Engagement-based adaptive behaviors for laboratory guide in human-robot dialogue

Koji Inoue, Divesh Lala, Kenta Yamamoto, Katsuya Takanashi, and Tatsuya Kawahara

Abstract We address an application of engagement recognition in human-robot dialogue. Engagement is defined as how much a user is interested in the current dialogue, and keeping users engaged is important for spoken dialogue systems. In this study, we apply a real-time engagement recognition model to laboratory guide by autonomous android ERICA which plays the role of the guide. According to an engagement score of a user, ERICA generates adaptive behaviors consisting of feedback utterances and additional explanations. A subject experiment showed that the adaptive behaviors increased both the engagement score and related subjective scores such as interest and empathy.

1 Introduction

Spoken dialogue systems are expected to realize social interaction with real users in more varied scenarios. Conventional systems were applied to scenarios such as museum guide [21] and mental diagnosis [3]. We have developed a spoken dialogue system for the autonomous android ERICA [11, 12]. Giving specific social roles to ERICA, we aim to realize natural dialogue between ERICA and users. We have considered several social roles so far by taking into account two factors of ERICA in dialogue: speaking and listening as depicted in Fig. 1. Focusing on the role of listening, we implemented an attentive listening system [14] and a job interview dialogue system [7] for ERICA. In this study, we focus on the role of speaking and implement a spoken dialogue system for laboratory guide where ERICA explains about a laboratory to users. In this scenario, the majority of dialogue is explanations from guides. However, it is needed to not only just explain but also recognize the listening attitude of visitors. A human-like laboratory guide is expected to dynami-

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

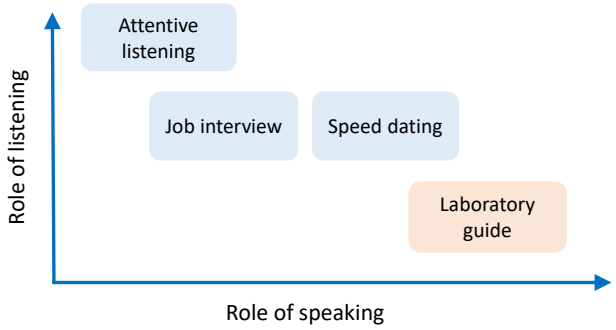

Fig. 1 Social roles expected to ERICA

cally change the explanation and its behaviors according to states of users so that it increases the quality of user experience in the dialogue.

We address engagement which represents the process by which dialogue participants establish, maintain, and end their perceived connection to one another [19]. This concept is practically defined as how much a user is interested in the current dialogue [24]. Therefore, it is important for spoken dialogue systems to make users engaged in dialogue. Engagement recognition has been widely studied using mainly non-verbal behaviors [1, 17, 15, 18, 23, 26, 2]. We also studied engagement recognition by utilizing listener behaviors such as backchannels, laughing, head nodding, and eye gaze [10, 9]. Furthermore, we implemented real-time engagement recognition by detecting the above listener behaviors automatically [8]. On the other hand, fewer studies have been made on how to manage system behaviors after the system recognizes user engagement [23, 25, 26, 20].

In this study, we utilize the real-time engagement recognition model in the laboratory guide by ERICA. According to the engagement level of a user, ERICA generates adaptive behaviors to keep or increase the engagement level itself. Furthermore, we investigate the subjective evaluation of engagement together with rapport that would be affected by the engagement-based adaptive behaviors. This study aims to confirm the effectiveness of engagement recognition in a social scenario of human-robot dialogue.

This paper is organized as follows. The real-time engagement recognition model is introduced in Section 2. The adaptive behaviors in the context of a laboratory guide are explained in Section 3. A user experiment is conducted in Section 4. Section 5 concludes this paper with future direction.

