Automated Social Skills Trainer

Hiroki Tanaka[1], Sakriani Sakti[1], Graham Neubig[1], Tomoki Toda[1], Hideki Negoro[2], Hidemi Iwasaka[2], Satoshi Nakamura[1]
[1]Nara Institute of Science and Technology, Japan, [2]Nara University of Education, Japan
[1]{hiroki-tan,ssakti,neubig,tomoki,s-nakamura}@is.naist.jp, [2]{gorosan,hiwasaka}@nara-edu.ac.jp

ABSTRACT

Social skills training is a well-established method to decrease human anxiety and discomfort in social interaction, and acquire social skills. In this paper, we attempt to automate the process of social skills training by developing a dialogue system named "automated social skills trainer," which provides social skills training through human-computer interaction. The system includes a virtual avatar that recognizes user speech and language information and gives feedback to users to improve their social skills. Its design is based on conventional social skills training performed by human participants, including defining target skills, modeling, role-play, feedback, reinforcement, and homework. An experimental evaluation measuring the relationship between social skill and speech and language features shows that these features have a relationship with autistic traits. Additional experiments measuring the effect of performing social skills training with the proposed application show that most participants improve their skill by using the system for 50 minutes.

Author Keywords

Social skills training (SST); behavior detection; dialogue system; embodied conversational avatar; computer-based training.

ACM Classification Keywords

H.5.2. Information Interfaces and Presentation: User Interfaces – User-centered design; K.3.1. Computing Milieux: Computers and Education – Computer-assisted instruction (CAI)

INTRODUCTION

Many people have difficulties or are anxious in social interactions such as presentations and job interviews. The extreme example of people with these difficulties are those with autism spectrum disorders (ASD) [1]. Persistent social skill deficits impede those afflicted with them from forming relationships or succeeding in social situations.

IUI'15, March 29–April 01, 2015, Atlanta, GA, USA.
Copyright © 2015 ACM 978-1-4503-3306-1/15/03...$15.00.
http://dx.doi.org/10.1145/2678025.2701368

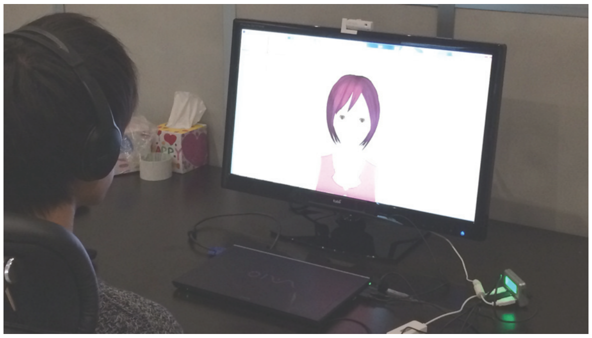


Figure 1. SST with the automated social skills trainer.

Social skills training (SST) is a general cognitive behavior therapy to train these social skills for people who have difficulties in social interaction, and is widely used by teachers, therapists, and trainers [4]. However, SST requires well-trained teachers, so the number of participants joining SST program is restricted and applications are competitive.

On the other hand, if part or all of the SST process could be automated, it would become easier for those requiring SST to receive it anywhere and anytime. In addition, it may be easier for those with social difficulties to use computers than interact directly. Donna Williams, who has autism, wrote a book entitled "Nobody nowhere" (1992), in which she stated

> *"The comprehension of words works as a progression, depending on the amount of stress caused from fear and the stress of relating directly. At best, words are understood with meaning, as with the indirect teaching of facts by a teacher or, better still, a record, television, or book. In my first three years in the special class at primary school, the teacher often left the room and the pupils responded to the lessons broadcast through an overhead speaker. I remember responding to it without the distraction of coping with the teacher. In this sense, computers would probably be beneficial for autistic children once they had the skills to use one."*

We propose a novel tool that tries to replicate conventional SST using a systematic and computer-based design. We develop a dialogue system named "automated social skills trainer," which is an application including video modeling of human behavior and real-time behavior detection as well as data visualization to help people improve their social skills (Figure 1). We investigate whether it is possible to help people who have difficulties in social interaction improve their social skills using an automated system which can be used anywhere, anytime.

RELATED WORK

The design of an automated social skills trainer for social-skills training brings together several fields, including research into computers in education, intelligent virtual agents, and affective computing. The following paragraphs briefly outline work in these areas.

The use of computers in SST is motivated by the fact that while individuals with social impairments have difficulty in social interaction, they also show good and sometimes even superior skills in "systemizing" [2]. Systemizing is the drive to analyze or build systems, and to understand and predict behaviour in terms of underlying rules and regularities. The use of systematic computer-based training for people who need to train social skills can take advantage of the following: 1) they favor the computerized environments because they are predictable, consistent, and free from social demands, 2) they can work at their own speed and level of understanding, 3) training can be repeated over and over again until the goal is achieved, 4) interest and motivation can be maintained through computerized rewards [5, 17, 18].

There has been one previous work on automated conversational coaches [15], which are dialogue systems aimed to train people for improving interview skills through real-time feature detection and feedback. They achieved 1) a realistic task involving training real users, 2) formative affective feedback that provides the user with useful feedback on the behaviors that need improvement, and 3) the interpretation or recognition of user utterances to drive the selection of backchannels or formative feedback. While this work is an excellent first step, it did not faithfully follow the traditional SST framework, omitting steps such as modeling of human behavior [9].

As conventional SST is a well-established method to improve human behavior and has clear goals and definitions for each module in the framework, our motivation is to follow the traditional SST framework as closely as possible. This paper presents the system design, modeling, and evaluation of the automated social skills trainer. Two experimental evaluations show that social skill is related to automatically extracted features and has a relationship to autistic traits, and that some participants improve in social skill using the automated social skills trainer.

SOCIAL SKILLS TRAINING

Conventional SST is an established method developed by Liberman for schizophrenics to reduce their anxiety and discomfort in social interaction [25]. SST is often performed with multiple sessions, and each session focuses on the training of one target skill for one or two hours. It is well known that SST can be used to effectively improve social skills for people with social disorders and ASD [4].

