Career Interview Readiness in Virtual Reality (CIRVR): A Platform for Simulated Interview Training for Autistic Individuals and Their Employers

Article *in* ACM Transactions on Accessible Computing · March 2022

DOI: 10.1145/3505560

CITATIONS
7

READS
204

10 authors, including:

Deeksha Adiani
Vanderbilt University
10 PUBLICATIONS **15** CITATIONS
SEE PROFILE

Dayi Bian
Biofourmis Inc.
32 PUBLICATIONS **575** CITATIONS
SEE PROFILE

Amy Swanson
Vanderbilt University
73 PUBLICATIONS **1,744** CITATIONS
SEE PROFILE

Timothy J Vogus
Vanderbilt University
123 PUBLICATIONS **6,819** CITATIONS
SEE PROFILE

Career Interview Readiness in Virtual Reality (CIRVR): A Platform for Simulated Interview Training for Autistic Individuals and Their Employers

2

DEEKSHA ADIANI, Department of Electrical Engineering and Computer Science, Vanderbilt University, USA
AARON ITZKOVITZ, Robotics and Autonomous Systems Lab, Vanderbilt University, USA
DAYI BIAN, Department of Electrical Engineering and Computer Science, Vanderbilt University, USA
HARRISON KATZ, MICHAEL BREEN, and SPENCER HUNT, Robotics and Autonomous Systems Lab, Vanderbilt University, USA
AMY SWANSON, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University Medical Center, USA
TIMOTHY J. VOGUS, Owen Graduate School of Management, Vanderbilt University, USA
JOSHUA WADE and NILANJAN SARKAR, Department of Mechanical Engineering, Vanderbilt University, USA

Employment outcomes for autistic[1] individuals are often poorer relative to their neurotypical (NT) peers, resulting in a greater need for other forms of financial and social support. While a great deal of work has focused on developing interventions for autistic children, relatively less attention has been paid to directly addressing the employment challenges faced by autistic adults. One key impediment to autistic individuals securing employment is the job interview. Autistic individuals often experience anxiety in interview situations, particularly with open-ended questions and unexpected interruptions. They also exhibit atypical gaze patterns that may be perceived as, but not necessarily indicative of, disinterest or inattention. In response, we developed a closed-loop adaptive virtual reality (VR)–based job interview training platform, which we have named Career Interview Readiness in VR (CIRVR). CIRVR is designed to provide an engaging, adaptive, and individualized experience to practice and refine interviewing skills in a less anxiety-inducing virtual context. CIRVR

[1]Due to a preference of autistic individuals and their families to use identity-first language [43], we have chosen this form of disability-related terminology.

This project was funded by a Microsoft AI for Accessibility grant and by the National Science Foundation under grant number 1936970.
Authors' addresses: D. Adiani, Department of Computer Science, Vanderbilt University, 400 24th Ave S, Nashville, TN 37212, USA; email: deeksha.m.adiani@vanderbilt.edu; D. Bian, Department of Electrical and Computer Engineering, Vanderbilt University, 400 24th Ave S, Nashville, TN 37212, USA; email: biandayi@gmail.com; A. Itzkovitz, H. Katz, M. Breen, and S. Hunt, Robotics and Autonomous Systems Lab, Vanderbilt University, 2400 Highland Avenue, Nashville, TN 37212, USA; emails: itzkovitza@gmail.com, harrisonmkatz@gmail.com, {michael.breen, spencer.hunt}@vanderbilt.edu; A. Swanson, Treatment and Research Institute for Autism Spectrum Disorders (TRIAD), Vanderbilt University Medical Center, 110 Magnolia Cir, Nashville, TN 37203, USA; T. J. Vogus, Owen Graduate School of Management, Vanderbilt University, 401 21st Ave S, Nashville, TN 37203, USA; J. Wade and N. Sarkar, Department of Mechanical Engineering, Vanderbilt University, 2400 Highland Ave, Nashville, TN 37212, USA; emails: joshua.w.wade@gmail.com, nilanjan.sarkar@vanderbilt.edu.

1936-7228/2022/02-ART2 $15.00
https://doi.org/10.1145/3505560

contains a real-time physiology-based stress detection module, as well as a real-time gaze detection module, to permit individualized adaptation. We also present the first prototype of the CIRVR Dashboard, which provides visualizations of data to help autistic individuals as well as potential employers and job coaches make sense of the data gathered from interview sessions. We conducted a feasibility study with 9 autistic and 8 NT individuals to assess the preliminary usability and feasibility of CIRVR. Results showed differences in perceived usability of the system between autistic and NT participants, and higher levels of stress in autistic individuals during interviews. Participants across both groups reported satisfaction with CIRVR and the structure of the interview. These findings and feedback will support future work in improving CIRVR's features in hopes for it to be a valuable tool to support autistic job candidates as well as their potential employers.

CCS Concepts: • **Human-centered computing** → **Virtual reality**; *Usability testing*; • **Computing methodologies** → Machine learning;

Additional Key Words and Phrases: Autism Spectrum Disorder, virtual job interview

ACM Reference format:
Deeksha Adiani, Aaron Itzkovitz, Dayi Bian, Harrison Katz, Michael Breen, Spencer Hunt, Amy Swanson, Timothy J. Vogus, Joshua Wade, and Nilanjan Sarkar. 2022. Career Interview Readiness in Virtual Reality (CIRVR): A Platform for Simulated Interview Training for Autistic Individuals and Their Employers. *ACM Trans. Access. Comput.* 15, 1, Article 2 (February 2022), 28 pages.
https://doi.org/10.1145/3505560

1 INTRODUCTION

Autism Spectrum Disorder (ASD) is a lifelong neurodevelopmental condition characterized by differences in social communication as well as restrictive or repetitive patterns of behavior or interests, with a current prevalence of 1 in 54 children (approximately 2% of the entire population) in the United States [1, 50]. Although significant attention has been paid to the development of early intervention programs for autistic children, it is only recently that industry, researchers, and policy makers have begun to broaden their focus and efforts to address a major challenge facing autistic adults —meaningful employment [55]. Finding and keeping meaningful employment is essential for independence as well as for personal fulfillment and well-being [38]. Unfortunately, estimates indicate that only 10% to 50% of autistic individuals are employed; even at the most optimistic end of this wide range, only about 15% to 20% of employed individuals are in full-time roles that are capable of supporting independence [3, 55].

Most autistic individuals report a desire to work but cite difficulties in finding employment and maintaining employment due to the severe challenges of the work environment [80]. Barriers to employment for autistic individuals have been attributed to multiple factors, including acute challenges of the job interview process; differences or deficits in social behavior and communication in the workplace; reduced access to post-secondary education; often co-occurring conditions, such as intellectual disability or attention-deficit hyperactivity disorder (ADHD); and workplace discrimination of autistic individuals by both employers and co-workers [14, 55, 80]. Amongst these, the job interview poses an initial barrier to employment in which strong verbal communication and overall demeanor as expected from neurotypical (NT) individuals are considered necessary traits for a qualifying candidate [2, 14]. Smith et al. [72] conducted a study to evaluate vocational interviewing outcomes among 656 transition-age youth who received special education pre-employment transition services. They found that 21% of the participants were employed and 89% of the employed individuals interviewed prior to obtaining employment. Their results suggest the importance of job interviewing in obtaining employment, thus, supplementing the need for job interview interventions for the target population.

Researchers have explored a range of methods, not limited to computing and technology, to support the transition to employment for people with disabilities. Some of these approaches include programs in which individuals are immersed in actual work environments [34, 87], the creation of structured interview questions to improve recall during job interviews [57], the use of audio prompting during task training to promote independence in task completion [53], and the use of assistive technology such as tablet computers to provide support in job training and task completion [39]. For example, Wehman et al. [87] modified an existing job intervention model by adding support from job coaches for autistic individuals. A job coach provided individualized support by creating a goal-specific job profile, by working with the individuals on job applications and completing interviews, by monitoring progress in learning the work, and by providing long-term support. Their intervention yielded improved employment outcomes. However, Wehman et al. mention that despite participants' good work during internships, to an employer who is uninformed of their skills, participants' performance during interviews might limit their chances of employment. They give an example of an autistic participant whose performance in the interview may have limited chances of gaining employment had the employer not had the opportunity of observing the participant in three different internships that highlighted the individual's work ethic and skills. Gilson et al. [34] conducted a pilot study in which a small number of college students with intellectual disability (ID) and/or ASD were given unpaid internships to encourage task engagement and social interaction in proximity with a job coach. They defined proximity as being within 5 feet of a person and in a position that would allow the opportunity for in-person interaction. They found that when job coaches eventually reduced proximity and delivered prompts via walkie-talkie in-ear devices (that the participants reported as unobtrusive and rather beneficial), task engagement maintained and interactions increased. Norris et al. [57] conducted a study to show that interviewing autistic individuals using structured questions, compared with open questions, was more effective for eliciting personal facts and memories. More specifically, autistic individuals found recalling information about themselves during interviews easier when they were given visual-verbal prompts in which a question about a particular memory was split into the following: when it happened, the people who were involved, the setting, the actions that took place, and the objects that were involved. Receiving semantic prompts to specify personal characteristics were also effective. Another approach of interviewing, not directly related to job employment, was conducted by Harrington et al. [36] in which the interviewers adapted their questioning strategies to fit each autistic participant's communication capabilities. For example, consideration was paid to autistic participants who needed substantial time to respond to answers or needed visual support, such as a written schedule with images to indicate the sequence of questions that would be asked during the interview. Although the interview data had not been completed at the time, the author reported that the use of adapted strategies enabled participants to express their views better.

All of these interventions point to ways to improve interview and employment performance using structured, realistic coaching and training related to social communication. These studies, however, emphasize support in the context of existing jobs and internships typically part of a pilot program [34, 39, 53, 87]. Although useful, most autistic individuals will not always have the aid of structured internships to showcase their skills and instead will need to secure employment through traditional channels (i.e., interviews). Thus, interview training is often a crucial first step to gaining meaningful employment. There are a few computer-based approaches that include interview simulation tools allowing candidates to practice interview skills in a comfortable and controlled environment [69, 70, 78], and Virtual-Reality (VR)–based tools to practice social communication and social interaction skills [6–8, 16, 22, 23, 68, 85, 88]. Strickland et al. [78] evaluated a job interview training program for autistic individuals between the ages 16 and 19 that involved video modelling via a web-based platform called JobTIPS, followed by a practice session in VR. They created a rating

instrument to assess interview skills with the help of human resources experts as well as a scale to measure social responsiveness in autistic individuals. The results from their study suggested that participants who completed the employment program demonstrated improvement in verbal responses to interview questions. However, they did not improve in delivery of responses that include non-verbal features such as posture. Burke et al. [16] evaluated Virtual Interactive Training Agents (ViTAs) with autistic adults and individuals with other disabilities. Their study included four sessions with ViTAs, followed by a face-to-face interview. The study showed that ViTA had influenced improvement in interviewing skills in the participants; their recent efficacy study with a larger number of participants confirms the same [17]. SIMmersion LLC created PeopleSim, an online VR-based conversation skill training platform [68]. Bell et al. [8] in partnership with SIMmersion LLC, adapted PeopleSIM into an interview skill training platform. The system included multiple scripts to choose from and a feedback system to help users reflect on their performance. Smith et al. [70] then joined Bell et al. and SIMmersion LLC to create Virtual Reality Job Interview Training (VR-JIT) for adults with psychiatric disabilities. The system was composed of more than 1,000 interactive prerecorded videos coupled with speech recognition software that showed evidence of increasing employment success among autistic individuals [69]. Recently, Smith et al. [70] further adapted VR-JIT to meet the needs of autistic transitioning-age youth (between ages 16 to 21) and created Virtual Interview Training for Transition Age Youth (VIT-TAY) [71]. Their study suggested that participants (between ages 16 and 26) who received pre-employment services in school as well as training with VIT-TAY showed better job interviewing skills, lower anxiety during interviewing when compared with those participants who did not train on VIT-TAY, and had better employment outcomes. Thus, the study demonstrated the effectiveness of job interview training prior to entering the job market [73]. A VR-based application called VirtualSpeech [85] was initially created for people to practice public speaking and, possibly, to reduce their anxieties of glossophobia (the fear of public speaking) [6]. It has now been extended to practicing social communication more generally. It allows users to choose from multiple interview environments and scripts, providing feedback such as eye contact and pace of speaking, and uses speech analysis to identify possible hesitation in speech [85]. Xu et al. [88] created LittleHelper, a Google Glass application aimed at providing real-time feedback to interviewees about appropriate levels of eye contact and speech volume. Baur et al. [7] created a virtual job interview simulation with the ability to detect and respond to social cues relevant to interviews, such as leaning back, smiling, looking away, and so on, that shows the importance of conscious or unconscious non-verbal communication in overall evaluation of candidates. DeVault et al. [22] created the SimSensei Kiosk, a VR system presenting a semi-structured interview designed to elicit distress indicators to capture measures complementing tests used by clinicians to assess depression, anxiety, and post-traumatic stress disorder. While the two aforementioned systems were not tested on autistic individuals, they demonstrated reliability for their respective target populations and represent a step forward in the design of simulated interview technology. Preliminary use of such VR-based social skills intervention systems, including interviewing systems, have shown initial promise in enhancing underlying social cognitive skills [23, 91].

