# Discourse Behavior of Older Adults Interacting with a Dialogue Agent Competent in Multiple Topics

S. ZAHRA RAZAVI, LENHART K. SCHUBERT, KIMBERLY VAN ORDEN, MOHAMMAD RAFAYET ALI, BENJAMIN KANE, and EHSAN HOQUE, University of Rochester, USA

We present a conversational agent designed to provide realistic conversational practice to older adults at risk of isolation or social anxiety, and show the results of a content analysis on a corpus of data collected from experiments with elderly patients interacting with our system. The conversational agent, represented by a virtual avatar, is designed to hold multiple sessions of casual conversation with older adults. Throughout each interaction, the system analyzes the prosodic and nonverbal behavior of users and provides feedback to the user in the form of periodic comments and suggestions on how to improve. Our avatar is unique in its ability to hold natural dialogues on a wide range of everyday topics—27 topics in three groups, developed in collaboration with a team of gerontologists. The three groups vary in "degrees of intimacy," and as such in degrees of cognitive difficulty for the user. After collecting data from nine participants who interacted with the avatar for seven to nine sessions over a period of 3 to 4 weeks, we present results concerning dialogue behavior and inferred sentiment of the users. Analysis of the dialogues reveals correlations such as greater elaborateness for more difficult topics, increasing elaborateness with successive sessions, stronger sentiments in topics concerned with life goals rather than routine activities, and stronger self-disclosure for more intimate topics. In addition to their intrinsic interest, these results also reflect positively on the sophistication and practical applicability of our dialogue system.

CCS Concepts: • **Human-centered computing** → **Human computer interaction (HCI);**

Additional Key Words and Phrases: Animated conversational agents, social skills training, older adults social isolation, nonverbal behavior, schema-based dialogue management

**ACM Reference format:**
S. Zahra Razavi, Lenhart K. Schubert, Kimberly van Orden, Mohammad Rafayet Ali, Benjamin Kane, and Ehsan Hoque. 2022. Discourse Behavior of Older Adults Interacting with a Dialogue Agent Competent in Multiple Topics. *ACM Trans. Interact. Intell. Syst.* 12, 2, Article 14 (July 2022), 21 pages.
https://doi.org/10.1145/3484510



## 1 INTRODUCTION

The population of senior adults is growing, in part as a result of advances in healthcare. According to United Nations studies on world population, the number of people aged 60 years and over is

The reviewing of this article was managed by special issue associate editors Baki Kocaballi, Liliana Laranjo, Leigh Clark, Robert J. Moore, Rafal Ko-cielnik, Q. Vera Liao, and Timothy W. Bickmore.

Authors' address: S. Z. Razavi, L. K. Schubert, K. Van Orden, M. R. Ali, B. Kane, and E. Hoque, University of Rochester, Elmwood Avenue, Rochester, New York, USA, 14627; emails: {srazavi, schubert}@cs.rochester.edu, kimberly_vanorden@urmc.rochester.edu, mali7@cs.rochester.edu, bkane2@u.rochester.edu, mehoque@cs.rochester.edu.



2160-6455/2022/07-ART14 $15.00
https://doi.org/10.1145/3484510

predicted to rise from 962 million in 2017 to 2.1 billion in 2050 [2]. Alarmingly, a huge number of elderly people end up living alone. Almost 28% of noninstitutionalized older persons live alone according to 2017 Profile of Older Americans [1]. Among those, older people who lack high-quality social relationships are at the risk of experiencing social isolation which in turn can affect their mood and well-being [33]. They might lose connection with their friends and find it difficult to initiate new friendships. This can affect their quality of life as it has been shown that social contacts may be as valued as health status in quality of life [27]. On the other hand, social communication skills have been shown to make a substantial contribution to social integration outcomes [82]. Studies show that impaired nonverbal communication can decrease social desirability, the ability to maintain positive and supportive relationships, and as a result decreases the overall quality of life [20, 41, 76].

Spoken dialogue systems have proven beneficial for helping older people with their needs, including social companionship [3, 54], health advice [59], palliative care [87], reminiscence therapy [10], and many other applications. However, many older people find it hard to learn and interact with new technology, which in turn can result in hesitation to try beneficial platforms [78]. With the increasing accuracy of **automatic speech recognition** (**ASR**) and **text to speech** (**TTS**) programs over the past few years, the popularity of voice assistants and conversational agents has grown as well. Moreover, advances in the design of natural virtual agents and social robots can lead to higher acceptance of conversational agents by elderly populations. The ability to have conversations with a virtual agent can help elderly people feel less lonely through providing casual chats, entertainment, story telling, news, and so on. These systems can also provide friendly support to people in managing their everyday tasks and healthcare needs; for example, reminding them about their plans, schedules, and medications. Some research also used conversational systems to help people with dementia and memory disorders, by helping to elicit their memories and achievements [10]. Last by not least, dialogue systems can help older people practice skills, such as communication skills [8]. Such systems provide users with everyday practices and comments and have been demonstrated to help improve social communication in adults [62]. Continued interactions with such systems can be incentivized by providing a reduced cost for continued interactions, a lightened burden of formal appointments, and establishing a setting perceived by users as neither stimagtizing nor judgmental [16]. In many of the mentioned applications, it is crucial to keep users involved in the task for multiple sessions through a human-like virtual agent and engaging conversations.

In this article, we present features of a web-based communication coach adapted for interaction with older adults. This system has been designed and implemented in a collaborative effort between computer scientists and geriatric mental health professionals, called the **Aging and Engaging Program** (**AEP**). The system features a virtual agent named **LISSA** (**Live Interactive Social Skills Assistance**) that holds several sessions of casual conversation with users. Our general approach to the design of this system and its predecessors has been to begin experimentation with **Wizard-of-Oz** (**WOZ**) versions (i.e., with an experimenter remotely choosing verbal system outputs), and then to fully automate the interaction. Previous WOZ and fully automated versions of LISSA have shown success among college students both in terms of improving nonverbal communication [5] and acceptance by users [71].

Our current LISSA system began with a single session WOZ study of 25 elderly participants, which showed 72% accuracy in nonverbal feedback, while user surveys revealed that users generally found the program useful and easy to use [7]. Based on these results and conversation transcripts from the WOZ phase, we designed an extended fully automatic version capable of handling multi-session in-home interactions. We planned for a 10-session intervention, where in each session the participants have 10–20 minutes of interaction with the avatar. Each session consists of three segments where the avatar leads a friendly conversation on different casual topics such as family, leisure activities, life goals, and so on.[1] Throughout the segments, the system processes the audio and video, which it uses to generate feedback on how the user can improve different communication skills. The feedback on eye contact, smiling, speaking volume, and content are displayed and narrated by the avatar at breaks between the segments. At the end of each session, a summary of the user's strengths and weaknesses is displayed to the user; this summary is also provided as a reminder during the start of the subsequent session. The first session was held in a lab, where experts collected information on users and rated their communication skills. We ran the study with 9 participants interacting with the avatar and 10 participants in a control group. The pre- and post-intervention assessment was done by two independent psychologists through role-play settings. Based on the communication skills ratings by the two psychologists, we found that the participants in the AEP group improved their eye contact and facial expressivity. By contrast, the participants in the control group, who also received online communication skills training, did not show significant improvement in their ratings. The dialogue manager was designed using the framework introduced in [71]. Topics and dialogue flow were sketched in a tight collaboration with geriatrician experts based on their experience in elderly therapy. We also used the collected data from the WOZ study to improve the avatar's contribution to the interaction. A comparison of the evaluation results of users' conversations with the automated version of LISSA, with the evaluation results of conversations from the WOZ study, showed no significant difference in quality, implying that a conversation with the automated avatar is of comparable quality to one with a WOZ-based system [72].

In this article, we also study some communicative behaviors of users through the course of the conversations. We focused on content analysis of both user and avatar mostly through language features including elaborateness, sentiment, and self-disclosure. Studying content analysis has proven helpful in increasing conversational agents' effectiveness in a variety of tasks. Content analysis of users' inputs can help in detecting problems such as schizophrenia [22] and dementia [86], can help in improving user satisfaction and perception of dialogue [56, 68], and can let the system show adapting behavior toward users [17]. In order to analyze the conversations' content, we discuss some linguistic aspects of human-machine casual dialogues using the data collected from the experiment. The data comes from 80 sessions of interaction with nine participants over a course of 3–4 weeks. The topics suggested by gerontological experts are categorized into three groups based on their degree of intimacy. We investigate how users' elaborateness, sentiment, and self-disclosure behavior depend on the topic under discussion and the avatar's tone, and how they evolve with time. More specifically, we first look at the level of elaborateness for different users in different sessions and investigate if elaborateness depends on the user's personality and the topic under discussion. We then measure user sentiment over the course of conversation and track its variation over time, and its relation with the topic and user's personality. Third, we evaluate self-disclosure and study its correlation with user mood and personality. We measure elaborateness in terms of users' turn lengths, sentiment in terms of the Vader sentiment analysis tool, and self-disclosure in terms of cues suggested by the literature, and extracted using **Linguistic Inquiry and Word Count** (**LIWC**) categories. We present the insights obtained from the analysis and discuss the results.