2 Engagement recognition based on listener behaviors

We addressed engagement recognition based on listener behaviors such as backchannels, laughing, head nodding, and eye gaze. The listener behaviors are non-linguistic

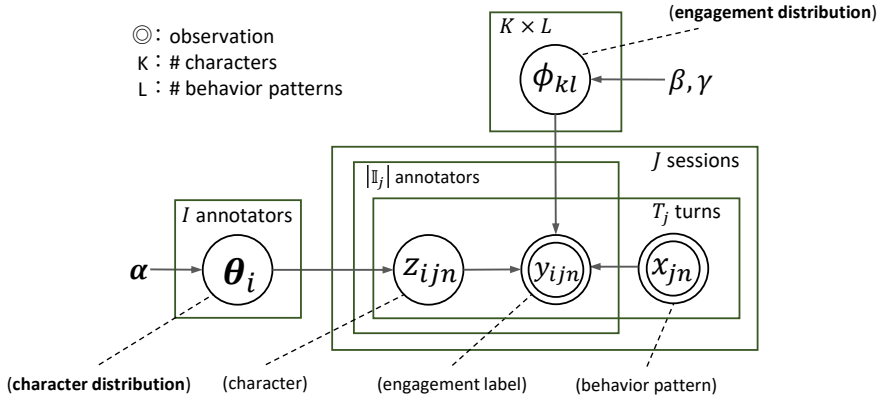

Fig. 2 Graphical model of latent character model

cues so that engagement recognition can be independent of dialogue content, which makes it robust and applicable to any scenario. The engagement recognition was done during each system's dialogue turn when a user was being a listener. Each listener behavior was defined as an occurrence of the behavior, and the input feature was represented as the combination of the binary values, termed a *behavior pattern*. Note that the eye-gaze behavior was defined as an event if the user looked at the robot's face longer than a certain time (10 seconds). Finally, the engagement recognition model outputs an engagement score for each system's turn so we would be able to utilize the score to decide a system action for the next system's turn.

In previous works, we proposed a latent character model for engagement recognition [10, 8]. Since the perception of engagement is subjective, oracle labels of engagement depend on perceivers (annotators). Our approach is based on a hierarchical Bayes model and introduces latent variables, called character, to represent the difference among annotators. Fig 2 illustrates the graphical model of the proposed model that contains two kinds of parameters to be estimated in the training phase: a character distribution of each annotator and an engagement distribution. In the test phase, we calculate the probability of the engaged label of a target annotator by using both the engagement distribution and the target annotator's character distribution. The advantage is that our model can simulate each annotator's perception by using the corresponding character distribution. Therefore, our model can recognize each annotator's label more precisely. We conducted a manual annotation where each annotator gave a binary label (engaged or not) in each dialogue turn. Experimental results showed that our model achieved an accuracy of 71.1% which was higher than those of other methods that did not take into account the character variables. Furthermore, we examined the relationship between the estimated character distribution and a personality trait (Big Five) of each annotator. We calculated regression coefficients from Big Five scores to parameters of the character distribution. Using this regression result, if we specify a personality trait score expected to a conversational agent or robot, corresponding character distribution is determined.

Fig. 3 Real-time engagement recognition

For example, if we specify an *extrovert* personality for ERICA, we can simulate the perception of engagement of extroverted people.

In order to use the engagement recognition model in live spoken dialogue systems, it is needed to detect listener behaviors in real time. We examined how to detect the listener behaviors with deep learning approaches [8]. Backchannels and laughing were detected from an audio signal using bi-directional long short-term memory with connectionist temporal classification (BLSTM-CTC). Head nodding was detected from a visual signal of the Kinect v2 sensor with a simpler LSTM model. Eye gaze behavior was detected also by the Kinect v2 sensor with a heuristic rule. Results of these automatic detection were used as the input to the engagement recognition model. We confirmed that the accuracy of engagement recognition was not so degraded (70.0%) even with the automatic detection of the listener behaviors. Finally, we implemented the real-time engagement recognition in the system of ERICA as shown in Fig. 3. In this study, we utilize the result of engagement recognition for generation of adaptive behaviors of ERICA.

3 Engagement-based adaptive behaviors for laboratory guide

We implement a spoken dialogue system of ERICA where ERICA plays the role of the laboratory guide, utilizing the real-time engagement recognition model[1]. The dialogue contents of the laboratory guide are hand-crafted and consist of several research topics. A structure of the dialogue on each topic is illustrated in Fig. 4.