Although the existing VR-based systems have demonstrated potential benefits, they have not fully leveraged recent advances in Artificial Intelligence (AI)–based technologies — such as affective computing, real-time eye-gaze monitoring, and closed-loop adaptivity —that would allow delivering dynamic, individually tailored training for job interviews (e.g., by managing social anxiety). In response, we developed, based on stakeholder inputs, a novel platform, Career Interview Readiness in VR (CIRVR). CIRVR is a VR-based job interview training platform intended for individuals to practice important interviewing skills through a bidirectional, flexible conversation mechanism. In addition to providing a virtual platform that builds on prior work

[7, 8, 16, 17, 22, 23, 68–71, 73, 78, 85, 88], CIRVR conducts real-time stress detection from physiological responses, eye-gaze monitoring, and emotion detection via facial image capture of the participant. Studies have associated both social anxiety and atypical eye contact as common phenomena in autistic individuals [19, 67, 74]. Although the literature is clear that inconsistent eye contact in autistic individuals is not indicative of lack of attention [1, 14], there is no systematic effort, to our knowledge, that captures gaze patterns of autistic individuals during interviews that can provide insight to job coaches, employers, and autistic individuals themselves. In fact, perceived eye contact is known to elicit arousal in certain regions of the brain of autistic individuals, which, in turn, causes anxiety [67]. Thus, measuring stress response and eye gaze, as well as capturing emotions through facial expressions (e.g., happy, sad, or frustrated), will likely provide insight into why and when autistic individuals may struggle during traditional job interviews and how these challenges can be overcome.

The overarching long-term goal of this research and this system is to provide insight into the behavior of autistic individuals, not to ask them to adjust to the "norm" and behave like their NT peers (e.g., by changing eye-gaze patterns) but rather to provide information that may lead to the adoption of more inclusive hiring practices. However, the scope of the current work is to present the design, feasibility, and usability of CIRVR. We believe that this first study will provide quantitative and qualitative data on how autistic individuals interact with CIRVR, which will then be used in the future to design appropriate feedback as well as create individualized practice opportunities to enhance their interviewing skills. CIRVR also provides a Dashboard for the employers and job coaches to visualize the data collected during practice interviews with CIRVR and analytics to help identify refinements to existing interview protocols that better suit autistic candidates. However, we understand that such changes take time. In the meantime, autistic individuals will need to successfully navigate interviews that often rely on multiple forms of open-ended questions. We expect that CIRVR will provide individualized interview practice opportunities to autistic individuals to improve interview performance.

The contributions of this work are twofold. First, we present the design and implementation of a novel virtual interview system, CIRVR, based on a closed-loop conversation management system, and real-time stress and eye-gaze detection. By closed-loop, we mean that CIRVR will have a mechanism to dynamically adapt the conversation based on real-time response patterns of the participants and their stress level. We also present the design of a Dashboard for CIRVR data visualization for employers and job coaches. Second, we present the results of a feasibility study to demonstrate the acceptability of CIRVR by the target population. It should be noted that our work presents only an initial feasibility assessment without any system-generated feedback to the participants. As needed for any intervention for autistic individuals, an efficacy study will be conducted in the future that will incorporate improvements based on insights learned from the current study.

This article is organized as follows. Sections 2, 3, and 4 highlight the first contribution of our work. In Section 2, we discuss the stakeholder input that informed CIRVR's design and the major components of the CIRVR design, including a discussion of the Dashboard. In Section 3, we discuss the physiology-based stress detection module, gaze detection module, and emotion detection module. In Section 4, we discuss the Conversation Management System and its components. Section 5 presents the results and discussion of the feasibility study, comprising the second contribution of our work. Section 6 summarizes the work, its current limitations, and our plan for the future of CIRVR.

2 DESIGN OF CIRVR

CIRVR aims to provide a platform on which interviewees may practice core interviewing skills, such as communicating effectively about their strengths, past work experience, and educational

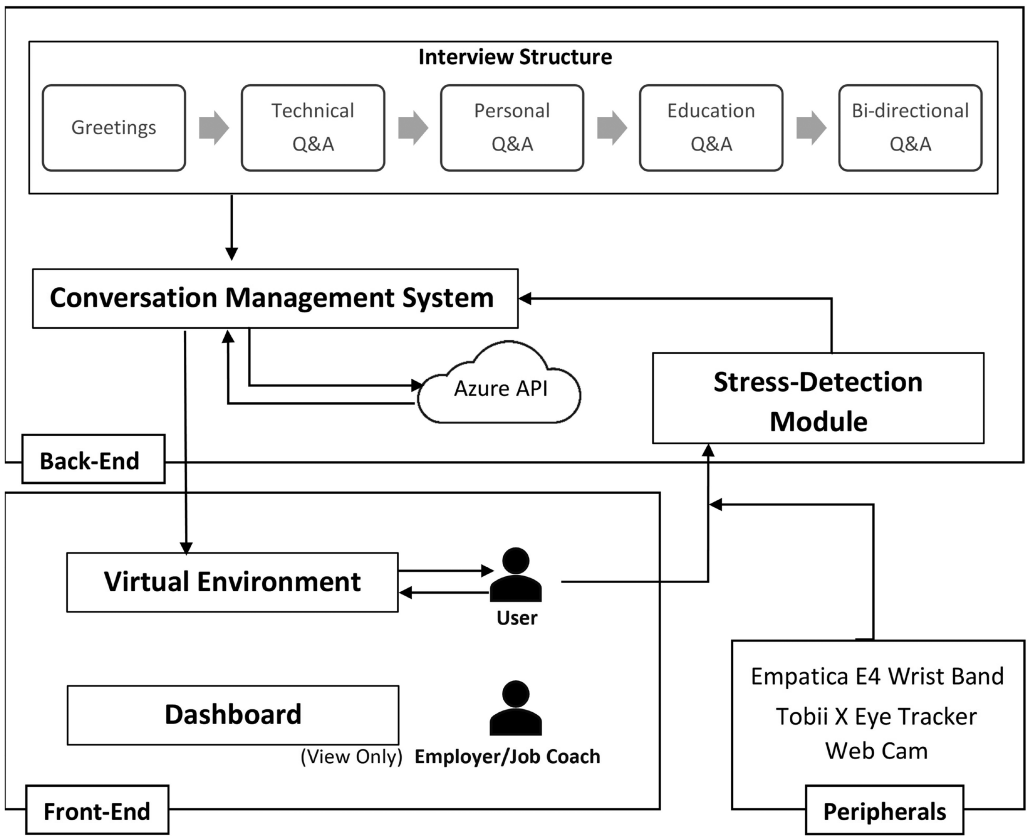

Fig. 1. CIRVR's design comprises four major components: (1) the virtual environment, (2) the Dashboard, (3) the Conversation Management System, and (4) the physiology-based stress detection, eye-gaze detection, and emotion detection modules via peripherals.

background. Additionally, it assesses the interviewee's stress and eye gaze in real-time, with the aim of using this information to guide support and training to improve interview performance. CIRVR's design is based on feedback received from several stakeholders, the details of which are discussed in Section 2.1. The design consists of four major components (Figure 1). On the front end, we have (1) the virtual environment and (2) the Dashboard. The virtual environment is where the interview simulation takes place. The Dashboard displays data collected by CIRVR during interviews, such as eye-gaze patterns, affective information from the interviewee's psychophysiology, and key metrics from facial expressions. These multimodal signals are fundamental to CIRVR's capacity for potential individualized skill training and are discussed in Section 3. On the back end, we have (3) the Conversation Management System (CMS) and (4) the physiology-based stress detection module, gaze detection module using an eye tracker, and emotion detection module using facial image capture. The CMS is responsible for processing the interviewee's speech and for transitioning from one question to another in the interview script. Participatory design with stakeholders, VR design of CIRVR, and the Dashboard are described in Section 2. The multimodal data capture from physiology-based stress detection, eye-gaze detection, and emotion detection modules are discussed in Section 3. Section 4 describes the CMS and interview script design.

To accelerate the development of CIRVR, we decided to leverage existing AI services for speech recognition, speech synthesis, and facial image processing. Specifically, we employed several

modules from Microsoft Azure Cloud Computing Services [4], including their Cognitive Services and Machine Learning Services within the CIRVR architecture. The Face API has been benchmarked in [89], and the Text Analytics API has been benchmarked and proven effective in previous works [20, 24, 61]. Both services have been further used for autistic individuals in [33].

2.1 Participatory Design and Stakeholder Input

The design of CIRVR was both motivated and improved by stakeholder input throughout its design, development, and testing. There were three primary phases in the systematic involvement of stakeholders into the process. During the conceptualization phase, we sought input from several stakeholders — autistic self-advocates, service providers such as job coaches and career counselors, and employers — to understand the typical barriers to interview success for autistic job candidates as well as how a VR-based interview simulator could be designed to be useful for the target population. This was achieved using a semi-structured interview protocol (that itself was developed in tandem with autistic self-advocates) and collecting data from these stakeholders to specifically inform the design and implementation of the VR-based interview technology. We interviewed 23 individuals (including autistic individuals, job coaches and support professionals, and employers) either individually or as part of a focus group. All interviews were recorded, transcribed, and coded for recurring themes regarding barriers and facilitators for job acquisition and retention. During this phase, we also collected survey data from 27 other autistic individuals at the "Autism Inclusion Summit" hosted by the College Autism Network at Vanderbilt University. We obtained several important insights during this phase. For example, stakeholders consistently identified that interviewees often struggled with the interview process in specific ways: incompletely answering questions, struggling to understand the context of unstructured questions, and showing excessive negative emotion in response to unexpected events or questions. Employers, service providers, and autistic individuals all described the need for more coaching and practice with both emotion regulation and stress management (i.e., being able to identify the specific triggers of emotional responses and manage through them) and developing focused, structured, and timely substantive responses to interview questions that demonstrated the requisite skills, knowledge, and thought processes for the job. These insights led to specific modifications in the VR interviewing system by focusing on two specific ways in which autistic individuals tended to struggle: with interruptions and with vague or open-ended questions. We added functionality, in which the interviewer is interrupted by a phone call or another employee while the interviewee is responding to a question, which, in turn, interrupts the interviewee. Initial data show that these interruptions, even in a virtual setting, increase stress and affect performance. Autistic individuals indicated that having a mechanism to express themselves nonverbally would be beneficial, which led to the concept of the Whiteboard. The service providers (i.e., job coaches) and employers indicated the need for quantitative data during interviews so that they could modify interview protocols to accommodate the needs of autistic individuals (e.g., specific questions that elicit stress and lower quality of substantive responses), which led to the development of the Dashboard.

During the development phase, we involved 4 autistic self-advocate interns, 2 males and 2 females, to test various features of CIRVR and provide this critical perspective. We implemented multiple changes as a result. For example, we introduced more text-based instructions to help familiarize interviewees with system functionality and we improved the vocal tone and animation of the interviewer avatar to make the system more engaging. We also refined the graphical representation of the data in the Dashboard. Finally, during the testing phase, we ran the system with 9 autistic participants to obtain their feedback in terms of usability and interaction. During this study, which is described in detail in this article, we further learned what worked and what needed improvement. For example, 7 out of 9 autistic individuals showed preference for the

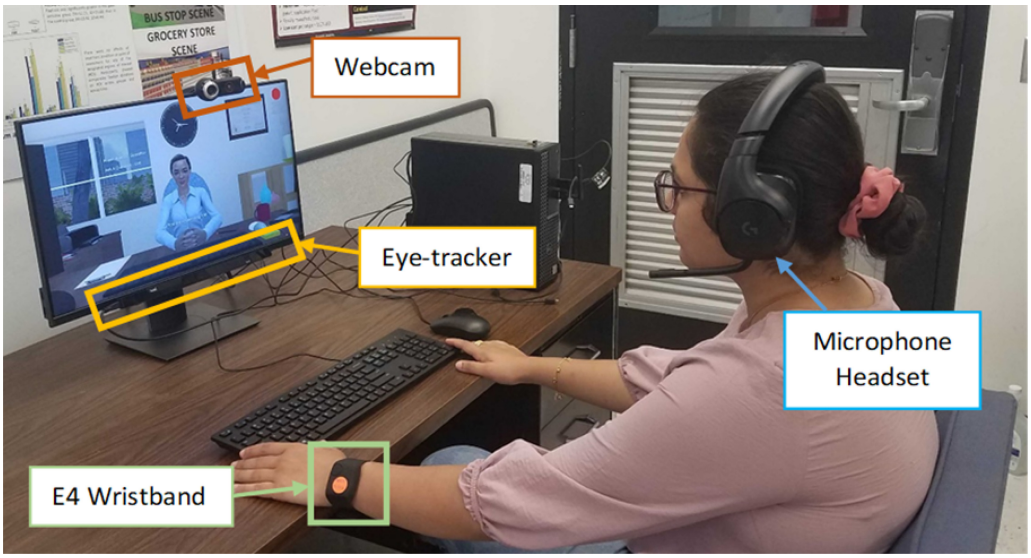

Fig. 2. The interviewee interacts with CIRVR running on a desktop computer, with a webcam for facial image capture and an eye tracker for gaze detection, wearing a pair of headphones with a microphone. The HMD version, not shown here, excludes facial image capture but allows eye tracking through a built-in eye tracker.

non-HMD version of CIRVR, which motivated us to conduct the pilot study without a head-mounted display (HMD).