[1]Our terminology here is intended to distinguish the types of conversations in the study from others common in AI dialogue systems such as task-oriented dialogues, or question/command-based interaction with systems like Alexa, Siri, and so on. "Casual" conversation (e.g., in social settings) is also a recognized category in sociolinguistics (e.g., Joos, M. (1961), *The Five Clocks*, New York: Harcourt, Brace and World).

## 2 BACKGROUND

Conversational agents have proven beneficial for helping people with their health needs in many different applications [57]. **Spoken dialogue systems (SDS)** are generally considered as the most natural way for social human-robot interaction [29], and the continuous advances of speech technologies have made it possible for people with little technical knowledge to interact with systems using natural language. Employing conversational agents in healthcare can lower the barriers imposed by stigmas associated with psychiatric conditions; for instance, people might be willing to report more symptoms to virtual humans rather than human interviewers [49]. Moreover, virtual agents can be autonomously accessible on a range of devices at users' preferred locations, while people currently often deal with long waiting times due to a shortage of psychiatrists [39]. They also reduce waiting costs for a continued interaction [16].

**Conversational Agents for Older Adults.** Elderly people are among communities who can benefit most from SDS as they provide smooth speech-based communication at the desired time and location of the user. In fact, some studies have shown that older people who are unfamiliar with technology prefer to interact with assistive systems in natural language [25]. Many applications have already been proposed aiming to assist older people using conversational agents. Some systems are designed to play the role of a companion for lonely older users. For instance, "Ryan" [3], a life-like robot, talks with older adults about different topics of interest to make them engage. [54] provides single-session interactions with a robot that reads newspapers for users, asks them some personal questions about their past, and shows nonverbal reactions to the response such as nodding and maintaining eye contact. Such systems report high acceptance among users [90]; users enjoy the companionship and are open to having more sessions with the robot [54, 90], although they don't always believe that the interaction resembles an interaction with a real human [3]. Some older adults prefer keeping robotic pets to alleviate their loneliness. Such pets react to user speech and petting by producing sounds and eye and body movements. Studies proved that such interactions can improve users' communication and interaction skills [83].

Some other conversational agents help older adults in ameliorating loneliness and depression by reminiscence therapy [10, 58]. The agent collects and organizes memories and stories in an engaging manner. In order to model the conversation for appropriate responses, such systems need to possess generic knowledge about events, habits, values, relationships, and so on [58].

Spoken agents can also be virtual home assistants who help users organize their daily schedule. For instance, "Billie" [44] focuses on the specific task of managing the user's daily calendar. The system shows some nonverbal behavior such as gesture, facial expression, and head movement; however, it does not track users' nonverbal behavior. It also cannot manage casual conversation. "Mary" [85] is another virtual agent that assists older adults with daily tasks such as reminders, guidance with household activities, and locating objects using its 3D camera. Although users show high acceptance of such assistants, occasional verbal misunderstanding and errors could cause user frustration [44]. Spoken dialogue systems can also help older patients with dementia [74]. As regular involvement in conversations can help dementia patients, [11] designed chatbots for open-ended conversation on a limited number of topics and received positive feedback from users. However, their systems provided a text interface only, with no human-like avatar.

Virtual assistants have also been proposed to gather users' health information (such as blood pressure and exercise regimen) and provide health advice based on the collected data [60]. An approach to helping users maintain their mental acuity is to develop dialogue systems that engage users in skill-dependent activities such as cognitive games. Other practices can help people improve their communication skills, potentially alleviating social isolation [79]. Yet another interesting application of dialogue systems is in palliative care—helping people with terminal illnesses reduce stress by steering them toward topics such as spirituality, or their life story. In [87], an experiment was performed with 44 participants, where users interacted with an avatar by selecting their response from multiple options. They interacted about death-related topics such as last will, funeral preparation, and spirituality. Although the avatar displayed variable posture and hand gestures, it did not process any user behavior and did not actually attempt to extract meaning from user inputs. Nevertheless, users were satisfied with their interaction and were ready to have more sessions with the virtual agent. Chatbots are also suggested to help people with self-diagnosis. A data-driven analysis [26], from more than 47K consultation sessions in the course of 6 months, showed success in helping people and led to suggestions on how to make the chatbot more trustworthy and easy to use.

Past studies have helped to identify important features for effective interaction of virtual agents with older users. A system designed for individuals with reduced cognitive abilities needs to be easy and intuitive to use. It also needs to be able to recover from mistakes and misunderstandings and resume normal interaction [78]. As older people might be more prone to pauses and hesitations, the system should allow for such intermittency, and rather than issuing error messages, should ask for repeating or rephrasing [43]. Virtual agents also should have a likable appearance and persona. Ideally, systems should also be able to personalize their behavior according to the specific needs and preferences of each user [78]. In order to act more appropriately, some studies introduced methods to infer users' personality traits from their input utterances [4, 92]. Users' perception of a virtual agent and their willingness to confide in it has been shown to be influenced by their personality [92].

The system we present in this work is designed to help older people who are at risk of losing their connections because of impaired communication skills. Studies show that social isolation can affect not only mood and well-being, but also can influence health outcomes such as worsening disability, illnesses such as chronic lung disease, arthritis, impaired mobility, depressive symptoms, and death [33, 66, 81]. By contrast, favorable social conditions such as high support from family, friends, and social groups are predictors of sustained health [84].

One way to overcome the feeling of loneliness is to improve social integration by boosting social communication abilities [82]. Some studies show that social skills deficits correlate with both physical and mental health [77]. One of the key aspects of communication skills is nonverbal behavior. Appropriate facial expressions, smiling, eye contact, and other forms of nonverbal behavior can yield a positive first impression, help the flow of conversation, and provide meta-communication [35, 51]. Thus, many researchers have examined the use of computer-based interventions to help people improve their social skills. For instance, [9, 36] provided a simulated job interview environment using a virtual agent; EmpaT [31] is a scenario-based, serious game simulation platform in the context of job interviews allowing users to confront a variety of uncomfortable situations such as talking to unfriendly people. [67] let medical students practice their communication skills with a virtual agent that acts as a patient. Virtual agents have also proved useful and acceptable in helping people with **Autism Spectrum Disorder** (**ASD**). ASD participants who practiced with virtual agents and coaches showed improvements in verbal and nonverbal behavior, and also indicated interest in continuing [13, 67, 70, 88]. To the best of our knowledge, not much work has been done addressing nonverbal behavior of older adults. One prior study [62] presented a psychosocial intervention to improve everyday living skills of older patients with chronic psychotic disorders, which led to improvements in communication. However, the program was focused on basic skills for adults with severe mental illness and as such is not extendable to a broader community of elderly users. Also, the program was not accessible for use ubiquitously. The AEP presented here uses a virtual agent not as a companion but as a communication skills coach. We present the program and results of the analysis of experiments.

**Content Analysis.** Content analyses of dialogues with non-task–oriented **conversational agents** (**CAs**) has proven helpful in increasing CAs' effectiveness in a variety of tasks. For instance, detecting the main themes of a dialogue, or speech and language features, could assist in detecting schizophrenia [22] and dementia [86], and preventing suicide [52]. In the context of intelligent tutoring systems such as AutoTutor [24], affective states can be predicted using temporal and lexical cues from natural language interaction in the learning environment. Studies showed that users' affect and emotion can impact their interest, self-efficacy and understanding during interactive learning sessions [18, 42, 47]. Studies using spoken dialogue systems also show that different aspects of communication style can affect users' satisfaction and perception of dialogue [56, 68]. Moreover, models of speakers learned from their language, interests, opinions, and communication styles can enable dynamic adaptation of system behavior and content to users, as is done in [17].

Our research investigates the detection of affective states that arise during interactive dialogue in natural language. The use of dialogue to detect affect in learning environments is a reasonable information source to explore initially.

In a system like ours, understanding users' communication behavior in social conversation can be a benefit in multiple ways. Firstly, we need to find ways to increase user engagement and satisfaction. Studies show that some communication behaviors such as reciprocity and self-disclosure are strong predictors of relationship building and user satisfaction [45]. Secondly, sentiment-adaptive responses have shown to be effective in users' perception of virtual agents. [23] show that users find agents with sentiment-adaptive responses as being more empathetic and human-like, and as a result more satisfying to communicate with. In general, personalizing the interaction between users and virtual agents has been one of the most significant engagement strategies [19]. This has been done both through adapting nonverbal behaviors such as voice and gestures [14, 21, 53] and adapting the content based on users' preferences [50] or emotions [28]. In this work, we focus on analyzing some users' communication styles and language features including elaborateness, sentiment, and self-disclosure. Elaborateness refers to the amount of additional information provided by a speaker [68]. Experiments with a spoken dialogue system in a healthcare application show that a user's elaborateness can be classified using just the dialogue act category of an utterance and the number of words it contains [55]. We also study the role of sentiment, since the use of sentiment features has been observed to increase the quality of conversational agent output [73], and allows for more human-like social conversation [15] and more appropriate recommendations [89].

