Virtual Agent Design for Social Skills Training Considering Autistic Traits

Hiroki Tanaka and Satoshi Nakamura

Abstract— **Social skills training by human coaches is a well-established method to obtain appropriate social interaction skills and strengthen social self-efficacy. Our previous works automated social skills training by developing a virtual agent that teaches social skills through interaction. This study attempts to investigate the effect of virtual agent design on automated social skills training. We prepared images and videos of a virtual agent, and a total of 912 crowdsourced workers rated the virtual agents by answering questions. We investigated the acceptability, likeability, and other impressions of the virtual agents and their relationship to the individuals' characteristics to design personalized virtual agents. As a result, a female anime-type virtual agent was rated as the most likable. We also confirmed that participants' gender, age, and autistic traits are related to the ratings. We believe our findings are important in designing a personalized virtual trainer.**

Clinical relevance— **This study examines the effect of virtual agent design on social skills training. Our findings are important in designing a personalized virtual trainer.**

I. INTRODUCTION

Social Skills Training (SST) is a method widely applied to people seeking to improve their social skills. This training is used in medical hospitals, employment support facilities, workplaces, schools, etc. [1]. SST is generally conducted by a human trainer to promote appropriate social interaction skills and strengthen social self-efficacy. We have been conducting studies to automate SST using virtual agents and have developed an automatic SST that resembles human SST. Our system includes video modeling of human behavior, real-time behavior recognition, and feedback. We previously confirmed the training effect in children and adults with autism spectrum disorder and in the general population [2], [3], [4]. The automated SST plays two roles: as a trainer and a listener. We confirmed that talking to a virtual agent is more comfortable and less tense than talking to a human [5]. Automatic SST targets various populations, from pediatric to adult males and females, and those with autism spectrum disorders and schizophrenia. However, for virtual agent design, what kind of virtual agents are more favored or more accepted has not been investigated. For automatic SST to be adapted and accepted by the individual, a detailed investigation is necessary. In this study, we focus on comparing virtual agent designs rather than humans and robots because they are easier to create.

This work was funded by JST CREST Grant Number JPMJCR19M5, and JSPS KAKENHI Grant Numbers JP17H06101 and JP18K11437.

Hiroki Tanaka and Satoshi Nakamura are with the Nara Institute of Science and Technology, 8916-5 Takayama-cho, Ikoma, Nara, Japan. hiroki-tan@is.naist.jp

Previous studies have compared various virtual agent designs from the viewpoint of appearance and behavior [6], [7], [8], realism [9], severity of dialogue scenarios, appropriateness when comparing proportions of the body and eye parts [10], as a voice for the elderly [11], gender, and racial impact on users' self-efficacy [12]. This study applies these findings and rating measures to investigate the design of our virtual agents for the creation of more favorable and acceptable automated SSTs.

In this study, we newly prepared various virtual agent designs for training social skills and evaluated their likeability, acceptability, realism, familiarity, etc. We also investigated differences in preference for virtual agent designs by user gender, age, and autistic traits in order to create personalized virtual agents. The purpose of this research is summarized below.

1) To investigate virtual agent design in terms of automated SST.
2) To investigate the relationship between likeability and an individual's characteristics (gender, age, and autistic traits).

II. METHODS AND MATERIALS

A. Design of Virtual Agents

We first prepared an illustration of a virtual agent as shown in Fig. 1. The virtual agent was designed by a design company specializing in Japanese animation. All characters face front with no emotional expression. Characters (a) and (b) (female) and (c) and (d) (male) were designed to unify their age and gender and to change only their realism. An inanimate object was created for (e) assuming usage by children. (f) was also created as a non-human organism (a dog) for usage by children. For (g), we chose a realistic 3D model virtual agent (`www.renderhub.com/3d-models`) similar in appearance to (a) and (b) and took a snapshot from the front. The virtual agent in (h) is the default agent provided by the Greta platform [13], which is an embodied conversation agent that can be created with the Autodesk character generator (`https://charactergenerator.autodesk.com/`). (h) is intended to be used mainly in French and English-speaking cultures. The virtual agent in (f) was designed for Japanese females. In the current automated SST, (f) was selected as the virtual agent [4].