2.2 Overview of CIRVR

Before we describe the details of the individual components of CIRVR, we present an overview of the flow of an interview simulation and the role of the components mentioned earlier. The interviewee interacts with CIRVR while seated using a desktop computer and standard mouse and keyboard controls, wearing a pair of headphones with a microphone, with or without a VR HMD (Figure 2). While CIRVR is agnostic, we are currently using the FOVE headset by Fove, Inc. [31], and the Vive Eye Pro by HTC Corporation [21], both of which have eye-tracking capability. Note that we have designed two versions of CIRVR —one with an HMD and one without an HMD. The version of CIRVR reported in this article is one without the HMD in order to capture facial image data. The VR HMD version with eye-tracking capabilities has been implemented in parallel and is fully functional. However, its limitation is the lack of facial image capture. At runtime, the interviewee enters the reception area(Figure 3) where the individual is directed by a virtual receptionist (Figure 4) to enter the office where the interviewer is present. The current interview simulation lasts about 12 to 15 minutes and is structured in four categories similar to the in-person mock interviews described by Baur et al. [7]: Greetings (e.g., "Good morning!"), Technical questions and answers (QA) (e.g., work experience and skill set), Education QA (e.g., favorite subjects in school), and Personal QA (e.g., behavior-based questions regarding specific work experiences). The interviewee is given a chance to answer through spoken words and the sentiment associated with the interviewee's utterance (whether positive or negative) or the type of response (e.g., a number such as "40 hours") determines to which parts of the technical questions the CMS will move next. For example, if the interviewer avatar asks, "Will you be comfortable coding for 30 hours a week?" and the interviewee responds with "yes," which is a positive sentiment labelled by the Microsoft Azure Text Analytics API, the interview transitions to the next appropriate dialog. On the other hand,

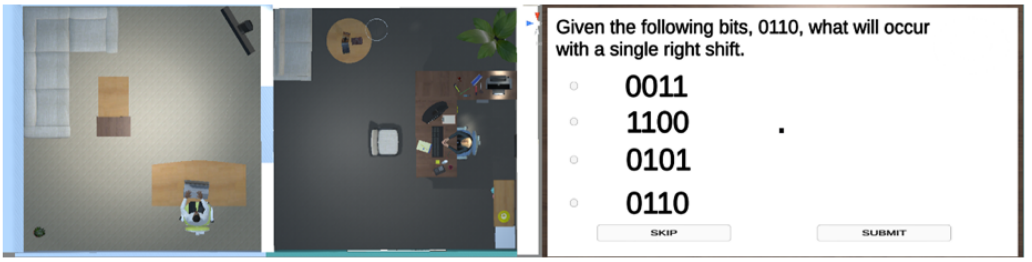

Fig. 3. CIRVR's virtual environment features a reception area and office in which interviews take place (left) and a virtual "whiteboard" for problem-solving tasks (right).

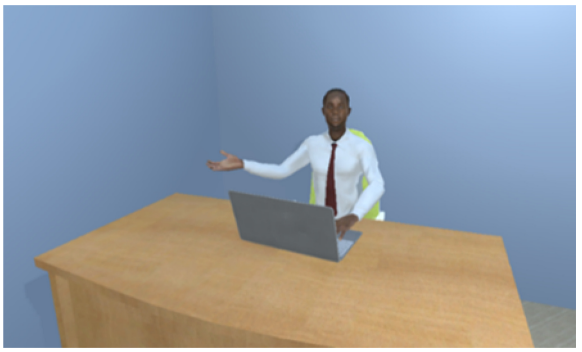

Fig. 4. When the interviewer is ready to begin the interview, the receptionist gestures for the interviewee to enter the office area.

if the response is negative, then the interview may continue to another dialog (see Appendix A.2 for an example). The Education portion of the interview focuses on the interviewee's education, including favorite subjects and academic performance. Interviews often contain specific tests of domain knowledge [40] involving a whiteboard test or paper and pencil to illustrate (diagram) a response. To simulate that in a virtual environment, a set of multiple-choice questions with or without images has been included in the Technical QA and in the Education QA sections in the form of a virtual whiteboard test. Finally, the Personal QA section of the interview is focused on behavioral questions such as asking about the interviewee's reaction to a difficult work situation (e.g., conflict) and the individual's ideal work environment. At the end of the interview, the interviewee can ask basic questions about the job, such as salary, the work environment, the number of vacation days, and so on. The Dashboard is intended for post-interview use, in which the job coach, career counselor, or a potential employer can view data gathered during an interview session through visualization tools, whose details are discussed in Section 2.4. For our current feasibility study, we used a single interview script based on one job domain. In future work, we will be adding more interview content for different types of jobs.

2.3 The Virtual Environment

A virtual office environment was created to mimic a real-world interview setting, including a reception area where an interviewee waits for the interview to begin and an office area in which interviews takes place (see Figure 3). The environment was created using the Unity 3D Game Engine (version 2019.3) [82] with a combination of purchased and custom-built elements, such as

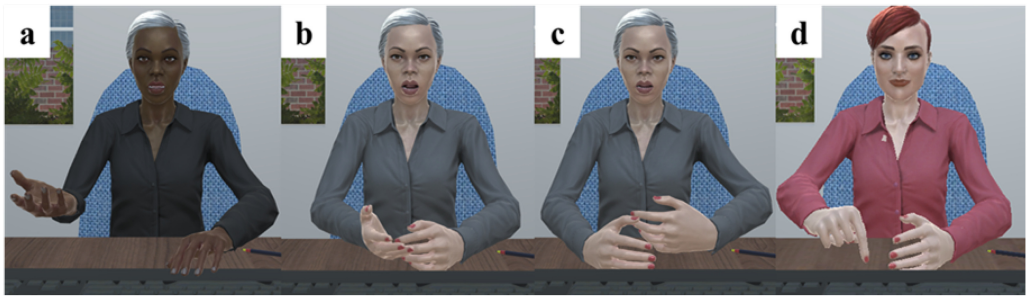

Fig. 5. During interviews, the interviewer avatar features a variety of facial expressions, lip movements, and deictic hand gestures; (a) nominative-deictic, (b) person-deictic, (c) self-deictic, and (d) time-deictic.

models of virtual characters and objects (e.g., furniture), and textures of visual details (e.g., wood, carpet, glass) for an office setting. The current environment has two characters, or avatars: a receptionist and a virtual interviewer. The selection of the avatar models was based on their ability to vary demographic characteristics, including hair, eye, and skin color, as well as hairstyle and clothing (Figure 5). We created animations for the characters to perform certain facial expressions and gestures, and movements that were both relevant to the job interview and would add a high degree of realism to the simulated experience.

To enhance naturalism, the interviewer avatar performs both random and idle animations, such as blinking and making eye movements [77] with slight head movements, and contextually relevant animations, such as lip movements along with socially directed gestures [75]. In real life, people tend to use hand movements during communication to convey a specific meaning [90]. For example, Figure 4 shows a receptionist gesturing towards the office area to direct the interviewee to enter. The type of gestures used in CIRVR are called *deictic gestures,* which are triggered based on specific words (i.e., deictic words) spoken by the avatar [36]. Figure 5 shows examples from each of the four categories of deictic gestures we have used for the avatar: (a) nominative (Figure 5(a)), (b) specific or general person (Figure 5(b)), (c) specific or general self (Figure 5(c)), and (d) past/present or time (Figure 5(d)). Based on this concept of deictic gestures linked to deictic words by Kendon [42], we have programmed the avatar to perform certain gestures based on deictic words that are in the interviewer's dialogue from these four categories. For example, if the avatar asks a question such as "tell me about your work experience," the word *your* will be considered a deictic word and the gesture in Figure 5(b) will be triggered. Words used in the nominative category were chosen based on the content of the interview. For example, our current interview script includes questions about the interviewee's experience with certain tools (e.g., programming languages) and words such as *Java, Python, C, and C++* were labelled as nominative deictic words. For the other three categories, the words chosen were different pronouns: *him*, *her*, *they* for general person, and *you*, *let's* for specific person; *I'm, my, mine* for specific self, and *we, ours, us*, for specific general; and *before, after* for past/present deictic.

2.4 The Dashboard

In an effort to effectively summarize the data captured by CIRVR, we developed a tool for visualization of the captured data, which we call the CIRVR Dashboard. The Dashboard currently displays three categories of data visualizations: (1) graphs of emotional valence, eye-gaze graph, and real-time stress, where emotional valence is obtained from facial expression data, stress is inferred from physiological data, and gaze from eye-gaze data; (2) a scatter plot of the interviewee's eye gaze on a fixed perspective of the interviewer and the environment, to visualize and understand the

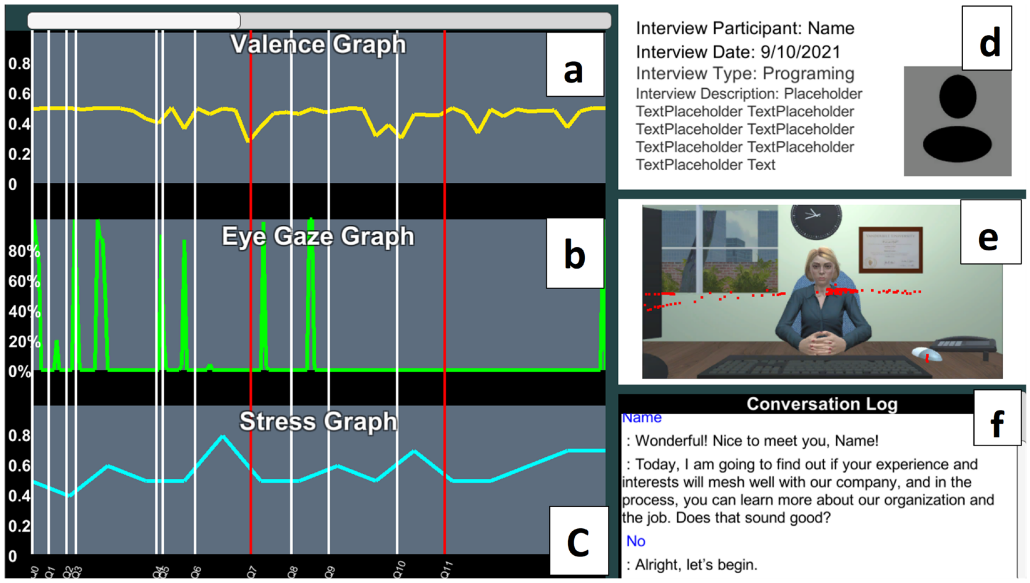

Fig. 6. The Dashboard: (a) valence plot, (b) eye-gaze plot, (c) stress plot, (d) participant (interviewee) details, (e) eye-gaze data, and (f) conversation log.

interviewee's focus areas throughout the interview; and (3) the sequence of dialogues between the interviewee and the virtual interviewer, presented as a chat log. Figure 6 displays an example of the prototype Dashboard. The "valence graph" shown in Figure 6(a) is a visualization of the data collected using the Azure Face API. The Face API processes an image of a face and returns a set of probabilities associated with a particular facial emotional expression from the eight emotions — joy, surprise, fear, anger, disgust, contempt, neutral, and sadness — which are regarded as universally recognized expressions [12, 25]. A detailed discussion on emotion recognition is presented in Section 3.3. The eye gaze graph shown in Figure 6(b) is another visualization of the proportion of time that the interviewee looks at the objects in the virtual environment that are regarded as explicitly task relevant (e.g., the interviewer, the whiteboard) relative to other objects in the environment (e.g., furniture, windows, etc.). The purpose of this visualization is to quantitatively measure and understand gaze patterns of the individuals who use this system. The stress graph shown in Figure 6(c) is a plot of stress data inferred from the physiological data recorded every 5 seconds from the E4 wristband (a detailed description of the stress detection module is presented in Section 3.1). Each of the three graphs in Figure 6 features an overlay of vertical markers, indicated by white and red lines representing specific events that occur during the interview, such as the exact moment that each question is asked by the avatar.

