Measuring Social Modulation of Gaze in Autism Spectrum Condition With Virtual Reality Interviews

Saygin Artiran, Raghav Ravisankar, Sarah Luo, Leanne Chukoskie, *Member, IEEE*,
and Pamela Cosman, *Fellow, IEEE*

***Abstract*— Gaze behavior in dyadic conversations can indicate active listening and attention. However, gaze behavior that is different from the engagement expected during neurotypical social interaction cues may be interpreted as uninterested or inattentive, which can be problematic in both personal and professional situations. Neurodivergent individuals, such as those with autism spectrum conditions, often exhibit social communication differences broadly including via gaze behavior. This project aims to support situational social gaze practice through a virtual reality (VR) mock job interview practice using the HTC Vive Pro Eye VR headset. We show how gaze behavior varies in the mock job interview between neurodivergent and neurotypical participants. We also investigate the social modulation of gaze behavior based on conversational role (speaking and listening). Our three main contributions are: (i) a system for fully-automatic analysis of social modulation of gaze behavior using a portable VR headset with a novel realistic mock job interview, (ii) a signal processing pipeline, which employs Kalman filtering and spatial-temporal density-based clustering techniques, that can improve the accuracy of the headset's built-in eye-tracker, and (iii) being the first to investigate social modulation of gaze behavior among neurotypical/divergent individuals in the realm of immersive VR.**

***Index Terms*— Gaze behavior, job interview practice, signal processing, neurodivergence, virtual reality, social modulation.**

Manuscript received 12 May 2022; revised 13 July 2022; accepted 10 August 2022. Date of publication 16 August 2022; date of current version 1 September 2022. This work was supported by the National Science Foundation under Grant DUE-1928604. *(Corresponding author: Saygin Artiran.)*

This work involved human subjects or animals in its research. Approval of all ethical and experimental procedures and protocols was granted by the UC San Diego Institutional Review Board under IRB Protocol No. 210775 on 7/1/2021.

Saygin Artiran and Pamela Cosman are with the Department of Electrical and Computer Engineering, UC San Diego, La Jolla, CA 92093 USA (e-mail: sartiran@eng.ucsd.edu; pcosman@eng.ucsd.edu).

Raghav Ravisankar was with the Computer Science and Engineering Department, UC San Diego, La Jolla, CA 92093 USA. He is now with the Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia (e-mail: ravisankarraghav@gmail.com).

Sarah Luo is with the Canyon Crest Academy, San Diego, CA 92130 USA (e-mail: sarahwluo@gmail.com).

Leanne Chukoskie is with the Department of Physical Therapy, Movement and Rehabilitation Sciences, and the Art + Design Department, Northeastern University, Boston, MA 02115 USA (e-mail: l.chukoskie@northeastern.edu).

Digital Object Identifier 10.1109/TNSRE.2022.3198933

I. Introduction

SOCIAL communication skills are important aspects of cognitive development, and technologies that support assessment and learning of such skills are in demand. Gaze behavior is a form of active sensing; individuals shift their gaze 3-4 times per second both consciously and unconsciously to gain high resolution information about aspects of a visual scene [1], [2], [3], [4], [5]. Auditory stimuli and internal factors related to personal experience, such as memory, have also been shown to affect eye movements and visual perception [6], [7], [8], [9], [10]. Many researchers have analyzed gaze behaviors of individuals who identify as neurodivergent, and particularly with an autism spectrum condition (ASC) [11] noting differences in gaze behavior particularly in social settings [12], [13], [14], [15], [16], [17], [18]. Numerous tools have been developed to support situational practice of gaze [19], [20], [21]. Commonly used technologies involve 2D computer screens accompanied by video cameras or eye-tracking glasses, and head mounted displays (HMDs) for augmented and virtual reality (AR/VR). Such studies have involved both neurotypical individuals and the neurodiverse.

Medical professionals often refer to ASC as autism spectrum disorder (ASD). ASD prevalence is 1 in 44 according to the most recent Autism and Developmental Disabilities Monitoring (ADDM) Network results monitoring 8 year olds across 11 ADDM network sites [22]. ASD is associated with a high rate of unemployment; studies report that 69% of individuals with ASD want to work [23] while merely 15% are employed [24]. Social communication deficits together with society's workplace communication norms and recruiting practices may be preventing many individuals with ASD from acquiring or retaining jobs [25], [26], [27], [28]. We are motivated to develop technologies to support autistic individuals in practicing conversational engagement skills, especially those which relate to job interviews and workplace communications [29], [30].

