# A Character Expression Model Affecting Spoken Dialogue Behaviors

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

**Abstract** We address character (personality) expression for a spoken dialogue system in order to accommodate it in particular dialogue tasks and social roles. While conventional studies investigated controlling the linguistic expressions, we focus on spoken dialogue behaviors to express systems' characters. Specifically, we investigate spoken dialogue behaviors such as utterance amount, backchannel frequency, filler frequency, and switching pause length in order to express three character traits: extroversion, emotional instability, and politeness. In this study, we evaluate this model with a natural spoken dialogue corpus. The results reveal that this model expresses reasonable characters according to the dialogue tasks and the participant roles. Furthermore, it is also shown that this model is able to express different characters among participants given the same role. A subjective experiment demonstrated that subjects could perceive the characters expressed by the model.

## 1 Introduction

Character (personality) expression is a desired function for spoken dialogue systems. Recently, spoken dialogue systems have been applied to many social roles such as a psychological counselor [6], a museum guide [27], and an attentive listener [17]. In order to realize natural dialogue in these social scenarios, it is important to assign proper characters to the spoken dialogue systems. For example, it is expected that a guide is extrovert, and a psychological counselor is introvert and emotionally stable. Other studies pointed out that proper character expression for spoken dialogue systems leads to an increase of user engagement and naturalness of dialogue [21, 11, 9, 23, 24]. It is also reported that setting a character is important for embodied virtual agents to make them empathetic and realistic [1, 28].

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

In this study, we focus on spoken dialogue behaviors that have not yet been studied well in character expression. Related works on character expression have mainly focused on content of utterance sentences. Earlier studies investigated control methods for linguistic patterns of system utterances [22, 18, 19]. Recently, collections of large-scale dialogue text corpora together with character information have been conducted towards neural-network-based text generation considering systems' character (personality) [15, 26, 14, 32].

When we consider character expression on spoken dialogue systems, it is important to control not only content of utterance sentences but also spoken dialogue behaviors such as utterance amount, backchannel frequency, filler frequency, and switching pause length. Other studies suggested that the dialogue behaviors were related to the impression on an interlocutor in dialogue [4, 25, 20].

We aim to realize a character expression model affecting the spoken dialogue behaviors. In our previous study [31], we proposed a model that expresses three character traits and showed it is possible to properly express extroversion and politeness by conducting a subjective evaluation with speech samples artificially generated. However, the validity of the model in natural dialogue was not investigated. In this study, we analyze and enhance the model with a human-robot dialogue corpus. Specifically, we estimate characters from dialogue behaviors observed in the corpus by using the character expression model. Then, we analyze the validity of the identified characters by comparing with characteristic of dialogue tasks in the corpus. This study contributes to the realization of spoken dialogue systems that express those proper characters in social scenarios by controlling not only utterance sentences but also the dialogue behaviors.

## 2 Character traits and spoken dialogue behaviors

In this study, we use three character traits: extroversion (extrovert vs. introvert), emotional instability (stable vs. instable), and politeness (polite vs. casual). Extroversion and emotional instability [8] are selected from the Big Five traits [7, 16, 3, 29], whereas politeness is additionally considered in this study. Politeness is important when considering the practical use of dialogue systems. Although an earlier study pointed out that these character traits are sometimes mutually related [3], in this study, we deal with them individually for the simplicity of the character expression model.

To express the character traits, we design a model affecting spoken dialogue behaviors such as utterance amount, backchannel frequency, backchannel variety, filler frequency, and switching pause length. Utterance amount represents the ratio of utterance between dialogue participants. Backchannels are interjections expressed by listeners such as "*Yeah*" in English and "*Un*" in Japanese [5]. Fillers are short phrases to fill the silence to hold the conversational floor such as "*Well*" in English and "*E-*" in Japanese [30]. Switching pause length is defined as the time length be-

tween the end of the preceding turn and the start of the following turn. Note that these utterances are occasionally overlapped in natural human-human dialogue.