3 HUMAN BEHAVIOR MEASUREMENT

A major aspect of this novel platform is its capacity to measure, interpret, and respond to real-time human signals (i.e., measurements collected from the interviewee through sensors). In the current work, we have incorporated several human behavior measurement modalities: (a) physiological signal measurement to predict high or low stress, (b) eye gaze measurement to infer visual gaze patterns, and (c) facial image capture to measure head orientation and to predict emotional expression. These modalities were selected for their potential insight into important elements of the human–computer interaction, such as eye gaze, mental workload, and emotional states [46], all of

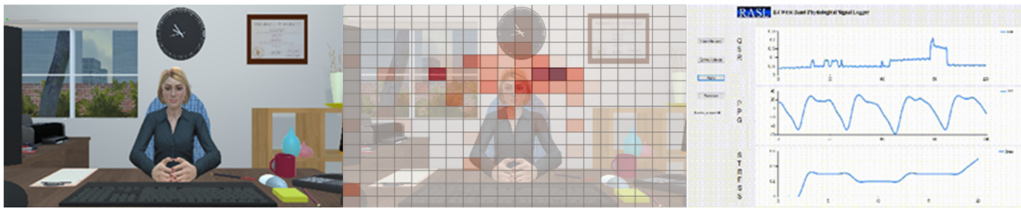

Fig. 7. The interviewer avatar as seen from the perspective of the interviewee (left); visualization of eye-gaze data overlaid onto the first-person perspective (middle); visualization of real-time stress measures that are captured using the Empatica E4 wristband sensor (right).

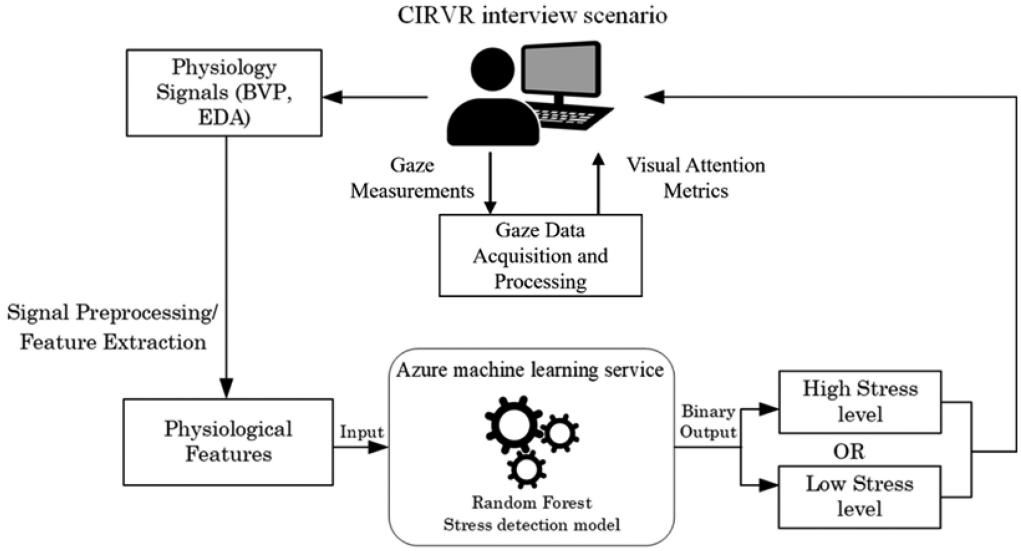

Fig. 8. Real-time physiology-based stress detection and gaze data acquisition modules.

which are relevant in the context of a face-to-face job interview. Once CIRVR is deployed in the target population, these measurements will facilitate a greater understanding of the state of the interviewee as well as how the interviewee interacts with CIRVR and with the interviewer. The details of each of these measurement modalities are given in the following sub-sections.

3.1 Physiology-Based Stress Detection

We endow the virtual interviewer with the capacity to "empathize" with the interviewee by means of affective computing [60]. In this work, physiological responses were used to detect the stress level of participants. Literature has shown that affective states, including stress, can be derived from physiological responses [37, 64]. Physiological responses have several advantages over other modalities such as facial expression, body gesture, and voice. For example, physiological responses come from the Autonomous Nervous System (ANS) and are, thus, generally involuntary responses reflecting the true state of the participant [10]. This is particularly beneficial for the current work because autistic individuals often show atypical social behavior in the form of facial expressions, gaze, and body gestures [1].

3.1.1 Data Collection. To build a stress detection model for CIRVR, we conducted a study with NT adults to collect labelled data for training. This research was approved by the Institutional Review Board (IRB) of the institution where this work was performed. Specifically, we designed a task based on the Modified Computerized Paces Auditory Serial Addition Test (PASAT-C) [47] to elicit mild stress among participants.

3.1.2 Feature Extraction. We collected blood volume pulse (BVP) and electrodermal activity (EDA) using the E4 wristband by Empatica Inc. [27]. The sampling rates for BVP and EDA were 64 Hz and 4 Hz, respectively. These two signals have strong correlations with one's stress level [18, 58, 63]. Ten adults volunteered to participate in this stress elicitation study. Data from level 1 of the stress task were labeled as "low stress" and the data from level 3 were labelled as "high stress." The data were preprocessed and segmented into 1-minute samples. Next, several features — heart rate, heart rate variability, and skin conductance response rate —were extracted from these data segments. Previous studies have demonstrated that these features are reliable indicators of emotional states, including stress [44, 58, 63, 84].

3.1.3 Model Training and Testing. After feature extraction, a binary stress detection model was built using the Random Forest algorithm, which has been shown to be an effective algorithm for this application in our previous work [11]. In order to ensure the generalizability of the model, the leave-one-subject-out cross-validation method was used to determine the optimal hyperparameters [9]. The overall accuracy for this model was 84%, which is on par with the state-of-the-art results in the field [79]. Finally, the stress detection module was deployed to the cloud through the Azure Machine Learning service. From CIRVR, (i.e., the Unity application), the model is called as a web service once per minute and the model outputs the current stress level of the participant (Figure 8). Figure 7(c) displays real-time logging of signals from the E4 sensor. The first two graphs represent two physiological signals — galvanic skin response (GSR) and the photoplethysmogram (PPG). GSR provides a measure of the resistance of the skin and the PPG measures the blood volume in a participant's wrist. From these two signals, the stress level is computed using the Random Forest algorithm and the stress graph is generated (third from the top in Figure 7(c)).

3.2 Eye-Gaze Detection

Eye-gaze data have been shown to provide direct insight into the interviewee's visual patterns and mental workload and are typically interpreted based on metrics such as fixations, saccades, and smooth pursuits [46]. In CIRVR, eye-gaze data are continuously collected throughout the interview in Unity (Figure 8) via the Tobii Eye Tracking Software Development Kit on Unity [76]. CIRVR currently supports two eye trackers — the Tobii EyeX (60 Hz) and the Tobii 4C (90 Hz) — that boast high accuracy and small form factors [83], and Tobii eye trackers have been benchmarked in previous technologies for autistic children and individuals [52, 65]. The coordinates of the interviewee's focus position are detected and recorded to a comma-separated value (CSV) file with a timestamp so that gaze data can be cross-referenced with key events such as the start and end times of the individual interview questions. Figure 6(e) in Section 2.4 shows a feature of CIRVR that visualizes this gaze data by plotting the gaze points onto the coordinate frame or view seen during the interview that can be interpreted as a plot with an x-axis and y-axis on which (x, y) gaze coordinates are marked. Salient regions of interest (ROI) are drawn onto this coordinate frame and the resulting plot can be used to determine how many points fall within specific ROI, a visualization method discussed in [13]. These ROI currently include areas such as the interviewer's face and eyes to determine levels of eye contact and focus on the interviewer that were achieved during the interview. Figure 7(b) shows an upgraded version of this visualization that is part of

our Dashboard web application. It highlights the ROI in the form of a heat map created using Nivo [56], a web application library for data visualization.

3.3 Facial Image Capture–Based Emotion Detection and Head Orientation

A wide variety of measures can be collected from images taken of the face. Facial expressions can reveal neutral affect as well as emotions such as joy, surprise, fear, anger, disgust, contempt, neutral and sadness [12, 25], as mentioned in Section 2.4. For every image, the facial expressions or emotions have a value between 0 and 1 returned by the Face API, representing the probability. These probability values are then assigned a positive (>0), negative (<0), or neutral (=0) emotion value, from which the weighted sum is presented as a time series signal on the Dashboard. Other measures that can be inferred from facial images include head orientation (i.e., roll, pitch, and yaw angles of the head) and demographic information such as age and gender. In the current work, we used facial image capture for facial emotional expression, head orientation, and demographics.

Facial image capture is performed using a Logitech C920 HD Pro web camera at a sampling rate of once every 5.5 seconds, although this parameter can be adjusted as needed. When an image is captured, the Face API is called from within the Unity application and the metrics — eight emotional expression probabilities, gender, the presence of glasses, and the head pose, including pitch, yaw, and roll — are returned by the Face API as a JSON-encoded object. A CSV file is produced, which includes the returned metrics as row entries in a time series log. In future work, this time series data can be analyzed either online or offline to infer meaningful changes in the state of the interviewee during the course of the interview.

Note that although CIRVR is endowed with the capability of facial expression recognition, the exactness of recognition will be dependent on training datasets that include images of autistic individual's faces. Currently, Azure training datasets are not trained with those images; thus, the detected facial expressions should be treated with caution.

4 CONVERSATION MANAGEMENT SYSTEM

In order to have more flexibility in terms of content development, maintenance, and future expansion of CIRVR, we needed to create a "bidirectional" chatbot that would initiate interview questions and get responses and, at the same time, would allow an interviewee to interrupt the bot at any time. Thus, we created our own CMS that would allow us the flexibility to add interview content from different domains for a variety of job types and that would allow the interviewee to ask the interviewer questions, as well. The design of the CMS is adapted from a task-oriented spoken dialogue system [41]. For programming the CMS, we used a mediator design pattern [32], a type of object-oriented software design that allows objects to communicate with one another without having to know the details of each other's implementations. This approach to object-oriented software development reduces the complexity of maintaining and extending code, which is ideal for the purpose of ongoing development and expansion of each component independently. The basic components of a dialogue management system are (a) speech recognition, (b) language understanding, (c) dialogue state tracking, (d) dialog policy determination, (e) natural language generation, and (f) text-to-speech conversion [41].

Figure 9 illustrates the two parts of the system. The Conversation Manager incorporates the aforementioned components and the Conversation Context contains the *Dialog* class, which incorporates a single verbal exchange between the interviewer and the interviewee from the interview script (see Section 4.6), and the *UserState*, which tracks the physiological information such as the stress level and the emotion analysis, which will be used in future work for dialog state changes as an additional feature to enhance the real-time closed-loop interaction. Both parts work together in the order (a) to (f) to form the dialogue system.

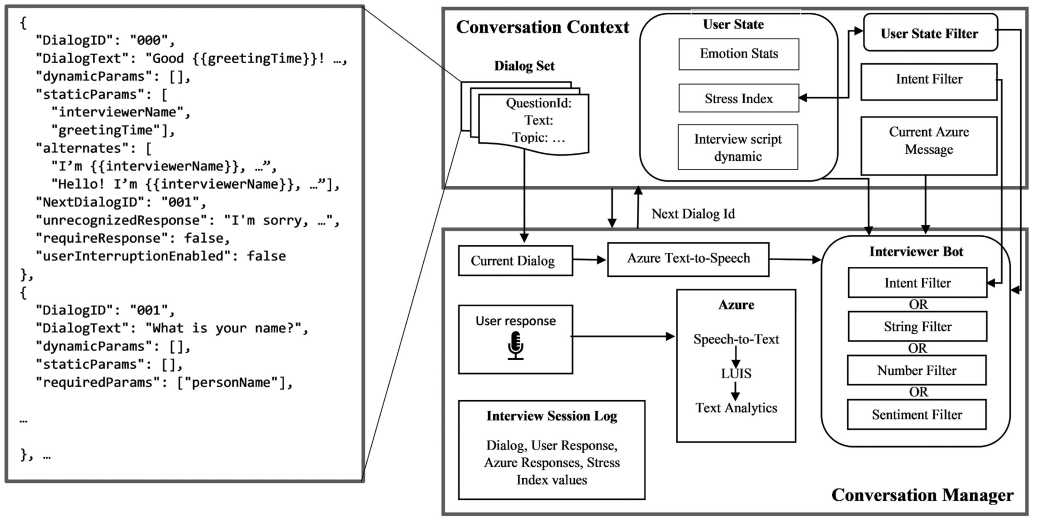

Fig. 9. The major elements of the Conversation Management System include the Conversation Manager and Conversation Context, which includes a Dialog Set (a collection of *Dialog* objects that represent a semi-scripted interview) and a *UserState* (a collection of attributes about the interviewee, including verbal response values and physiological state measures).