The sentence "Hello, let's practice communication together" was embedded in the image as both a male and female voice. The utterance is 5 seconds in length and spoken by Google Text-to-Speech (`https://cloud.google.`

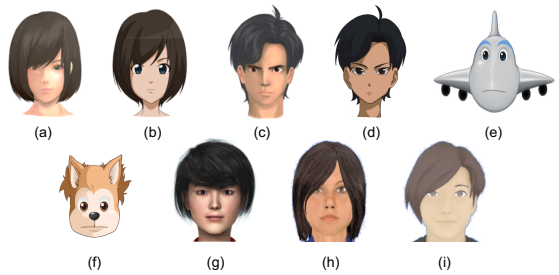

Fig. 1. Images of the nine created virtual agents.

Fig. 2. Snapshots of the videos of two of the created virtual agents.

com/text-to-speech?hl=ja). The virtual agents (e) and (f) were created with a higher pitch than normal female speech synthesis, assuming children's voices.

Since the 3D model can be created on virtual agents (h) and (i), videos were also recorded while operating on Greta (see Fig. 2). In the movement, gestures (raising hands, putting hands on one's chest) were added in synchronization with the speech content. The same behavior was generated in (h) and (i), and Japanese speech synthesis and its lip-sync were created with the same utterance content (8 seconds of utterance length) as above. For the speech synthesis, we used Yuki's voice from CereProc (https://www.cereproc.com/).

B. Participants

For data collection, we recruited participants from a crowdsourcing service (https://crowdworks.jp/). We followed the principles outlined in the Helsinki Declaration of 1975, as revised in 2000. The job description for recruiting participants sought people 18 years of age or older with Japanese nationality. In order to divide the effort of each participant, data was collected in three parts with different participants. In dataset 1), n=305 (males: n=148, females: n=157); in dataset 2), n=305; and, in dataset 3), n=302. Dataset 1 is for investigating image acceptability, likeability, familiarity, likability of certain parts (eyes, face, hair, voice, perceived age), autistic traits, and alexithymia. Dataset 2 was collected to investigate the realism, trustworthiness, and eeriness of the images. Dataset 3 was collected to investigate the videos with (h) and (i). In this study, we performed grouped analysis using 45 years as the threshold for high and low age (high age: n=84, low age: n=21).

C. Autistic Traits and Alexithymia

In dataset 1, we measured the adult version of the Social Responsiveness Scale-2 (SRS) [14] to assess autistic traits and the Toronto Alexithymia Scale-20 (TAS) [15] to assess alexithymia. In both cases, we calculated the total score. We did not calculate its subscales in this study. As for the relationship between the two questionnaires, the Spearman's correlation coefficient was 0.67 ($p < 0.05$), which means a high correlation. The later analysis will use SRS as a measure of autistic traits. The mean and standard deviation of SRS was 72.0 (SD: 29.3). In this study, we used a cut-off value of 81 points [16] as the threshold for high and low SRS (high SRS: n=113, low SRS: n=192).

D. Measures

Question items and scales were made with reference to the studies by Esposito et al. and Ring et al. [6], [7], [11], [10]. Question items consisted of acceptability as a trainer, acceptability as a listener, realism, familiarity, trustworthiness, eeriness, and the likeability of the face, eyes, perceived ages, voice, and overall. Each question was answered via a Google form. In dataset 1, each question item was answered after answering SRS and TAS. In Dataset 3, in addition to the above, we added the likeability of the clothes the agent wore because the video includes the entire upper body of the virtual agent. Participants first took a look at the list of all images (Fig. 1) to get an impression of all the virtual agents, then watched the individual virtual agents and answered each question. The questions were evaluated with a 5-point Likert scale (1: I don't think so at all, 5: I think so very much).

Statistical software R was used for the analysis. Since the normality could not be confirmed in the ratings of the questions (Kolmogorov-Smirnov test, $p < 0.05$), the Kruskal-Wallis test was used for the difference in virtual agents. In the analysis for each group of gender, age, and SRS, we calculated the effect size r. We report the top three combinations of r from all combinations of virtual agents and question items. We performed the Wilcoxon signed-rank test between two factors (e.g., images and videos). In this study, the significance level α value of the statistical hypothesis test was set to 0.05.

III. RESULTS

First, we report on the differences in ratings between the virtual agents. The Kruskal-Wallis test confirms significant differences between the virtual agents in terms of likability and familiarity ($p < 0.05$). Regarding realism, the distribution is as expected by the original design (in the order of $(b) < (a) < (g)$). The most preferred virtual agent among the participants was (b), averaging 3.29 (SD: 1.0) (see Fig. 3). We can see that (b) is also highly evaluated in other question items. We also found that the male virtual agents (c) and (d) and the non-human virtual agents (e) and (f) have lower likeability than (b), and (h) has less likeability and less familiarity. Comparing the images and videos with respect to (h) and (i), making the video did not change the likeability significantly ($p > 0.05$). However, virtual agent (h) was found to have significantly increased familiarity by making the video, as shown in Fig. 4 ($p < 0.05$).

