

A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons

Mohammad Rafayet Ali¹, Seyedeh Zahra Razavi¹, Raina Langevin², Abdullah Al Mamun¹

¹University of Rochester, ²University of Washington
{mali7,srazavi,amamun}@cs.rochester.edu
rlangevi@uw.edu

Benjamin Kane¹, Reza Rawassizadeh², Lenhart K Schubert¹, Ehsan Hoque¹

¹University of Rochester, ²Metropolitan College, Boston
University
{bkane2,schubert,mehoque}@cs.rochester.edu
rrawassizadeh@acm.org

ABSTRACT

We present the design of an online social skills development interface for teenagers with autism spectrum disorder (ASD). The interface is intended to enable private conversation practice anywhere, anytime using a web-browser. Users converse informally with a virtual agent, receiving feedback on nonverbal cues in real-time, and summary feedback. The prototype was developed in consultation with an expert UX designer, two psychologists, and a pediatrician. Using the data from 47 individuals, feedback and dialogue generation were automated using a hidden Markov model and a schema-driven dialogue manager capable of handling multi-topic conversations. We conducted a study with nine high-functioning ASD teenagers. Through a thematic analysis of post-experiment interviews, identified several key design considerations, notably: 1) Users should be fully briefed at the outset about the purpose and limitations of the system, to avoid unrealistic expectations. 2) An interface should incorporate positive acknowledgment of behavior change. 3) Realistic appearance of a virtual agent and responsiveness are important in engaging users. 4) Conversation personalization, for instance in prompting laconic users for more input and reciprocal questions, would help the teenagers engage for longer terms and increase the system's utility.

CCS CONCEPTS

• **Human-centered computing** → *Empirical studies in HCI*.

KEYWORDS

Animated conversational agent design, social skills training, real-time feedback, autism intervention, nonverbal behavior, schema-based dialogue management

ACM Reference Format:

Mohammad Rafayet Ali, Seyedeh Zahra Razavi, Raina Langevin, Abdullah Al Mamun, Benjamin Kane, Reza Rawassizadeh, Lenhart K Schubert, and Ehsan Hoque. 2020. A Virtual Conversational Agent for Teens with Autism Spectrum Disorder: Experimental Results and Design Lessons. In *IVA '20*.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for components of this work owned by others than ACM must be honored. Abstracting with credit is permitted. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee. Request permissions from permissions@acm.org.

IVA '20, October 19–23, 2020, Virtual Event, Scotland UK

© 2020 Association for Computing Machinery.

ACM ISBN 978-1-4503-7586-3/20/09...\$15.00

<https://doi.org/10.1145/3383652.3423900>

Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents (IVA '20), October 19–23, 2020, Virtual Event, Scotland UK. ACM, New York, NY, USA, 8 pages. <https://doi.org/10.1145/3383652.3423900>

1 INTRODUCTION

Autism spectrum disorder (ASD) is a developmental disorder which affects one in 59 individuals in the US alone [1]. Almost all the individuals with ASD show deficits in nonverbal communication [2]. Current practices for improving the communication skills deficit involve therapy sessions with behavior experts. There is a significant shortage of behavioral experts resulting in therapy time being limited and inaccessible. A computer-mediated online social skills intervention has the potential to enable individuals with ASD to practice conversation frequently with a standardized and repeatable stimulus. Additionally, computer-mediated tools can enable a therapist to serve more individuals by monitoring their progress remotely while continuing their weekly face-to-face therapy. As an added benefit, such technology will enable remote therapy which is particularly useful when social distancing is recommended due to virus outbreaks (e.g., COVID-19). As we write this paper, we face a global pandemic that forced 91,000 U.S. public and private schools to close and affected more than 50 million school students¹ including people on autism spectrum. The sudden lockdown and extreme mitigation did not allow the therapists to prepare, leaving the individuals who are in need of behavioral therapy to look for online solutions. We believe online technologies such as LISSA has the potential to augment the traditional therapy for social skills development.

In this paper, we review the design of an online social skills development program, LISSA - Live Interactive Social Skills Assistance. We focus on design lessons learned, based on trials with teens with ASD, particularly on analysis of post-session interviews with participants. LISSA features a virtual agent capable of engaging the user in a multi-topic conversation and giving real-time feedback on the nonverbal cues, by analyzing the spoken language and the facial expressions of the user in real-time. In addition, LISSA provides a summary feedback after the conversation. The motivating idea is that users can interact with LISSA repeatedly, and potentially learn and observe their skills improvement from the summary feedback in multiple conversation sessions, held in a private and safe environment. The design of the interface was guided by an expert UX designer, psychologists, and a pediatrician. The initial design of the interface was evaluated in a Wizard of Oz setting with 47

¹<https://www.educationnext.org/covid-19-closed-schools-when-should-they-reopen-coronavirus>

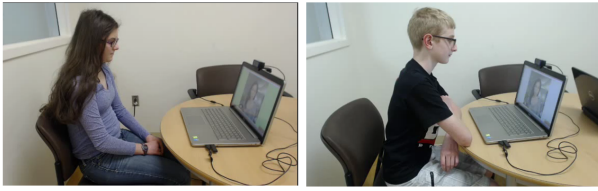


Figure 1: Teenagers with Autism Spectrum Disorder interacting with LISSA.

participants. The feedback system and the dialogue manager were then automated using the data. Subsequently, we focus on adapting the system design to teenagers with ASD to inform further development. The study, whose lessons we explore here, involved nine teenagers with ASD. Figure 1 shows two teenagers with ASD interacting with LISSA in the lab. The teens with ASD interacted with LISSA and were interviewed about the experience. Through a thematic analysis of the interview transcripts we identified several key design guidelines, 1) Users should be fully briefed about the purpose of an interface, and its capabilities and limitations. 2) An interface should incorporate positive acknowledgment of behavior change. 3) Realistic appearance of a virtual agent and responsiveness are significant factors in engaging users. 4) Conversation personalization, e.g., user-sensitive turn-taking and prompting laconic users, would help the teenagers engage for longer terms, and thus have them benefit from the interaction.

