |
| Backchannel variety | Large | 4 types |
| | Small | 1 type |
| Filler frequency | High | At all clause boundaries and beginning of sentences |
| | Low | Delete all robot's fillers |
| Switching pause length | Long | 3 seconds |
| | Short | 0.5 seconds overlap |

TABLE II
QUESTIONNAIRE ITEMS

| Character traits | Items |
|-----------------------|-------------|
| Extroversion | Talkative |
| | Reticent* |
| | Frisky |
| Emotional instability | Extrovert |
| | Worrisome |
| | Insecure |
| Politeness | Overanxious |
| | Precautious |
| | Polite |
| | Gracious |

(* reverse scale)

in Japanese. It is also used for the impression evaluation of speeches [17]. We added two items for politeness. The items we used are summarized in Table II. In addition, the naturalness of the dialogue was also evaluated. Finally, we compute the average value for the corresponding items of each character trait.

C. Results

A variance analysis is performed among the three groups (Low (small, short) condition, Baseline, and High (large, long) condition) for each dialogue behavior feature. Character traits that showed significant differences ($p < 0.05$) between High and Low conditions by multiple comparisons in both scenarios are described below.

The analysis results for the extroversion trait are shown in Table III. The higher the backchannel frequency, the lower the filler frequency, and the shorter the switching pause length are, the more the robot is considered to be extroverted. The analysis results on the emotional instability trait are shown in Table IV. The higher the filler frequency is and the longer the switching pause length is, the more emotionally unstable the robot is deemed. The analysis results on the politeness trait are shown in Table V. The lower the backchannel frequency is and the longer switching pause length is, the politer the robot is deemed. The analysis results on the utterance amount are shown in Table VI. The robot is felt to be extroverted when the speaking time is longer, and introvert when the speaking time is shorter. No significant difference is found regarding the

emotional instability. These results confirmed the hypotheses made in this section, except for the backchannel type.

D. Discussions

The evaluation score of naturalness except for the switching pause condition got equal to or higher than that of the baseline dialogue. Therefore, there was no problem of naturalness due to the adjustment of the features.

There were few traits where the backchannel variety had an effect. Since the number of backchannels is not so large in the one-minute dialogue, the participants may not have noticed the difference in the backchannel variety. On the other hand, the filler frequency had effects in many traits. In multiple comparisons, many character traits showed the tendency in the order of Low condition, Baseline condition, High condition. The switching pause length had a large effect on the impressions of all traits. The difference in the switching pause length is easily recognized. The utterance amount had an effect on extroversion and politeness. These results showed the feasibility of expressing characters by controlling the utterance amount, the backchannel frequency, the fillers frequency, and the switching pause length.

IV. DIALOGUE BEHAVIOR CONTROL MODEL

We construct a model that controls the dialogue behavior features based on a given character using a logistic regression model. Then, we conduct an evaluation experiment of expressing characters by the dialogue behavior control model.

A. Logistic regression model from character to control parameter of dialogue behavior features

The control model is constructed for each dialogue behavior feature. The backchannel variety is not used because it had little effect on any character traits in the experiment in Section III. The inputs of the models are variables of character traits: extroversion, emotional instability, and politeness, taking a value from 1 to 7. The output of the models is control parameters [0 - 1] of a corresponding dialogue behavior feature. We used the data (features and traits) which showed a significant effect on the impression in Section III (bold items in Table III, IV, V, and VI) for model training. The total amount of the samples was 920. The Low condition of each dialogue behavior feature is labeled as 0 and the High condition is labeled as 1. The model learns the mapping from