## 3 A character expression model

We briefly explain our proposed model [31]. At first, we conducted a subjective evaluation experiments to find the relationship between the character traits and the spoken dialogue behaviors. Using the evaluation results, we trained a character expression model controlling the spoken dialogue behaviors.

### *3.1 Impression evaluation of character traits on varied samples of dialogue behaviors*

In this experiment, each subject was asked to listen to speech samples and then to evaluate his/her impression on the character traits of the speaker in each sample. For the evaluation, we used several adjectives with the 7-point scale. For extroversion and emotional instability, we used 8 adjectives (4 for each) from a short version of Big Five scale [29] such as *talkative* for extroversion and *anxious* for emotional instability. We also used two adjectives, *polite* and *courteous*, for politeness. The subjects were 46 university students (18 females and 28 males, from 18 to 23 years old). Note that the experiment hereafter was done in Japanese.

The speech samples used in this evaluation experiment were generated as follows. We selected two dialogue scenarios from our human-robot dialogue corpus, which is described in Section 4.1. For each dialogue scenario, we generated several speech samples by controlling the dialogue behaviors. The speaking parts of the robot were replaced with different voice samples generated by text-to-speech software. At first, we generated a baseline speech sample where backchannel and filler tokens are kept as the original dialogue and the switching pause length is set to 0.5 seconds. Using the baseline sample, we changed each dialogue behavior one by one [31]. We used these generated speech samples to compare the perceived character traits between different conditions on each dialogue behavior (e.g. high backchannel frequency vs. low backchannel frequency).

We analyzed the evaluation scores with ANOVA and multiple comparisons. The relationship between the condition of dialogue behaviors and perceived character traits are summarized as below.

- The more extrovert was perceived with the larger utterance amounts, the more frequent backchannels, the fewer frequent fillers, and the shorter switching pauses.
- The more emotionally unstable was perceived with the more frequent fillers and the longer switching pauses.

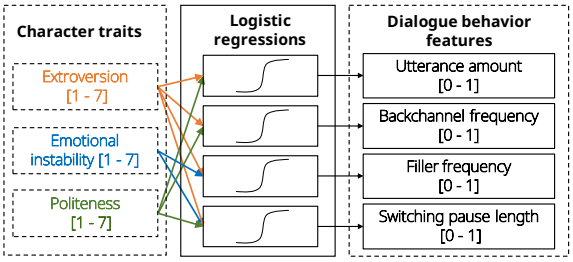


**Fig. 1** Character expression model affecting spoken dialogue behaviors

**Table 1** Correlation coefficients between characters given to the model and characters evaluated by subjects

| Character trait | Correlation coefficient | $t$-ratio |
|---|---|---|
| Extroversion | 0.570 | 9.163* |
| Emotional instability | −0.004 | −0.056 |
| Politeness | 0.235 | 3.185* |

(* $p < 0.01$)

- The more polite was perceived with the smaller utterance amounts, the less frequent backchannels, and the longer switching pauses.

It was shown that the three character traits are related to the different set of the dialogue behaviors.

## *3.2 Training of character expression model*

Using the scores obtained in the impression evaluation, we trained a character expression model to control the dialogue behaviors. The model is given scores of the character traits in the 7-point scale and then outputs control values for the dialogue behaviors, modeled by logistic regressions as shown in Fig. 1. Note that we trained each logistic regression for each dialogue behavior individually, and used only scores related to each behavior found in the previous evaluation, described as crossed lines in the left side of Fig. 1. We used a paired data of the behavior conditions (binary value) and the corresponding character trait scores obtained in the impression evaluation, as reference labels and input data, respectively. In order to evaluate the trained model, we generated other speech samples whose dialogue behaviors were controlled by the trained model and then asked different subjects to evaluate character trait scores on the samples. Table 1 reports Pearson's product-moment correlation coefficients between characters input to the trained model and characters evaluated by the subjects. As a result, for extroversion and politeness, correlations were confirmed between the input and the perceived characters.