4.1 Speech Recognition, Dialog State Tracking, and Language

When the interviewee sits across from the interviewer in the virtual office, the CMS initiates the first *Dialog* object (Figure 9) that contains information from the interview script in JavaScript Object Notation (JSON). This dialog text is converted from text to speech using Azure's Text-to-Speech service and is spoken by the interviewer avatar. The interviewee is given a chance to respond as desired; the spoken response is then transcribed into text by Azure's Speech-to-Text service. This text is then sent for analysis by Azure's Language Understanding (LUIS) and Text Analytics services. LUIS classifies parsed text as intents (i.e., interviewee-initiated requests, such asking the interviewer to repeat a question) and entities (i.e., discrete keywords or values) [45]. Together, entities and intents are used to construct a shallow, semantically meaningful representation of the interviewer's utterance. The Text Analytics service performs sentiment analysis, key phrase extraction, language detection, and named entity recognition on a document of text. Our CMS relies on sentiment analysis and entity recognition to capture data that is relevant in directing the flow of the conversation. For each interviewee response, Azure's Sentiment Analysis returns sentiment scores between 0 and 1 (close to 0 is negative and close to 1 is positive) and labels (such as "positive," "negative," or "neutral"). From this information, the CMS gauges "yes" or "no" responses if scores and the label imply positive or negative, respectively. Entity recognition is used to identify entities from the interviewee's response, for example, programming languages such as C# or Java. All interviewer and interviewee exchanges are logged in a CSV file.

From the current *Dialog* object (see example in Figure 9), based on the interviewee's response, the next dialog identification string is returned to the Conversation Context, and the next question content is sent to the Conversation Manager, which is then spoken by the virtual avatar, and the flow continues. The *UserState* class in the Conversation Context is responsible for keeping track of the important attributes of the interviewee, including the stress level, emotions based on facial expressions, gaze pattern, and information about the interviewee such as the individual's name and stored entities from previous responses (e.g., key words about the interviewee's skillset).

Upon receiving results from LUIS and Text Analytics, the *UserState* is updated along with any information gathered from the conversation transcript (e.g., length of utterance), any entities or intents, and the predicted text sentiment. The Interviewer Bot class then generates a response in text by passing the interviewee's last response through a series of filters to extract relevant information that needs to replace the placeholders in the text, which is then converted to speech using Azure's Text-to-Speech converter. The filters and their purpose are discussed in detail in the next subsection.

4.2 Filters for Dialog Transition

To determine the next interview dialog or dialog policy, CIRVR employs three filters to process the interviewee's responses: the intent filter, sentiment filter, and number filter. The intent filter checks to see whether the interviewee utterance contains registered intents such as "skip this question," "exit the interview," and "please speak up." If the checks are affirmative, the interviewer reacts with the associated predefined action for those intents. Whether the number or the sentiment filter is applied to the interviewee's response is determined by a predefined attribute in the *Dialog* object that specifies the type of expected responses. If an expected response has a number, then the number filter is applied where the value is checked against a threshold that determines the next response (e.g., an expected response could be "30 hours of work per week" in response to a question, "How many hours of coding per week will be okay for you?"). On the other hand, if the expected response is in text, then the sentiment filter is applied whose output determines the next dialog; the sentiment score is used to gauge positive or negative (e.g., "yes" or "no") responses (see Appendix A.2). For feasibility testing, the threshold values for the number filter are hard-coded in the system, which will be modified in future work to be flexible to accommodate interview content of different domains. In the case in which the system is unable to understand the interviewee (either the interviewee did not speak or expected information was not detected (via entities extracted), the interviewer says "I'm sorry, I don't understand" and repeats the question at most three times before moving on to the next question.

4.3 Interruptions

Another feature of CIRVR includes the addition of interruptions during the interview. Interruptions are often linked to physical stress, emotional exhaustion, and anxiety [48]. Although there are several ways to interrupt an interview, CIRVR implements interruptions where the unembodied voice of the receptionist is heard after a "knock on the door" sound. The receptionist informs the interviewer of something that halts the interview for approximately 10 seconds. This type of interruption is an intrusion, which requires disengagement from a current task before engaging with another [29]. An intrusion is intended to indirectly interrupt the user and disrupt the user's thought processes or sense of what can be expected during the interview. Such intrusions can be especially difficult for autistic individuals since they often require longer time to process information [36, 80].

4.4 Customizable Vocal and Facial Characteristics of the Interviewer

A feature that contributes to CIRVR's closed-loop interaction is the adjustment of facial expressions and vocal tone of the avatar. We developed a range of vocal tones, such as cheerful, formal, and neutral, using Speech Synthesis Markup Language (SSML) to make the avatar sound as human as possible [81]. These vocal tone variations are based on specific SSML configurations that are predefined using the neural voice option in Azure's SSML implementation by adjusting acoustic parameters such as pitch, volume, rate, and prosody. Each parameter impacts both the sound of the avatar's voice and the experience of the user. For example, if the speaking rate of the avatar is

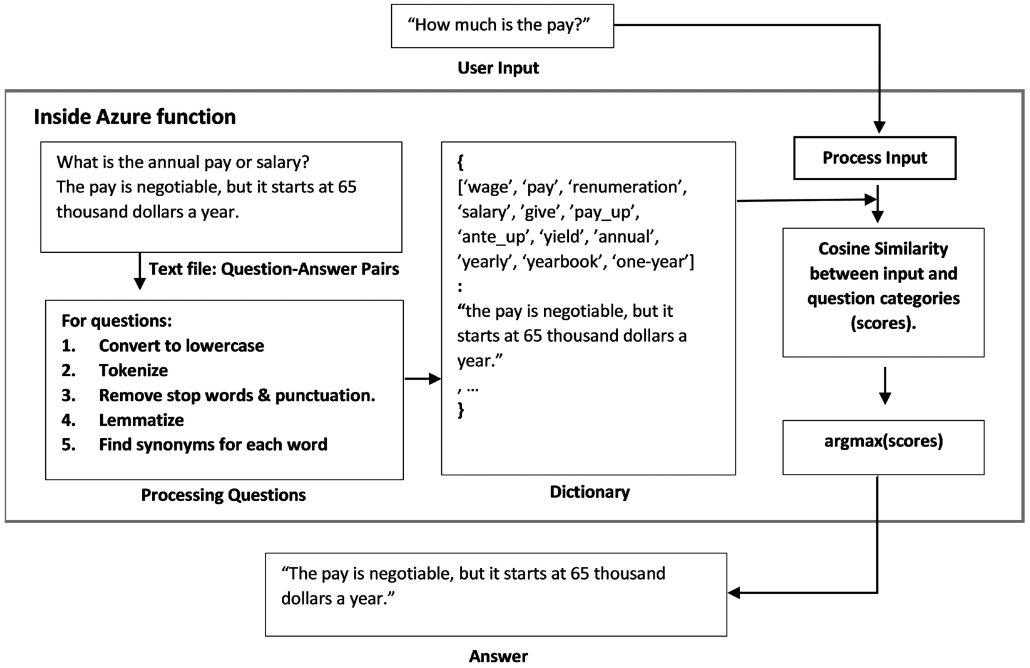

Fig. 10. Structure of the response generator for the bidirectional questioning using a sample example.

increased, that is, the avatar talks quickly, it might make it challenging for the autistic interviewee, who may need to hear the question repeated.

In addition to the vocal characteristics, facial animation has been used to achieve limited realism in conversation. The interviewer avatar model has facial animation blend shapes, which are essentially predefined poses for the mouth region that can be iterated to give the illusion of animation for the mouth region. Specifically, there are three different blend shapes: one that affects the intensity of an open-mouth smile, one that affects how open the mouth is, and another that affects the intensity of a closed-mouth smile. They allow changes in the facial expressions throughout the interview during times of low, medium, or high stress.

4.5 The Bidirectional Component of the Interview

As discussed earlier, our intention was to allow interviewees to ask about the job, the salary, or about the company where they are interviewing, as in a real-world interview [30], and provide flexibility to a developer or job coach to modify these questions and corresponding answers while including interview content for another job domain. Figure 10 illustrates how this component is implemented. A set of question-answer pairs are defined in a sequential format in a text file that is stored in the Unity application directory where it is read in the event of an interviewee-initiated question. The program to generate a response to the question is written in Python and stored as an Azure Function to which an API call is made on the Unity side that sends the contents of the text file and the user input to the Azure service. The process from user input to final answer is described as follows. The Python program reads the question-answer pairs from the text file and stores them in a Python dictionary or associative-array as key-value pairs. For each key-value pair, the key (question) is processed using tools from Python's Natural Language Toolkit (NLTK) [49].

The NLTK provides a tokenizer that is used to split the sentence into a list of words and punctuation. After tokenization, stop words, or commonly used words — such as "the," "an," and "a" — are removed from the list of tokens. After this step, only keywords remain in the list. The NLTK provides a WordNet [28] lemmatizer that uses vocabulary and morphological analysis to reduce a word to its base form, such as converting plural to singular (e.g., "studies" becomes "study"). After lemmatization, synonyms are extracted from WordNet for each word in the list and stored in another. This list of synonyms of all of the keywords replaces the key in the dictionary and each key-value pair is now composed of the list of synonyms of relevant keywords along with the corresponding answer as the value. The user input question is processed in the same manner using the four steps outlined earlier and the final input sentence is a list of synonyms of keywords extracted from the input. This list is compared with each key (i.e., the synonym list) in the dictionary and a similarity score is computed using the cosine similarity algorithm [35]. Finally, the key that is similar to the user input is assigned the maximum similarity score. The corresponding value (answer) of this key is sent back to the Unity application, which is converted to speech and is output via the avatar. If there is no similar question to the user input, then a general response is returned such as "Sorry, I don't have an answer to that."

A machine-learning approach for this type of response generation would have required a large dataset for training and testing of a model. Our approach allows a potential interview content creator the flexibility to alter and add questions in the text file for different interview contexts. It does not matter what the interview domain might be, as the approach relies on a similarity score for the output. This implementation is an initial version of this bidirectional conversation component. In some cases, depending on how the interviewee worded the question, there could be ambiguity, possibly due to overlapping words in between questions in the text file, which could lead to incorrect output. In future enhancement of this implementation, it might be useful to incorporate parts-of-speech tags and/or word embeddings such as Glove [59] along with keyword extraction to calculate the cosine similarity score that considers context.

4.6 Interview Script Design

Another key objective in the design of the CMS, and the CIRVR platform more generally, was the inclusion of support for customizable interview content, enabling content creators to define as much of the interview structure and data capture as possible. To this end, we used JSON to create interview scripts based on a simple file format. These files are loaded into the CMS before the interview begins. Each dialog or *Dialog* object in the script contains a set of attributes that include the question, the response timeout variable, alternate versions of the question, criteria for advancing to the next dialog, and a set of dynamic and static parameters to be inserted into the question text where appropriate (Figure 10).

To provide a starting point for the development of CIRVR's CMS and interview design, we chose to draft an interview script for a very specific job — a software developer for a video game company — to serve as a guiding template throughout development and feasibility testing. This initial choice was based on the fact that technology-related jobs such as this are often sought by autistic individuals, who quite often have a strong affinity for technology [54]. The general approach to designing the interview structure was based on the literature describing a basic progression through topics covering work experience, educational background, and behavioral information followed by an opportunity for the interviewee to ask questions as described in [81]. Aiming to create an interview lasting approximately 15 minutes, we created a script containing over 50 dialog exchanges. Appendix A.1 shows two examples of interviewee-interviewer exchanges showcasing two different kinds of entity recognition used by CIRVR.