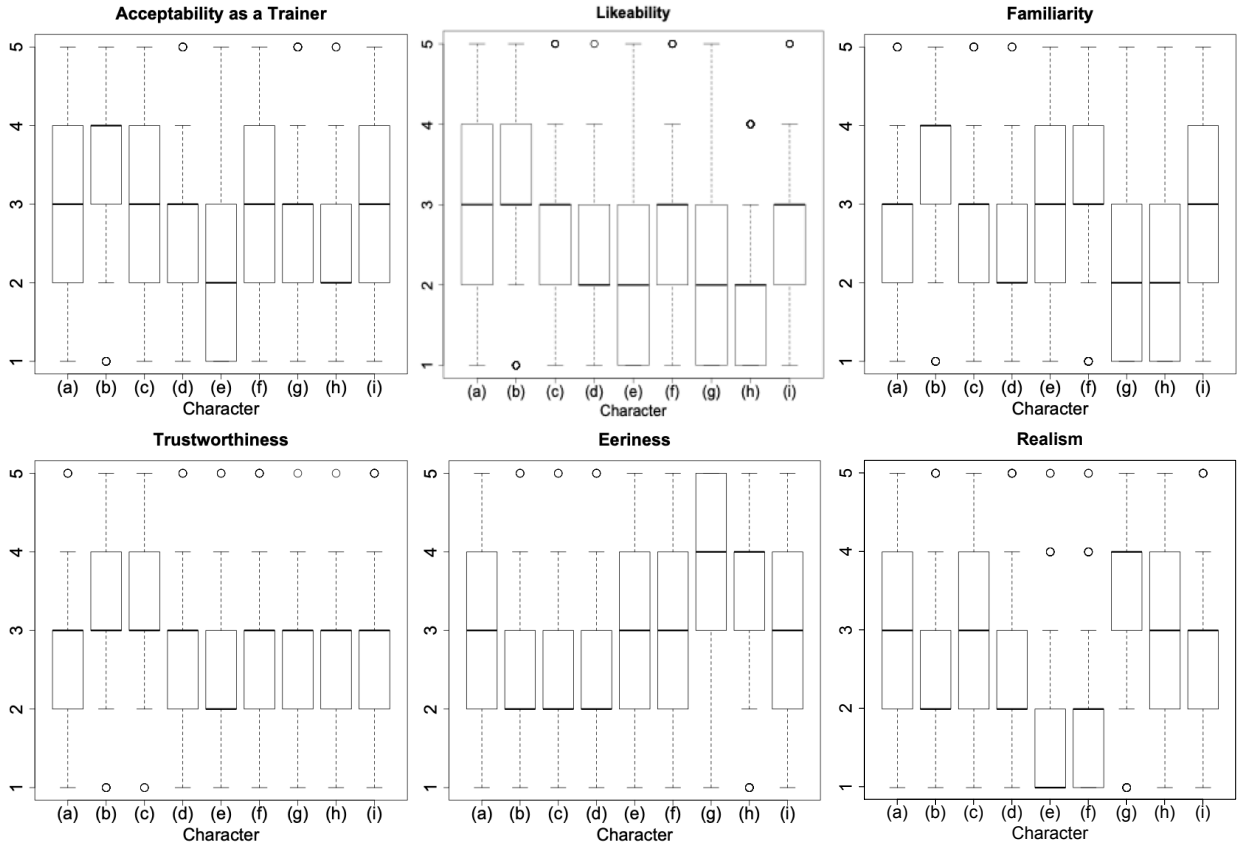

Fig. 3. Ratings of images.