Another important aspect of conversational content we investigate in this article is the degree of self-disclosure. Encouraging self-disclosure can increase rapport in user-CA interaction [46, 64] and thereby the effectiveness of the virtual agent in different tasks (e.g., health coaching [48]). Self-disclosure by a user can also assist in evaluating mental health [40].

While sentiment can be classified using analysis tools such as Vader [38], the set of features indicative of self-disclosure remains ill-defined. [69] suggested utterance length, negation words, **part-of-speech** (**POS**) tags, and emotion-laden words as self-disclosure markers in an open-ended conversation with a chatbot. [37] identifies personal pronouns, word count, and family and sexual words as significant, based on comparing secret tweets with normal tweets. [12] observed that tweets with deeper self-disclosure are apt to mention secretive wishes or sensitive information, while moderately self-disclosing tweets convey general information about self such as family, education, and so on [80], shows that not only linguistic features but also some nonverbal behaviors such as nods and speech pauses are associated with self-disclosure. In our work, we tried several **Linguistic Inquiry and Word Count** (**LIWC**) categories, based on the cited literature.

## 3 THE LISSA VIRTUAL AGENT

LISSA is a virtual agent designed to help people improve their communication skills [5]. The avatar maintains open-ended conversations with users and offers feedback on their nonverbal behavior, so that users can try to improve their social skills. We have noted the desirability of human-like agents for such open-ended, realistic conversations. Our avatars were obtained from Sitepal (https://www.sitepal.com/), which are reasonably realistic in terms of appearance, motion, and speech output. The particular avatar chosen was agreed upon by a focus group of 10 older adults who considered various versions of the virtual agent varying in gender, age, and realism. As far as we can tell from observation of the interactions and feedback from participants, LISSA's degree of realism (also in previous versions) was satisfactory to the participants, without any issues arising such as the putative "uncanny valley." (In some of the earlier studies, there were some criticisms of insufficient realism, such as unnatural eye-blinking [6].)

### 3.1 Nonverbal Feedback

The feedback provided by LISSA on users' nonverbal behavior includes eye contact, smiling, speaking volume, and one verbal feature—the emotional valence of the user's utterances. These behavioral features are detected automatically, and feedback can be provided in real time through icons on the screen. The program also displays charts and texts summarizing the user's performance after each interaction session. More specifically, the automated feedback system was designed and trained using 506 minutes of labeled video collected from previous studies [5]. Each conversation consists of three sessions; during each, the system captures audio and video and uploads them to the server. Then smile intensity, volume, and eye gaze direction are extracted from the video and audio. Using the previously trained model, a Hidden Markov Model–based classifier classifies the nonverbal features into two (positive and negative) categories. The audio collected from each session is transcribed and a sentiment analysis model is performed to detect the valence of the text. Both audio and video feedback are given to the user after each session in the form of simple English sentences, designed to reinforce positive results and make suggestions for remedying weaknesses. More details on how the automated feedback system works can be found in [8]. For some user groups such as teens with autism and older adults, real-time feedback sometimes proved cognitively challenging. Divided attention is an aspect of cognition that is required when processing two or more sources of information. This aspect is shown to be associated with significant age-related declines in performance, particularly for complex tasks [32]. Other studies showed that the slowdown in performance is greater compared to younger ones, when attention must be switched between several tasks [91]. As a result, to alleviate the load on such users, we removed the screen feedback icons, which requires processing multiple icons and modifying behavior accordingly throughout the whole conversation with the program.

### 3.2 LISSA Dialogue Manager

In order to engage users in a smooth natural conversation, we designed a dialogue manager that can handle open-domain chatting on a number of everyday topics. The **dialogue manager (DM)** initiates topics and follows a coherent plan to lead the conversation and respond to users' inputs with appropriate comments and questions. The dialogue plan is a flexible, modifiable schema [75] consisting of a sequence of expected actions by the system and user, dynamically instantiated in the course of the conversation. At each user turn, the DM interprets the user's input in the context of LISSA's previous utterance, using pattern transduction hierarchies relevant to that context. The interpretation consists of one or more simple explicit context-independent English sentences called "gist-clauses." The gist-clauses are then used by response pattern transduction trees to construct

***LISSA***: ... Tell me a bit about your day today. For example, what did you have for breakfast?

***User***: Well I had a granola bar, a banana and a cheese stick. I didn't take the time for cereal and milk which is my usual breakfast. Then I came over here to this to this session and that's been my day so far. I didn't have any trouble with traffic.

***LISSA***: I try to have fruit for every breakfast. It's very healthy for me. Speaking of food, what is your favorite flavor of ice-cream?

***User***: Oh, gosh... I like all sorts of ice-cream. Maybe... maybe coffee ice-cream would be my favorite although chocolate is right up there too.

***LISSA***: I think coffee ice-cream is pretty tasty. Though I am big fan of mint chocolate chip. Do you have a favorite food? What is it and why do you like it?

***User***: Oh fruit. I eat all sorts of fruit. Umm... I suppose the favorite is grapefruit. Uh but I try to eat a banana, banana every day or almost every day.

***LISSA***: I like grapefruit. It's nice when food is healthy and tastes good.

How did get here today? Did someone drive you or did you take the bus?

***User***: No, I drove my car here today. It's a good day to be out driving.

***LISSA***: That's good that you drive. It must be nice to have the freedom to go somewhere when you want to.

***User***: It is.

***LISSA***: Let's pause here so I can give you feedback.

Fig. 1. An example of dialogue between LISSA and a user.

an appropriate response. An example of a conversation between a user and LISSA is shown in Figure 1. More details on the dialogue manager can be found in [71].

## 4 PREVIOUS LISSA STUDIES

A previous WOZ version of the system demonstrated success in helping college students improve their nonverbal behavior [5]. This system provided real-time feedback through icons signaling changes in different aspects of nonverbal behavior. Participants were also shown charts and figures summarizing their performance. An experiment with 47 students with this prototype system showed improvement in the nonverbal behavior of those who practiced with the system. Another experimental study, involving teens with ASD interacting with a fully automated system, showed

Fig. 2. "Aging and Engaging Program" Virtual Agent.

that some users with more severe cognitive impairment might face challenges with real-time icons [6], as already noted. However, participants found the system interesting and were willing to have further sessions [70]. The AEP was designed on the basis of lessons learned from previous versions of LISSA, and in tight collaboration with an expert advisory panel of professionals (at our affiliated medical research center). The goal of the AEP was to help older patients who are at risk of isolation and depression, by providing them a means to practice their social skills. In order to study the feasibility and acceptability of such system on the target community, a WOZ experiment was run with 25 older adults. The results showed that users found the program useful, intuitive, and fairly accurate [7].

During interviews with users after the experiment, participants talked about their experience with the program. The experience was novel for all of them; some reported uneasiness at the beginning but they got comfortable soon after. They reported the ways in which the program was easy or difficult to use. While only half of the participants found the feedback accurate, most thought the program could help improve conversational skills. The effectiveness of the system (i.e., improving social skills) was not assessed through the WOZ study, but was instead left for the automated multi-sessions setup. We used both verbal and nonverbal data collected from the WOZ experience to improve the program for the automated version.

## 5 LISSA MULTI-SESSION AGING AND ENGAGING PROGRAM

Based on the feedback and content collected from users during the first AEP experiment with WOZ, we designed a program for longer-term practice by users. The AEP program is fully automatic and was available to users on their personal devices so that they could practice at their convenience.

### 5.1 Setup

We designed the Web-based automated version of the program to conduct multiple sessions of conversation with a user. Participants access the AEP system via their web browser that can be accessed with minimal assistance in the home (Figure 2). Web-based technology allows individuals to access the program from anywhere using internet enabled computers. This is extremely powerful considering the fact that many older adults are unable to drive to those places that offer in-person training. Additionally, web-based technologies are scalable to a wider population, while avoiding social stigma and minimizing burdens associated with in-person interventions.

We asked participants to use a laptop or personal computer. We also offered a lab laptop to those who had no personal computer. They needed to have a microphone and webcam in order to be able to complete the sessions. They could complete the sessions at their time of convenience. They used a link to access the program web page, where they start interacting with our interface by pressing the start button. The first interactions were held in the lab and the rest were self-initiated by users at home. The issues raised include not being able to start the Web browser (Google Chrome), allowing the Webcam to access the browser, and enabling flash player. They could contact the technical team through phone or email at any time, and the team would help them deal with the issue. Each session consisted of three conversation topics and feedback was delivered between the three segments. Much as in previous versions, the feedback contained positive and negative reinforcement concerning the usage of eye contact, smiling, speaking volume, and content. Each session ended with a summary feedback where the performance in the three segments was discussed. Specifically, the feedback indicated where the user made improvements, which areas needed more focus, and ways to improve them. The summary was presented using simple text read aloud by the avatar. Moreover, some suggestions were delivered to users at the end of each session on how to overcome their weaknesses and maintain their strengths.