VARIANCE ANALYSIS OF SCORES BASED ON SUBJECT EVALUATION

TABLE III
EXTROVERSION

| | Dialogue behavior features ⁱ | High (Large / Long) | | Baseline | | Low (Small / Short) | | F-measure | Multiple comparison conditions ⁱⁱ |
|----|---|------------------------|------|----------|------|------------------------|------|-----------|--|
| | | Mean | SD | Mean | SD | Mean | SD | | |
| S1 | Backchannel frequency | 5.44 | 1.08 | 3.99 | 1.16 | 4.08 | 1.07 | 27.174** | High>Low, High>Baseline |
| | Backchannel variety | 4.64 | 1.22 | 3.99 | 1.16 | 4.76 | 0.90 | 9.342** | Large>Baseline, Small>Baseline |
| | Filler frequency | 3.51 | 1.02 | 3.99 | 1.16 | 4.84 | 1.07 | 18.963** | Low>Baseline>High |
| | Switching pause length | 2.56 | 0.91 | 3.99 | 1.16 | 5.20 | 1.06 | 70.495** | Short>Baseline>Long |
| S2 | Backchannel frequency | 5.35 | 0.89 | 4.67 | 1.17 | 4.73 | 0.83 | 10.757** | High>Low, High>Baseline |
| | Backchannel variety | 4.88 | 0.94 | 4.67 | 1.17 | 4.80 | 0.88 | 0.289 | |
| | Filler frequency | 3.53 | 1.06 | 4.67 | 1.17 | 5.17 | 0.89 | 41.892** | Low>Baseline>High |
| | Switching pause length | 2.70 | 0.97 | 4.67 | 1.17 | 4.64 | 1.19 | 83.329** | Short>Long, Baseline>Long |

(* p < 0.05, ** p < 0.01)

TABLE IV
EMOTIONAL INSTABILITY

| | Dialogue behavior features ⁱ | High (Large / Long) | | Baseline | | Low (Small / Short) | | F-measure | Multiple comparison conditions ⁱⁱ |
|----|---|------------------------|------|----------|------|------------------------|------|-----------|--|
| | | Mean | SD | Mean | SD | Mean | SD | | |
| S1 | Backchannel frequency | 2.55 | 1.05 | 3.38 | 1.17 | 3.10 | 1.10 | 9.149** | Low>High, Baseline>High |
| | Backchannel variety | 3.08 | 1.15 | 3.38 | 1.17 | 2.73 | 0.94 | 7.060** | Baseline>Small |
| | Filler frequency | 4.71 | 1.42 | 3.38 | 1.17 | 2.42 | 1.09 | 56.378** | High>Baseline>Low |
| | Switching pause length | 4.76 | 1.16 | 3.38 | 1.17 | 2.38 | 1.03 | 62.988** | Long>Baseline>Short |
| S2 | Backchannel frequency | 2.91 | 1.21 | 3.92 | 1.45 | 3.22 | 1.16 | 1.921 | |
| | Backchannel variety | 3.05 | 1.24 | 3.92 | 1.45 | 3.16 | 1.11 | 0.811 | |
| | Filler frequency | 5.07 | 1.50 | 3.92 | 1.45 | 2.37 | 1.03 | 70.268** | High>Baseline>Low |
| | Switching pause length | 4.84 | 1.26 | 3.92 | 1.45 | 3.09 | 1.31 | 48.321** | Long>Short, Long>Baseline |

(* p < 0.05, ** p < 0.01)

TABLE V
POLITENESS

| | Dialogue behavior features ⁱ | High (Large / Long) | | Baseline | | Low (Small / Short) | | F-measure | Multiple comparison conditions ⁱⁱ |
|----|---|------------------------|------|----------|------|------------------------|------|-----------|--|
| | | Mean | SD | Mean | SD | Mean | SD | | |
| S1 | Backchannel frequency | 4.37 | 1.62 | 5.38 | 1.19 | 5.21 | 1.05 | 9.512** | Low>High, Baseline>High |
| | Backchannel variety | 5.10 | 1.17 | 5.38 | 1.19 | 5.46 | 1.05 | 1.508 | |
| | Filler frequency | 4.77 | 1.33 | 5.38 | 1.19 | 5.21 | 1.00 | 4.078* | Baseline>High |
| | Switching pause length | 4.60 | 1.18 | 5.38 | 1.19 | 2.84 | 1.09 | 63.181** | Baseline>Long>Short |
| S2 | Backchannel frequency | 4.12 | 1.62 | 4.05 | 1.41 | 4.85 | 0.96 | 13.939** | Low>High>Baseline |
| | Backchannel variety | 5.01 | 0.96 | 4.05 | 1.41 | 4.88 | 0.94 | 31.274** | Large>Baseline, Small>Baseline |
| | Filler frequency | 4.54 | 1.02 | 4.05 | 1.41 | 4.75 | 1.21 | 17.613** | High>Baseline, Low>Baseline |
| | Switching pause length | 4.41 | 1.12 | 4.05 | 1.41 | 3.10 | 1.59 | 18.896** | Long>Baseline>Short |