TABLE I

QUESTIONS RELATED TO GENDER $(all, p < 0.05)$

Character	Question	r	Trend
(g)	Face	0.29	Male $>$ Female
(g)	Likeablity	0.25	Male $>$ Female
(g)	Trainer	0.25	Male $>$ Female

TABLE II

QUESTIONS RELATED TO AGE GROUPS $(all, p < 0.05)$

Character	Question	r	Trend
(i)	Eyes	0.21	High $>$ Low
(i)	Face	0.19	High $>$ Low
(a)	Listener	0.17	High $<$ Low

TABLE III

QUESTIONS RELATED TO SRS GROUPS $(all, p < 0.05)$

Character	Question	r	Trend
(g)	Eyes	0.19	High $>$ Low
(g)	Hair	0.18	High $>$ Low
(h)	Face	0.16	High $>$ Low

Table I lists the top three combinations of virtual agents and question items that have a high effect size r in terms of gender. Similarly, Table II lists the top three combinations with respect to age and Table III lists with respect to the level of the SRS. Regarding gender, males showed higher values than females in terms of likeability of the face and overall likeability, and acceptability as a trainer for character (g). For ages, the higher age group showed higher values than the lower age group in terms of likeability of eyes and face regarding (i). For SRS, the high SRS value group showed higher values than the low SRS value group in terms of likeability of eyes and hair for character (g). Fig. 5 shows the difference between the SRS levels for character (g) ($p < 0.05$).

IV. DISCUSSION

In this study, we investigated the acceptability, likeability, and various measures of virtual agents in designing automated SSTs. We were able to confirm the difference in the ratings of the virtual agents. First, it appears that the realism of the virtual agent design can be controlled by (a), (b), and (g). We found that virtual agent (b) was the most likable. (b) was originally designed as an anime-like teenage female character. As Japanese people are rather accustomed to watching anime-like videos, familiarity with such characters is high. On the other hand, virtual agents, such as inanimate objects (e) and animals (f), as well as (g) and (h), were less preferred. However, for children with autism spectrum disorders, virtual agents such as trains may be preferred [17]. We must consider the effect for younger

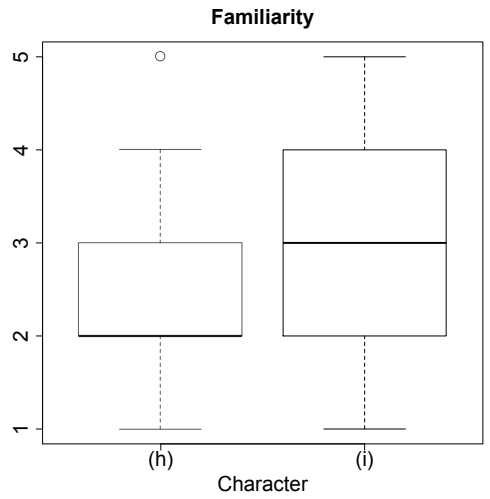

Fig. 4. Ratings of videos.

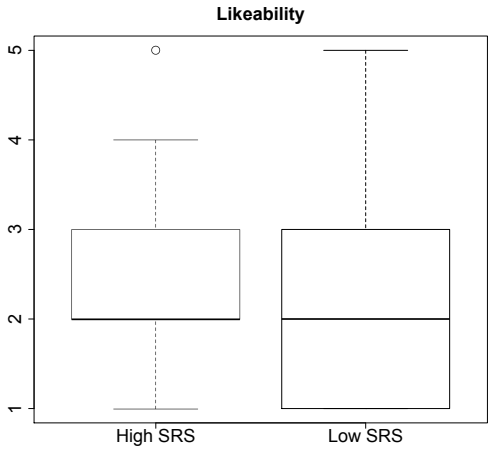

Fig. 5. Effects of SRS levels to rate likeability of character (g).

ages in the future. We showed that almost the same tendency can be seen in videos. However, in terms of familiarity, we confirmed that the rating for the (h) video increased compared to its image due to adding naturalistic movement.

We found that the female virtual agent of (g) was rated as more preferred by male participants. In addition, since we confirmed that virtual agent (b) was also significantly highly rated by males, it appears that the male participants rate female virtual agents as more preferable. Regarding age, we found that virtual agent (i) was preferred by older age groups. Since (i) was made relatively old (originally designed for those in their 40s), it seems that the older group rated agents close in age to themselves as trustworthy. We also confirmed that the group with a high autistic trait showed a high rating for virtual agents (g) and (h). This virtual environment could be explored in other domains such as cognitive behavioral therapy [18].