SST can be classified into individual (one to one training) and group (one to many or many to many training) settings. One advantage of group SST is that it enables participants to observe other participants' behaviour and also receive feedback from others. On the other hand, the advantage of individual SST is that the training can be relaxed and comfortable for participants, and that lessons can be tailored to the individual's needs.

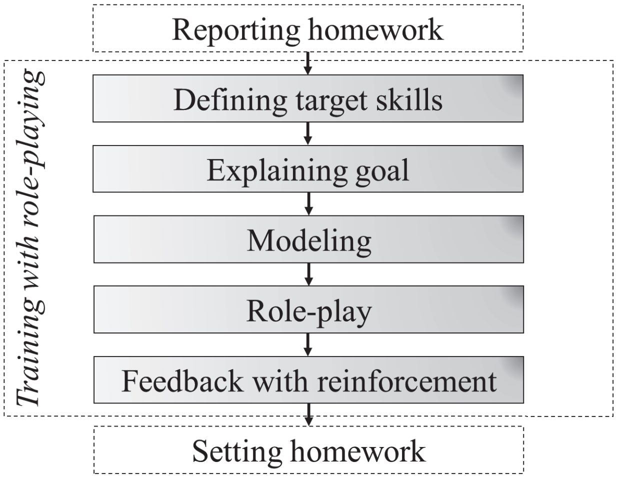


Figure 2. Human-to-human SST framework.

As shown in Figure 2, SST is generally based on the following steps: defining target skills, modeling, role-play, feedback, reinforcement, and homework. We briefly describe these steps as follows:

- **Defining target skills:** The major social problem is identified, and the skills to be trained are decided based on this problem. In order to figure out the major problems, the participants and trainers work together through discussion or trainers decide the target skills. Once the target skills are decided the trainers decide the goal after intervention. In this step, trainers sometimes use related books to help participants understand target skills and the goal of SST. Examples of target skills include presentation skills, job interview skills, self-introduction skills, or skills regarding how to decline another's offer or request.

- **Modeling:** Before participants are asked to perform an interaction, trainers act as a model, demonstrating the skill that the participants are focusing on so that participants can see what they need to do before attempting to do it themselves. For example, trainers may show a good story telling example using appropriate verbal and non-verbal cues.

- **Role-play:** Participants are asked to role-play. For example, participants tell their experience to the trainer. This allows the participants to practice their own skills in the target situation. Trainers observe participants' social skills subjectively, but mainly focus on voice quality, amplitude, facial expression, eye-gaze and other factors. This practice is a very important aspect of SST.

- **Feedback:** Trainers provide feedback at the end of role-playing (in the case of group SST, participants also receive feedback from other participants). This feedback helps participants to identify their strengths and weaknesses. For example, trainers may tell the participant that the role-play was very good because he/she used appropriate voice amplitude.

- **Reinforcement:** Trainers give positive reinforcement, praising the participant about their achievement of targeted behaviors, in addition to feedback. This is a very important aspect of the SST, because the participants often do not have confidence in social interaction, and tend to have low self-esteem. Therefore, positive encouragement helps build confidence in social environments.
- **Homework:** Trainers set little homework challenges that participants are required to do in their own time throughout the week. For example, trainers may ask the participant to tell their story to friends or family, and let the trainer know about the result.

By performing this training, participants can learn better social skills in a number of different ways, a core aspect contributing to the effectiveness of training. However, it should be noted that the human trainer plays a very involved role in the majority of these steps. As a consequence, SST requires professional or at least well-trained trainers satisfying the above abilities (e.g. being able to perform modeling and give appropriate feedback comments). The number of skilled trainers is small, and thus the number of participants joining SST program is restricted and applications are competitive.

AUTOMATED SOCIAL SKILLS TRAINING

In this section, we describe our proposed automated social skills trainer following the conventional individual SST framework. We replicate human-to-human individual SST by using a spoken dialogue system. While one disadvantage of individual SST is that there is no chance to see other participants' behaviour, we can provide a surrogate by playing video of others on the computer screen.

Table 1 shows the correspondence between conventional SST and our proposed method. In the following few pages, we describe each proposed module corresponding to a step in conventional SST.

Table 1. Correspondence of conventional SST and our proposed method.

Conventional SST	Proposed Method
Defining target skills	Story telling/narrative
Modeling	Recorded model video
Role-play	Conversation with an avatar
Feedback	Generation of visual feedback
Reinforcement	Generation of positive comments
Homework	Spoken instructions

- **Defining target skills:** Ideally, we would define specific target skills for each user through an initial interaction with the system. We plan to tackle this in future work, but for the time being we focused on a single target skill that has been shown to be widely applicable: story-telling or narrative ability. Story telling/narrative is a task of telling memorable stories, and is related to social interaction skills such as presentations and job interviews [8]. It has also been shown useful to distinguish children with ASD and children with typical development [22]. In the step of defining this goal, the system tells users that "This application will help you learn to tell stories well, and after training you will have more fun telling stories."

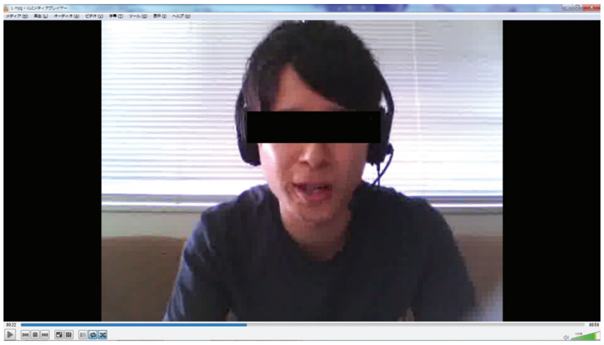

[redacted]

Figure 3. An example of video modeling.