Our paper makes the following contributions -

- We motivate and explain the design of the LISSA system, and describe the automation of the system that makes it suitable for private, ubiquitous use.
- Based on post-session interviews, we discuss key design guidelines that we identify as important for future interface design for individuals with ASD.

2 RELATED WORK

Individuals with autism spectrum disorder (ASD) are characterized in varying degrees by their social interaction, difficulties in verbal and nonverbal communication, and repetitive behaviors [3]. The existing treatments for high functioning ASD do not address the condition as a whole, rather they focus on individual symptoms. Thus there is no single intervention for such individuals. In this section, we discuss the existing computer-based social skills interventions for individuals with ASD. Additionally, we discuss the work that employed virtual agents for social skills training.

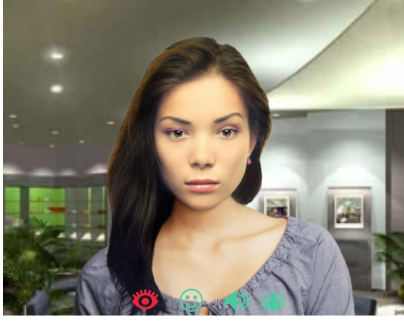
2.1 Social Skills Intervention for ASD

In the past, interventions that focused on children and teenagers with ASD touched a variety of technologies, including virtual agents. Tartaro and Cassell [4] developed a virtual peer to engage children with ASD in a collaborative storytelling task. The results showed that children with autism approach the collaborative activities with virtual peer with excitement and they get engaged and improve in contingent discourse. These works prove the applicability of interactive technologies in social skills training for individuals with ASD. Feedback on nonverbal and verbal behavior was also applied as communication skills training. Foster et al. [5] designed

ECHOES as a multimodal learning environment intended for children with ASD. The system features a virtual agent that engages a child in a collaborative learning activity and provides feedback based on sensed features including gaze direction and gesture. In a subsequent work with ECHOES, Bernardini et al. [6] designed a virtual agent as a credible social partner for children with ASD, that engages them in interactive learning activities. Milne et al. [7] designed a virtual agent as an educational tool for children with ASD, which helps improve their conversational skills and ability to deal with bullying. In a study with ten participants, the authors showed that participants who received the intervention had gained higher conversational skills and more knowledge about bullying than the control participants. Mower et al. [8] developed Rachel, an embodied conversational agent designed to elicit and analyze naturalistic interactions. This tool was designed for children with autism to encourage their affective and social behavior. Boujarwah et al. [9] presented a tool to enable non-expert humans to generate conversational scenarios, which can be used to teach children with ASD, appropriate behaviors in different social scenarios. DeVault et al. [10] developed SimSensei - a virtual agent in the context of the healthcare decision support system. The goal of this system is different than our work as it aims to identify psychological distress indicators through a conversation with a patient in which the patient feels comfortable sharing information. This system has both nonverbal sensing and a dialogue manager. The dialogue manager uses four classifiers to categorize the users' speech, and hence to generate a relevant response. Hopkins et al. [11] developed a social skills training system for children with ASD. In a study with 49 individuals, the authors demonstrated that the program helped the participants improve their emotion recognition ability and social interactions. Razavi et al. [12] developed a conversational agent capable of conversing with teenagers with ASD. The authors employed a script-like schema for guiding the dialogues, and generated appropriate responses using hierarchical pattern transductions.

2.2 Dialogue Management

Smartphone-based conversational agents can be impressive in their question-answering tasks but have been found to provide incomplete and inconsistent responses in areas like mental health, interpersonal violence, and physical health [13]. Woebot [14], a chat-bot recently developed at Stanford for delivering CBT to young adults with depression and anxiety, was able to significantly reduce symptoms of depression; however, its inability to hold a natural conversation was reported as the main issue users complained about. Popular methods in dialogue managements include frame-based methods such as [14] and [15] where the user's responses are used to fill frame slots and to perform a task according to the contents of the frames. Another method, employed in SimCoach [16] and ASST [17] controls dialogue via transitions between predetermined states based on the user's input. While the method shows good performance in limited dialogues, scalability remains an issue for more open-ended conversation. Some other end-to-end systems achieve meaningful single-turn responses through data-driven methods (e.g., [18]), but cannot conduct a coherent longer dialogue; also, data-driven methods are susceptible to bias and privacy issues, among others [19].



(a) LISSA Interface. The virtual agent engages the users in a conversation and provides real-time feedback by flashing the icons at the bottom.