## 4 Corpus-based analysis

Since the generated speech samples used in the impression evaluation were artificial data. We investigate the validity and generality of our character expression model by using a natural spoken dialogue corpus. The dialogue corpus consists of multiple dialogue tasks, and each task has corresponding suitable characters. The character expression model shown in Fig. 1 is applied in the backward direction (right-to-left) in order to calculate characters from spoken dialogue behaviors observed in the corpus. Finally, we examine the tendency of the identified characters for each dialogue task to confirm whether our character expression model can express characters that match each task.

### *4.1 Human-robot dialogue corpus*

We used a human-robot dialogue corpus where a human subject, called a subject hereafter, talked with the android robot ERICA [10, 12] which was remotely operated by another person, called an operator. In this corpus, three types of dialogue tasks are designed: speed dating, job interview, and attentive listening. The roles of ERICA (operators) in these tasks are a practice partner in a first-time-meeting conversation, a job interviewer, and an attentive listener to encourage a subject's talk, in the above order. In this study, we analyze ERICA's (operators') characters by our character expression model because the subjects were different people in each session so that it was difficult to reliably analyze their characteristic. The number of used dialogue sessions are 33 for speed dating, 31 for job interviews, and 19 for attentive listening. Each dialogue session lasted about 10 minutes. In each dialogue session, whereas the human subject was a different person, the operator was randomly assigned from four persons who are amateur actresses. Besides the transcription of the dialogue, we annotated the dialogue with the events and timing of backchannels [5], fillers [30], conversational turns, and dialogue act [2].

### *4.2 Analysis method*

We calculated the ERICA's (operators') characters from her spoken dialogue behaviors observed in the corpus. At first, we divided each dialogue session into two-minute segments in order to ensure a sufficient amount of data for this analysis. We also empirically confirmed that two minutes is enough duration to observe the spoken dialogue behaviors to calculate the character trait scores. For each segment sample, the corresponding character trait scores were calculated by using our character expression model as below. We calculated feature values of the four spoken dialogue behaviors. Then, the feature values were converted to control amounts corresponding to the outputs of the logistic regression models. The amount of speech

was classified into large or small (1 or 0) by using the median value of the entire corpus. The number of backchannels was normalized by the number of inter-pausal units (IPUs) [13], namely pause segmentation, of the interlocutor who is the current speaker. The number of fillers was normalized by the number of IPUs of herself. Switching pause length was linearly converted from the range of $[-0.5, 3]$ seconds to the range of $[0, 1]$. If the length was shorter than -0.5 or larger than 3 seconds, the converted value was clipped at 0 or 1, respectively. Meanwhile, we enter all possible combinations of character trait scores ($7^3 = 343$ ways) to our character expression model and then calculated the corresponding control amount of the spoken dialogue behaviors. Finally, we compared the control amounts observed from the corpus behaviors with those from each combination of character trait scores. We identified the corresponding character trait scores by the minimal Euclid distance between the control amounts.

### *4.3 Analysis result among dialogue tasks*

We analyzed the distribution of the estimated ERICA's (operators') characters for each dialogue task. Fig. 2 reports the distributions in the speed dating task. Our character expression model indicates that extroversion and politeness varied from middle to high and emotional instability was low (stable). In this dialogue task, the participants met each other for the first time, and the purpose of the dialogue is to build a relationship between them. Therefore, they should exhibit extrovert and polite behaviors. At the same time, they could show their own individual characters on their behaviors because this dialogue is not too much constrained by their participant roles. This is the reason why the distribution is varied for middle to high on extroversion and politeness.