- **Modeling:** Users can watch a recorded model video (Figure 3) before the role-playing. The recorded models are people who have relatively good narrative skills according to subjective evaluation. Users can watch and imitate the good examples.
- **Role-play:** The main part of the proposed system is the role-playing, which is performed through interaction with an avatar. When the user says "start role-playing," the system says "Please tell me your recent memorable story" The role-playing starts after the avatar's question, and continues for one minute. During this time, the avatar nods its head, and the system detects and analyzes language and speech features automatically. In this work, we focus on features that could differentiate between people with and without ASD as described in [22]: F0 variation, amplitude, voice quality, pauses, words per minute, words more than 6 letters, and fillers. We show a list of these features as follows:

  F0 variation: F0 indicates fundamental frequency, or pitch of voice. F0 variation therefore corresponds to the amount of variety in pitch, with less variety corresponding to a more monotone voice. We use these features because it has been widely noted that people with ASD have prosody that differs from that of their peers [12, 16, 6, 13, 23]. For instance, Kiss *et al.,* [14] found several differences in the fundamental frequency characteristics of people with ASD.

  Amplitude: In human-to-human SST, trainers often focus on volume of voice because both overly small and loud voices are not appropriate for many social situations. A previous study investigating the amplitude (power) of people with ASD [22] confirmed the importance of amplitude to identify children with ASD.

  Voice quality: People with ASD often exhibit abnormal voice quality, often described as more clear than their peers. Bonneh *et al.,* [7] quantified speech abnormalities in terms of the properties of the voice quality

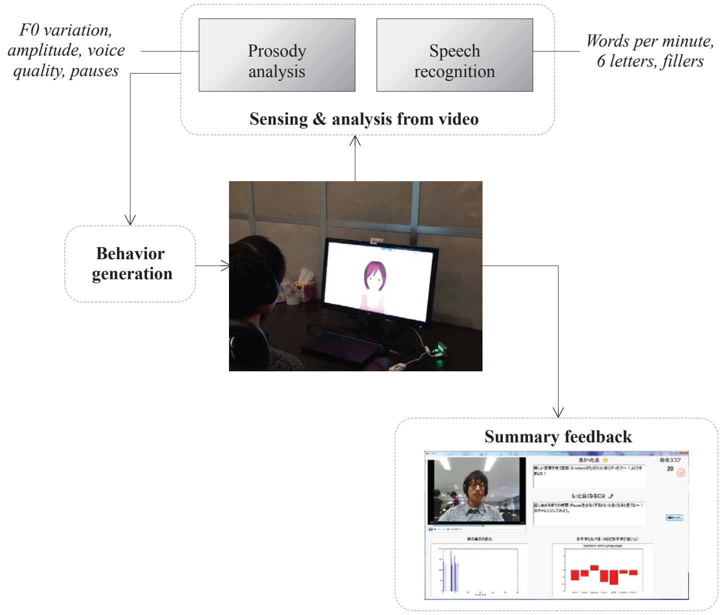

F0 variation, amplitude, voice quality, pauses
Prosody analysis
Speech recognition
Words per minute, 6 letters, fillers
Sensing & analysis from video
Behavior generation
Summary feedback

Figure 4. The automated social skills trainer framework.

and was able to identify children with ASD with more than 80% accuracy.

Pauses: There are reports finding that children with ASD tend to delay responses to their parent more than children with typical development in natural conversation [11].

Words per minute (WPM): There is a report that speaking rate was strongly correlated to interview skills [15]. WPM is related to frequency of speech.

Words more than 6 letters: Children with ASD use more complicated or unexpected words than typically developing children, and deficits of ASD affect inappropriate usage of words [21]. Words more than six letters may be related to complicated words [19].

Fillers: The frequency of filler usage is important in story telling or presentation. Too frequent use of fillers disturbs listener focus on the contents of speech.

- **Feedback:** The system displays summary feedback according to detected features. The feedback includes comments, the user's video, the parameters compared to model speaker, and the overall narrative score. The user can objectively confirm their strengths and weaknesses.
- **Reinforcement:** In addition to simply listing scores, the system also chooses good parts of the interaction, and gives positive feedback encouraging the targeted behavior.
- **Homework:** The system tells users to "Please tell your story to others throughout the week, and let me know about it." However, it should be noted that we did not evaluate SST across sessions in this work, though we plan to do so in the future.

Through this framework, we can replicate to some extent conventional individual SST with the spoken dialogue system replicating each module in the framework.

IMPLEMENTATION DETAILS

The automated social skills trainer system works on a regular laptop, which processes the audio input in role-playing. The processed data is used to generate the behaviors of an avatar that interacts with and provides feedback to users.

The role-playing, feedback, and reinforcement consist of three modules: behavior generation, sensing & analysis from video, and summary feedback as shown in Figure 4. The following subsections describe the modules in detail. It should be noted that the target language of our system is Japanese, and all data creation and experiments are performed with native Japanese speakers.

Data creation and subjective evaluation

As a first step towards building our system, we collected model data of people with relatively high levels of social skills. This video data is used both in the modeling module and for predicting scores in the feedback module. We collected data from a total of 19 people. Using this data, we assigned each dialogue an overall narrative skill score based on subjective evaluation, and the top five people were selected as models. Subjective evaluation was performed by having two

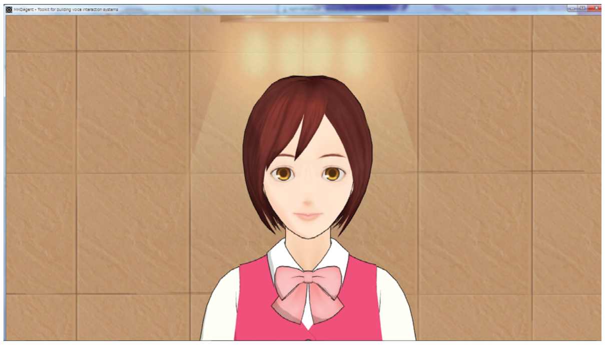


Figure 5. The avatar used in the automated social skills training system.

raters watch the recorded participants' narrative and answer a questionnaire. This process is described in more detail in the following Experiment 1.