[1] Demo video (in Japanese language) is available at `https://youtu.be/53I3lhJ6aUw`

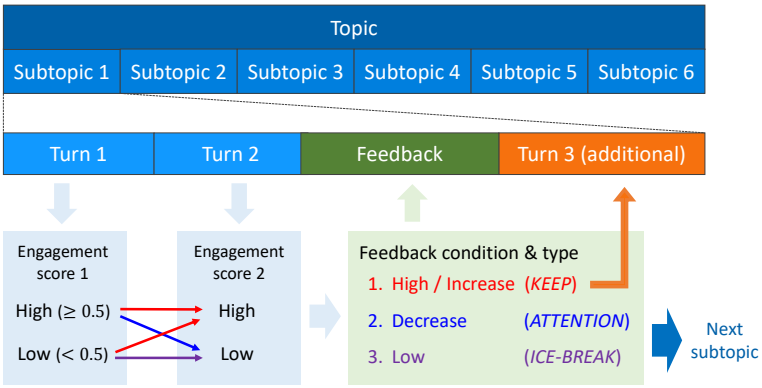

Fig. 4 Dialogue structure of laboratory guide and engagement-based adaptive behaviors

Each topic consists of 6 subtopics where each subtopic corresponds to each research theme. Each subtopic includes two of ERICA's turns. In each turn, an engagement score is measured by the real-time recognition model, and the result is regarded as a binary: high or low, with a threshold of 0.5 for the posterior probability of engaged label. After ERICA's two turns, ERICA generates feedback utterances according to the combination of the two engagement scores in the same subtopic. We define three kinds of feedbacks as follows.

- **KEEP feedbacks**:
  When both scores are high or the scores change from low to high (increase), ERICA says feedbacks like "*You seems to be interested in my talk. I am delighted.*" in order to keep the current engagement.
- **ATTENTION feedbacks**:
  When the engagement score changes from high to low (decrease), ERICA says a different type of feedbacks such as "*Are you tired? I will explain it in easier words.*" to gain attention from the user.
- **ICE-BREAK feedbacks**:
  When both scores are low, ERICA says another type of feedbacks such as "*Are you nervous? Please relax.*" to ease the tension of the user like ice-breaking. ERICA also says like "*It would be easy to explain if you show a reaction.*" to implicitly tell a user that ERICA is monitoring their listener behaviors.

It is expected that these feedbacks make the user more engaged in the laboratory guide. In the case where a KEEP feedback is triggered, ERICA introduces an additional content of the current topic. This additional content would be beneficial for users who are potentially interested in the content of the laboratory guide. For users who are not engaged, this additional content would be difficult to understand and makes these users more disengaged. Therefore, engagement recognition needs to be accurate for precise and effective information providing.