After completing at least eight sessions at home, the participants were invited back to the lab for a follow-up session. In the lab, participants were asked to fill out surveys and were evaluated for their communication skills by experts.

To assess the viability of the AEP system, we conducted a randomized control study with 20 participants recruited through community advertisement and outpatient geriatric psychiatry clinics. All the participants experienced at least mild difficulties on the social skills. We excluded applications with no access to an email address or Internet in a private location. One participant in the AEP group withdrew from the study after the first visit because she was not interested in the study. Participants were all at least 60 years old (average age = 71), including 13 females and 6 males, where 8 were married and almost half the participants lived alone. All participants in this study possessed a home computer or a laptop. Participants had a wide range of depression and anxiety symptoms and nonverbal impairments. Prior to the study, the participants were assessed by two clinical psychologists using **Social Skills Performance Assessment** (**SSPA**) measurement. SSPA is an observer-rated assessment of verbal and nonverbal behavior in social communication, obtained through completing roles in standardized role-plays [61, 63]. The results indicated that among 19 participants, the most common impairments were facial expressivity ($n = 10$), eye contact ($n = 8$), and lack of gesture ($n = 12$). All participants also completed self-report assessments to characterize the sample: PROMIS computerized adaptive tests for depressive and anxiety symptoms [30]. The reports showed significant variability in severity of depression symptoms (T-score range 34.20-65.80) and anxiety symptoms (T-score range 32.90-65.40), indicating our sample included individuals with moderate symptomatology. More details can be found in [8].

Participants were randomly divided into two groups: a control group ($n = 10$), and an AEP group ($n = 9$). The control condition involved reading information about improving conversation skills provided on the Web with videos (with no feedback or engagement with AEP). The pre- and post-intervention assessment of communication skills was performed in the lab where each participant interacted with a psychologist (blind to participants' condition) in a role-play setting. An independent psychologist observed the entire role-play session and both psychologists then rated the participants on their eye contact, facial expressivity, speaking volume, and content. During the first session in the lab, we also ensured that the participants could complete AEP and/or control at home smoothly. The AEP group completed at least eight sessions of the AEP, while the control

Table 1. Dialogue Topics and their Emotional Intensity Level
(E: Easy, M: Medium, H: Hard)

| | Subsessions Topics | EI |
|---|---|---|
| S1 | Getting to know each other (I, II), Activity | E,E,E |
| S2 | City you live in (I, II), Pets | E,E,M |
| S3 | Family, Gathering with family and friends, Yourself | E,M,H |
| S4 | Weather, Driving, Cooking | E,H,E |
| S5 | Outdoors, Travel, Plan for today | M,M,E |
| S6 | Chores, Money, Growing older | E,M,H |
| S7 | Education, Job, Life goals | M,M,H |
| S8 | Technology, Books, Arts | M,M,M |
| S9 | Sleep, Health, Exercise | M,M,M |

group was directed to educational materials on the Web to improve conversation skills, both over the course of 4–6 weeks.

To evaluate the impact of the system on users' verbal and nonverbal behavior based on SSPA ratings, we utilized two linear regression models (for verbal impairment and nonverbal impairment), with the condition (treatment and/or control) as the primary predictor and baseline scores on the SSPA as a covariate. Comparing the SSPA scores indicated that the participants randomized to AEP improved their eye contact and facial expressivity significantly, while results were nonsignificant for verbal impairment. Participants in the control group did not show improvement in any of the skill ratings. More details on how we evaluate the impact can be found in a previous paper [8].

### 5.2 Dialogue Content

To allow users to remain engaged in an interaction with the program, the virtual agent needs to have smooth, meaningful conversations. The agent leads users in casual conversations controlled by an automatic dialogue manager. As mentioned, each interaction consists of three subsessions, along with sporadic self-disclosures by LISSA.

The character of LISSA has been carefully designed to represent a 65-year-old widow who moved to the city a few years ago to live with her daughter. We followed a participatory design methodology while selecting the persona of the virtual agent. Specifically, we had a focus group of 10 older adults who looked at various versions of the virtual agent including different gender, age, and realism. From their opinion on different characters and features of virtual agents, we selected the one presented in the article. The choice of topics, their ordering, and potential questions and comments and other relevant statements provided by LISSA, were all designed meticulously during several sessions of consultation with gerontologists with expertise in interventions. In each subsession, LISSA opens the topic by mentioning some thoughts and experiences of her own and asks relevant questions from the user. Upon receipt of an input from the user, LISSA makes relevant comments, responds to user's questions (if applicable), and smoothly moves to the next question.

The automated version of the AEP system was designed to support nine sessions of interaction with users (Table 1). Each of the three segments of a session covers a specific topic, usually led by LISSA by asking three to five questions. Each complete session takes 15–20 minutes depending on the number of questions and the user's elaborateness. In collaboration with geriatric mental health professionals, we collected 27 topics from everyday life known to be of interest for the target community. A list of the chosen topics is shown in 1. To better plan for a long-term interaction, the geriatric experts divided the topics into three groups based on their emotional intensity or degree of intimacy: easy, medium, and hard. According to [34], emotional intensity refers to

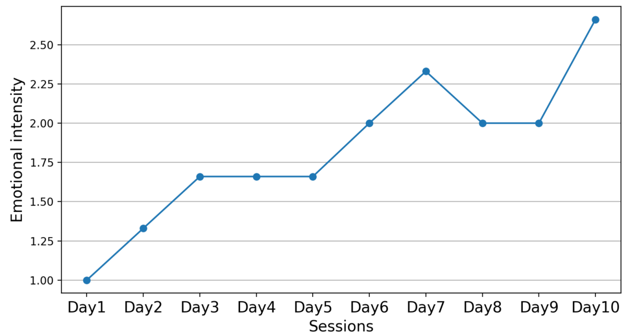


Fig. 3. Change in emotional intensity of the sessions over time.

variations in the magnitude of emotional content. We assign a lower emotional intensity score to less intimate topics (easy = 1) and a higher score to those topics which are more emotionally intense and intimate (medium = 2, hard = 3). We calculate the average value for different days. Easy topics such as "Getting to know each other" and "family" are ones likely to be broached in making someone's acquaintance, while the harder ones such as "life goals" and "getting older" are more emotionally evocative and call for more self-disclosure. We designed the sessions so that LISSA starts with easier topics at the earlier sessions and gradually moves to more intimate ones, as the user becomes better acquainted with the avatar and feels more comfortable to share thoughts and interests with her. Figure 3 displays graphically how the level of intimacy changes throughout the study. During all sessions, LISSA's interjection of her own beliefs and life experiences helps to make the conversation natural and to evoke user responses.

## 6 DIALOGUE CONTENT ANALYSIS

As mentioned before, content analysis of open-domain dialogues with conversational agents can be helpful in increasing the system's effectiveness and performing other health-related evaluations. We collected data from the multiple-session experiment with AEP explained in 5. Each of the nine participants who interacted with LISSA had seven to nine sessions. In total, we collected dialogues from 72 interaction sessions, most of them held at users' homes. In this section, we analyze three aspects of the dialogue content. The first concerns elaborateness, where we looked for differences in elaborateness across different sessions, users, and topic classes; we also analyzed changes in elaborateness over time.

The second concerns the results of sentiment analysis for different sessions and the tone change over time. The final aspect concerns the kinds of self-disclosure cues we gleaned from the literature.

### 6.1 Data Statistics

We collected the transcripts from the nine users interacting with the system over 7–9 days. The transcripts are taken from the output of the system's ASR software. A few subsessions were omitted due to technical issues. Table 2 summarizes the collected data. Each interaction session consists of three subsessions, each on a topic. Verbal and nonverbal feedback is provided after each subsession and also at the end of a whole session.

### 6.2 Elaborateness

Following [55], our metric for utterance length is the word count of the user utterance. The results show that users on average tend to provide longer responses as they proceed in a conversation.