Dialogue agent

The automated social skills trainer was developed using MMDAgent[1], which is a Japanese spoken dialogue system integrating speech recognition, dialogue management, text-to-speech, and behavior generation. MMDAgent works as a Windows application. We selected a character who is similar to an actual human, as we hope that this will make it easier for the user to generalize learned skills in a real situation. The avatar is displayed from the front, and there are no distractions in background (Figure 5). The user can operate and interact with the avatar by using speech throughout the training. All dialogue system utterances were created using templates written by the first author.

In addition, the system performs a number of behaviors to keep the user engaged. It blinks its eyes once every three seconds, and reacts to users during the role-playing. When the system recognizes an utterance, after a few seconds the system nods its head. The blinking and nodding behaviour motions were created by MikuMikuDance[2].

Sensing and analysis from video

To calculate the linguistic features, we performed automatic speech recognition (ASR) using the Julius dictation kit[3]. We used Mecab[4] for part-of-speech tagging in Japanese utterances. For speech feature extraction, we used the Snack sound toolkit[5].

The implementation of features is as follows: 1) F0 variation: We used the coefficient of variation for fundamental frequency with a minimum pitch of 100 Hz. We did not use mean, maximum and minimum values because there are individual and gender differences in terms of these features, 2) Amplitude: We used the mean value of amplitude, 3) Voice quality: We extracted the spectral tilt by calculating the difference between the first harmonic and the third formant "h1a3" as a feature expressing voice quality [10], 4) Pauses: We calculated values of pauses before new turns as time between the end of the avatar's utterance and the start of the user's utterance, 5) WPM: In the automated social skills trainer, the narrative continues for one minute, and we counted the number of words in one narrative, 6) Words more than 6 letters: We extracted percentage of words more than 6 letters as a feature, 7) Fillers: We calculated percentage of fillers such as "umm," or "eh" in Japanese. This feature automatically extracted by using output of the Mecab.

Summary feedback

Based on the calculated features, we provide feedback to the users about their social skills (Figure 6). Our goal was to design visualizations so that it would be easy for users to understand and interpret their narrative skill. The summary feedback provides following information.

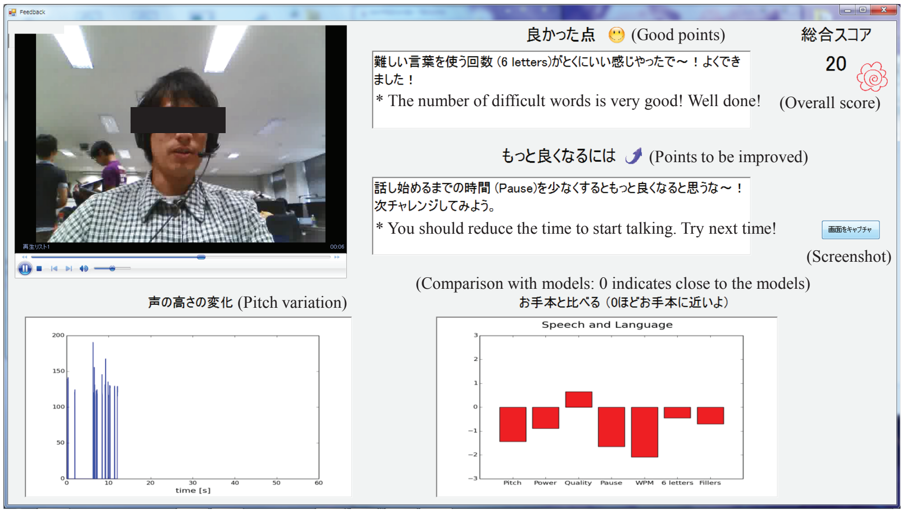
- **User video:** Participants can watch the recorded video and audio in the narrative. In doing so, the user can confirm their speech contents, facial expression, posture and so on [15], which are not analyzed automatically in the current version of the automated social trainer.
- **Overall score:** People with ASD prefer to check their improvement quantitatively [2]. The system displays the predicted overall score, which motivates the user to practice more and improve their score. We predict the overall score using the multiple regression method on a scale of 0 to 100.
- **Pitch variation:** Participants can see their pitch movement corresponding to the time. This also shows a visualization of how frequently they spoke.
- **Comparison with models:** The system visualizes the comparison of extracted features between the user's current narrative and model persons' narratives in terms of z-score, which is a statistical measurement of a score's relationship to the mean in a group of scores. The users are informed that they should attempt to emulate the model in all aspects.
- **Good points:** The system generates positive comments that reinforce the user's motivation with encouraging words [4]. The comments are generated based on the features that have values close to those of the models.
- **Points to be improved:** The system generates comments about points to be improved for next trial. The comments are generated based on the features that have values far from the models.
- **Screenshot:** Participants can save the feedback by clicking a button, and this is used for checking improvements over the course of training.

EXPERIMENT 1: DEFINING MODEL PERSONS

To evaluate the effectiveness of the proposed automated social skills trainer, we performed two experiments. In the first experiment, we sought to answer the following questions:

1) Does narrative skill relate to linguistic, acoustic, and other information?

[1] http://www.mmdagent.jp/
[2] http://www.geocities.jp/higuchuu4/
[3] http://julius.sourceforge.jp/index.php
[4] https://code.google.com/p/mecab/
[5] http://www.speech.kth.se/snack



[redacted]

難しい言葉を使う回数 (6 letters)がとくにいい感じやったで～！よくできました！

* The number of difficult words is very good! Well done!

話し始めるまでの時間 (Pause)を少なくするともっと良くなると思うな～！次チャレンジしてみよう。

* You should reduce the time to start talking. Try next time!

Figure 6. The summary feedback provided by the automated social skills trainer.

2) Is there a difference between talking to humans (human-human interaction: HHI) and talking to avatars (human-computer interaction: HCI) in terms of narrative?

3) Does narrative skill relate to autistic traits?

4) Are the extracted features effective for identifying narrative skill?

The result of the first experiment is used in data collection and summary feedback of the automated social skills trainer.