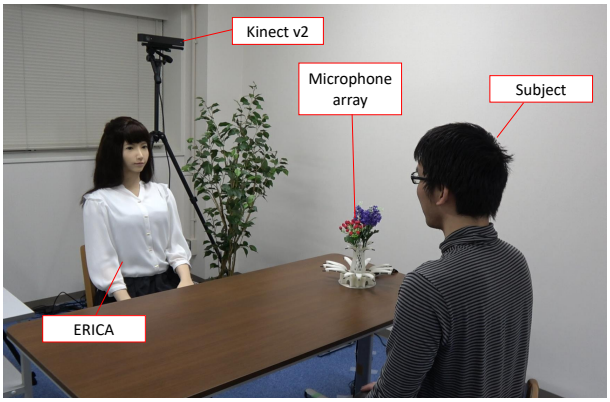

Fig. 5 Snapshot of dialogue experiment

4 User experiment

A human subject experiment was conducted in order to confirm the effectiveness of engagement recognition and also the engagement-based adaptive behaviors. Fig. 5 is the snapshot of the dialogue experiment. For speech processing, we used a 16-channel microphone array to localize sound sources and to enhance the user's speech [11]. We also used end-to-end acoustic-to-word automatic speech recognition [22]. In this experiment, a subject and ERICA sat on chairs to face each other. To elicit listener behaviors of the subjects, ERICA generated backchannels and head nodding automatically during the subjects' turns [13]. The subjects were 11 persons (6 males and 5 females) who were recruited in our university. They are all native Japanese speakers. The experiment procedure is as follows. At first, each subject had a practice dialogue to get used to talking with ERICA. They practiced a simple interaction consisting of several turns. After this, ERICA explained about two research topics: automatic speech recognition and spoken dialogue systems. During the explanation, ERICA sometime gave questions toward the subject. The order of the dialogue topics was randomized among the subjects. Two experiment conditions were prepared: *engagement* and *control*. In the *engagement* condition, ERICA measured engagement scores and generated adaptive behaviors mentioned above. In the *control* condition, ERICA did not generate any adaptive behaviors, which meant only the turn 1 and 2 were explained. The dialogue of the first topic was conducted with the *control* condition, and the second topic was done with the *engagement* condition. The order of these conditions was not randomized because it was thought that the engagement-based adaptive behaviors in the *engagement* condition would affect the subjects' behaviors and impressions in the subsequent topic. Among the two conditions, measured engagement scores were compared.

Additionally, we also investigated subjective evaluations of users after they talked about each topic. To measure subjective evaluations, we used more specific concepts related to engagement by referring to previous works [4]. We selected related

Table 1 Evaluation scale for other concepts related to engagement (* represents invert scale.)

concept	question
interest	(1) I felt the dialogue was boring. *
	(2) I wanted to listen to other topics more.
	(3) I was fascinated by the dialogue content.
	(4) I was not concerned with the dialogue content. *
continuing	(5) I wanted to quit the dialogue during that. *
	(6) I think the dialogue should finish earlier. *
	(7) I wanted to continue the dialogue more.
willingness	(8) I wanted to make the dialogue fulfilling.
	(9) I could not feel like talking. *
	(10) I participated in the dialogue by concentrating on that.
	(11) I felt that what I needed to do was just being there. *
	(12) I actively answered the questions from the robot.
	(13) I actively responded to the robot talk.
rapport	(14) I liked the robot.
	(15) I felt the robot was friendly.
	(16) I was relieved when I was having the dialogue with the robot.
	(17) I could trust the dialogue content the robot talked.
empathy	(18) I could understand what emotion the robot was having.
	(19) I could agree with the idea the robot had.
	(20) I could advance the dialogue by considering the viewpoint from the robot.
	(21) I could understand the reason why the robot had that kind of emotion.

concepts as *interest* [24, 25], *continuing* [16], *willingness* [23], *rapport* [5], and *empathy* [6]. We designed question items for each concept as listed in in Table 1. Two related researchers independently validated each question item by considering both the relevance to the subordinate concept and also the correctness of the question sentence. We used the 7 point scale to evaluate each item because we observed that evaluation scores tended to high and dense in a preliminary experiment. Finally, the evaluated scores were averaged for each subordinate concept. We hypothesized that the scores of the subordinate concepts would be improved in the *engagement* condition.

Average engagement scores are reported in Fig.6. Since each topic consisted of 6 subtopics and each subtopic included two turns, the total number of turns was 12. Note that scores of the additional turns in the *engagement* condition are not included in this result. A t-test was conducted between the two conditions on all engagement scores except those of the first and second turns which were before the first-time feedback utterance. As a result, it turned out that the *engagement* condition significantly increased the engagement scores ($p = 8.06 \times 10^{-4}$). The difference between the two conditions was observed in the latter part of the dialogue. This result suggests that the adaptive behaviors in the *engagement* condition made the subjects more engaged in the dialogue. Accordingly, engagement recognition is the important function in social human-robot dialogue such as the laboratory guide.

Average subjective scores on the related concepts are reported in Fig. 7. For each concept, a paired t-test was conducted between the two conditions. The results show that the scores of *interest* and *empathy* significantly increased in the *engagement*