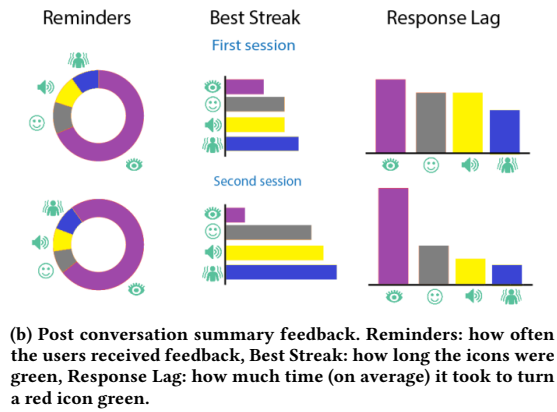


Figure 2: LISSA system and post conversation feedback interface.

Taken together, prior studies in computer-based social skills training and dialogue management suggest that dialogue-based systems hold considerable promise for helping individuals with ASD (or certain others with special needs). In our work, we have focused on improving users' nonverbal behavior in communication through real-time feedback. Through participatory studies, and post-session interviews, we have identified significant design guidelines for such virtual agent based intervention.

3 SYSTEM DESIGN

In this section, we first describe the LISSA interface and the initial data collection process using a Wizard-of-Oz prototype. Then we describe the feedback and dialogue generation of the system. The automation of the system, prepared the way for the study with ASD teens, on which we focused later on.

3.1 Interface

Our design team includes two psychologists, a psychiatrist, and a user experience designer. The interface consists of two major parts: a conversation interface, and a post-conversation feedback interface. The conversation interface features a virtual agent (see Figure 2a), able to hold conversations with users, and provide real-time feedback on their nonverbal behavior. The feedback is presented

through four flashing icons placed at the bottom of the interface. The four icons represent eye-contact, smile, speaking volume, and body movement. The icons are green by default but turn to flashing red as a prompt to the user to adjust their behavior. We kept the simple red-green color scheme to reduce the cognitive load on the user, since real-time feedback can be distracting [20]. The persona of the virtual agent was selected from the recommendation of the pediatric psychiatrist of our team. The four nonverbal cues were selected because of their known applicability in improving social skills [21]. The post-conversation feedback interface (see Figure 2b) shows a summary of the feedback provided during the conversation. This interface shows how many times the user received feedback (Reminder), how long the user kept the icons green (Best Streak), and how much time the user took to adjust nonverbal behavior (Response Lag). In the past, similar interface has been used for communication skills training [12, 22].

3.2 Data Collection

We first developed a Wizard of Oz prototype of the interface. There were two human operators for driving the system. One was responsible for the dialogue management and the other was responsible for giving feedback by flashing icons. Both operators were able to monitor the user remotely and control the interface through a web interface. This design allowed us to collect data from users and learn about their user experience without as yet applying machine learning techniques. Subsequently, we used the collected data to train machine learning models for automation.

We collected 46 videos of 23 neurotypical young adults (age between 18 and 23) interacting with LISSA. The videos were captured through a camera attached on top of the computer monitor. The Wizard of Oz operators were driving the LISSA program while we recorded the videos. We then employed six research assistants majoring in Psychology to label the collected videos. The psychologists in our team provided multiple training sessions to the research assistants. During the labeling the research assistants were additionally given a code book to follow. The code book includes rules for marking the moments where the participants should receive feedback. An example of these rule could be marking the moment as eye contact feedback when the participant is looking down. The research assistants watched each of the video recordings and marked those moments where the participants should receive red icon flash. The initial average inter-rater reliability was 0.42 (Krippendorff's alpha). We then decided to consider those moments for feedback labels where more than two research assistants marked it as a feedback moment due to the low overall agreement.

3.3 Automated System

We extracted facial and prosodic features from the recorded videos, including head pose, smile, facial action units, volume, and voice pitch. For this, we used off-the-shelf software tools, namely OpenFace [23], and Praat [24]. We then trained a hidden Markov model (HMM) to generate the flashing-icon feedback from the facial and prosodic features. The details and accuracy of this model can be found here [25]. In the past, HMMs have proven successful in modeling human behaviors and actions [26].

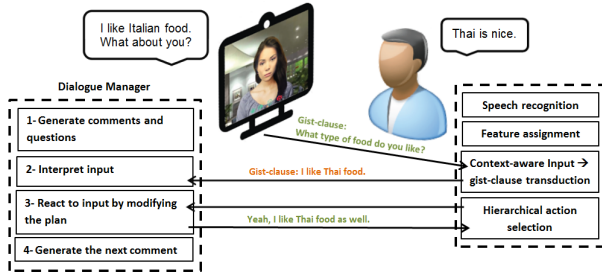


Figure 3: Overview of LISSA dialogue manager.

3.4 Dialogue Management

In order for LISSA to conduct a conversation, we developed an automated dialogue manager capable of handling multiple topics. An outline of the functionality of the dialogue module is presented in Figure 3. LISSA leads the conversation by asking questions on different topics (often after a "personal" remark about the topic), and making relevant comments on the user's responses – comments intended to show actual understanding of the user. The dialogue manager can also handle some questions asked by user. It continually updates the conversation plan, based on the user's responses. At the top level, LISSA uses a structure called a schema, which contains a list of expected successive events in a dialogue; it allows for actions by both interlocutors and can be dynamically modified by user responses. The schemas are hierarchically structured, allowing LISSA to insert subschemas into the dialogue plan, helping to make the conversation more spontaneous. In order to capture users' inputs, we use the Nuance [27] speech recognizer.