Procedure

We recruited 19 graduate students (16 males and 3 females), all of whom were native Japanese speakers[6]. All subjects used the proposed system and were told that their speech and video would be recorded. A webcam (ELECOM UCAM-DLY300TA) placed on top of the laptop and headset (ELECM HS-HP168K) recorded the video and audio of participants. We recorded not only HCI but also an HHI setting in which the first author listened to the speaker's story and nodded his head according to the speaker's utterances. The same 19 participants participate in the HCI and HHI. For HCI, participants interact with the same avatar.

To get a grasp of each subject's social skills independent of the proposed system, or the narrative setting in general, we also administered a social skills test for each subject. Specifically, we measured the sum of subarea scores for communication and social skills of Japanese version of the Autism-spectrum quotient (AQ) [3, 24] which is a standard tool to measure autistic traits with a total of 50 questions including 5 subareas.

Next, we had raters watch the interactions of each participant and rate their narrative skill. Although it would be ideal for raters to be professional social skills trainers, they are few and far between, so it is difficult to recruit them for the experiment. Thus, as a proxy, we selected raters from members of the general population. Because raters are required to have good social skills to recognize users' non-verbal expressiveness, we selected two people (male and female) with good social skills as annotators. Specifically, the annotators were selected to have low sums of the AQ subarea scores for communication and social skills (where lower indicates better social skills). The sums of both areas were 1 and 4, which is lower than the mean value of 7.6 for Japanese students [24]. The raters did not participate in the experiments as subjects. The raters did not know the recorded participants, and were trained by rating several examples prior to the evaluation. Two raters watched recorded participants' narrative for both HHI and HCI, and answered a questionnaire[7], which is based on [15]. The questionnaire included the following items related to the participant's overall narrative performance and use of non-verbal cues such as intonation, amplitude, and lexicon usage, rated on a scale of 1 (not good, not appropriate, or small (few)) to 7 (good, appropriate, or large (frequent)).

Q1. Overall narrative skill

Q2. Concentration

Q3. Friendliness

Q4. Attractiveness

Q5. Speaking rate

Q6. Usage of fillers

Q7. Intonation

Q8. Voice quality

Q9. Amplitude

Q10. Usage of easy words

[6] Note that the Research Ethic Committee of our institution has reviewed and approved both this and the following experiments. Written informed consent was obtained from all subjects before the experiments.

[7] https://docs.google.com/forms/d/1AQRc1sAQQooEt7zY89H7aJQzKFf8zqGH4u-nCCVwnGs/viewform?c=0&w=1&usp=mail_form_link

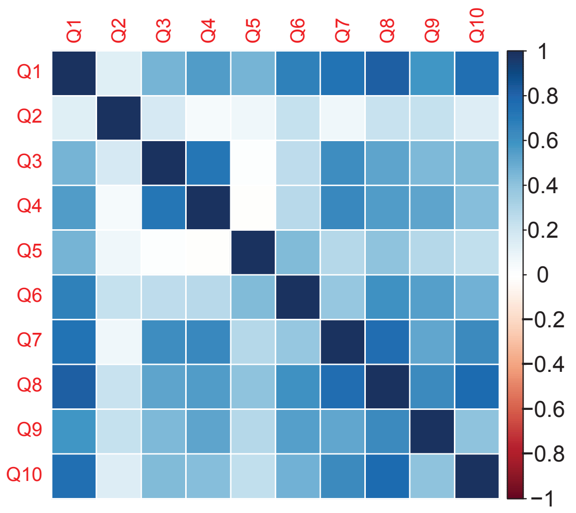


Figure 7. Pearson's r correlations between various questions. Color indicates the strength of statistically significant correlations, and white indicates zero. Rows and columns represent the questions in the same order, so the diagonal is self-correlation.

Agreement

The agreement of the two raters was measured by Cohen's kappa-coefficient, which calculates agreement beyond chance by distinguishing the observed agreement (A_{obs}) from the agreement by chance (A_{ch}), as follows:

$$\kappa = (A_{obs} - A_{ch}) / (1 - A_{ch}) . \qquad (1)$$

The two raters answered the ten questions for each speaker's narrative. We analyzed agreement for the question of overall narrative skills. For each rater, each subject was assigned a class of being either above or below the average score for the rater, and agreement between the classes was used to calculate the coefficient. The Kappa coefficient of two classes for two raters was 0.580, which corresponds to moderate agreement according to the scale proposed by [20].

Correlation between questions

Figure 7 shows the correlation matrix of each question. For Q6, because usage of fillers can be assumed to be inversely proportional to social skill, we inverted the ratings before measuring correlation. The result showed that questions especially asking about the speech and language features were significantly related to overall narrative skill. On the other hand, questions asking about concentration were not related to overall skills and other features. We can also see that questions related to speech and language features, excluding Q5, were correlated each other.

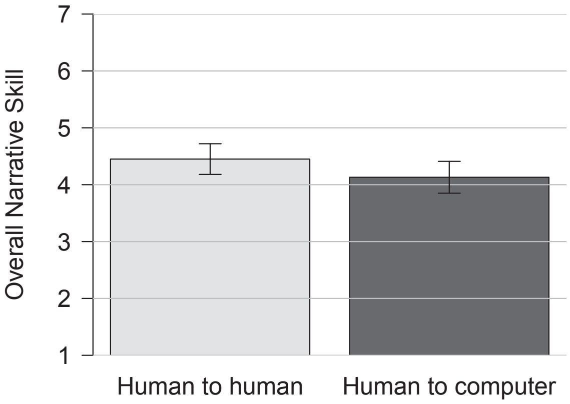


Figure 8. The difference of raters' scores between HHI and HCI. Error bars indicate standard error.

Differences between human and computer interaction

To examine the difference between HHI and HCI, we show averaged rater scores for HHI and HCI in Figure 8. We can see that there were differences between HHI and HCI, and raters' scores of narrative skill in HHI were slightly higher than HCI. However, we did not find a statistical difference ($p > .05$) by Student's t-test. It is likely that if differences exist between interaction with our proposed system and interaction with an actual human in terms of overall narrative skills, they are small.