In this paper, we report on a system that uses a VR headset in the context of a virtual job interview which involves interacting with a virtual interviewer. Computer-aided instruction has numerous advantages from a practice point of view. Scenarios are reproducible, controllable and individualizable [31]. HMD systems evoke presence and immersion, and more recently, wearable tools have furnished novel ways of measuring and understanding the impacts of virtual content on practice outcomes. The capability of these media to replicate the exterior world allows researchers to emulate real-world experiences, which makes this technology widely applicable for practicing skills, including social skills such as interpersonal communication and interview skills. In a review, Bonaccio *et al.* [32] investigated nonverbal behavior in the workplace and pinpointed many "codes" of nonverbal behavior: communication via body movement, touch, voice, and physical space. For interview-specific skills, strong influence of nonverbal cues (e.g., gaze and body movement) on how an interviewer perceives an interviewee's performance was shown in [33]. These perceptions bias the selection of new employees which motivates our project in utilizing an objective physical input (gaze) from HMDs to assess social appropriateness and suitability for a professional environment.

For these assessments to be reliable, high-quality gaze-tracking is essential. HMDs with eye-trackers commonly require repeated eye-tracking calibration. Moreover, current eye-tracking technologies experience drops in data quality caused by factors such as glasses, contact lenses, eye color, eyelashes, and mascara [34], [35]. To improve gaze analysis, we develop a signal processing-based algorithm which boosts the accuracy of the HTC Vive Pro Eye VR headset built-in eye-tracker. The algorithm uses Kalman filtering to smooth out the spurs in gaze data arising from actions such as blinking or abrupt head movements, and applies a spatial-temporal density-based clustering to gaze samples.

VR comes with some limitations, including nausea, dizziness, and headaches; results in [36], [37], and [38] suggested that such side effects may present an obstacle to regular lengthy VR use. However, no special connection between autism and the level of such side effects has yet been established. Although [39] reported that some participants with low-functioning ASD were too overwhelmed to complete the experiments, participants with ASD enjoyed VR headsets without developing any side effects in many studies [39], [40], [41]. In addition, methods to mitigate such side effects have been developed, including accurate position tracking so as not to break the sense of 3D space [42], minimal delay between user inputs and corresponding in-game actions [43], and avoiding sudden loud sounds and visual clutter [39].

The contributions of this work are three-fold: (1) Although gaze behavior is widely studied, we are the first to provide an analysis of social modulation of gaze behavior via a portable VR headset and a realistic job interview simulation. Because the gaze analysis is fully automated, it is suitable for inexpensive solo practice. (2) We make an algorithmic contribution to the task of eye-tracking using HMDs. In studies such as [44], [45], [46], and [47], researchers report concerns about time consuming repetitive calibrations, and eye-tracking inaccuracies due to headset slippage, blinking, abrupt and extreme shifts in gaze, and large head movements. The algorithm we developed offers a signal processing-based solution to many of these tracking issues. (3) This work is the first to study social modulation of gaze in the context of immersive VR. The results provide insight, for neurotypical (NT) and autistic individuals, of what people most commonly look at during a job interview setting, and how the populations differ in the impact of conversational role on gaze.

The rest of this paper is organized as follows. Section II summarizes related work, while Section III describes the VR mock job interview application. In Section IV, we present our algorithm for increasing eye-tracking accuracy. Section V explains the demographics of our participants and the steps in a VR mock job interview session. Section VI showcases our gaze behavior analyses for NT and autistic participants, and we conclude with Section VII.

II. Related Work

Section II-A summarizes studies that examine the relation between gaze behavior and autism in dyadic conversations, as well as research that focuses on the impact of conversational role (listener versus speaker) on gaze, which is known as social modulation. Section II-B reviews papers that analyze gaze behavior in job interview settings, as well as research on technologies and procedures for practicing gaze behavior in social or professional settings.

A. Autism's Impact on Gaze Behavior and Social Modulation of Gaze

While there is a large literature on gaze behavior for autistic individuals, we are primarily interested in studies that involve VR or eye-tracking technologies, rather than involving a trained professional who observes and analyzes the behavior. In [45], [48], [49], and [50], the authors analyzed differences in gaze behavior between NT individuals and those with social disorders, concluding that autistic individuals or those with high social anxiety looked at the other conversationalist less often compared to NT participants and those with low social anxiety. Participants were also assigned an emotion recognition task in [45], interacting with virtual agents in a PC-based platform. NT participants recognized emotions more accurately, and usually in less time. Yoshikawa *et al.* [48] designed a social robot that converses by playing pre-recorded sounds in sync with its mouth. The study also included in-person conversations with a real person. Both populations looked at the android more and the autistic individuals looked at the android's eyes more often than at the person's eyes. The HTC Vive VR headset was used in [50].