Table 2. Collected Data Statistics

| Feature | Number |
|---|---|
| users interacted with the avatar | 9 |
| sessions | 72 |
| subsessions | 198 |
| total users' turns | 668 |
| total users' words | 29,054 |
| total avatars' words | 24,296 |

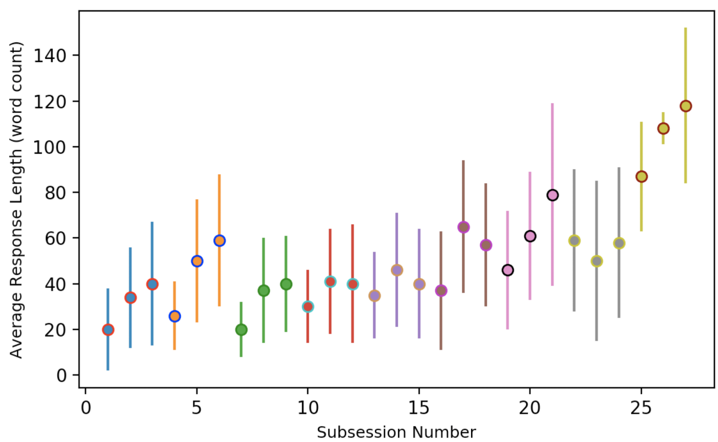


Fig. 4. Users' average turn length in each subsession.

Figure 4 shows the average response length among all users in different subsessions. We also observe a strong, significant correlation (a) between the average word count and the particular subsession (Pearson $r = 0.76$, $p < 10^{-5}$); (b) between the average word count and the user's turn number in the complete interaction ($r = 0.68$, $p < 10^{-12}$); and (c) between the average word count and the interaction number ($r = 0.81$, $p = 0.008$). Trends (b) and (c), however, are not the same for all individuals. For five out of the nine users, the turn length correlation with time is significantly strong, while for the rest we cannot see any significant correlation.

*6.2.1 Users' Turn Length and Topic Classes.* The various topic classes significantly affect users' response length. The average among all users shows that users provide longer responses to "hard" questions, where the average is 57.60($\sigma = 33.92$) words, while responses to "medium" and "easy" questions contain an average number of 51.41($\sigma = 30.39$) words and 32.73($\sigma = 24.03$) words, respectively. Interestingly, the response length change over time is not significant for easy topics but it is significantly strong for medium ($r = 0.81$, $p < 10^{-3}$) and hard ($r = 0.94$, $p = 0.05$) topics.

*6.2.2 User and Avatar Turn Length.* Some studies suggest that the utterance lengths of one speaker can influence the interlocutor's utterance lengths. We looked for any correlation between the avatar's input length and users' corresponding turn length, but did not observe any meaningful relation.

## 6.3 Sentiment

We used VADER [38] to quantify utterance sentiment for each avatar and user turn.

Table 3. Average Sentiment Score for Topic Classes

| **Topic class** | $ave(\|Sent_a\|)$ | $ave(\|Sent_u\|)$ |
|---|---|---|
| Easy | 0.3 ($\sigma = 0.31$) | 0.43 ($\sigma = 0.35$) |
| Medium | 0.36 ($\sigma = 0.3$) | 0.62 ($\sigma = 0.31$) |
| Hard | 0.38 ($\sigma = 0.32$) | 0.63 ($\sigma = 0.3$) |

*6.3.1 User vs. Avatar Turn Sentiment.* The correlation coefficient value shows a weak but significant correlation between a given user turn and the avatar's preceding turn ($r = 0.23$, $p < 10^{-8}$); this suggests a slight dependence of the user's sentiment on the avatar's tone (though both might be derivative from the particular question content). In order to compensate for the possible influence of the avatar's tone on the user, we studied sentiment difference over time ($Sentiment_{user} - Sentiment_{avatar}$). We observed a significant weak increase in positive tone over time ($r = 0.35$, $p < 10^{-3}$).

*6.3.2 Sentiment for Different Topics.* A more careful look into different interaction sessions provides some insight into the relation between user sentiment and dialogue topics. We should first note that the avatar is designed to convey a positive, friendly tone in its interactions, thereby encouraging a generally positive tone on the user's side. However, we find user sentiment to be significantly more positive for some topics than others. Among them are "Travel," with sentiment score = $0.74(\sigma = 0.22)$, "Health," with sentiment score = $0.76(\sigma = 0.13)$, "Education," with sentiment score = $0.74(\sigma = 0.07)$, and "Outdoor," with sentiment score = $0.68(\sigma = 0.28)$. On the other hand, in talking about subjects such as "Family," "Getting to know each other," and "Managing money," participants tended to be more neutral, with respective average sentiment scores of $0.045(\sigma = 0.18)$, $0.19(\sigma = 0.38)$, and $0.32(\sigma = 0.39)$.

We infer that topics concerned with life goals evoke stronger emotions than those concerned with routine activities of daily life. As well, discussion of eventualities such as the death of a partner or living alone after others have moved out naturally leads to a more negative emotional tone. There are other themes that evoke both negative and positive user comments, and hence sentiment fluctuations resulting in a high standard deviation and no meaningful average. An example is the topic "Growing older" with sentiment score = $0.37(\sigma = 0.50)$.

*6.3.3 Sentiment for Different Topic Classes.* We also studied the average sentiment for the three topic classes introduced in Section 5.2. Our hypothesis was that emotionally evocative topics produce stronger user sentiment than more neutral ones. We therefore evaluated the average absolute sentiment value across all users for different topic classes. The results can be seen in Table 3.

The results show that although the avatar's tone remains almost the same for all classes, users tend to use stronger tones when they talk about "medium" and "hard" topics compared to "easy" ones.

## 6.4 Self-Disclosure

Under this heading, we focus on sessions mainly concerning users' lives, beliefs, interests, and so on, expected to elicit some degree of self-disclosure. The goal is to gain insight into the dependence of self-disclosure on different topics. As mentioned earlier, there is no well-defined set of cues for measuring self-disclosure, but various studies have suggested some potentially significant ones (recall Section 2). We instantiated these as follows, relying on LIWC features [65]: We employed LIWC features in measuring word counts in the following feature categories: (1) word count per turn, (2) first person pronoun, (3) family and friends, (4) negative emotions (anxiety, anger,

Table 4. LIWC Score of SD Cues for Three Topic Classes

| Potential SD Cues | Easy | Medium | Hard |
|---|---|---|---|
| Word count per turn | 31.97 | 49.61 | 55.45 |
| First-person pron. | 9.91 | 9.29 | 9.46 |
| Family and friend | 1.02 | 1.08 | 1.03 |
| Negative emotion | 0.51 | 0.54 | 0.84 |
| Positive emotion | 4.91 | 4.9 | 5.54 |
| Drives | 5.71 | 6.82 | 7.26 |
| Personal concerns | 5.2 | 7.07 | 4.15 |

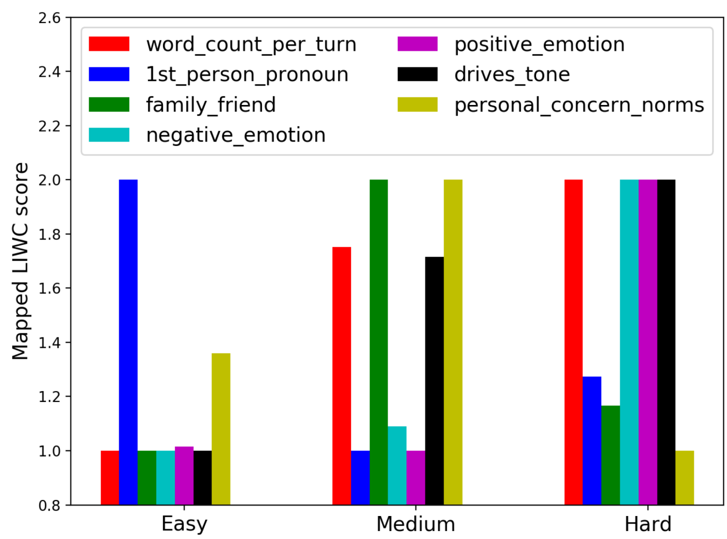


Fig. 5. Self-disclosure cues for three topic classes, mapped to interval [1, 2] for better visualization.

and sadness), (5) positive emotions, (6) drives (affiliation, achievement, power, reward, risk), and (7) personal concerns (work, leisure, home, money, religion, and death).

We first report the LIWC-based scores of the above features in the three topic classes in Table 4. To make the comparison more vivid, we linearly map the scores to [1, 2] for each category independently and plot a bar graph (Figure 5). It can be seen that the "hard" topics contain more words per turn, and more negative and positive emotions and drives. On the other hand, people use personal pronouns more often in easy topics such as when they introduce themselves or talk about their activities. Conversation about family, friends, and personal concerns, though somewhat intimate, need not involve high self-disclosure.

We also make a list of topics with the highest LIWC category scores. As can be observed in Table 5, participants used the most first-person pronouns in the initial greeting session and in talking about themselves and their families. Family and friend words not surprisingly were used in "Family" and "Gathering" sessions but also when the topic concerned "Cooking." "Growing older" is among the topics where people use the most negative emotion and personal concern words.