Model people and autistic traits

Based on the raters' scores, we determined the top 5 of 19 subjects to be our models for additional experiments including the modeling step of SST. As shown in Figure 9, the median value of the AQ was 1 in the case of model persons, and 13 in the case of the others. This indicates that there is also a strong relationship between the raters' assessment of narrative skill and the subjects' answers on the AQ test.

Regression

We calculated the statistical differences of the automatically extracted features between model persons and others using Student's t-test. We found that WPM, words of more than 6 letters, and amplitude were significantly different between the groups ($p < .05$), and other features were not significantly different ($p > .05$). Thus we used these three features to predict overall narrative skill using the multiple regression method. Finally we found that the correlation between the predicted value and actual value was 0.51 ($p < .05$). This regression model was integrated into the system as the feedback module's overall score.

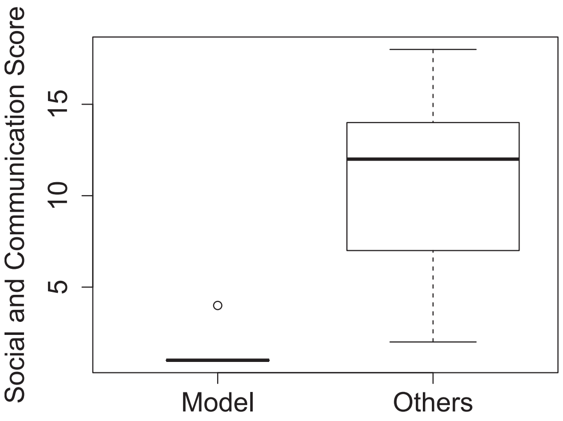


Figure 9. The ranges of the AQ for model persons and others. Zero indicates high social and communication skills, and 20 indicates low social and communication skills.

EXPERIMENT 2: SOCIAL SKILLS TRAINING

In the second experiment, we examined whether the automated social skills trainer is effective to train social skills, specifically:

1) How effective is the automated social skills trainer in helping users improve their narrative skills?

2) Do users find the automated social skills trainer easy to use and helpful?

Procedure

We recruited a total of 30 graduate students (22 males and 8 females) all of whom were native Japanese speakers, different from those who participated in the first experiment. Participants first entered the experiment room, and were given instructions by the first author. All subjects were told that their speech and video would be recorded. The webcam (ELECOM UCAM-DLY300TA) placed on top of the laptop and headset (ELECM HS-HP168K) recorded the video and audio of participants.

We separated participants into 3 groups: the reading book group (10 males), the video modeling group (6 males and 4 females), and the feedback group (6 males and 4 females). The reading book group, which serves as a control, read two types of social skills books which were related to story telling/narrative skills. The book titles were "Social skills training: collection of cases" and "The easiest guide to presentations" (in Japanese). The video modeling group and the feedback group used the automated social skills trainer for their training. The video modeling group only watched the model videos, while the feedback group performed role-play and received automated feedback. The feedback group can also watch the model videos. As shown in Figure 10, all subjects spoke their narratives to the agent (pre), received training for 50 minutes, and spoke their narratives to the agent again (post).

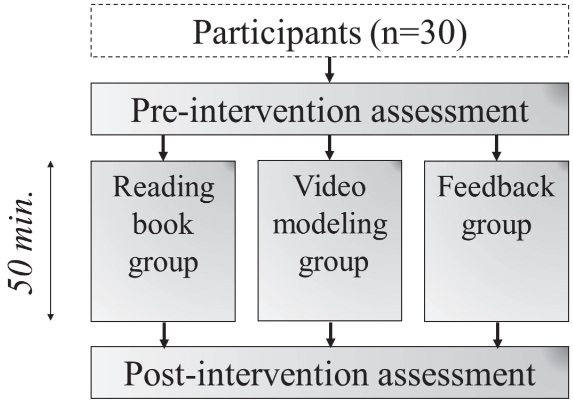


Figure 10. Study design and participant assignment to experimental groups in the second experiment.

The same two raters from the first experiment evaluated the subjects' narrative skill by answering overall narrative skill (Q1 of the first experiment) rated on a scale of 1 to 7. Raters did not know subjects, and the order of the pre- and post-training narratives was randomized to prevent bias. We averaged the two raters' scores and calculated improvement in score (post - pre) for each group[8]. Note that the initial scores of each group were not significantly different (F[2,25]=0.90, $p > .05$).

The effect of intervention type was analyzed using one-way analysis of variance (ANOVA). Post-hoc comparisons between the feedback group and the reading book group, and the feedback group and the video modeling group involved Bonferoni's method.

After using the automated social skills trainer, the feedback group answered a questionnaire to evaluate usability and effectiveness of the system[9]. The questionnaire included the following items related to the system usability and training effect, rated on a scale of 1 (disagree) to 7 (agree). The users were also asked to provide comments about each question.

Q1. The system was easy to use.

Q2. I would like to use this system frequently.

Q3. The trainer looks like a human.

Q4. Watching my own video and feedback were useful.

Q5. Watching model video was useful.

Agreement

We calculated agreement according to the same procedure described in the previous section. The Kappa coefficient of two classes for two raters was 0.638, which indicates good agreement based on [20]. The agreement of the two raters was almost the same as the first experiment.

[8] Among the 30 participants, one subject each in the reading book and feedback group did not have sufficient time to train for 50 min, so we omitted these subjects.