Several papers analyzed the social modulation of gaze, that is, changes in eye movements based on a person's conversational role as speaker or listener. Using in-person conversations involving NT participants, researchers found that listeners gazed more at the speaker than speakers looked at listeners [46], [51]. Twenty eight undergraduate students participated in [46] and 13 adults took roles in [51] (the latter is a foundational study that used expert observation rather than technologically-based eye-tracking).

Similar studies have included neurodivergent participants. During in-person conversations of 13 autistic and 13 NT adults [47], the NT participants were found to look at the other person's eyes more often. Similarly, the effect of conversational role on gaze behavior during dyadic video calls was studied in [52] with 68 undergraduate students with varying autistic traits; the authors concluded that individuals with fewer autistic traits made more eye contact with the other party. Both studies concluded that regardless of neurodivergence, people spent more time looking at the other party's face while listening. The authors of [47] additionally reported that the autistic participants spent more time looking at the experimenter's mouth, and when the experimenter looked directly at them, they tended to avert their gaze more than did NT participants. This between-group difference was significantly reduced when the experimenter's gaze was averted. These studies did not involve immersive VR.

Interpersonal distance also affects gaze behavior, but research shows varying findings for both regular two-way conversations and interviews. Examining gaze behavior with interpersonal distances between interviewee and interviewer of 2.5, 6.5, and 10.5ft, Aiello [53] reported that male (female) participants made more eye contact at 6.5ft (10.5ft). Using a similar setup, Russo [54] found male participants made more eye contact seated at 6ft from the other party instead of 3ft, whereas the results were reversed for the female participants. Furthermore, the effect of autism on interpersonal distance tends to be inconsistent. Asada *et al.* [55] investigated the impact of eye contact on social proximity, finding that participants maintained a larger interpersonal distance when there was eye contact, regardless of having an ASC. In contrast, Perry *et al.* [56] reported significant variations in interpersonal distances preferred by autistic individuals.

B. Gaze Behavior in Job Interviews and Its Practice

A few studies investigated gaze behavior in interview contexts. The authors of [44] developed a tool to enable individuals to practice making eye contact during job interviews. Twelve adults took part in virtual conversations; if their gaze fell within the virtual character's bounds, the virtual character looked interested in the participant. Otherwise, it moved as if it were not paying attention. The participants were expected to make eye contact to receive attention. With both computer-mediated and face-to-face interview settings, 171 undergraduate students were asked in [57] to act as interviewers listening to applicants. Some applicants had scar-like facial features; an eye-tracker was used to study how these affected interviewer gaze patterns. Tian *et al.* [58] employed the FOVE headset to display a virtual room with multiple interviewers and conducted mock interviews with 10 graduate students. The researchers categorized the gaze data based on the fixation location (e.g., eyes, face, neck) and scored interviewee performance.

Wang *et al.* [59] designed a system that displays the external world inside the Oculus Rift VR headset. A teacher manually tracked participant gaze and overlaid a temporary prompt (e.g., an apple image) on top of their own eyes to stimulate eye contact, if needed. Although the system was intended to remind autistic children to make eye contact, it was tested on 4 NT college students who retained eye contact following the visual reminders as expected.

Several researchers designed VR games to improve eye contact, gaze sharing, and gaze following skills of children in social settings (e.g., [60], [61], [62], [63], [64], [65], the latter two used a VR headset and the rest used a PC). Mostly small numbers of subjects attended the studies: while 20 pairs of autistic and NT children participated in [61] and 9 pairs in [62], the other studies involved between 2 and 8 children in total. Although all six studies analyzed and aimed to improve gaze behavior, [61] and [65] did not employ any explicit eye-tracking. In those papers, participants played games that involved various stages, differing in difficulty, and gaze behavior practice proceeded based on the completion of stages.

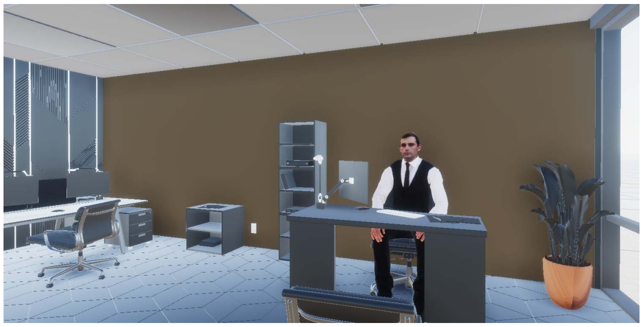

Fig. 1. Office space we designed for the VR-interview application. The setup is identical between version A and B except for the interviewer.