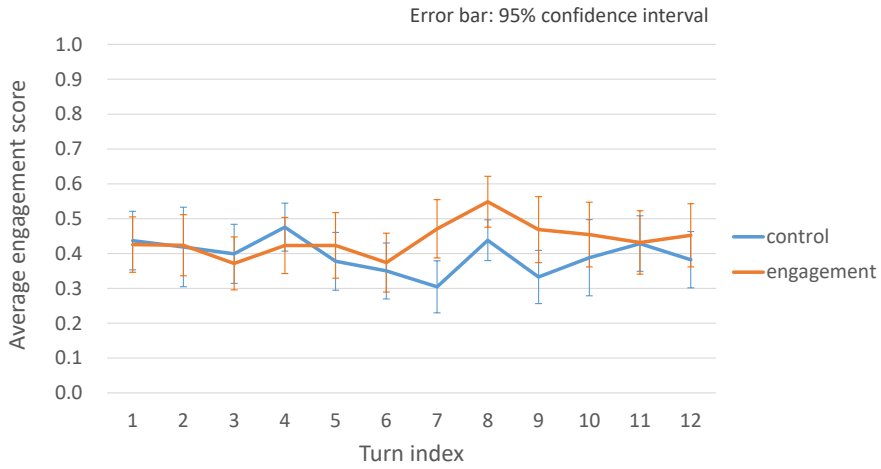

Fig. 6 Engagement scores during dialogue

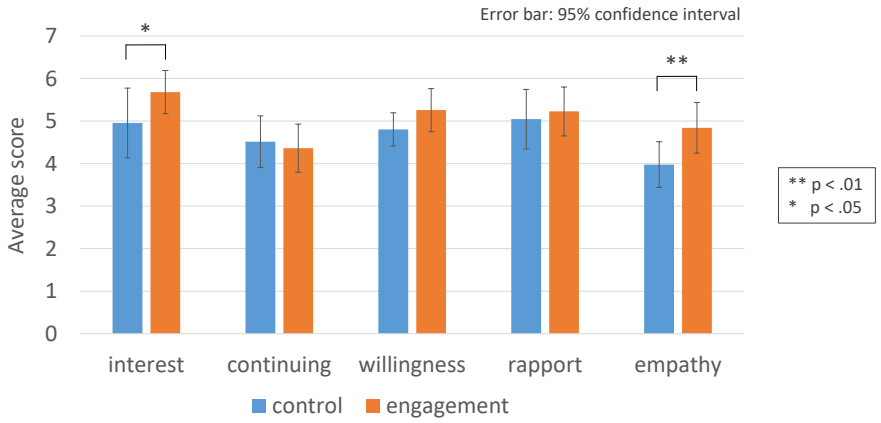

Fig. 7 Subjective evaluation on other concepts related to engagement

condition ($p < .05$ for *interest* and $p < .01$ for *empathy*). One possible reason is that the additional explanation made the subjects more interested in the research topic. Besides, the feedback responses were perceived as emotional expressions of ERICA so that they perceived higher empathy scores.

5 Conclusions

We have addressed applications of engagement recognition in order to realize social dialogue with autonomous android ERICA. In this study, the real-time engagement recognition model was applied to the dialogue of laboratory guide where ERICA

plays the role of the guide. In the laboratory guide, since ERICA talks most of the time, ERICA needs to track user engagement based on listener behaviors while ERICA is speaking. ERICA was implemented to adaptively generate feedback utterances and additional explanations by measuring user engagement. The experimental results showed that the adaptive behaviors increased both the measured engagement scores and subjective evaluations of interest and empathy. Although the adaptive behaviors of ERICA were handcrafted in the current study, we will investigate how to obtain the adaptive behaviors from dialogue data in the manner of machine learning in future work.

Acknowledgments

This work was supported by JST ERATO Ishiguro Symbiotic Human-Robot Interaction program (Grant Number JPMJER1401), Japan.