[9] https://docs.google.com/forms/d/1Qhe1UvXrZlHvOfY5YewQD1d2pwue5APhrNiLHl3Iasc/viewform?c=0&w=1

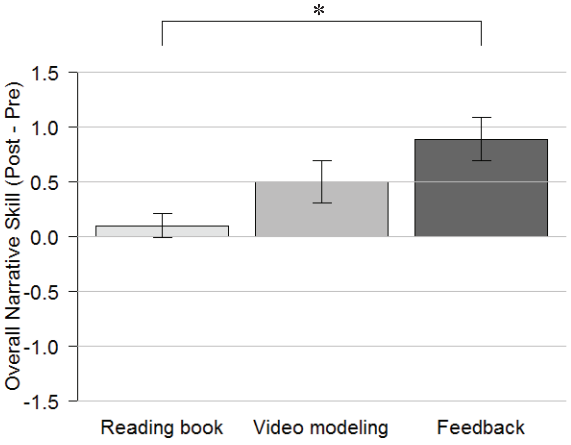


Figure 11. The overall narrative score of each group. Error bars indicate standard error (*: p < .05).

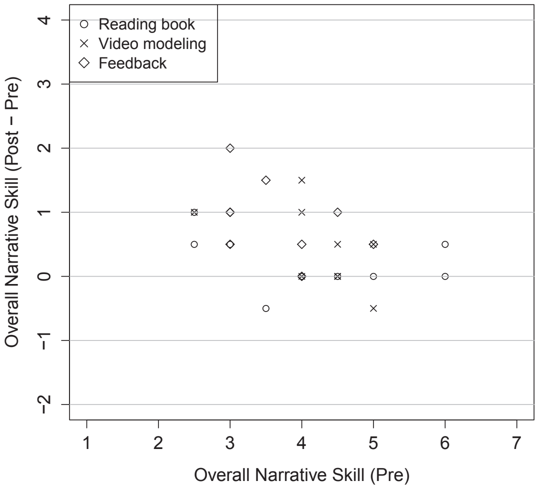


Figure 12. The relationship between initial and improvement in scores.

Training effect

Figure 11 shows the improvement of overall narrative skills in each group. These results show that intervention type significantly affected the change in raters' scores (F[2,25]=4.67, $p < .05$) with $\eta^2 = .26$ according to ANOVA. Comparisons showed that the change in raters' scores of participants in the feedback group who used the automated social skills trainer was significantly higher than the reading book group ($p < .05$). The difference between the video modeling group and the feedback group, and the video modeling group and the reading book group was not judged as statistically significant ($p > .05$).

Figure 12 shows the improvement of overall narrative skills and initial scores in each group. The correlation coefficient between overall narrative skills prior to training and improvement was -0.438 ($p < .05$) showing a weak negative correlation. This is a natural result, because people who have difficulties in social interaction have more space to improve.

Subjective evaluations

The paragraphs below describe findings from the participants' subjective evaluations of the automated social skills trainer and their feedback on their experience. We analyzed qualitative and quantitative results to represent user experience and system usability.

- **The system was easy to use:** The usability of the automated social skills trainer was rated an average of 5.4 (SD = 0.9). Most participants found the system is easy to use.

  "It is easy to operate the system using only speech. My voice was recognized and I felt comfortable."

  "The content of training was separated according to purpose (e.g. modeling, feedback, and homework), and it was easy for me."

- **I would like to use this system frequently:** The question regarding whether the user would like to use the system again was rated an average of 5.0 (SD = 0.7). Most participants would like to use the system frequently.

  "I would like to talk to system with more variation. I want to use this system every day, and also record a life log."

  "It is interesting to watch my score with helpful comments."

- **The avatar looks like a human:** The question about whether the avatar looks like a human was rated an average of 4.8 (SD = 0.4). Some participants thought the character looks like a human.

  "I thought avatar's behavior is natural, and I did not feel unnaturalness in interaction."

  "I felt like I spoke to real human."

  However, some participants thought the avatar did not seem like a human specifically in terms of speech synthesis.

  "I felt the synthesized speech is robot-like, the intonation was unnatural."

- **Watching my own video and feedback were useful:** According to the participant's responses to the questionnaire, the feedback and watching the user's own video was rated an average of 5.6 (SD = 1.1). Most participants thought the feedback and video were useful.

  "It was easy to train my skill because the system indicated the points to be improved."

  "I was happy to be encouraged."

  "Conversation is abstract, but the system displayed the concrete values. It is very interesting and helpful."

- **Watching model video was useful:** Overall, participants rated their preference toward watching the models' video an average of 5.2 (SD=1.5), suggesting the usefulness of model video.

  > *"After watching the role model, I easily started to talk because of the good reference."*
  >
  > *"I was interested in the variation of the good examples."*

  However, the result also showed the SD value was large. Some participants did not say that model video was helpful.

  > *"I think I already had good skill so the modeling is not useful for me."*
  >
  > *"I would like to see the good points of the model persons."*

DISCUSSION

In this paper, we developed a dialogue system named "automated social skills trainer" which provides social skills training in the context of human-computer interaction. The automated social skills trainer is based on conventional SST including defining target skills, modeling, role-play, feedback, reinforcement, and homework. We focus on story telling/narrative skill as a target skill. The system includes several modules: behaviour generation, sensing & analysis from recorded video, and summary feedback. In this study, our focus was to assess the effectiveness of an automated social skills trainer that follows human-to-human SST as closely as possible. To evaluate effectiveness of the automated social skills trainer, we performed two experiments.

In our first experiment, we confirmed the relationship between overall narrative skills and speech and language information, confirmed that there was no significant difference between HHI and HCI, set model persons according to the evaluation of two raters, and found a relationship between observed narrative ability and AQ. Baron-Cohen and their colleagues reported that the AQ value was widely distributed among members of the general population and that it is related to autistic traits [3]. Our result showing a relationship between AQ and overall narrative skills is consistent with the above report.

In our second experiment, we confirmed a training effect particularly for participants in the case of the feedback group rather than the reading book group. It showed that the system could help people who have difficulties in social interaction improve their social skills. The video modeling group also improved in their scores, which is consistent to the previous work [9]. The video modeling of others was also helpful in social skills training. In this experiment, we did not set a group that watched their own video and did not watch the feedback. There is previous work reporting that subjects dislike looking at their own video during interview skill training and the skills did not change [15], so we plan to investigate these elements separately in the future. We also confirmed a weak negative correlation between initial narrative skills and improvements in scores. This shows that training effects are found more strongly in people who have difficulties in social interaction than others. In subjective evaluation, we confirmed most participants of the feedback group were satisfied with the system in terms of usability and the feedback.