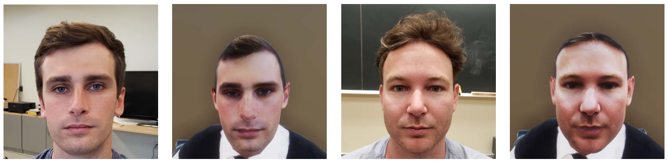

Fig. 2. Reference photos and corresponding virtual interviewers for version A (left pair) and B (right pair) of the VR-interview application.

III. VR Interview Application

We designed a virtual office space in Unity, with office objects such as a desk, computer, cabinet, and plant (see Fig. 1). We created two virtual interviewers (versions A and B); each version of the application shows a single interviewer sitting across from the interviewee. The interviewers ask questions and describe the job positions following a pre-defined script, performing naturalistic facial expressions and gestures. We used the Live Link Face app [66] to record audio and facial animations performed by a person. We transferred the recorded facial animation data to Blender, adding them on virtual head models as key points. Using photos of the recorded person from different perspectives, we created facial textures. We finalized the interviewer head models by adding animations such as head nodding and shaking. We also generated head models with no facial animations or audio, but with small head movements such as head tilts, which would be displayed while the interviewees were talking, so that the virtual interviewers would look engaged. Because the system is not yet processing verbal responses from the subject, the avatar does not yet have the capability of smiling and nodding at moments appropriate for the subject's speech. However, the ongoing small head tilts and other minor position adjustments make it appear that the avatar is paying attention. Fig. 2 displays a reference photo for each interviewer, and

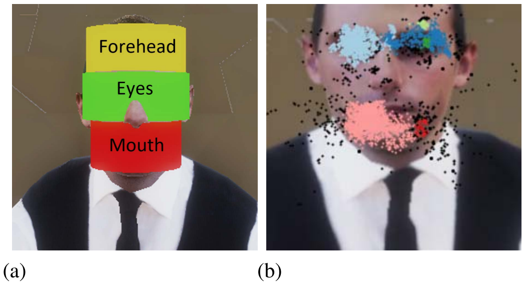

Fig. 3. (a) Facial regions: Forehead, Eyes, Mouth. (b) Example ST-DBSCAN output. Black dots represent points deemed noise by the algorithm. Six clusters are detected– three large (pink, light blue, darker blue), and three small (red, light green, darker green). The small regions differ from nearby large ones in that the gaze occurred at substantially different points in time.

its corresponding virtual interviewer. Voice recordings were added in Unity.

Version A represents a job interview at a research center at UC San Diego. The 45-question script has general questions (“Can you describe a project that you worked on and tell me something that you learned from it?”) and more focused ones (“Do you have any experience with project management software, such as Jira, Asana or Trello?”, “Have you ever used sensor technology, like eye-tracking before?”). The version B interview, for a position at a gaming company, has 43 questions more and less specific to the job position (“Can you think of some ways that games can promote positive play?”, “Can you tell me about your strengths?”).

The interviews proceed linearly, with the virtual interviewer asking questions one by one, not referring back to previous questions. The subjects are unscripted, and they respond as they deem appropriate with a mixture of short yes/no answers and much longer answers, including pauses for thinking. The system did not record or analyze the subject’s verbal responses. To signal that they provided their full answer to a question, the subjects use the trigger button on a controller. If a question is not clear, or if the subject is not able to completely hear a question, they can press the controller’s trackpad, and the interviewer will repeat the question. This study was approved by the UC San Diego Institutional Review Board under IRB Protocol 210775 (Date of approval 7/1/2021).

IV. Gaze Processing

A. Algorithm Design

By means of the Unity-compatible Vive SRanipal SDK, we use the HTC Vive’s built-in eye-tracking to track our users’ eye movements in real time. As shown in Fig. 3a, we divided the virtual interviewer’s face into three regions: forehead, eyes, and mouth. We aim to determine the distribution of gaze over the regions, and the connection between gaze location and conversational role. In each time step, after obtaining the gaze origin and direction, the system finds the virtual object hit by the gaze ray, which we refer to as the “hit object”, as well as the 3D location of the hit point. As the application does not run at a fixed frames-per-second rate, it tracks the time spent in each time step. For each time step, the system records the hit object label, the 3D in-game hit location, and the time spent. After each session, the total time spent looking at each facial region is obtained by summing the corresponding durations.