## 7 DISCUSSION

Analysis of users' conversational behavior in the course of multiple social dialogues with a virtual agent led us to some interesting observations, which we briefly summarize in this section. First,

Table 5. Topics with the Highest LIWC Category Scores

| Feature | Highest score sessions |
|---|---|
| First-person pron. | Getting to know, Yourself, Family |
| Fmly/Frnd | Gathering, Family, Cooking |
| Neg. emot. | Driving, Growing older, Money |
| Pos. emot. | Yourself, Weather, Outdoors |
| Drives | Gathering, Life goals, Arts |
| Pers. concern | Growing older, Activity, Family |

concerning the effect of topical choice on users' conversational behavior, we observed that when topics are more intimate, as in the case of life goals and the challenges of getting older, users tend to elaborate more. They also use stronger emotion words—both positive and negative—when talking about more intimate and emotionally intense topics. Furthermore, we observed that the average response length increases as users progress through the series of interactions. This suggests that users feel increasingly comfortable in extended interactions with our virtual agent. We also observed a slight dependence of the user's sentiment and the avatar's sentiment. Furthermore, our study of self-disclosure features in the interactions showed that users' comments on emotionally less intense topics such as their everyday life correlated with more use of personal pronouns.

These results from the conversational content analysis support the use of dialogue agents for dialogue practice with older adults, even when touching on potentially difficult conversation topics. Indeed, our participants were more engaged with the agent when the conversation topics were more emotionally intense and intimate. Given the importance of effective communication during challenging conversations in later life—driving cessation, healthcare, and end-of-life decision-making—our findings suggest that dialogue agents like LISSA could provide valuable practice and coaching to help older adults successfully navigate these challenging conversations and thereby improve both health and quality of life. Thus, our studies have provided some insights into users' behavior in multi-session casual interaction with automated virtual agents, and some potential guidance in further implementation of effective social agents.

## 8 CONCLUSION

We have presented a conversational agent designed to provide realistic conversational practice to older adults at risk of isolation or social anxiety. The system can handle multiple sessions of interaction with users, engage users in casual conversation, and provide feedback on their nonverbal behavior. Users can access the program through a web browser on their personal computers at their place and time of convenience. The topics of the dialogues and the avatar's contribution to the dialogues have been meticulously designed in collaboration with geriatric mental health professionals based on their experience in therapy sessions. We summarized the data collected from 72 sessions of interaction with nine individuals, studied their dialogue behavior, and inferred sentiment focusing on three aspects: elaborateness, sentiment, and self-disclosure. The naturalness of the interactions, generally attested by the users [6, 72], indicates that our results are meaningful. During the COVID-19 pandemic it became very challenging for older adults to engage in face-to-face interaction with others. Our tool serves as a crucial alternative for continuing behavioral therapy and maintaining social connections.

Although the AEP showed significant impact in changing nonverbal behaviors, there might be adverse consequences when such a system provides incorrect feedback. For example, the AEP can erroneously give negative feedback even when the users have adjusted their behavior. This could demotivate users and create cognitive pressure. We understand that these types of errors are unavoidable for AI-driven systems, and we made it clear to each of our participants before starting the study not to take the feedback too personally. However, it is a distinctive feature of our system that its behavior is generally appropriate as judged by users. This is achieved by the techniques we employ to ensure robust handling of a considerable variety of topics. For example, **Speech-to-Text** (**STT**) transcription inaccuracies are dealt with robustly with our pattern transduction methods, and our use of "gist-clausesâĂİ derived from inputs as a basis for formulating responses largely ensures that system outputs will be relevant and appropriate.

Larger studies, and branching out to other age and culture groups, will be needed to gain a fuller understanding of user behavior in such settings, and to make inferences going beyond correlations to causal analyses.

## REFERENCES


[1] 2017. Profile of Older Americans. (2017). https://acl.gov/sites/default/files/AgingandDisabilityinAmerica/2017OlderAmericansProfile.pdf.

[2] 2017. World Population Prospects: The 2017 Revision. (2017). https://esa.un.org/unpd/wpp/Publications/Files/WPP2017_KeyFindings.pdf.

[3] Hojjat Abdollahi, Ali Mollahosseini, Josh T. Lane, and Mohammad H. Mahoor. 2017. A pilot study on using an intelligent life-like robot as a companion for elderly individuals with dementia and depression. In *Proceedings of the 2017 IEEE-RAS 17th International Conference on Humanoid Robotics (Humanoids)*. IEEE, 541–546.

[4] Sibel Adali and Jennifer Golbeck. 2012. Predicting personality with social behavior. In *Proceedingsl of the 2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining*. IEEE, 302–309.

[5] Mohammad Rafayet Ali, Dev Crasta, Li Jin, Agustin Baretto, Joshua Pachter, Ronald D. Rogge, and Mohammed Ehsan Hoque. 2015. LISS-live interactive social skill assistance. In *Proceedings of the 2015 International Conference on Affective Computing and Intelligent Interaction (ACII'15)*. IEEE, 173–179.

[6] 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 *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.

[7] Mohammad Rafayet Ali, Kimberly Van Orden, Kimberly Parkhurst, Shuyang Liu, Viet-Duy Nguyen, Paul Duberstein, and M. Ehsan Hoque. 2018. Aging and engaging: A social conversational skills training program for older adults. In *Proceedings of the 23rd International Conference on Intelligent User Interfaces*. 55–66.

[8] Rafayet Ali, Ehsan Hoque, Paul Duberstein, Lenhart Schubert, Seyedeh Zahra Razavi, Benjamin Kane, Caroline Silva, Jennifer S Daks, Meghan Huang, and Kim Van Orden. 2020. Aging and engaging: A pilot randomized controlled trial of an online conversational skills coach for older adults. *The American Journal of Geriatric Psychiatry* (2020), 804–815.

[9] Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The TARDIS framework: Intelligent virtual agents for social coaching in job interviews. In *International Conference on Advances in Computer Entertainment Technology*. Springer, 476–491.

[10] Patricia A. Arean, Michael G. Perri, Arthur M. Nezu, Rebecca L. Schein, Frima Christopher, and Thomas X. Joseph. 1993. Comparative effectiveness of social problem-solving therapy and reminiscence therapy as treatments for depression in older adults. *Journal of Consulting and Clinical Psychology* 61, 6 (1993), 1003.

[11] Christina Atay, David Ireland, Jacki Liddle, Janet Wiles, Adam Vogel, Dan Angus, Dana Bradford, Alana Campbell, Olivia Rushin, and Helen J. Chenery. 2016. Can a smartphone-based chatbot engage older community group members? The impact of specialised content. *Alzheimer's & Dementia* 12 (2016), P1005–P1006.

[12] JinYeong Bak, Chin-Yew Lin, and Alice Oh. 2014. Self-disclosure topic model for classifying and analyzing Twitter conversations. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. 1986–1996.

[13] Sara Bernardini, Kaśka Porayska-Pomsta, and Tim J. Smith. 2014. ECHOES: An intelligent serious game for fostering social communication in children with autism. *Information Sciences* 264 (2014), 41–60.

[14] Nicholas Bott, Sharon Wexler, Lin Drury, Chava Pollak, Victor Wang, Kathleen Scher, and Sharon Narducci. 2019. A protocol-driven, bedside digital conversational agent to support nurse teams and mitigate risks of hospitalization in older adults: Case control pre-post study. *Journal of Medical Internet Research* 21, 10 (2019), e13440.

[15] Kevin K. Bowden, Shereen Oraby, Amita Misra, Jiaqi Wu, Stephanie Lukin, and Marilyn Walker. 2019. Data-driven dialogue systems for social agents. *Advanced Social Interaction with Agents*. Springer, 53–56.

[16] Zoraida Callejas and David Griol. 2021. Conversational agents for mental health and wellbeing. *Logic, Argumentation and Reasoning* 22 (2021), 219–244.

[17] Hao Cheng, Hao Fang, and Mari Ostendorf. 2019. A dynamic speaker model for conversational interactions. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*. 2772–2785.

[18] Scotty Craig, Arthur Graesser, Jeremiah Sullins, and Barry Gholson. 2004. Affect and learning: An exploratory look into the role of affect in learning with autotutor. *Journal of Educational Media* 29, 3 (2004), 241–250.

[19] Noemi da Paixão Pinto, Juliana Baptista dos Santos França, Henrique Prado de Sá Sousa, Adriana Santarosa Vivacqua, and Ana Cristina Bicharra Garcia. 2021. Conversational agents for elderly interaction. In *Proceedings of the 2021 IEEE 24th International Conference on Computer Supported Cooperative Work in Design (CSCWD'21)*. IEEE, 1–6.

[20] Mark L. Knapp and John Augustine Daly. 2002. *Handbook of Interpersonal Communication*. Sage.

[21] Berardina De Carolis, Stefano Ferilli, and Giuseppe Palestra. 2017. Simulating empathic behavior in a social assistive robot. *Multimedia Tools and Applications* 76, 4 (2017), 5073–5094.

[22] Laura Dellazizzo, Olivier Percie du Sert, Kingsada Phraxayavong, Stéphane Potvin, Kieron O'Connor, and Alexandre Dumais. 2018. Exploration of the dialogue components in avatar therapy for schizophrenia patients with refractory auditory hallucinations: A content analysis. *Clinical Psychology & Psychotherapy* 25, 6 (2018), 878–885.