(* p < 0.05, ** p < 0.01)

S1 : Scenario 1, S2 : Scenario 2

i : Dialogue behavior features (**Bold letters**) showed significant difference in both scenarios.ii : Multiple comparison conditions (**Bold letters**) showed significant difference (p < 0.05) between high and low conditions.TABLE VI
T-TEST ON UTTERANCE AMOUNT

| Character traits | Large | | Small | | T ratio |
|-----------------------|-------|------|-------|------|---------|
| | Mean | SD | Mean | SD | |
| Extroversion | 5.74 | 0.82 | 5.03 | 0.84 | 4.991** |
| Emotional instability | 2.74 | 0.93 | 2.83 | 1.04 | 0.545 |
| Politeness | 4.76 | 1.22 | 5.91 | 1.04 | 5.688** |

(* p < 0.05, ** p < 0.01)

the evaluation score [1 - 7] of the character traits to this target label.

We conducted a cross-validation using three-quarters of each dataset for training and one-quarter for evaluation. The data of the same subject are not included in both the training data and the test data. The results of binary prediction in which

the threshold is set to 0.5 are shown in Table VII. We achieve F-measure scores of 0.73 to 0.91, which is reasonable for the behavior control model.

B. Method for controlling dialogue behavior features

The dialogue behavior feature is controlled using the output of the logistic regression model. To control the utterance amount, we prepare two utterance patterns: many utterances and fewer utterances. One of the utterance pattern is selected. Backchannels can be generated according to a statistical model [18], [19], [20]. In order to simplify the model, it is assumed that the probabilities of occurrence of a backchannel at all clause boundaries in the user's utterance are equal, and the output of the behavior control model is used for a threshold. A value sampled from a uniform distribution in the range

TABLE VII
PREDICTION ACCURACY OF CONTROL OF EACH DIALOGUE BEHAVIOR FEATURE

| Dialogue behavior features | Used traits | | | Class | Precision | Recall | F-measure |
|----------------------------|-------------|----|----|-------|-----------|--------|-----------|
| | Ex | Em | Po | | | | |
| Utterance amount | ✓ | ✓ | | Long | 0.86 | 0.78 | 0.81 |
| | | | | Short | 0.81 | 0.87 | 0.83 |
| Backchannel frequency | ✓ | ✓ | | High | 0.74 | 0.73 | 0.73 |
| | | | | Low | 0.75 | 0.74 | 0.74 |
| Filler frequency | ✓ | ✓ | | High | 0.81 | 0.75 | 0.77 |
| | | | | Low | 0.76 | 0.81 | 0.78 |
| Switching pause length | ✓ | ✓ | ✓ | Long | 0.91 | 0.90 | 0.91 |
| | | | | Short | 0.91 | 0.91 | 0.91 |

Ex: Extroversion, Em: Emotional instability, Po: Politeness

TABLE VIII
PEARSON'S PRODUCT-MOMENT CORRELATION COEFFICIENT OF A GIVEN CHARACTER AND THE NORMALIZED IMPRESSION EVALUATION POINT

| Character traits | Correlation coefficient | T ratio |
|-----------------------|-------------------------|---------|
| Extroversion | 0.570 | 9.163** |
| Emotional instability | -0.004 | -0.056 |
| Politeness | 0.235 | 3.185** |

(* p < 0.05, ** p < 0.01)

Bold letters indicate the dialogue behavior features was correlated

of [0, 1] at each clause boundary is set as the occurrence probability of the backchannel, and when the probability exceeds the threshold, a backchannel is generated. Similarly, the filler frequency is controlled by adjusting the threshold of the statistical model according to the filler control model. The output of the switching pause length in the behavior control model [0, 1] is normalized in accordance with the range [-0.5, 3] of the experimental condition in Section III. The value is set as the switching pause length. In the case of a negative value, it overlaps the end of the utterance of the user.