Fig. 3 reports the distributions in the job interview task. Our character expression model also showed the similar tendency as in the speed dating task. Extroversion and politeness varied from middle to high. This variation can be interpreted by that the operators (interviewers) held the dialogue initiative in this dialogue so that there was more chance to control their behaviors expressing their characters. Compared to the speed dating task, extroversion relatively tended to be neutral. This can be interpreted by the style of this dialogue which is more formal. Thus, it is expected that extroversion was restricted by the style of this dialogue.

Fig. 4 reports the distributions in the attentive listening task. Our character expression model showed the biased distributions on extroversion and politeness. In this dialogue, the operators (attentive listeners) needed to encourage and elicit the subjects' talk. Therefore, they should behave as extrovert and polite. Moreover, the dialogue initiative was held by the subjects (story tellers) in this dialogue and the behaviors of the operators were constrained by the dialogue role (attentive listener). This is the reason that the distributions are more biased than those of other tasks.

In summary, it is shown that our character expression model can represent reasonable characters according to the scenario of each dialogue task. As a whole, the

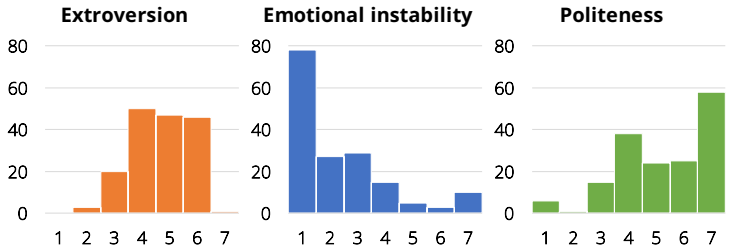


**Fig. 2** Estimated character distributions in speed dating task

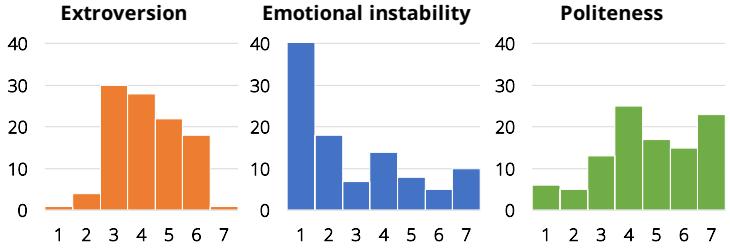


**Fig. 3** Estimated character distributions in job interview task

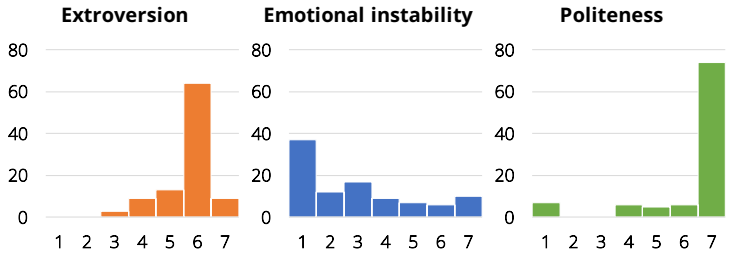


**Fig. 4** Estimated character distributions in attentive listening task

model shows that extroversion and politeness tended to be middle or high and emotional instability was low (stable) in this corpus. These character traits are expected in dialogue where participants meet each other for the first time.

## *4.4 Analysis result among operators*

Since there were four robot operators in this corpus, we further analyzed the distribution within each operator to find the individual difference among the operators the same dialogue task. Since emotional instability was low (stable) among the entire corpus, we analyzed only extroversion and politeness in this section. We also analyzed only the speed dating task where the number of samples is the largest for each operator.