5 FEASIBILITY STUDY

Seventeen participants, both autistic and NT, were recruited for a study of feasibility of the prototype job interview simulation system. It is to be noted that the scope of this study was to assess whether CIRVR was acceptable to users and functioned as designed. Feasibility was operationalized as the successful capture of multimodal data (i.e., speech, gaze, physiology, and facial expressions), and favorable levels of user-reported system usability and overall user experience. All study procedures were conducted with approval from the IRB at the university where the research was conducted, and informed consent was collected as applicable for all participants. The only inclusion criterion for this study was age eligibility for full-time employment in the state where this research was conducted (16+ years). Each of 9 autistic individuals (mean age = 22.11; SD = 9.10) and 8 NT individuals (mean age = 23.00; SD = 6.68) participated in a single session. All autistic participants had a clinical diagnosis of ASD from licensed clinical psychologists based upon *DSM-5* criteria. Each session lasted approximately 1 hour from the moment participants arrived at the research facilities. The session consisted of informed consent/assent, sensor calibration, the virtual interview, post-interview survey items, and a discussion between the researcher and participants regarding user experience and qualitative feedback. Possible adverse effects of the study include the possibility that participants might experience increased anxiety given the nature of the tasks (i.e., conversing with a virtual interviewer) and that wearing the sensor and microphone headset may also cause some level of discomfort. No other adverse events were expected. One autistic participant requested to end the virtual interview portion of the study early due to reported discomfort during the simulation, but still wished to take part in the post-interview components of the session. No specific reason was given by the participant to explain the discomfort and the general response was "I don't know."

Given the small sample size, non-parametric statistical analyses were applied to the gathered feedback using the biostatistical analysis software MedCalc, version 19.3.1 [66]. Inferential tests were conducted using independent sample Mann-Whitney U tests and effect sizes were estimated using Cohen's d. Statistical significance was benchmarked based on a critical alpha of .05 and effect sizes were interpreted based on widely used benchmarks in the literature [26]. Based on a Mann-Whitney U test, participant age did not differ significantly between ASD (median = 18 years) and NT participants (median = 22 years; U = 30.5, p = .592, Cohen's d = 0.11). Table 1 gives a detailed comparison of both groups across a number of variables.

Regarding the feasibility of multimodal data capture, data were successfully captured for all participants across all channels except in two cases in which both real-time stress and gaze data were lost due to recording errors. Fortunately, these issues were detected at the beginning of the study; after being addressed, no further data were lost. All speech transcripts were successfully captured in log files that contained the sequence of dialogs between the interviewer and the interviewee. For all participants, the virtual interviews were completed with all dialogs delivered successfully. CIRVR uses an event-logging system that records information about the interviewee's vocalizations. Preliminary testing demonstrated that CIRVR successfully captured interviewee's vocalization data, which included a transcript of the speech, intent type, detected entities, pause length, and response length. Facial image capture was achieved at a rate of one image taken every 5.5 seconds, which was sufficiently fast for the purposes of this feasibility evaluation. As previously noted in Section 3.3, this frequency can be adjusted based on research objectives. Finally, physiological and gaze data were successfully captured at frequencies specified by the manufacturers, with no data lost apart from the aforementioned instances. Further evaluation was conducted on the CMS. The first author, in the presence of a behavioral analyst, analyzed the interview log files of autistic participants to evaluate how well the CMS was able to understand interviewee responses and

Table 1. Group Comparison Across Variables

Independent Variable	Participants	
	ASD (N = 9) M (SD)	NT (N = 8) M (SD)
Participant age	22.11 (9.10)	23.00 (6.68)
System Usability Scale (SUS) Composite Score[2]	53.61 (17.24)	76.25 (15.00)
Item 1[3]**:** *I think that I would like to use this system frequently*	2.56 (1.13)	3.25 (1.16)
Item 2: *I found the system unnecessarily complex*	3.00 (1.41)	1.88 (1.46)
Item 3: *I thought the system was easy to use*	3.56 (1.33)	4.50 (1.07)
Item 4: *I think that I would need the support of a technical person to be able to use this system*	3.11 (1.17)	2.88 (1.25)
Item 5: *I found the various functions in this system were well integrated*	3.22 (1.09)	4.38 (0.74)
Item 6: *I thought there was too much inconsistency in the system*	2.67 (1.32)	1.88 (0.83)
Item 7: *I would imagine that most people would learn to use this system very quickly*	3.44 (1.13)	4.63 (0.74)
Item 8: *I found the system very cumbersome to use*	3.33 (1.12)	2.13 (1.13)
Item 9: *I felt very confident using the system*	3.00 (1.22)	3.75 (1.04)
Item 10: *I needed to learn a lot of things before going with this system*	2.22 (1.20)	1.25 (0.46)
Self-reported level of comfort during the interview[4]	3.00 (1.31)	3.88 (0.99)
Self-reported level of confidence during the interview	3.00 (1.07)	3.75 (1.28)

[2]For the computation of the SUS composite score, see [15].
[3]All individual SUS items were scored on a 5-point Likert scale (1 = strongly disagree, 5 = strongly agree) [5].
[4]Self-reported comfort and confidence levels were scored on a 5-point Likert scale (1 = "very uncomfortable/very low confidence, 5 = very comfortable/very high confidence).

transition from question to question, accordingly. It was found that the CMS was able to understand the interviewee responses 93% of the time. The stress values received from the autistic participants during the study were labelled by a trained expert to form the ground truth dataset. The Random Forest Model accuracy for stress detection was 84.5%.

Following each virtual interview, participants completed two short surveys and then spoke with the researchers about their experiences using the system. The first survey was the System Usability Scale (SUS; see Table 1 for details). The SUS is a widely used measure that relates to both technology acceptance and usability and is scored on a 100-point scale [15]. Participants were asked to carefully read each of the SUS prompts before providing responses. Participants were also free to ask clarifying questions about the instrument as needed. The second survey was a researcher-defined pair of questions relating to self-reported comfort and confidence based on a 5-point Likert scale in which 1 indicates very low comfort/confidence and 5 indicates very high comfort/confidence. As with the SUS, participants were free to ask clarifying questions before providing responses to the prompts. Following each session, the researchers entered the captured survey data into a primary spreadsheet — accessible only to key study personnel — for offline processing.

A Mann-Whitney U test revealed that ASD participants reported significantly lower overall perceived usability of the system (median SUS = 57.5) in comparison with NT participants (median SUS = 77.5) with U = 9.5, p = .011, Cohen's d = 1.40. An empirical analysis by Bangor et al. [5] provides benchmarks for contextualizing and interpreting the SUS scores. Based on these benchmarks, autistic individuals and NT individuals regarded system usability as "Okay" and "Good,"

respectively. This disparity in perceived usability is consistent with our past work (e.g., [86]) and may be emblematic of greater levels of experienced stress by autistic individuals during the virtual interviews.

A Mann-Whitney U test showed ASD participants' self-reported comfort during the virtual interview (median = 3), in comparison with NT participants (median = 4), was nominally lower with medium effect size ($U = 19.00$, $p = .157$, Cohen's $d = 0.75$). Similarly, ASD participants' self-reported confidence in their performance during the virtual interview (median = 3), in comparison with NT participants (median = 4), was nominally lower with medium effect size ($U = 18.50$, $p = .145$, Cohen's $d = 0.64$). As described in [36, 57, 62], there exist differences in comfort regarding social interaction that are characteristic of autistic individuals [1]. Maras et al. [51] conducted a study to compare the performance of autistic versus NT individuals in two types of mock interview scenarios in which one phase of the study included a conventional employment interview and another was based on an adapted version of the initial interview questions based on feedback from employers as well as interviewees. They found a similar difference in the first phase of the study in which autistic individuals reported more difficulty than their NT peers in communication challenges as well as cognitive difficulties such as processing questions and recalling a specific memory. The lower scores in levels of comfort and confidence from our study suggest the need for more practice opportunities to familiarize autistic job candidates with real interview questions and to gather additional information about their communication abilities to not only help them understand and improve their performance but also for potential employers to understand how they can help their autistic job candidates by adapting interview strategies. For example, CIRVR allows users to repeat a question if they find it unclear, and the repeated question can be worded differently than the first for more clarity. This feature can help interviewees practice asking help from the interviewer in a real-life interview scenario when they need a question reworded or clarified.

Following the surveys, participants engaged in a brief discussion with researchers about their experience with the system, which allowed us to obtain qualitative feedback to contextualize the survey responses. Among ASD participants, a commonly cited point of frustration was with the interruption that takes place during the simulated interview (i.e., when the receptionist knocks on the interviewer's door to announce that a call is waiting for her). As described earlier, differences in cognitive abilities, such as processing time and memory recollection, are characteristic of autistic individuals and have been observed and reported in interviews that are not only job related [36, 57]. This feedback from the participants provides validation that the interruption feature is, in fact, perceived as disruptive. We believe that CIRVR can be useful to autistic participants who may seek to practice resilience to disruptive situations. Despite reporting frustration with this particular aspect of the system, participants across both groups reported being satisfied and impressed with CIRVR's conversation management system and the whiteboard feature, noting that the structure of the interview transitioned smoothly from one topic to the next. Participants also noted areas in which the system could be improved in the future. For instance, some participants noted and reported that the virtual interviewer's voice was a bit "robotic" and one participant felt as though she were "talking to Siri." This feedback is valuable, as in future work we will address this by adjusting voice characteristics of the interviewer avatar through SSML. Based on our cumulative findings, however, we believe that CIRVR is functionally robust and ready for a larger evaluation with the target population.

6 CONCLUSION

Presently, the under- and unemployment of autistic people is a major challenge. Research and innovation aimed at addressing this challenge, such as our novel CIRVR platform, may offer support to help this population enter the workforce. The VR-based platforms that we surveyed

[7, 8, 16, 17, 22, 23, 68–71, 73, 78, 85, 88] were aimed at training individuals to perform better in social interactions and demonstrated reliability in VR as a medium. The eye-gaze data can be visually represented for the employer to understand the candidate's nonverbal response patterns. The analysis of stress levels and emotion data during specific questions can give employers an understanding of participants' emotional and mental states during different sections of the interview, which, in turn, could support employers in altering and adapting their job interview process to support the needs of autistic individuals. Our findings from this feasibility study demonstrated the technical feasibility of the novel platform as well as preliminary evidence of acceptability with participants, setting the stage for a large-scale evaluation of the system with the target population in the future.

While implementation of the prototype technology was a success, there are a few key limitations of this work that must be addressed before CIRVR can be applied to addressing the employment barriers faced by autistic individuals. First, while we demonstrated the usability and features of CIRVR in the current study, the efficacy of the CIRVR prototype has not yet been demonstrated and remains the most important next step. In our next study, we will be testing the feasibility of our Dashboard with the interview system with employers, job coaches, and autistic participants. We will then seek to demonstrate the efficacy of CIRVR by measuring any improvements in participant stress levels and interview performance. Second, the validity of the physiology-based stress detection module must be further tested as the work progresses due to the small dataset that was used to train the predictive model. Also, this initial model was trained on data from NT participants. We plan to address this issue by incorporating data collected in the feasibility study to produce a refined model that includes data collected from autistic participants. In addition, a web-based Dashboard is currently under development that will include enhanced visualizations of the data collected. We are currently developing a mechanism to utilize the metrics that are continuously logged during each interview — stress, gaze patterns, and facial expressions — that would determine the transition of questions based on the interviewee's state during the interview, such as using the stress level to transition to a different set of questions. The stress-induced dialog transition will be implemented as a filter that will work in a manner similar to other filters discussed in Section 4.2. If the user's state attributes meet a specific criterion, for example, significantly elevated stress, then the interviewer will be programmed to react appropriately to this information. If the detected stress level is judged to be "high," the interviewer will transition to a different set of dialogs or questions that may be less stress inducing in the hope that the interviewee will calm down and then return to the more challenging questions later. This will be part of CIRVR's closed-loop adaptivity in the future development of CIRVR. We believe that employers, job coaches, and other professionals involved in supporting employment for autistic individuals could use the Dashboard as a tool to gain insight into how autistic individuals feel during an interview as it can help identify interview questions that cause increases in stress or frustration. It might help understand that the nonconformity of members of this population to normative social behavior may not be an indication of their lack of knowledge but rather due to differences or deficits in social interaction. A third limitation is the current lack of feedback provided by CIRVR to the interviewees. Although we have incorporated a feedback mechanism within our software development, the actual content of the feedback needs to be carefully developed in consultation with job coaches when more interaction data are available. CIRVR is still in its initial phases; more data from participants will be collected that will include their responses to questions. The responses will be analyzed by job coaches to help define measures of performance that will then be used to provide feedback to the interviewees and will be part of future work. Now that we have demonstrated the feasibility of our system, another area of future work will be to begin the implementation of a custom content-authoring tool to facilitate the task of creating new content for a wide range of job

interview scenarios beyond the software developer role examined in this research. Currently, CIRVR can be deployed as a desktop application. However, we will be creating a web-based version of CIRVR in future work to allow for accessibility and scalability of the system. This scalable version will allow for deployment at vocational agencies for a longitudinal study on the efficacy of CIRVR. Last, we would like to report that a limitation of the Face API performance of the Azure Face API model on autistic individuals is not known. In order to evaluate that, we will need a trained behavioral psychologist to annotate a subset of images, which is a time-consuming, costly, and laborious process. Evaluation of the Azure facial emotion recognition model will be part of future work.