References

1. Castellano, G., Pereira, A., Leite, I., Paiva, A., McOwan, P.W.: Detecting user engagement with a robot companion using task and social interaction-based features. In: ICMI, pp. 119–126 (2009)
2. Chiba, Y., Nose, T., Ito, A.: Analysis of efficient multimodal features for estimating user's willingness to talk: Comparison of human-machine and human-human dialog. In: APSIPA ASC (2017)
3. 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., Xu, Y., Rizzo, A., Morency, L.P.: SimSensei kiosk: A virtual human interviewer for healthcare decision support. In: AAMAS, pp. 1061–1068 (2014)
4. Glas, N., Pelachaud, C.: Definitions of engagement in human-agent interaction. In: International Workshop on Engagement in Human Computer Interaction, pp. 944–949 (2015)
5. Gratch, J., Wang, N., Gerten, J., Fast, E., Duffy, R.: Creating rapport with virtual agents. In: IVA, pp. 125–138 (2007)
6. Hall, L., Woods, S., Aylett, R., Newall, L., Paiva, A.: Achieving empathic engagement through affective interaction with synthetic characters. In: ICACII, pp. 731–738 (2005)
7. Inoue, K., Hara, K., Lala, D., Nakamura, S., Takanashi, K., Kawahara, T.: A job interview dialogue system with autonomous android ERICA. In: IWSDS (2019). Submitted
8. Inoue, K., Lala, D., Takanashi, K., Kawahara, T.: Engagement recognition by a latent character model based on multimodal listener behaviors in spoken dialogue. APSIPA Trans. Signal & Information Processing **7**(e9), 1–16 (2018)
9. Inoue, K., Lala, D., Takanashi, K., Kawahara, T.: Engagement recognition in spoken dialogue via neural network by aggregating different annotators' models. In: Interspeech, pp. 616–620 (2018)
10. Inoue, K., Lala, D., Takanashi, K., Kawahara, T.: Latent character model for engagement recognition based on multimodal behaviors. In: IWSDS (2018)
11. Inoue, K., Milhorat, P., Lala, D., Zhao, T., Kawahara, T.: Talking with ERICA, an autonomous android. In: SIGDIAL, pp. 212–215 (2016)

12. Kawahara, T.: Spoken dialogue system for a human-like conversational robot ERICA. In: IWSDS (2018)
13. Kawahara, T., Yamaguchi, T., Inoue, K., Takanashi, K., Ward, N.G.: Prediction and generation of backchannel form for attentive listening systems. In: Interspeech, pp. 2890–2894 (2016)
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, pp. 127–136 (2017)
15. Nakano, Y.I., Ishii, R.: Estimating user's engagement from eye-gaze behaviors in human-agent conversations. In: IUI, pp. 139–148 (2010)
16. Poggi, I.: Mind, hands, face and body: A goal and belief view of multimodal communication. Weidler (2007)
17. Rich, C., Ponsler, B., Holroyd, A., Sidner, C.L.: Recognizing engagement in human-robot interaction. In: HRI, pp. 375–382 (2010)
18. Sanghvi, J., Castellano, G., Leite, I., Pereira, A., McOwan, P.W., Paiva, A.: Automatic analysis of affective postures and body motion to detect engagement with a game companion. In: HRI, pp. 305–311 (2011)
19. Sidner, C.L., Lee, C., Kidd, C.D., Lesh, N., Rich, C.: Explorations in engagement for humans and robots. Artificial Intelligence **166**(1-2), 140–164 (2005)
20. Sun, M., Zhao, Z., Ma, X.: Sensing and handling engagement dynamics in human-robot interaction involving peripheral computing devices. In: CHI, pp. 556–567 (2017)
21. Swartout, W., Traum, D., Artstein, R., Noren, D., Debevec, P., Bronnenkant, K., Williams, J., Leuski, A., Narayanan, S., Piepol, D., Lane, C., Morie, J., Aggarwal, P., Liewer, M., Yuan-Jen, C., Gerten, J., Chu, S., White, K.: Ada and grace: Toward realistic and engaging virtual museum guides. In: IVA, pp. 286–300 (2010)
22. Ueno, S., Moriya, T., Mimura, M., Sakai, S., Shinohara, Y., Yamaguchi, Y., Aono, Y., Kawahara, T.: Encoder transfer for attention-based acoustic-to-word speech recognition. In: Interspeech, pp. 2424–2428 (2018)
23. Xu, Q., Li, L., Wang, G.: Designing engagement-aware agents for multiparty conversations. In: CHI, pp. 2233–2242 (2013)
24. Yu, C., Aoki, P.M., Woodruff, A.: Detecting user engagement in everyday conversations. In: ICSLP, pp. 1329–1332 (2004)
25. Yu, Z., Nicolich-Henkin, L., Black, A.W., Rudnicky, A.I.: A Wizard-of-Oz study on a non-task-oriented dialog systems that reacts to user engagement. In: SIGDIAL, pp. 55–63 (2016)
26. Yu, Z., Ramanarayanan, V., Lange, P., Suendermann-Oeft, D.: An open-source dialog system with real-time engagement tracking for job interview training applications. In: IWSDS (2017)