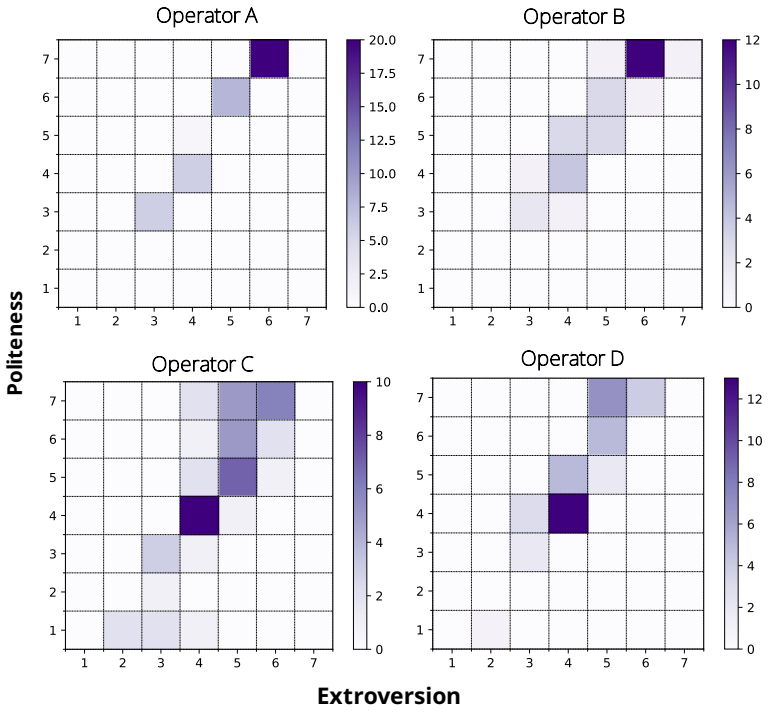


**Fig. 5** Character distribution (extroversion and politeness) among operators in speed dating task

Fig. 5 shows a two-dimensional distribution of extroversion and politeness for each operator in the speed dating task. Our character expression model showed the different characters among the operators. The operators A and B showed high scores whereas the operators C and D showed the neutral scores. This result suggests that the operators could have their different characters in this dialogue task.

## 5 Subjective evaluation

Finally, we conducted a subjective evaluation to confirm if third-party persons can perceive the calculated characters from the spoken dialogue behaviors in the corpus. We extracted 15 samples from the analysis (5 samples from each task). Note that these samples were balanced in terms of the variation of the calculated characters. The samples were taken from one operator (the operator B in Fig. 5) to avoid the effect of individual differences, but this difference should be verified in future work. We asked 13 subjects (3 females and 10 males, from 23 to 60 years old) to listen to each dialogue sample and then evaluate if they could agree with the calculated character. The calculated character was shown as a sentence such as “The robot was extrovert and little casual.” when the character scores were 7, 4, and 3

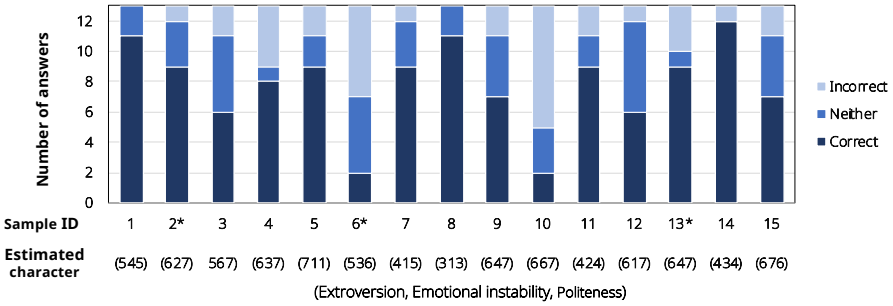


**Fig. 6** Ratio of correct answers per sample in subjective experiment (* Character scores were flipped.)

on extroversion, emotional instability, and politeness. Note that the adjective *little* was added to the sentence when the score was 3 or 5, and nothing was added for the scores of 1, 2, 6, and 7. If the score was 4 which was neutral, the corresponding character trait was not mentioned in the sentence. The subjects were asked to choose one from *Agree*, *Disagree*, and *Neither*. We regarded *Agree* as correct answers, but to avoid biased answering, character scores of three samples were flipped. In this case, the correct answer was *Disagree*.