When LISSA's dialogue manager receives input from the speech recognizer, it generates a high-level interpretation in the form of short, explicit, context-independent English sentences which we call "gist clauses". Gist clauses are extracted by applying several pattern transduction trees to the user's input, taking the current LISSA question as the context. In order to facilitate input matching, input words are automatically annotated with syntactic and semantic features before extraction of any gist clauses. Features are recursively attached to input words, such as GOODPRED for words like "happy", and ones like SOCIAL-SCIENCE and (by recursion) ACADEMIC-SUBJECT for "linguistics".

After gist clauses have been derived from the user's input, they undergo a second stage of transduction, producing a reaction for LISSA to output. If the user's input answers a question by LISSA, the reaction will usually be a relevant reaction to that answer; or, if there is a question among the extracted gist clauses (typically at the end of a user's input), LISSA is likely to answer the question. Every gist clause obtained is stored in LISSA's memory so that LISSA won't ask for information already provided by the user in previous turns. Also, the collected gist-clauses could be used for future enhancements that allow inference during the conversation and reference to previous contributions of the participants. Figure 4 provides an overview of the functioning of the system. More details on the dialogue manager can be found in [15]

4 STUDY

Over the past three years, we have been recruiting participants for our studies through the developmental/behavioral pediatric research center at the local medical center. Our experimental sessions have been conducted with nine teenage participants (i.e., between 13 and 17 years old) with high functioning ASD. There were one female and eight male participants. All participants were white and native English speakers. The pediatric research center allowed us to contact the participants who were already receiving behavioral therapy and expressed their willingness to participate in research studies. The diagnosis of the participants were also done by the research center. We made sure that our participants had the ability to read and able to have a conversation.

The goal of this study has been to learn what aspects of LISSA are judged to be useful, and what adjustments need to be made in order to make LISSA a useful tool for iterative conversation training in the lives of teens with ASD anywhere. The sessions with the teens (whose parents were also invited) were scheduled on separate days. Each participant first interacted with LISSA for five minutes, then took a break for two minutes, followed by a second conversation with LISSA for another four minutes. As dialogue topics, we picked ones common in casual conversations such as "getting to know each other", "living in the current city", "crazy room" (aimed at eliciting imaginative responses), "city I want to move to in future", "free time", and "movies". During the conversations, the participants received real-time feedback through the flashing icons. After each conversation, the participants received post-session summary feedback. We conducted an interview with both the accompanying parent and the participant right after the LISSA session. The interview included survey questions on LISSA's usability and capacity for open-ended discussion. The interview was audio recorded and then transcribed by professional transcribers.

5 SURVEY RESULTS

We presented 12 statements to the participants and asked them to specify their opinion ('strongly disagree' to 'strongly agree'). The questions were targeted to understand the usefulness of the feedback and the dialogues. Figure 5 shows the specific questions and percentage of participants' answers in each category. The questions marked with a star (*) were answered significantly more ($p < 0.05$) with options 'agree' or 'strongly agree' compared to other options. We performed a single sampled non-parametric significance test [28] with Bonferroni correction [29] against the option 'neutral'. Participants felt that they were being understood by LISSA. Participants also expressed that they could continue the conversation and pay attention to the icons without any trouble. This indicates that real-time feedback might be applicable to the teenage population. Additionally, participants felt that the feedback they received from LISSA was useful. As can be seen in the figure 5, participants had mixed opinions about several questions, such as whether the conversational experience felt real, and whether LISSA's movements seemed natural. Three participants responded positively about the latter question, four responded negatively, and two were neutral. In the interview session, several mentioned that the lip movement and eye gaze were unnatural. Additionally, LISSA was not responding immediately. This was due to the fact that LISSA processes the

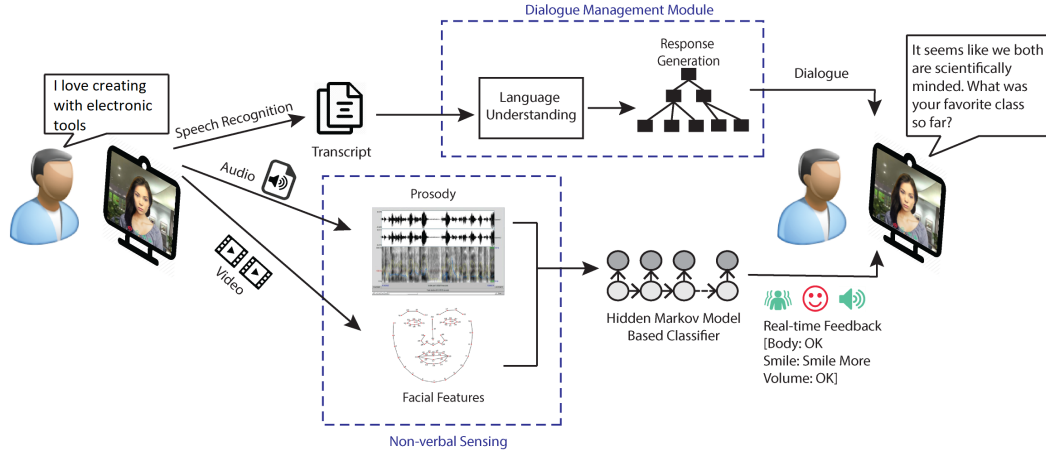


Figure 4: (From left to right) User having conversation with LISSA. System captures the speech, audio and video. Dialogue Management Module generates response. Nonverbal sensing extracts features and feeds to a hidden Markov Model based classifier, which produces real-time feedback in the form of flashing icon.