For future directions, we would like to confirm the training effect over a longer period, and recruit special-need populations such as people with ASD. We also plan to add other target social skills in the automated social skills trainer, and compare with human-to-human SST. In order to do so, we will more thoroughly examine SST from the viewpoint of HHI and HCI including types of agent (e.g. the effect of human-like avatars). Multi-modal feature analysis is also important. For example, we plan to incorporate visual image processing and eye-gaze analysis into the automated social skills trainer.

ACKNOWLEDGMENTS

We would like to thank Kazuyo Iida for her helpful comments regarding the system. This study is supported by JSPS KAKEN 26540117 and Foundation for Nara Institute of Science and Technology.

REFERENCES

1. American Psychiatric Association. *Diagnostic and statistical manual of mental disorders* (5th ed.). Washington, DC, (2013).
2. Baron-Cohen, S., Richler, J., Bisarya, D., Gurunathan, N., Wheelwright, S. The systemizing quotient: an investigation of adults with Asperger syndrome or high-functioning autism, and normal sex differences. *Philosophical Trans. the Royal Society of London Series B: Biological Sciences 358*, 361-374 (2003).
3. Baron-Cohen, S., Wheelwright, S., Skinner, R., Martin, J., Clubley, E. The Autism-Spectrum Quotient (AQ): evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians. *J. Autism and Developmental Disorders 31*, 5-17 (2001).
4. Bauminger, N. The facilitation of social-emotional understanding and social interaction in high-functioning children with autism: Intervention outcomes. *J. Autism and Developmental Disorders 32*, 283-298 (2002).
5. Bishop, J. The Internet for educating individuals with social impairments. *J. of Computer Assisted Learning 19*, 546-556 (2003).
6. Bone, D., Black, P., Lee, C., Williams, E., Levitt, P., Lee, S., Narayanan, S. Spontaneous-Speech Acoustic-Prosodic Features of Children with Autism and the Interacting Psychologist. *Proc. Interspeech*, (2002).
7. Bonneh, Y., Levanon, Y., Dean Pardo, O., Lossos, L., Adini, Y. Abnormal speech spectrum and increased pitch variability in young autistic children. *Frontiers in Human Neuroscience 4*, (2011)
8. Davis, M., Dautenhahn, K., Nehaniv, C., Powell, S. Towards an Interactive System Facilitating Therapeutic Narrative Elicitation. *Proc. 3rd Conf. on NILE*, (2004).

9. Essau, C. A., Olaya, B., Sasagawa, S., Pithia, J., Bray, D., Ollendick, T. H. Integrating video-feedback and cognitive preparation, social skills training and behavioural activation in a cognitive behavioural therapy in the treatment of childhood anxiety. *J. Affect Disorders 167*, 261-267 (2014).

10. Hanson, M. H. Glottal characteristics of female speakers. *Harvard University, Ph. D. dissertation*, (1995).

11. Heeman, P. A., Lunsford, R., Selfridge, E., Black, L., Van Santen, J. Autism and interactional aspects of dialogue. *Proc. 11th SIGDIAL*, 249-252 (2010).

12. Kanner, L. Autistic disturbances of affective contact. *Nervous Child 2*, 217-250 (1943).

13. Kiss, G., Van Santen, H. Estimating Speaker-Specific Intonation Patterns Using the Linear Alignment Model. *Proc. Interspeech*, 354-358 (2013).

14. Kiss, G., Van Santen, H., Prud'hommeaux, T., Black, M. Quantitative Analysis of Pitch in Speech of Children with Neurodevelopmental Disorders. *Proc. Interspeech*, (2012).

15. Hoque, E., Courgeon, M., Mutlu, B., Martin, C., Picard W. MACH: my automated conversation coach. *Proc. 15th Conf. on UbiComp*, 697-706 (2013).

16. McCann, J., Sue, P. Prosody in autism spectrum disorders: a critical review. *International J Language & Communication Disorders 38*, 325-350 (2003).

17. Moore, D., Mcgrath, P., Thorpe, J. Computer-aided learning for people with autism - a framework for research and development. *Innovations in Education and Teaching International 37*, 218-228 (2000).

18. Parsons, S., Mitchell, P. The potential of virtual reality in social skills training for people with autistic spectrum disorders. *J. Intellectual Disability Research 46*, 430-443 (2002).

19. Pennebaker, W., Martha, F., Roger, B. Linguistic inquiry and word count (LIWC). LIWC [Computer software], (2005).

20. Rietveld, T., Van Hout, R. Statistical techniques for the study of language behaviour. *Mouton de Gruyter*, (1993).

21. Rouhizadeh, M., Prud'hommeaux, E., Roark, B., Van Santen, H. Distributional semantic models for the evaluation of disordered language. *Proc. NAACL-HLT*, 709-714 (2013).

22. Tanaka, H., Sakriani, S., Neubig, G., Toda, T., Nakamura, S. Linguistic and Acoustic Features for Automatic Identification of Autism Spectrum Disorders in Children's Narrative. *ACL2014 Workshop on Computational Linguistics and Clinical Psychology*, 88-96 (2014).

23. Van Santen, H., Richard, S., Alison, H. Quantifying repetitive speech in autism spectrum disorders and language impairment. *Autism Research 6*, 372-383 (2013).

24. Wakabayashi, A. Baron-Cohen, S., Wheelwright, S., Tojo, Y. The Autism-Spectrum Quotient (AQ) in Japan: a cross-cultural comparison. *J. Autism and Developmental Disorders 36*, 263-270 (2006).

25. Wallace, C. J., Nelson, C. J., Liberman, R. P., Aitchison, R. A., Lukoff, D., Elder, J. P., Ferris, C. A review and critique of social skills training with schizophrenic patients. *Schizophr Bull 6*, 42-63 (1980).