Despite the limitations just discussed, we believe that CIRVR could potentially be an effective job interview training platform for autistic individuals as well as the employers who are interested in exploring and benefitting from the talents of these individuals. As mentioned in Section 1, the development of CIRVR does not aim to ask autistic individuals to conform to the norm and behave like their NT peers. With CIRVR, we aim to offer a virtual interview platform to autistic individuals to practice and improve their interviewing skills. The unique features of CIRVR, along with the metrics it generates, may impart insight and awareness to autistic individuals regarding their own performance, provide more precise guidance for how job coaches can intervene and best support them, and inform employers of ways that they may revise their assessments and methods (e.g., specific questions) to make their hiring process more inclusive.

A APPENDICES

A.1 Example Interviewee-Interviewer Exchanges

The following are two examples of exchanges between the interviewer and the interviewee that shows the two types of entity recognition used by CIRVR. The first demonstrates binary input from the user in terms of yes or no responses to yes or no types of questions. The second is an example of numerical entity recognition.

Example Exchange 1 – Binary Input

(1) **Interviewer:** "Do you have any experience using a game engine such as Unity or Unreal? And if so, which ones?"
(2) **Interviewee - Affirmative (Negative):** "Yes, I have experience with Unity." ("I do not.")
(3) **Interviewer - Affirmative (Negative):** "Excellent. How would you rate your skill with Unity?" ("That's okay. We can provide training in these tools.")

[Exchange continues]

Example Exchange 2 – Numerical Input

(1) **Interviewer** - "How many hours per week on average do you spend writing code?"
(2) **Interviewee - Above Threshold (Below Threshold):** "40." ("15.")
(3) **Interviewer - Above Threshold (Below Threshold):** "40 hours per week is great." ("Would you be comfortable programming for at least 30 hours per week?")

[Exchange continues]

A.2 Example of Dialog Flow

CIRVR's dialog flow can be pictured as a directed graph with numbered nodes as shown in Figure 11. Suppose that Node 1 represents the question "How many hours per week on average do you normally spend entering data into spreadsheets?". If the interviewee responds with "About 30 hours.", Azure's Language Understanding recognizes 30 as a built-in number type and a check

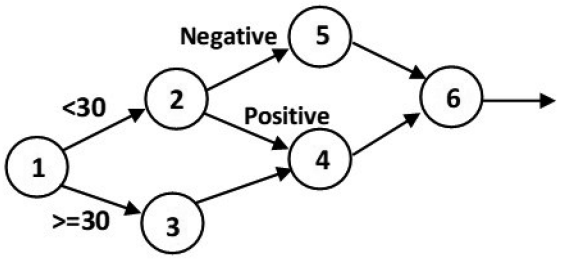

Fig. 11. Directed graph for dialog flow.

is made with a pivot value of 30. Since 30 >= 30, the conversation management system moves on to Node 3. If the number is <30, then the system moves on to Node 2. Suppose that the question in Node 2 is "Will you be comfortable working with spreadsheets for at least 30 hours a week?". If the user responds with "yes…", Azure's Text Analytics returns this utterance as having a positive sentiment and the interview moves to Node 4. If a negative sentiment is returned, then the interview moves to Node 5 and the interview proceeds as such. When there is no filter, the interviewer receives the interviewee response and moves linearly from question to question (from Node 6 to 7, then to 8, and so forth).

ACKNOWLEDGMENTS

The authors would like to thank members of the Vanderbilt Kennedy Centre's Treatment & Research Institute for Autism Spectrum Disorders (TRIAD), especially Zachary Warren and Amy Weitlauf, for their feedback in the design of the prototype technology. We would also like to thank Keivan Stassun and Dave Caudel, Director and Associate Director, respectively, of the Frist Centre for Autism and Innovation at Vanderbilt University, for their ongoing support for this project.

REFERENCES

[1] American Psychiatric Association. 2013. *Diagnostic and Statistical Manual of Mental Disorders: DSM-5* (5th ed.). American Psychiatric Association, Arlington, VA. xliv, 947 pages.

[2] Robert D. Austin and Gary P. Pisano. 2017. Neurodiversity as a competitive advantage. *Harvard Business Review* 95, 3 (2017), 96–103.

[3] Robert D. Austin and Thorkil Sonne. 2014. The case for hiring "outlier" employees. *Harvard Business Review Digital Articles* (2014), 2–3.

[4] Microsoft Azure. 2021. Azure. (2021). Retrieved September 11, 2021 from https://azure.microsoft.com/en-us/.

[5] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[6] Dom Barnard. 2017. VirtualSpeech on BBC Business Live. (2017). Retrieved September 11, 2021 from https://virtualspeech.com/blog/virtualspeech-interview-on-bbc-business-live-world-news.

[7] Tobias Baur, Ionut Damian, Patrick Gebhard, Kaska Porayska-Pomsta, and Elisabeth André. 2013. A job interview simulation: Social cue-based interaction with a virtual character. In *2013 International Conference on Social Computing*, (Alexandria, Virginia). IEEE, 220–227.

[8] Morris D. Bell and Andrea Weinstein. 2011. Simulated job interview skill training for people with psychiatric disability: Feasibility and tolerability of virtual reality training. *Schizophrenia Bulletin* 37, suppl_2 (2011), S91–S97.

[9] James Bergstra and Yoshua Bengio. 2012. Random search for hyper-parameter optimization. *Journal of Machine Learning Research* 13, 10 (2012), 281–305.

[10] Jennifer Romano Bergstrom, Sabrina Duda, David Hawkins, and Mike McGill. 2014. Physiological response measurements. In *Eye Tracking in User Experience Design.* Morgan Kaufmann, 81–108.

[11] Dayi Bian, Joshua Wade, Amy Swanson, Amy Weitlauf, Zachary Warren, and Nilanjan Sarkar. 2019. Design of a physiology-based adaptive virtual reality driving platform for individuals with ASD. *ACM Transactions on Accessible Computing* 12, 1 (2019), 1–24.

[12] Michael Biehl, David Matsumoto, Paul Ekman, Valerie Hearn, Karl Heider, Tsutomu Kudoh, and Veronica Ton. 1997. Matsumoto and ekman's japanese and caucasian facial expressions of emotion (JACFEE): Reliability data and cross-national differences. *Journal of Nonverbal Behavior* 21, 1 (1997), 3–21.
[13] Tanja Blascheck, Kuno Kurzhals, Michael Raschke, Michael Burch, Daniel Weiskopf, and Thomas Ertl. 2017. Visualization of eye tracking data: A taxonomy and survey. In *Computer Graphics Forum*, Vol. 36. Wiley Online Library, 260–284.
[14] Janine Booth. 2016. *Autism Equality in the Workplace: Removing Barriers and Challenging Discrimination*. Jessica Kingsley Publishers.
[15] John Brooke. 1996. SUS-A quick and dirty usability scale. *Usability Evaluation in Industry* 189, 194 (1996), 4–7.
[16] Shanna L. Burke, Tammy Bresnahan, Tan Li, Katrina Epnere, Albert Rizzo, Mary Partin, Robert M. Ahlness, and Matthew Trimmer. 2018. Using virtual interactive training agents (ViTA) with adults with autism and other developmental disabilities. *Journal of Autism and Developmental Disorders* 48, 3 (2018), 905–912.
[17] Shanna L. Burke, Tan Li, Adrienne Grudzien, and Stephanie Garcia. 2021. Brief report: Improving employment interview self-efficacy among adults with autism and other developmental disabilities using virtual interactive training agents (ViTA). *Journal of Autism and Developmental Disorders* 51, 2 (2021), 741–748.
[18] John T. Cacioppo, Louis G. Tassinary, and Gary Berntson. 2007. *Handbook of Psychophysiology*. Cambridge University Press.
[19] Romuald Carette, Mahmoud Elbattah, Gilles Dequen, Jean-Luc Guérin, and Federica Cilia. 2018. Visualization of eye-tracking patterns in autism spectrum disorder: Method and dataset. In *13th International Conference on Digital Information Management (ICDIM'18)*, (Berlin, Germany). IEEE, 248–253.
[20] Arijit Chatterjee and William Perrizo. 2016. Investor classification and sentiment analysis. In *IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM'16)*. 1177–1180.
[21] HTC Corporation. 2021. VIVE Pro Eye. (2021). Retrieved September 11, 2021 from https://www.vive.com/eu/product/vive-pro-eye/overview/.
[22] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, Gale Lucas, Stacy Marsella, Fabrizio Morbini, Angela Nazarian, Stefan Scherer, Giota Stratou, Apar Suri, David Traum, Rachel Wood, Yuyu Xu, Albert Rizzo, and Louis-Philippe Morency. 2014. SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems* (Paris, France).
[23] Nyaz Didehbani, Tandra Allen, Michelle Kandalaft, Daniel Krawczyk, and Sandra Chapman. 2016. Virtual reality social cognition training for children with high functioning autism. *Computers in Human Behavior* 62 (2016), 703–711.
[24] Foteini S. Dolianiti, Dimitrios Iakovakis, Sofia B. Dias, Sofia J. Hadjileontiadou, Jose A. Diniz, Georgia Natsiou, Melpomeni Tsitouridou, Panagiotis D. Bamidis, and Leontios J. Hadjileontiadis. 2019. Sentiment analysis on educational datasets: A comparative evaluation of commercial tools. *Educational Journal of the University of Patras UNESCO Chair* 6, 1 (2019), 262–273.
[25] Paul Ekman. 1993. Facial expression and emotion. *American Psychologist* 48, 4 (1993), 384.
[26] Paul D. Ellis. 2010. *The Essential Guide to Effect Sizes: Statistical Power, Meta-analysis, and the Interpretation of Research Results*. Cambridge University Press.
[27] Empatica. 2021. E4 Wristband: Real-time physiological data streaming and visualization. Retrieved September 11, 2021 from https://www.empatica.com/research/e4/.
[28] Christiane Fellbaum. 2010. WordNet. In *Theory and Applications of Ontology: Computer Applications*. Springer, 231–243.
[29] Keaton A. Fletcher, Sean M. Potter, and Britany N. Telford. 2018. Stress outcomes of four types of perceived interruptions. *Human Factors* 60, 2 (2018), 222–235.
[30] Credit Research Foundation. 2018. Job interviewer techniques and script. (2018). Retrieved September 12, 2020 from https://www.crfonline.org/orc/ca/ca-14.html.
[31] Fove. 2021. Fove VR Platform. Retrieved September 11, 2021 from https://fove-inc.com/developers/.
[32] Erich Gamma, Ralph Johnson, Richard Helm, Ralph E. Johnson, John Vlissides, et al. 1995. *Design Patterns: Elements of Reusable Object-oriented Software*. Pearson Deutschland GmbH.
[33] Cristina Gena, Claudio Mattutino, Stefania Laura Brighenti, Nazzario Matteo, Buratto Federico, and Fernando Vito Falcone. 2020. Social assistive robotics for autistic children. In *Workshop on Adapted Interaction with SociAl Robots (cAESAR'20)*, Vol. 2724. CEUR, 7–10.
[34] Carly B. Gilson and Erik W. Carter. 2016. Promoting social interactions and job independence for college students with autism or intellectual disability: A pilot study. *Journal of Autism and Developmental Disorders* 46, 11 (2016), 3583–3596.
[35] Jiawei Han, Jian Pei, and Micheline Kamber. 2011. *Data Mining: Concepts and Techniques*. Elsevier.
[36] Caitlin Harrington, Michele Foster, Sylvia Rodger, and Jill Ashburner. 2014. Engaging young people with autism spectrum disorder in research interviews. *British Journal of Learning Disabilities* 42, 2 (2014), 153–161.