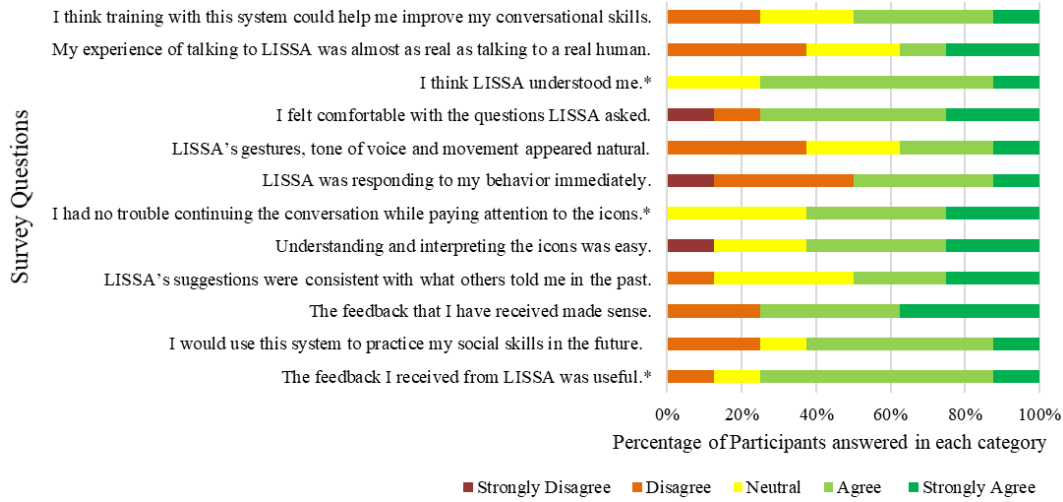


Figure 5: Survey questions and responses.

dialogues and facial features in real-time and the processing takes place on a remote server. In our future versions of LISSA, we will make it more responsive by performing most of the computing locally.

6 QUALITATIVE ANALYSIS

We performed a thematic analysis [30] on the interview transcripts. In the past, thematic analysis was successfully used for user-centered design [31] and rapid online interface prototyping [32]. Additionally, thematic analysis was used for identifying the design guidelines for developing computer and phone-based technology for children with ASD [33]. In our analysis, three researchers performed thematic analysis and then the themes were merged to produce the final analysis report. As a basis for a qualitative thematic analysis of the interview transcripts, we considered individual interviews from

the perspective of the following 14 labels: usefulness, perceptiveness, related systems, accuracy, familiarity, curiosity, realism, speed, appearance, improvements, social, multitasking, uncertainty, and adult identity. Five themes (closely related to some of these labels) stood out to us as relevant to summarizing the experiences of the participants. These are elaborated in the following. We believe it is at least as important to focus on weaknesses as on strengths, as a guide to further development.

6.1 Utility for Practicing Conversation in Private

Participants generally found LISSA useful for practicing conversations. Additionally, they thought that the program was not hard to navigate. One participant said,

"Wasn't that hard to use. It could definitely be used for somebody who really needs help with conversations or for somebody who is not really social or for somebody who is not really the kind of person to be talkative."

Participants liked the fact that the feedback was not coming from a human and they could use it in their private space without being observed. When we asked if they prefer human or a computer for giving feedback, some were ambivalent, some preferred the computer (*"I would rather have that (LISSA) for feedback,"*), and some were skeptical about LISSA.

Some participants with favorable reactions added that the automatic facial feature detection and the accuracy of the feedback were the main reason for endorsing LISSA for conversational skills training. One participant said,

"The fact that it was able to actually detect the facial features and everything being so accurate, I would consider that is good enough to actually train on."

The caregivers, as well, liked the fact that LISSA allows users to have a conversation with a virtual agent instead of (for example) a stranger online. They felt that LISSA was realistic and appreciated that it provided quite complex, and private interactions.

6.2 Self-awareness

When conversing with LISSA, participants were often evaluating the experience in relation to real, social settings, for example, whether the feedback they received was truly appropriate. One participant noted that smiling broadly enough for LISSA to notice might be perceived as inappropriate in a public setting:

"If they can make it so that she can respond a little quicker and that she is able to pick up the smile little better because I was smiling a little bit she just didn't think it was good enough and if I go too big then people are gonna think I am creepy."

Occasionally the self-aware evaluation of LISSA took the form of push-back against the presumption that they needed to improve their behavior, implicit in LISSA's feedback. One participant indicated that any inadequacy in their behavior with LISSA was due to the awkwardness of interacting with a virtual agent:

"Well, I am actually social. I do make good eye contact with other people but it's just kinda awkward you know."

From this it appears that some participants would need some time to become more familiar with LISSA. After multiple interactions they might well feel more comfortable with the system and be more accepting if the purpose of the system were more fully explained.

6.3 Multi-tasking and Feedback

When they were asked about multitasking between the conversations with LISSA and looking at the icons, they wanted more prominent and bigger icons, that is easy to focus on.

"Maybe the icons at the bottom make it a little bigger and maybe make it a little stand out a little bit more."