[23] Stephan Diederich, Max Janßen-Müller, Alfred Benedikt Brendel, and Stefan Morana. 2019. Emulating empathetic behavior in online service encounters with sentiment-adaptive responses: Insights from an experiment with a conversational agent. In *Proceedings of the International Conference on Information Systems*.

[24] Sidney K. D'Mello, Scotty D. Craig, Jeremiah Sullins, and Arthur C. Graesser. 2006. Predicting affective states expressed through an emote-aloud procedure from AutoTutor's mixed-initiative dialogue. *International Journal of Artificial Intelligence in Education* 16, 1 (2006), 3–28.

[25] Mohammad El-Ramly, Eleni Stroulia, and Paul Sorenson. 2002. Recovering software requirements from system-user interaction traces. In *Proceedings of the 14th International Conference on Software Engineering and Knowledge Engineering*. 447–454.

[26] Xiangmin Fan, Daren Chao, Zhan Zhang, Dakuo Wang, Xiaohua Li, and Feng Tian. 2020. Utilization of self-diagnosis health chatbots in real-world settings: Case study. *Journal of Medical Internet Research* 23, 1 (2020), e19928.

[27] Morag Farquhar. 1995. Elderly people's definitions of quality of life. *Social Science & Medicine* 41, 10 (1995), 1439–1446.

[28] Ben Fielding, Philip Kinghorn, Kamlesh Mistry, and Li Zhang. 2016. An enhanced intelligent agent with image description generation. *International Conference on Intelligent Virtual Agents*. Springer, 110–119.

[29] Terrence Fong, Illah Nourbakhsh, and Kerstin Dautenhahn. 2003. A survey of socially interactive robots. *Robotics and Autonomous Systems* 42, 3-4 (2003), 143–166.

[30] Chiara Garattini, Joseph Wherton, and David Prendergast. 2012. Linking the lonely: An exploration of a communication technology designed to support social interaction among older adults. *Universal Access in the Information Society* 11, 2 (2012), 211–222.

[31] Patrick Gebhard, Tanja Schneeberger, Elisabeth André, Tobias Baur, Ionut Damian, Gregor Mehlmann, Cornelius König, and Markus Langer. 2018. Serious games for training social skills in job interviews. *IEEE Transactions on Games* 11, 4 (2018), 340–351.

[32] Elizabeth L. Glisky. 2007. Changes in cognitive function in human aging. *Brain Aging: Models, Methods, and Mechanisms* 1 (2007), 3–20.

[33] Jeannette Golden, Ronán M. Conroy, Irene Bruce, Aisling Denihan, Elaine Greene, Michael Kirby, and Brian A. Lawlor. 2009. Loneliness, social support networks, mood and wellbeing in community-dwelling elderly. *International Journal of Geriatric Psychiatry: A Journal of the Psychiatry of Late Life and Allied Sciences* 24, 7 (2009), 694–700.

[34] Nobuhiko Goto and Alexandre Schaefer. 2020. Emotional intensity. *Encyclopedia of Personality and Individual Differences* (2020), 1311–1319.

[35] Amanda R. Hemmesch. 2014. The detrimental effects of atypical nonverbal behavior on older adults' first impressions of individuals with Parkinson's disease. *Psychology and Aging* 29, 3 (2014), 521.

[36] Mohammed Hoque, Matthieu Courgeon, Jean-Claude Martin, Bilge Mutlu, and Rosalind W. Picard. 2013. MACH: My automated conversation coach. In *Proceedings of the 2013 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. 697–706.

[37] David J. Houghton and Adam N. Joinson. 2012. Linguistic markers of secrets and sensitive self-disclosure in Twitter. In *Proceedings of the 2012 45th Hawaii International Conference on System Sciences*. IEEE, 3480–3489.

[38] Clayton J. Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the 8th International AAAI Conference on Weblogs and Social Media*.

[39] Becky Inkster, Shubhankar Sarda, and Vinod Subramanian. 2018. An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth* 6, 11 (2018), e12106.

[40] Masamune Kawasaki, Naomi Yamashita, Yi-Chieh Lee, and Kayoko Nohara. 2020. Assessing users' mental status from their journaling behavior through chatbots. In *Proceedings of the 20th ACM International Conference on Intelligent Virtual Agents*. 1–8.

[41] Caroline F. Keating. 2016. The developmental arc of nonverbal communication: Capacity and consequence for human social bonds. In *APA Handbook of Nonverbal Communication*, D. Matsumoto, H. C. Hwang, and M. G. Frank (Eds.). American Psychological Association, 103–138.

[42] Yanghee Kim. 2005. Empathetic virtual peers enhanced learner interest and self-efficacy. In *Workshop on Motivation and Affect in Educational Software, in Conjunction with the 12th International Conference on Artificial Intelligence in Education*. 9–16.

[43] Masatomo Kobayashi, Akihiro Kosugi, Hironobu Takagi, Miyuki Nemoto, Kiyotaka Nemoto, Tetsuaki Arai, and Yasunori Yamada. 2019. Effects of age-related cognitive decline on elderly user interactions with voice-based dialogue systems. In *IFIP Conference on Human-Computer Interaction*. Springer, 53–74.

[44] Stefan Kopp, Mara Brandt, Hendrik Buschmeier, Katharina Cyra, Farina Freigang, Nicole Krämer, Franz Kummert, Christiane Opfermann, Karola Pitsch, Lars Schillingmann, et al. 2018. Conversational assistants for elderly users—The importance of socially cooperative dialogue. In *Proceedings of the AAMAS Workshop on Intelligent Conversation Agents in Home and Geriatric Care Applications Co-located with the Federated AI Meeting*, Vol. 2338.

[45] SeoYoung Lee and Junho Choi. 2017. Enhancing user experience with conversational agent for movie recommendation: Effects of self-disclosure and reciprocity. *International Journal of Human-Computer Studies* 103 (2017), 95–105.

[46] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–12.

[47] Elizabeth A. Linnenbrink and Paul R. Pintrich. 2002. The role of motivational beliefs in conceptual change. In *Reconsidering Conceptual Change: Issues in Theory and Practice*. Springer, 115–135.

[48] Christine Lisetti, Reza Amini, Ugan Yasavur, and Naphtali Rishe. 2013. I can help you change! An empathic virtual agent delivers behavior change health interventions. *ACM Transactions on Management Information Systems (TMIS'13)* 4, 4 (2013), 19.

[49] Gale M. Lucas, Albert Rizzo, Jonathan Gratch, Stefan Scherer, Giota Stratou, Jill Boberg, and Louis-Philippe Morency. 2017. Reporting mental health symptoms: Breaking down barriers to care with virtual human interviewers. *Frontiers in Robotics and AI* 4 (2017), 51.

[50] Galina Madjaroff, Helena Mentis, and Judah Ronch. 2016. Differences in perceived impact of person-centered technology on older adults' quality of life. In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2200–2208.

[51] Fatik Baran Mandal. 2014. Nonverbal communication in humans. *Journal of Human Behavior in the Social Environment* 24, 4 (2014), 417–421.

[52] Juan Martínez-Miranda. 2017. Embodied conversational agents for the detection and prevention of suicidal behaviour: Current applications and open challenges. *Journal of Medical Systems* 41, 9 (2017), 135.

[53] Sean A. McGlynn, Shawn Kemple, Tracy L. Mitzner, Chih-Hung Aaron King, and Wendy A. Rogers. 2017. Understanding the potential of PARO for healthy older adults. *International Journal of Human-computer Studies* 100 (2017), 33–47.

[54] Juliana Miehle, Ilker Bagci, Wolfgang Minker, and Stefan Ultes. 2019. A social companion and conversational partner for the elderly. *Advanced Social Interaction with Agents*. Springer, 103–109.

[55] Juliana Miehle, Isabel Feustel, Julia Hornauer, Wolfgang Minker, and Stefan Ultes. 2020. Estimating user communication styles for spoken dialogue systems. In *Proceedings of the 12th Language Resources and Evaluation Conference*. 540–548.

[56] Juliana Miehle, Wolfgang Minker, and Stefan Ultes. 2018. Exploring the impact of elaborateness and indirectness on user satisfaction in a spoken dialogue system. In *Adjunct Publication of the 26th Conference on User Modeling, Adaptation and Personalization*. 165–172.

[57] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2019. Survey of conversational agents in health. *Expert Systems with Applications* 129 (2019), 56–67.

[58] Svetlana Nikitina, Sara Callaioli, and Marcos Baez. 2018. Smart conversational agents for reminiscence. In *Proceedings of the 2018 IEEE/ACM 1st International Workshop on Software Engineering for Cognitive Services (SE4COG'18)*. IEEE, 52–57.

[59] Risako Ono, Yuki Nishizeki, and Masahiro Araki. 2019. Virtual Dialogue Agent for Supporting a Healthy Lifestyle of the Elderly. In *9th International Workshop on Spoken Dialogue System Technology*. Springer, 253–258.