Fig. 6 reports the ratio of correct answers. Among all samples except two samples 6 and 10, the majority of answers was the correct one. Even in the flipped character scores (sample 2 and 13), many subjects correctly answered as *Disagree*. When we regarded *Neither* as incorrect answers, the ratio of correct answers was 0.600 which was significantly higher than the chance level of 0.500 ($p = 0.004$ in the binomial test). When we did not use the *Neither* answers, the ratio became 0.770. This result demonstrates that characters represented by our model can be understandable by humans with high accuracy.

## 6 Conclusion

We have proposed the character expression model affecting certain spoken dialogue behaviors: utterance amount, backchannel frequency, filler frequency, and switching pause length. We validated this model with the human-robot dialogue where several dialogue tasks are designed. We confirmed that the model represents reasonable characters for each dialogue task. Furthermore, it was also found that the model represents the individual difference of the robot operators in the same dialogue task. Finally, we conducted the subjective evaluation and showed that the subjects perceived the calculated characters from the dialogue behaviors. We are now implementing the character expression model that works in the spoken dialogue system of the android robot ERICA. We will conduct a live dialogue experiment using ER-

ICA that expresses the proper character according to dialogue scenarios as a future work.

## Acknowledgement

This work was supported by JST ERATO Ishiguro Symbiotic Human-Robot Interaction program (Grant number JPMJER1401) and Grant-in-Aid for Scientific Research on Innovative Areas "Communicative intelligent systems towards a human-machine symbiotic society" (Grant number JP19H05691).

## References

1. Brave, S., Nass, C., Hutchinson, K.: Computers that care: investigating the effects of orientation of emotion exhibited by an embodied computer agent. International Journal of Human-Computer Studies **62**(2), 161–178 (2005)
2. Bunt, H., Alexandersson, J., Carletta, J., Choe, J.W., Fang, A.C., Hasida, K., Lee, K., Petukhova, V., Popescu-Belis, A., Romary, L., Soria, C., Traum, D.: Towards an iso standard for dialogue act annotation. In: LREC, pp. 2548–2555 (2010)
3. Costa, P.T., McCrae, R.R.: Normal personality assessment in clinical practice: The NEO personality inventory. Psychological Assessment **4**(1), 5–13 (1992)
4. D. Sevin, E., Hyniewska, S.J., Pelachaud, C.: Influence of personality traits on backchannel selection. In: IVA, pp. 187–193 (2010)
5. Den, Y., Yoshida, N., Takanashi, K., Koiso, H.: Annotation of Japanese response tokens and preliminary analysis on their distribution in three-party conversations. In: Oriental COCOSDA, pp. 168–173 (2011)
6. DeVault, D., Artstein, R., Benn, G., Dey, T., Fast, E., Gainer, A., Georgila, K., Gratch, J., Hartholt, A., Lhommet, M., Lucas, G., Marsella, S., Morbini, F., Nazarian, A., Scherer, S., Stratou, G., Suri, A., Traum, D., Wood, R., Morency, L.P.: Simsensei kiosk: A virtual human interviewer for healthcare decision support. In: AAMAS, pp. 1061–1068 (2014)
7. Digman, J.M.: Personality structure: Emergence of the five-factor model. Annual review of psychology **41**(1), 417–440 (1990)
8. Eysenck, H.: Dimensions of personality. Oxford (1947)
9. Fong, T., Nourbakhsh, I., Dautenhahn, K.: A survey of socially interactive robots. Robotics and Autonomous Systems **42**, 143–166 (2003)
10. Inoue, K., Milhorat, P., Lala, D., Zhao, T., Kawahara, T.: Talking with ERICA, an autonomous android. In: SIGDIAL, pp. 212–215 (2016)
11. 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)
12. Kawahara, T.: Spoken dialogue system for a human-like conversational robot ERICA. In: IWSDS (2018)
13. Koiso, H., Horiuchi, Y., Tutiya, S., Ichikawa, A., Den, Y.: An analysis of turn-taking and backchannels based on prosodic and syntactic features in japanese map task dialogs. Language and speech **41**(3-4), 295–321 (1998)
14. Li, J., Galley, M., Brockett, C., Spithourakis, G., Gao, J., Dolan, B.: A persona-based neural conversation model. In: ACL, pp. 994–1003 (2016)
15. Mairesse, F., Walker, M.A.: Controlling user perceptions of linguistic style: Trainable generation of personality traits. Computational Linguistics **37**(3), 455–488 (2011)