And when asked about if they preferred feedback from a computer or human, the participants had mixed opinions. For example, one participant said,

"The feedback I received from LISSA was useful. Well it was kind of a bit choppy and kind of pretty much prefer something a bit more realistic."

Interpreting feedback through LISSA's behavior while conducting a conversation may be overwhelming for many participants. Perhaps if LISSA herself provided the feedback verbally during the conversations, rather than relying entirely on icons for feedback, it would feel more like a real interaction to users. Regardless of their preferences in feedback delivery, all participants expressed that the feedback was consistent with what other people had said to them in the past face-to-face training sessions. An important observation was that users broadly agreed on the need for positive feedback. While the flashing red icons noticeably indicated the need for improvement, the reversion to static green icons was not sufficiently noticeable as positive feedback.

6.4 Related Systems

At the end of the interviews, participants tended to compare LISSA to other conversational agents. They were often familiar with Siri and Alexa, and discussed the standard set by these virtual assistants in terms of speed and knowledgeability. They liked that LISSA could detect behavioral features, but hoped that LISSA could better understand and recognize them in the future. For example, a participant said,

"In its current state yes I might mess around with it some but I don't believe I would use it as an actual social skills training and jump into actual conversations just yet."

It seemed that they had high expectations of LISSA which in some cases led to impatience during the conversation. One participant said they would feel more comfortable talking with systems like Siri, that they believe understands them.

"That's like my first time ever talking to a computer, well except for Siri that's different though. That one is fine."

6.5 The Desire for Adult Identity

One teenager, after being urged by a caregiver to be honest, admitted that they probably wouldn't choose to interact with LISSA, unless perhaps LISSA "popped up on their computer". Part of the reason seemed to be that LISSA straddles the boundary between fantasy and reality.

"I don't know if it was like a real person. Cause I do like fantasy things but I am also a kid who but when it comes to talking to people I like talking to real people."

These participants were ambivalent about the future use of LISSA because they were inclined to regard it as a conversational tool for children. One participant said their schedule was too busy, but others could benefit from LISSA.

"Just so you know I am already 17 years old, I am growing up and some of these little kids' things I have outgrown but not all of them."

These comments again suggest the need to make LISSA's purpose clearer to users. It is not intended to be a surrogate human, but rather a tool for repetitive, private practice of conversational behavior.

7 DISCUSSION

7.1 Lessons about Interface Design for ASD Teens

A majority of participants found that LISSA provided useful feedback. They also liked the fact that LISSA would allow them to converse in private. Our interviews helped shed light on how the interface design could be improved. The qualitative analysis of these interviews provide evidence for the soundness of our design so far and grounds for optimism about our further development plans. The analysis also allowed us to formulate several important guidelines for the design of LISSA-like interfaces for conversation practice.

7.1.1 Appropriate Prior Briefing of Users about LISSA's Purpose. Our experience with users made clear that users' assessment of LISSA's behavior and potential utility for conversation practice depended very much on their expectations. They generally acknowledged that LISSA seemed to understand them and responded appropriately to inputs, yet was not genuinely human-like. The perceived shortcomings concerned LISSA's physical behavior, accuracy of perception of user behavior, and depth of knowledge. In part, this perception arose from comparisons with commercial systems such as Alexa and Siri, which have been optimized for smooth functioning in targeted information retrieval and other assistive functions.

These reactions indicate the need for fuller preparation of users about LISSA's purpose: It is not a surrogate human, and it is not an app for access to useful knowledge or personal assistance. It is simply a tool for repeatedly practicing casual conversation for those who feel they could improve in that area. While LISSA has a range of verbal reactions to users depending on their particular inputs, and provides nonverbal feedback as a function of the user's behavior, the conversations are bound to be shallow, and to become more repetitive with multiple uses. Further, LISSA's physical behavior is not the focus; it is merely intended to be sufficiently human-like to make a casual conversation possible. All this should be made clear to potential users – along with a comment that there is no assumption that all users with ASD are lacking in the skills that LISSA is intended to help with. Such preparatory information could be provided both in advance of actual use of LISSA, and as part of LISSA's opening remarks (which already include *"I might sound a bit choppy, but I am still able to have a conversation with you"*).

7.1.2 Positive Acknowledgment of Behavior Change. As noted in the previous section, the participants wanted to be made aware of positive changes. Perhaps flashing green could be used for behavioral improvements. Better yet, the virtual agent could say, for example, "You have good eye contact now". The efficacy of positive feedback and acknowledgment has been observed in past research [34, 35] and our experience further confirms the desirability of positive acknowledgments for interventions aimed at social skills development.

7.1.3 Realistic appearance of a virtual agent and responsiveness. Notwithstanding disclaimers about LISSA's physical behavior, the

issue deserves further attention. A possible reaction to users' comments about insufficiently realistic smiling, eye blinking, and reaction speed might be to back away from the "uncanny valley" (e.g., Chattopadhyay et al. [36]) by using a more cartoon-like avatar. However, this would risk reducing LISSA to a toy in the eyes of potential users. Instead, we interpret the users' comments as urging further development of the avatar towards greater realism. Participants' comments suggest that the more life-like the character, the more likely they are to take it seriously. Realistic appearance of virtual characters has also been shown to be effective in other scenarios such as negotiation, tactical questioning etc. [37].