[60] Risako Ono, Yuki Nishizeki, and Masahiro Araki. 2019. Virtual dialogue agent for supporting a healthy lifestyle of the elderly. In *Proceedings of the 9th International Workshop on Spoken Dialogue System Technology*. Springer, 253–258.

[61] Thomas L. Patterson and Brent T. Mausbach. 2010. Measurement of functional capacity: A new approach to understanding functional differences and real-world behavioral adaptation in those with mental illness. *Annual Review of Clinical Psychology* 6 (2010), 139–154.
[62] Thomas L. Patterson, Brent T. Mausbach, Christine McKibbin, Sherry Goldman, Jesus Bucardo, and Dilip V. Jeste. 2006. Functional adaptation skills training (FAST): A randomized trial of a psychosocial intervention for middle-aged and older patients with chronic psychotic disorders. *Schizophrenia Research* 86, 1–3 (2006), 291–299.
[63] Thomas L. Patterson, Sherry Moscona, Christine L. McKibbin, Kevin Davidson, and Dilip V. Jeste. 2001. Social skills performance assessment among older patients with schizophrenia. *Schizophrenia Research* 48, 2–3 (2001), 351–360.
[64] Florian Pecune, Jingya Chen, Yoichi Matsuyama, and Justine Cassell. 2018. Field trial analysis of socially aware robot assistant. In *Proceedings of the 17th International Conference on Autonomous Agents and MultiAgent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, 1241–1249.
[65] James W. Pennebaker, Ryan L. Boyd, Kayla Jordan, and Kate Blackburn. 2015. The development and psychometric properties of LIWC2015.
[66] Carla M. Perissinotto, Irena Stijacic Cenzer, and Kenneth E. Covinsky. 2012. Loneliness in older persons: A predictor of functional decline and death. *Archives of Internal Medicine* 172, 14 (2012), 1078–1084.
[67] Kaśka Porayska-Pomsta, Alyssa M. Alcorn, Katerina Avramides, Sandra Beale, Sara Bernardini, Mary Ellen Foster, Christopher Frauenberger, Judith Good, Karen Guldberg, Wendy Keay-Bright, et al. 2018. Blending human and artificial intelligence to support autistic children's social communication skills. *ACM Transactions on Computer-Human Interaction (TOCHI'18)* 25, 6 (2018), 1–35.
[68] Louisa Pragst, Wolfgang Minker, and Stefan Ultes. 2019. Exploring the applicability of elaborateness and indirectness in dialogue management. *Advanced Social Interaction with Agents*. Springer, 189–198.
[69] Abhilasha Ravichander and Alan W. Black. 2018. An empirical study of self-disclosure in spoken dialogue systems. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*. 253–263.
[70] Seyedeh Zahra Razavi, Mohammad Rafayet Ali, Tristram H. Smith, Lenhart K. Schubert, and Mohammed Ehsan Hoque. 2016. The LISSA virtual human and ASD teens: An overview of initial experiments. In *International Conference on Intelligent Virtual Agents*. Springer, 460–463.
[71] Seyedeh Zahra Razavi, Lenhart K. Schubert, Mohammad Rafayet Ali, and Mohammed Ehsan Hoque. 2017. Managing casual spoken dialogue using flexible schemas, pattern transduction trees, and gist clauses. In *Proceedings of the Annual Conference on Advances in Cognitive Systems (ACS'17)*.
[72] S. Zahra Razavi, Lenhart K. Schubert, Benjamin Kane, Mohammad Rafayet Ali, Kimberly A. Van Orden, and Tianyi Ma. 2019. Dialogue design and management for multi-session casual conversation with older adults. In *IUI Workshops*.
[73] Alex Rinaldi, Omar Oseguera, Joann Tuazon, and Albert C. Cruz. 2017. End-to-end dialogue with sentiment analysis features. In *International Conference on Human-Computer Interaction*. Springer, 480–487.
[74] Alessandro Russo, Grazia D'Onofrio, Aldo Gangemi, Francesco Giuliani, Misael Mongiovi, Francesco Ricciardi, Francesca Greco, Filippo Cavallo, Paolo Dario, Daniele Sancarlo, et al. 2019. Dialogue systems and conversational agents for patients with dementia: The human–robot interaction. *Rejuvenation Research* 22, 2 (2019), 109–120.
[75] Roger C. Schank and Robert P. Abelson. 2013. *Scripts, Plans, Goals, and Understanding: An Inquiry Into Human Knowledge Structures*. Psychology Press.
[76] Rachel Schwartz and Marc D. Pell. 2017. When emotion and expression diverge: The social costs of Parkinson's disease. *Journal of Clinical and Experimental Neuropsychology* 39, 3 (2017), 211–230.
[77] Chris Segrin. 2019. Indirect effects of social skills on health through stress and loneliness. *Health Communication* 34, 1 (2019), 118–124.
[78] Nava A. Shaked. 2017. Avatars and virtual agents—Rrelationship interfaces for the elderly. *Healthcare Technology Letters* 4, 3 (2017), 83–87.
[79] Archana Singh and Nishi Misra. 2009. Loneliness, depression and sociability in old age. *Industrial Psychiatry Journal* 18, 1 (2009), 51.
[80] Mohammad Soleymani, Kalin Stefanov, Sin-Hwa Kang, Jan Ondras, and Jonathan Gratch. 2019. Multimodal analysis and estimation of intimate self-disclosure. In *Proceedings of the 2019 International Conference on Multimodal Interaction*. 59–68.
[81] Andrew Steptoe, Aparna Shankar, Panayotes Demakakos, and Jane Wardle. 2013. Social isolation, loneliness, and all-cause mortality in older men and women. *Proceedings of the National Academy of Sciences USA* 110, 15 (2013), 5797–5801.
[82] Margaret A. Struchen, Monique R. Pappadis, Angelle M. Sander, Christina S. Burrows, and Katherine A. Myszka. 2011. Examining the contribution of social communication abilities and affective/behavioral functioning to social integration outcomes for adults with traumatic brain injury. *The Journal of Head Trauma Rehabilitation* 26, 1 (2011), 30–42.

[83] Huei-Chuan Sung, Shu-Min Chang, Mau-Yu Chin, and Wen-Li Lee. 2015. Robot-assisted therapy for improving social interactions and activity participation among institutionalized older adults: A pilot study. *Asia-Pacific Psychiatry* 7, 1 (2015), 1–6.

[84] Joe Tomaka, Sharon Thompson, and Rebecca Palacios. 2006. The relation of social isolation, loneliness, and social support to disease outcomes among the elderly. *Journal of Aging and Health* 18, 3 (2006), 359–384.

[85] Christiana Tsiourti, Maher Ben Moussa, João Quintas, Ben Loke, Inge Jochem, Joana Albuquerque Lopes, and Dimitri Konstantas. 2016. A virtual assistive companion for older adults: Design implications for a real-world application. In *Proceedings of SAI Intelligent Systems Conference*. Springer, 1014–1033.

[86] Tsuyoki Ujiro, Hiroki Tanaka, Hiroyoshi Adachi, Hiroaki Kazui, Manabu Ikeda, Takashi Kudo, and Satoshi Nakamura. 2018. Detection of dementia from responses to atypical questions asked by embodied conversational agents. In *Proceedings of Interspeech* 2018, 1691–1695.

[87] Dina Utami, Timothy Bickmore, Asimina Nikolopoulou, and Michael Paasche-Orlow. 2017. Talk about death: End of life planning with a virtual agent. In *International Conference on Intelligent Virtual Agents*. Springer, 441–450.

[88] Katherine Valencia, Virginia Zaraza Rusu, Erick Jamet, Constanza Zúñiga, Eduardo Garrido, Cristian Rusu, and Daniela Quiñones. 2020. Technology-based social skills learning for people with autism spectrum disorder. In *International Conference on Human-Computer Interaction*. Springer, 598–615.

[89] Yves Vanrompay, Mario Cataldi, Marine Le Glouanec, and Myriam Lamolle. 2014. Sentiment Analysis for Dynamic User Preference Inference in Spoken Dialogue Systems. Paper *Presented at the First Workshop on Semantic Sentiment Analysis (SSA) at ESWC2014*.

[90] Laura Pfeifer Vardoulakis, Lazlo Ring, Barbara Barry, Candace L. Sidner, and Timothy Bickmore. 2012. Designing relational agents as long term social companions for older adults. In *International Conference on Intelligent Virtual Agents*. Springer, 289–302.

[91] Paul Verhaeghen and John Cerella. 2002. Aging, executive control, and attention: A review of meta-analyses. *Neuroscience & Biobehavioral Reviews* 26, 7 (2002), 849–857.

[92] Michelle X. Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TIIS'19)* 9, 2–3 (2019), 1–36.

Received January 2021; revised June 2021; accepted August 2021