REFERENCES

[1] Alan S. Bellack, Kim T. Mueser, Susan Gingerich, and Julie Agresta: Social Skills Training for Schizophrenia, Second Edition: A Step-by-Step Guide. Guilford Publications (2013)

[2] Hiroki Tanaka, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura: Embodied Conversational Agents for Multimodal Automated Social Skills Training in People with Autism Spectrum Disorders, Plos One, Vol. 12, No. 8, pp. 1–15 (2017)

[3] Hiroki Tanaka, Hidemi Iwasaka, Hideki Negoro, and Satoshi Nakamura: Analysis of Conversational Listening Skills toward Agent-based Social Skills Training, Journal on Multimodal User Interfaces, Vol. 14, No. 1, pp. 73–82 (2020)

[4] Hiroki Tanaka, Hidemi Iwasaka, Yasuhiro Matsuda, Kosuke Okazaki, and Satoshi Nakamura: Analyzing Self-efficacy and Summary Feedback in Automated Social Skills Training, IEEE Open Journal of Engineering in Medicine and Biology (2021)

[5] Hiroki Tanaka, Sakti Sakriani, Graham Neubig, Tomoki Toda, Hideki Negoro, Hidemi Iwasaka, and Satoshi Nakamura: Teaching Social Communication Skills Through Human-Agent Interaction. ACM Transactions on Interactive Intelligent Systems, No. 18, pp. 1–26 (2016)

[6] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Antonietta M. Esposito, Alda Troncone, Maria Ines Torres, Stephan Schlogl, and Gennaro Cordasco: Seniors' Acceptance of Virtual Humanoid Agents, Italian Forum of Ambient Assisted Living, pp. 429–443 (2018)

[7] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Ilaria Pica, Maria Teresa Riviello, Alda Troncone, Gennaro Cordasco, and Antonietta M. Esposito: Elders Prefer Female Robots with a High Degree of Human Likeness, IEEE 23rd International Symposium on Consumer Electronics, pp. 243–246 (2019)

[8] Kazunori Terada, Liang Jing, and Seiji Yamada: Effects of Agent Appearance on Customer Buying Motivations on Online Shopping Sites, In Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems, pp. 929–934 (2015)

[9] Rachel McDonnell, Martin Breidt, and Heinrich H. Bulthoff: Render Me Real? Investigating the Effect of Render Style on the Perception of Animated Virtual Humans, ACM Transactions on Graphics Vol. 31, No. 4, Article 91, pp. 1–11 (2012)

[10] Lazlo Ring, Dina Utami, and Timothy Bickmore: The Right Agent for the Job? The Effects of Agent Visual Appearance on Task Domain, International Conference on Intelligent Virtual Agents, pp. 374–384 (2014)

[11] Anna Esposito, Terry Amorese, Marialucia Cuciniello, Maria Teresa Riviello, Antonietta M. Esposito, Alda Troncone, and Gennaro Cordasco: The Dependability of Voice on Elders' Acceptance of Humanoid Agents, Interspeech, pp. 31–35 (2019)

[12] Amy L. Baylor, and Kim Yanghee: Pedagogical Agent Design: The Impact of Agent Realism, Gender, Ethnicity, and Instructional Role, Intelligent Tutoring Systems, pp. 592–603 (2004)

[13] Isabella Poggi, Catherine Pelachaud, F. de Rosis, Valeria Carofiglio, and Berardina De Carolis: Greta. A Believable Embodied Conversational Agent, Multimodal Intelligent Information Presentation. Text, Speech and Language Technology, Vol. 27. Springer, Dordrecht, pp. 3–25 (2005)

[14] John N. Constantino: Social Responsiveness Scale - Second Edition (SRS-2), WPS (2012)

[15] R. Michael Bagby, James D.A. Parker, and Graeme J. Taylor: The Twenty-item Toronto Alexithymia Scale—I. Item Selection and Cross-validation of the Factor Structure, Journal of Psychosomatic Research, Vol. 38, No. 1, pp. 23–32 (1994)

[16] M. L. Bezemer, Els Blijd-Hoogewys, and M. Meek-Heekelaar: The Predictive Value of the AQ and the SRS-A in the Diagnosis of ASD in Adults in Clinical Practice, Journal of Autism and Developmental Disorders (2020)

[17] Ofer Golan, and Simon Baron-Cohen: Systemizing Empathy: Teaching Adults with Asperger Syndrome or High-functioning Autism to Recognize Complex Emotions using Interactive Multimedia, Development and Psychopathology, Vol. 18, No. 2, pp. 591–617 (2006)

[18] Kazuhiro Shidara, Hiroki Tanaka, Hiroyoshi Adachi, Daisuke Kanayama, Yukako Sakagami, Takashi Kudo, and Satoshi Nakamura: Analysis of Mood Changes and Facial Expressions during Cognitive Behavior Therapy through a Virtual Agent. In Companion Publication of the 2020 International Conference on Multimodal Interaction, pp. 477–481 (2020)