7.1.4 Conversation personalization would help the teenagers engage for longer terms. Some participants expressed enthusiasm about home-use of LISSA. However, they varied in their opinions on the choices of topics. For example, while the "crazy room" topic struck a chord with some (e.g., they would fill it with video games), others termed it childish. Participants thought that it would be useful if LISSA could talk about topics of their own choosing. For example, one participant was interested in computer programming and wanted to talk about it. The LISSA program, at its current stage, is designed for initiating conversational topics, and treating a specialized topic like computer programming seriously would be a major challenge. However, adding further mundane topics is feasible. Thus we could personalize interactions to a considerable degree by having LISSA choose topics dynamically, skipping those that the user seems indifferent to. Also, choices could be made sensitive to the user's age or maturity. Some of our newly developed topics pertain only to seniors, just as some of LISSA's topics for teens, such as bullying at school, pertain only to school-age users.

Another opportunity for personalization lies in the verbosity of the user. Our experiments showed that while some users provide expansive responses to LISSA, others respond tersely. As the goal of the system is to help users improve their communication skills, the system could gauge users' verbosity and provide helpful feedback where appropriate. For instance, LISSA might encourage laconic users elaborate their answers, or conversely, provide gentle suggestions about curtailing rambling or off-topic inputs. Assessing users' verbosity throughout the conversation may improve turn-taking behavior. Users who tend towards longer answers should probably be allowed slightly longer silences before the turn is seized from them. The same applies to hesitant, slower speakers.

7.2 Limitations and Future Work

The current version of LISSA was not designed for immediate use in a randomized control intervention study. Rather, it is an exploratory system, which will enable a randomized control study after modification and enhancement based on the lessons learned from the trials with the teens with ASD. While our human-centric exploration adds important information on the design of interface and dialogue for a unique application such as autism, running a clinical trial remains part of our immediate future work. Additionally, in the future we will include color-blind friendly color palette for feedback.

Our use of neurotypical individuals to train LISSA's feedback may seem nonoptimal, but we thought that using wizards not trained

as therapists might have led to negative perceptions by ASD participants; furthermore, the wizards might have had difficulty with the cognitive load of managing the computer mediated platform, and providing live feedback to the ASD teens, taking account of the great individual differences in their reactions.

LISSA's dialogue manager was anticipating needs of teens with ASD, with advice from experts. It worked well, but the experiments have shown where improvements are most desirable, for example in topical adaptation to the user, inclusion of direct helpful hints in the verbal reactions to the user, and allowance for different turn-taking styles. Also, while the post-session interviews indicated that the users liked the appearance and voice of the avatar, they saw a need for improvements in the naturalness of the avatar's behavior. In our future work, we will design a customizable interface based on the knowledge we gathered through this study.

Data collected from teenagers with ASD using future versions of the system will help us to further improve the sensitivity and responsiveness of the system. In the current system, the dialogue and the feedback modules are independent. It clearly would be useful to tie the nonverbal feedback to the dialogue content, and to supplement nonverbal feedback signals with direct verbal ones. In the future, we will make the feedback dialogue aware.

8 CONCLUSION

In this paper, we described an interface capable of conducting a multi-topic conversation and provide feedback to help improve the user's overt behavior. The design benefited from the expertise of a pediatrician, psychologists, and a UX designer. We investigated further design desiderata through a study with nine teenagers with ASD. Using a thematic analysis we formulated several guidelines for improved interface design for teens with ASD. In future this knowledge will help guide interface design for this population as well as others with similar needs for social skills enhancement in casual dialogues.

ACKNOWLEDGEMENT

This research was funded by the following grants NSF NRT-DESE Award 1449828, NSF IIS-1543758, DARPA subcontract W911NF-15-1-0542-1.

The team is indebted to late Dr. Tristram Smith for his contributions to data collection for this project.