C. Evaluation experiment

We prepared dialogue samples by controlling the robot's behavior with the behavior control model. We used 16 kinds of characters based on the value of three character traits [1,7]. Eleven male university students evaluated the impression of the 16 dialogues where the generated behaviors are different. The questionnaire is same as that used in the experiment in Section III. The rating points were normalized by the mean and the variance for each participant. The Pearson's product moment correlation coefficients between the ratings (Z-score) and the original values of the given character traits are shown in Table VIII. From the results, significant correlations are confirmed with extroversion and politeness. This means that it is possible to express extroversion and politeness by using the model. There is no correlation for the emotional instability probably because the control parameter for the emotional instability is smaller than those of extroversion and politeness in the trained model.

V. CONCLUSIONS

In this paper, a model of character expression for humanoid robots is presented. This model generates behaviors matched to a given character by controlling the dialogue behavior features.

We investigated the effect on the character impression by controlling the dialogue behavior features. It is shown that the dialogue behavior features such as the amount of utterance, the frequency of backchannels, the frequency of filler, and the switching pause have effects on the impression of extroversion, emotional instability, and politeness. We constructed a dialogue behavior behavior control model based on this analysis. Experimental results showed that it is possible to express extroversion and politeness by using the model. Our behavior control model can be extended to various kinds of spoken dialogue systems. In this study, we used a simple model for generating dialogue behavior features. In future work, we will improve the feature generation and implement it in a humanoid robot that plays several social roles.

ACKNOWLEDGEMENT

This work was supported by JST ERATO Ishiguro Symbiotic Human-Robot Interaction program (Grant Number JPM-JER1401), Japan.

REFERENCES

- [1] C. Breazeal and B. Scassellati, "A context-dependent attention system for a social robot," *IJCAI*, pp. 1146–1153, 1999.
- [2] K. Inoue, P. Milhorat, D. Lala, T. Zhao, and T. Kawahara, "Talking with ERICA, an autonomous android," *SIGDIAL*, pp. 212–215, 2016.
- [3] H. Ishiguro, T. Ono, M. Imai, and T. Kanda, "Development of an interactive humanoid robot "Robovi" –an interdisciplinary approach–," *Robotics Research*, Springer, pp. 179–191, 2003.
- [4] F. Mairesse and M. A. Walker, "Controlling user perceptions of linguistic style: Trainable generation of personality traits," *Computational Linguistics*, vol. 37, no. 3, pp. 455–488, 2011.
- [5] C. Miyazaki, T. Hirano, R. Higashinaka, T. Makino, and Y. Matsuo, "Automatic conversion of sentence-end expressions for utterance characterization of dialogue systems," *PACLIC*, pp. 307–314, 2015.
- [6] S. Raymond, K. Hideaki, O. Katumi, and M. Takeshi, "Towards the characterization control in personalized vehicles by text-based style change," *Information Processing Society of Japan*, vol. 53, no. 4, pp. 1269–1276, 2012, in Japanese.
- [7] R. Levitan, S. Benus, R. H. Gálvez, A. Gravano, F. Savoretti, M. Trnka, A. Weise, and J. Hirschberg, "Implementing acoustic-prosodic entrainment in a conversational avatar," *INTERSPEECH*, vol. 16, pp. 1166–1170, 2016.
- [8] H. Eysenck, *The biological basis of personality*, Routledge, 2017.
- [9] P. T. Costa and R. R. McCrae, "Normal personality assessment in clinical practice: The neo personality inventory," *Psychological Assessment*, vol. 4, no. 1, pp. 5–13, 1992.
- [10] S. Wada, "Construction of the big five scales of personality trait terms and concurrent validity with NPI," *Japanese Journal of Psychology*, vol. 67, no. 1, pp. 61–67, 1996, in Japanese.
- [11] E. D. Sevin, S. J. Hyniewska, and C. Pelachaud, "Influence of personality traits on backchannel selection," *IVA*, pp. 187–193, 2010.