16. McCrae, R.R., John, O.P.: An introduction to the five-factor model and its applications. Journal of personality **60**(2), 175–215 (1992)
17. McKeown, G., Valstar, M., Pantic, M.: The SEMAINE database: Annotated multimodal records of emotionally colored conversations between a person and a limited agent. IEEE transactions on affective computing **3**(1), 5–17 (2012)
18. Miyazaki, C., Hirano, T., Higashinaka, R., Makino, T., Matsuo, Y.: Automatic conversion of sentence-end expressions for utterance characterization of dialogue systems. In: PACLIC, pp. 307–314 (2015)
19. Mizukami, M., Neubig, G., Sakti, S., Toda, T., Nakamura, S.: Linguistic individuality transformation for spoken language. In: IWSDS (2015)
20. Nagaoka, C., Komori, M., Nakamura, T., Draguna, M.R.: Effects of receptive listening on the congruence of speakers' response latencies in dialogues. Psychological Reports **97**(1), 265–274 (2005)
21. Nass, C., Moon, Y., B.J.Fogg, Reeves, B., Dryer, D.: Can computer personalities be human personalities? Human-Computer studies **43**, 223–239 (1955)
22. Ogawa, Y., Miyazawa, K., Kikuchi, H.: Assigning a personality to a spoken dialogue agent through self-disclosure of behavior. In: HAI, pp. 331–337 (2014)
23. Salem, M., Ziadee, M., Sakr, M.: Effects of politeness and interaction context on perception and experience of HRI. In: ICSR, pp. 531–541 (2013)
24. 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)
25. Shiwa, T., Kanda, T., Imai, M., Ishiguro, H., Hagita, N.: How quickly should communication robots respond? International Journal of Social Robotics **1**, 153–160 (2009)
26. Sugiyama, H., Meguro, T., Higashinaka, R., Minami, Y.: Large-scale collection and analysis of personal question-answer pairs for conversational agents. In: IVA, pp. 420–433 (2014)
27. Traum, D., Aggarwal, P., Artstein, R., Foutz, S., amd Athanasios Katsamanis, J.G., Leuski, A., Noren, D., Swartout, W.: Ada and Grace: Direct interaction with museum visitors. In: IVA, pp. 245–251 (2012)
28. van Vugt, H.C., Konijn, E.A., Hoorn, J.F., Keur, I., Eliëns, A.: Realism is not all! User engagement with task-related interface characters. Interacting with Computers **19**(2), 267–280 (2007)
29. Wada, S.: Construction of the Big Five scales of personality trait terms and concurrent validity with NPI. Japanese Journal of Psychology **67**(1), 61–67 (1996). In Japanese
30. Watanabe, M.: Features and Roles of Filled Pauses in Speech Communication: A corpus-based study of spontaneous speech. Hitsuji Syobo Publishing (2009)
31. Yamamoto, K., Inoue, K., Nakamura, S., Takanashi, K., Kawahara, T.: Dialogue behavior control model for expressing a character of humanoid robots. In: APSIPA ASC, pp. 1732–1737 (2018)
32. Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., Weston, J.: Personalizing dialogue agents: I have a dog, do you have pets too? In: ACL, pp. 2204–2213 (2018)