REFERENCES

- [1] J. Baio, L. Wiggins, D. Christensen, M. Maenner, J. Daniels, et al. Prevalence of autism spectrum disorder among children aged 8 years—autism and developmental disabilities monitoring network. *MMWR Surveillance Summaries*, 2018.
- [2] A. Georgescu, B. Kuzmanovic, D. Roth, G. Bente, and K. Vogeley. The use of virtual characters to assess and train non-verbal communication in high-functioning autism. *Frontiers in human neuroscience*, 8:807, 2014.
- [3] S. Ozonoff, A. Iosif, F. Baguio, I. Cook, M. Hill, T. Hutman, et al. A prospective study of the emergence of early behavioral signs of autism. *Journal of the American Academy of Child & Adolescent Psychiatry*, 49(3):256–266, 2010.
- [4] A. Tartaro and J. Cassell. Playing with virtual peers: bootstrapping contingent discourse in children with autism. In *ICLS (2)*. Citeseer, 2008.
- [5] M. Foster, K. Avramides, S. Bernardini, J. Chen, C. Frauenberger, et al. Supporting children's social communication skills through interactive narratives with virtual characters. In *ACM International Conference on Multimedia*. ACM, 2010.
- [6] S. Bernardini, K. Porayska-Pomsta, and H. Sampath. Designing an intelligent virtual agent for social communication in autism. In *AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment, AIIDE'13*, 2014.
- [7] M. Milne, D. Powers, and R. Leibbrandt. Development of a software-based social tutor for children with autism spectrum disorders. In *Australian Computer-Human Interaction Special Interest Group: Design: Open 24/7, OZCHI '09*, 2009.
- [8] E. Mower, M. Black, E. Flores, M. Williams, and S. Narayanan. Rachel: Design of an emotionally targeted interactive agent for children with autism. In *IEEE International Conference on Multimedia and Expo*, 2011.
- [9] F. Boujarwah, G. Abowd, and R. Arriaga. Socially computed scripts to support social problem solving skills. In *SIGCHI Conference on Human Factors in Computing Systems*. ACM, 2012.
- [10] D. DeVault, R. Artstein, G. Benn, T. Dey, E. Fast, et al. Sensesense kiosk: A virtual human interviewer for healthcare decision support. In *International conference on Autonomous agents and multi-agent systems*, 2014.
- [11] I. Hopkins, M. Gower, T. Perez, D. Smith, F. Amthor, et al. Avatar assistant: improving social skills in students with an asd through a computer-based intervention. *Journal of autism and developmental disorders*, 2011.
- [12] S. Razavi, M. Ali, L. Smith, T. Schubert, and M. Hoque. The lissa virtual human and asd teens: An overview of initial experiments. In *International Conference on Intelligent Virtual Agents*. Springer, 2016.
- [13] A. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, et al. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, 2016.
- [14] K. Fitzpatrick, A. Darcy, and M. Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): a randomized controlled trial. *JMIR Mental Health*, 2017.
- [15] S. Razavi, L. Schubert, M. Ali, and M. Hoque. Managing casual spoken dialogue using flexible schemas, pattern transduction trees, and gist clauses. *Annual Conference on Advances in Cognitive Systems (ACS)*, 2017.
- [16] F. Morbini, E. Forbell, D. DeVault, K. Sagae, D. Traum, and A. Rizzo. A mixed-initiative conversational dialogue system for healthcare. In *Annual meeting of the special interest group on discourse and dialogue*, 2012.
- [17] H. Tanaka, H. Negoro, H. Iwasaka, and S. Nakamura. Embodied conversational agents for multimodal automated social skills training in people with autism spectrum disorders. *PloS one*, 2017.
- [18] I. Serban, T. Klinger, G. Tesaro, K. Talamadupula, B. Zhou, et al. Multiresolution Recurrent Neural Networks: An Application to Dialogue Response Generation. 2016.
- [19] P. Henderson, K. Sinha, N. Angelard-Gontier, N. Ke, G. Fried, et al. Ethical Challenges in Data-Driven Dialogue Systems. 2017.
- [20] R. G. Heimberg. Cognitive-behavioral therapy for social anxiety disorder: Current status and future directions. *Biological Psychiatry*, 2002.
- [21] S. Siegman, A. W. Feldstein. *Nonverbal Behavior and Communication*. Psychology Press, 2014.
- [22] M. Ali, D. Crasta, L. Jin, A. Baretto, J. Pachter, et al. LISSA- Live Interactive Social Skill Assistance. In *Affective Computing and Intelligent Interaction (ACII)*, 2015.
- [23] T. Baltrušaitis and P. Robinson. OpenFace: an open source facial behavior analysis toolkit.
- [24] P. Boersma and D. Weenink. Praat: doing phonetics by computer.
- [25] M. R. Ali, K. V. Orden, K. Parkhurst, S. Liu, V. Nguyen, et al. Aging and engaging: A social conversational skills training program for older adults. In *Intelligent User Interfaces*, pages 55–66, 2018.
- [26] P. Chung and C. Liu. A daily behavior enabled hidden Markov model for human behavior understanding. *Pattern Recognition*, 2008.
- [27] NUANCE: Speech Recognition Solutions.
- [28] H. B. Mann and D. R. Whitney. On a test of whether one of two random variables is stochastically larger than the other. *Ann. Math. Statist.*, 18(1):50–60, 03 1947.
- [29] R. A. Armstrong. When to use the Bonferroni correction, 2014.
- [30] G. Guest, K. MacQueen, and E. Namey. Introduction to applied thematic analysis. *Applied Thematic Analysis*, 2012.
- [31] T. McCurdie, S. Taneva, M. Casselman, M. Yeung, C. McDaniel, et al. mHealth consumer apps: the case for user-centered design. *Association for the Advancement of Medical Instrumentation*, 2012.
- [32] M. Kinzie, W. Cohn, M. Julian, and W. Knaus. A user-centered model for web site design: needs assessment, user interface design, and rapid prototyping. *Journal of American Med Info Assoc*, 2002.
- [33] J. Mintz. Additional key factors mediating the use of a mobile technology tool designed to develop social and life skills in children with Autism Spectrum Disorders: Evaluation of the 2nd HANDS prototype. *Computers and Education*, 2013.
- [34] A. Zeeland, M. Dapretto, D. Ghahremani, R. Poldrack, and S. Bookheimer. Reward processing in autism. *Autism Research*, 2010.
- [35] C. Kasari, M. Sigman, P. Mundy, and N. Yirmiya. Caregiver interactions with autistic children. *Journal of Abnormal Child Psychology*, 1988.
- [36] D. Chattopadhyay and K. MacDorman. Familiar faces rendered strange: Why inconsistent realism drives characters into the uncanny valley. *Journal of Vision*, 2016.
- [37] P. Kenny, A. Hartholt, J. Gratch, W. Swartout, D. Traum, et al. Building Interactive Virtual Humans for Training Environments. (*IITSEC*), 2007.