- [12] S. Gupta, M. A. Walker, and D. M. Romano, "How rude are you?: Evaluating politeness and affect in interaction," *ACII*, pp. 203–217, 2007.
- [13] N. Wang, W. L. Johnson, R. E. Mayer, P. Rizzo, E. Shaw, and H. Collins, "The politeness effect: Pedagogical agents and learning outcomes," *International Journal of Human-Computer Studies*, vol. 66, no. 2, pp. 98–112, 2008.
- [14] C. Nagaoka, M. Draguna, M. Komori, and T. Nakamura, "The influence of switching pauses on interpersonal perception in dialogues," *Human Interface Symposium*, pp. 171–174, 2002, in Japanese.
- [15] K. Inoue, D. Lala, K. Takanashi, and T. Kawahara, "Latent character model for engagement recognition based on multimodal behaviors," *International Workshop Spoken Dialogue Systems*, 2018.
- [16] T. Kawahara, "Spoken dialogue system for a human-like conversational robot ERICA," *International Workshop Spoken Dialogue Systems*, 2018, keynote speech.
- [17] T. Uchida, "Effects of the speech rate on speakers' personality-trait impressions," *Japanese Journal of Psychology*, vol. 73, no. 2, pp. 131–139, 2002, in Japanese.
- [18] S. Fujie, K. Fukushima, and T. Kobayashi, "A conversation robot with back-channel feedback function based on linguistic and nonlinguistic information," *ICARA*, pp. 379–384, 2004.
- [19] T. Kawahara, T. Yamaguchi, K. Inoue, K. Takanashi, and N. G. Ward, "Prediction and generation of backchannel form for attentive listening systems," *INTERSPEECH*, pp. 2890–2894, 2016.
- [20] K. P. Truong, R. Poppe, and D. Heylen, "A rule-based backchannel prediction model using pitch and pause information," *INTERSPEECH*, pp. 3058–3061, 2010.

APPENDIX

Two dialogue samples used in the experiments in Section IV-C are shown below. These dialogues are spoken in Japanese.

"R:" stands for a robot utterance. "U:" stands for a user utterance. Utterances surrounded by parentheses are robot's fillers. Utterance surrounded by parentheses with the tag of "BC:" are robot's backchannels.

The first dialogue sample is the case where the model input of extroversion is 1, that of emotional instability is 4, and that of politeness is 4.

(Switching pause length : 3.0s)

- R: (Well,) where did you come from today?
- U: Today I came from Suita by train.
- R: (Oh!) Is your company located in Suita?
- U: Oh Hum. My company is in Suita, and I came here from there.
- R: (Oh!) I see. Thank you very much. Do you live in Osaka, too?
- U: Oh Hum. Well, I live in Fukushima, Osaka.
- R: (Um,) Fukushima, it is convenient.
- U: That's right. It is very convenient because there are many shops that offer delicious dishes near Osaka station.
- R: (Um,) That's right. (You know,) I occasionally go to Fukushima.
- U: Is that so?
- R: (Well,) since there are many shops crowded with businessmen, (you know,) I feel they have fun.
- U: Yes, that's right. There are a lot of shops I like, so I am satisfied with them.

The second dialogue sample is the case where the model input of extroversion is 4, that of emotional instability is 7, and that of politeness is 1.

(Switching pause length : 0s)

- R: (Well,) Mr. Kimura, where did you come from today?
- U: Today I came from Suita (BC:Hum.) by train.
- R: (Oh!) Is your company located in Suita?
- U: Oh Hum. (BC:Hum.) My company is in Suita, (BC:Hum.) and I came here from there. (BC:Hum.)
- R: Suita? Thank you very much for coming from the far place. (Well,) do you live in Osaka, too?
- U: Oh Hum. (BC:Hum.) Well, I live in Fukushima, Osaka. (BC:Hum.)
- R: (Well,) you live in Fukushima? It is convenient. (You know,) I occasionally go to Fukushima, it is very convenient.
- U: I agree. (BC:Hum.)
- R: (Well,) since there are many shops crowded with businessmen, I feel they have fun.
- U: Yes, that's right. (BC:Hum.) There are a lot of shops I like, (BC:Hum.) so I am satisfied with them. (BC:Hum.)
- R: (Well,) would you introduce recommended shops next time (, you know;) if you don't mind?