[37] Jennifer Anne Healey. 2000. *Wearable and Automotive Systems for Affect Recognition from Physiology*. Ph.D. Dissertation. Massachusetts Institute of Technology, Cambridge, MA. Advisor(s) Rosalind W. Picard.
[38] Darren Hedley, Mirko Uljarević, Lauren Cameron, Santoshi Halder, Amanda Richdale, and Cheryl Dissanayake. 2017. Employment programmes and interventions targeting adults with autism spectrum disorder: A systematic review of the literature. *Autism* 21, 8 (2017), 929–941.
[39] Doris Adams Hill, Leigh Belcher, Holly E. Brigman, Scott Renner, and Brooke Stephens. 2013. The Apple iPad™ as an innovative employment support for young adults with autism spectrum disorder and other developmental disabilities. *Journal of Applied Rehabilitation Counseling* 44, 1 (2013), 28–37.
[40] Allen I. Huffcutt. 2011. An empirical review of the employment interview construct literature. *International Journal of Selection and Assessment* 19, 1 (2011), 62–81.
[41] Dan Jurafsky. 2000. *Speech & Language Processing*. Pearson Education India.
[42] Adam Kendon. 2004. *Gesture: Visible Action as Utterance*. Cambridge University Press.
[43] Lorcan Kenny, Caroline Hattersley, Bonnie Molins, Carole Buckley, Carol Povey, and Elizabeth Pellicano. 2016. Which terms should be used to describe autism? Perspectives from the UK autism community. *Autism* 20, 4 (2016), 442–462.
[44] Jonghwa Kim and Elisabeth André. 2008. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 30, 12 (2008), 2067–2083.
[45] Leif Larsen. 2017. *Learning Microsoft Cognitive Services*. Packt Publishing Ltd.
[46] Jonathan Lazar, Jinjuan Heidi Feng, and Harry Hochheiser. 2017. *Research Methods in Human-computer Interaction*. Morgan Kaufmann.
[47] Carl W. Lejuez, Christopher W. Kahler, and Richard A. Brown. 2003. A modified computer version of the paced auditory serial addition task (PASAT) as a laboratory-based stressor. *The Behavior Therapist* 26, 4 (2003), 290–293.
[48] Bing C. Lin, Jason M. Kain, and Charlotte Fritz. 2013. Don't interrupt me! An examination of the relationship between intrusions at work and employee strain. *International Journal of Stress Management* 20, 2 (2013), 77.
[49] Steven Bird, Ewan Klein, and Edward Loper. 2009. *Natural language processing with Python: analyzing text with the natural language toolkit*. O'Reilly Media, Inc.
[50] Matthew J. Maenner, Kelly A. Shaw, Jon Baio, et al. 2020. Prevalence of autism spectrum disorder among children aged 8 years–autism and developmental disabilities monitoring network, 11 sites, United States, 2016. *MMWR Surveillance Summaries* 69, 4 (2020), 1.
[51] Katie Maras, Jade Eloise Norris, Jemma Nicholson, Brett Heasman, Anna Remington, and Laura Crane. 2021. Ameliorating the disadvantage for autistic job seekers: An initial evaluation of adapted employment interview questions. *Autism* 25, 4 (2021), 1060–1075.
[52] Katherine Martin. 2018. Differences aren't deficiencies: Eye tracking reveals the strengths of individuals with autism. Retrieved September 12, 2021 from https://www.tobiipro.com/blog/eye-tracking-reveals-strengths-of-people-with-autism/.
[53] Joyce Montgomery, Keith Storey, Michal Post, and Jacky Lemley. 2011. The use of auditory prompting systems for increasing independent performance of students with autism in employment training. *International Journal of Rehabilitation Research* 34, 4 (2011), 330–335.
[54] Meredith Ringel Morris, Andrew Begel, and Ben Wiedermann. 2015. Understanding the challenges faced by neurodiverse software engineering employees: Towards a more inclusive and productive technical workforce. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, (New York, New York). 173–184.
[55] David B. Nicholas, Sandra Hodgetts, Lonnie Zwaigenbaum, Leann E. Smith, Paul Shattuck, Jeremy R. Parr, Olivia Conlon, Tamara Germani, Wendy Mitchell, Lori Sacrey, et al. 2017. Research needs and priorities for transition and employment in autism: Considerations reflected in a "Special Interest Group" at the International Meeting for Autism Research. *Autism Research* 10, 1 (2017), 15–24.
[56] Nivo. 2020. Heatmap. Retrieved September 12, 2020 from https://nivo.rocks/heatmap/.
[57] Jade Eloise Norris, Laura Crane, and Katie Maras. 2020. Interviewing autistic adults: Adaptations to support recall in police, employment, and healthcare interviews. *Autism* 24, 6 (2020), 1506–1520.
[58] Anna Pecchinenda. 1996. The affective significance of skin conductance activity during a difficult problem-solving task. *Cognition & Emotion* 10, 5 (1996), 481–504.
[59] Jeffrey Pennington, Richard Socher, and Christopher D. Manning. 2014. GloVe: Global vectors for word representation. In *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP'14)*. (Doha, Qatar). Association for Computer Linguistics, 1532–1543.
[60] Rosalind W. Picard. 2000. *Affective Computing*. MIT Press.
[61] Alifia Revan Prananda and Irfandy Thalib. 2020. Sentiment analysis for customer review: Case study of GO-JEK expansion. *Journal of Information Systems Engineering and Business Intelligence* 6, 1 (2020), 1–8.
[62] David Preece. 2002. Consultation with children with autistic spectrum disorders about their experience of short-term residential care. *British Journal of Learning Disabilities* 30, 3 (2002), 97–104.

[63] Pierre Rainville, Antoine Bechara, Nasir Naqvi, and Antonio R. Damasio. 2006. Basic emotions are associated with distinct patterns of cardiorespiratory activity. *International Journal of Psychophysiology* 61, 1 (2006), 5–18.
[64] Nilanjan Sarkar. 2002. Psychophysiological control architecture for human-robot coordination-concepts and initial experiments. In *Proceedings of the 2002 IEEE International Conference on Robotics and Automation*, Vol. 4, (Washington, DC). IEEE, 3719–3724.
[65] Noah J. Sasson and Jed T. Elison. 2012. Eye tracking young children with autism. *Journal of Visualized Experiments: JoVE* 61 (2012), 3675.
[66] Frank Schoonjans, A. Zalata, C. E. Depuydt, and F. H. Comhaire. 1995. MedCalc: A new computer program for medical statistics. *Computer Methods and Programs in Biomedicine* 48, 3 (1995), 257–262.
[67] Atsushi Senju and Mark H. Johnson. 2009. Atypical eye contact in autism: Models, mechanisms and development. *Neuroscience & Biobehavioral Reviews* 33, 8 (2009), 1204–1214.
[68] Simmersion. 2021. PeopleSim. (2021). Retrieved September 11, 2021 from https://www.simmersion.com/peoplesim.
[69] Matthew J. Smith, Michael F. Fleming, Michael A. Wright, Molly Losh, Laura Boteler Humm, Dale Olsen, and Morris D. Bell. 2015. Brief report: Vocational outcomes for young adults with autism spectrum disorders at six months after virtual reality job interview training. *Journal of Autism and Developmental Disorders* 45, 10 (2015), 3364–3369.
[70] Matthew J. Smith, Emily J. Ginger, Katherine Wright, Michael A. Wright, Julie Lounds Taylor, Laura Boteler Humm, Dale E. Olsen, Morris D. Bell, and Michael F. Fleming. 2014. Virtual reality job interview training in adults with autism spectrum disorder. *Journal of Autism and Developmental Disorders* 44, 10 (2014), 2450–2463.
[71] Matthew J. Smith, Rogério M. Pinto, Leann Dawalt, J. D. Smith, Kari Sherwood, Rashun Miles, Julie Taylor, Kara Hume, Tamara Dawkins, Mary Baker-Ericzén, et al. 2020. Using community-engaged methods to adapt virtual reality job-interview training for transition-age youth on the autism spectrum. *Research in Autism Spectrum Disorders* 71 (2020), 101498.
[72] Matthew J. Smith, Kari Sherwood, Shannon Blajeski, Brittany Ross, Justin D. Smith, Neil Jordan, Leann Dawalt, Lauren Bishop, and Marc S. Atkins. 2021. Job interview and vocational outcomes among transition-age youth receiving special education pre-employment transition services. *Intellectual and Developmental Disabilities* 59, 5 (2021), 405–421.
[73] Matthew J. Smith, Kari Sherwood, Brittany Ross, Justin Dean Smith, Leann Smith DaWalt, Lauren Bishop, Laura Boteler Humm, Jeff Elkins, and Chris Steacy. 2021. Virtual interview training for autistic transition age youth: A randomized controlled feasibility and effectiveness trial. *Autism* 25 (2021), 1536–1552.
[74] Debbie Spain, Jacqueline Sin, Kai B. Linder, Johanna McMahon, and Francesca Happé. 2018. Social anxiety in autism spectrum disorder: A systematic review. *Research in Autism Spectrum Disorders* 52 (2018), 51–68.
[75] Unity Asset Store. 2019. Business Woman. Retrieved September 11, 2021 from https://assetstore.unity.com/packages/3d/characters/humanoids/humans/business-woman-143301.
[76] Unity Asset Store. 2020. Tobii Eye Tracking SDK. Retrieved September 12, 2020 from https://assetstore.unity.com/packages/tools/input-management/tobii-eye-tracking-sdk-90604.
[77] Unity Asset Store. 2021. Realistic Eye Movements. Retrieved September 11, 2021 from https://assetstore.unity.com/packages/tools/animation/realistic-eye-movements-29168.
[78] Dorothy C. Strickland, Claire D. Coles, and Louise B. Southern. 2013. JobTIPS: A transition to employment program for individuals with autism spectrum disorders. *Journal of Autism and Developmental Disorders* 43, 10 (2013), 2472–2483.
[79] Jianhua Tao and Tieniu Tan. 2005. Affective computing: A review. In *International Conference on Affective Computing and Intelligent Interaction*, (Beijing, China). Springer, 981–995.
[80] Julie Lounds Taylor, Leann Smith DaWalt, Alison R. Marvin, J. Kiely Law, and Paul Lipkin. 2019. Sex differences in employment and supports for adults with autism spectrum disorder. *Autism* 23, 7 (2019), 1711–1719.
[81] Paul Taylor and Amy Isard. 1997. SSML: A speech synthesis markup language. *Speech Communication* 21, 1-2 (1997), 123–133.
[82] Unity Technologies. 2021. Unity 2019.3. Retrieved September 11, 2021 from https://unity.com/releases/2019-3.
[83] Tobii. 2020. Powerful Eyetracking for PC Games. Retrieved September 12, 2020 from https://gaming.tobii.com/.
[84] Conny M. A. van Ravenswaaij-Arts, Louis A. A. Kollee, Jeroen C. W. Hopman, Gerard B. A. Stoelinga, and Herman P. van Geijn. 1993. Heart rate variability. *Annals of Internal Medicine* 118, 6 (1993), 436–447.
[85] VirtualSpeech. 2021. VirtualSpeech: Soft Skills Training with VR. Retrieved September 11, 2021 from https://virtualspeech.com/.
[86] Joshua Wade, Amy Weitlauf, Neill Broderick, Amy Swanson, Lian Zhang, Dayi Bian, Medha Sarkar, Zachary Warren, and Nilanjan Sarkar. 2017. A pilot study assessing performance and visual attention of teenagers with ASD in a novel adaptive driving simulator. *Journal of Autism and Developmental Disorders* 47, 11 (2017), 3405–3417.
[87] Paul Wehman, Carol M. Schall, Jennifer McDonough, Carolyn Graham, Valerie Brooke, J. Erin Riehle, Alissa Brooke, Whitney Ham, Stephanie Lau, Jaclyn Allen, et al. 2017. Effects of an employer-based intervention on employment outcomes for youth with significant support needs due to autism. *Autism* 21, 3 (2017), 276–290.

[88] Qingguo Xu, Sen-ching Samson Cheung, and Neelkamal Soares. 2015. LittleHelper: An augmented reality glass application to assist individuals with autism in job interview. In *2015 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA'15)*, (Hong Kong, China). IEEE, 1276–1279.

[89] Kangning Yang, Chaofan Wang, Zhanna Sarsenbayeva, Benjamin Tag, Tilman Dingler, Greg Wadley, and Jorge Goncalves. 2021. Benchmarking commercial emotion detection systems using realistic distortions of facial image datasets. *The Visual Computer* 37, 6 (2021), 1447–1466.

[90] Li Yang, Jin Huang, Tian Feng, Wang Hong-An, and Dai Guo-Zhong. 2019. Gesture interaction in virtual reality. *Virtual Reality & Intelligent Hardware* 1, 1 (2019), 84–112.

[91] Y. J. Daniel Yang, Tandra Allen, Sebiha M. Abdullahi, Kevin A. Pelphrey, Fred R. Volkmar, and Sandra B. Chapman. 2017. Brain responses to biological motion predict treatment outcome in young adults with autism receiving virtual reality social cognition training: Preliminary findings. *Behaviour Research and Therapy* 93 (2017), 55–66.

Received December 2020; revised September 2021; accepted December 2021