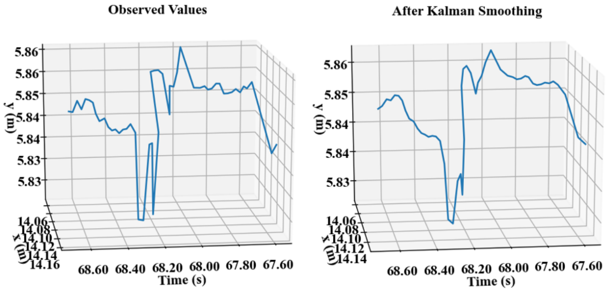

Fig. 4. Example usage of Kalman smoothing. Original gaze data stream on left, filtered gaze data on right.

We calibrate the eye-tracker following the standard steps accessible at the SteamVR dashboard while wearing the headset. Some jitter in the gaze rays arises due to blinking and to rapid large shifts in gaze; these are smoothed using Kalman filtering on the 3D gaze location stream. Fig. 4 displays an example of original gaze location data and the output after Kalman smoothing. The smoothed gaze locations are compared against the pre-defined region borders, and the hit object labels are updated correspondingly as needed.

As a further step in correcting the labels, the filtered gaze data goes through Spatial-Temporal Density-Based Clustering of Applications with Noise (ST-DBSCAN) [67]. This is an unsupervised clustering method based on the DBSCAN algorithm [68]. DBSCAN groups points that are spatially close. The algorithm is defined by two main parameters: $\epsilon 1$ is the maximum spatial distance for one point to be considered as in the neighborhood of another, and $minPts$ is the minimum number of points required to form a cluster. After a cluster is formed, the algorithm iteratively expands the cluster, if possible, by going through each data point in that cluster and adding neighboring external points. The cluster expands until there are no more points within $\epsilon 1$. Then, the algorithm starts looking for the next cluster. DBSCAN does not require a pre-defined number of clusters and determines the number based on the parameters and spatial distribution of the data. The algorithm produces a list of numeric labels; points that do not fall into any cluster are considered noise and labeled with -1.

As DBSCAN does not consider time consistency, two data points which are spatially close may be put in the same cluster even if there is a large temporal distance between them. ST-DBSCAN introduces a time constraint using parameter $\epsilon 2$, which is the maximum temporal distance between two points for them to be in the neighborhood of each other [67]. The algorithm starts by finding the points that are within a temporal range $\epsilon 2$ and a spatial distance $\epsilon 1$ of a gaze point in the data. If the number of points that satisfy the distance criteria is greater than $minPts$, a cluster is formed. Similar to DBSCAN’s cluster expansion, a cluster can expand if external points exist within a temporal distance $\epsilon 2$ and a spatial distance $\epsilon 1$ of a point in that cluster. ST-DBSCAN

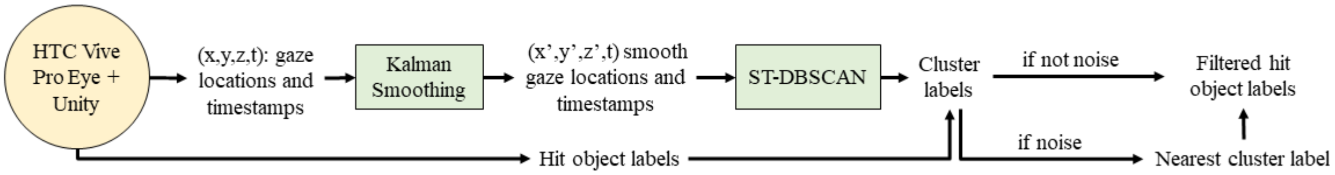

Fig. 5. Gaze processing system overview.

might form multiple clusters within the borders of the same hit object, which could represent looks to different parts of the same object, different time-separated looks to the same part of the object, or looks that involved both different times and different object parts. A resulting ST-DBSCAN cluster is defined by a 4D centroid c: $\{c_x, c_y, c_z, c_t\}$ where c_x, c_y, c_z are the spatial coordinates of the centroid, and c_t is the average temporal coordinate. An ST-DBSCAN output is displayed in Fig. 3b.

For each ST-DBSCAN cluster, the most common hit object label for that cluster is found, and is assigned as the object label for all gaze points in that cluster. Points labeled as noise are dealt with separately through a relabeling step. The number of points, and which specific points, are deemed by ST-DBSCAN to be noise is highly dependent on the ST-DBSCAN parameters; the noise relabeling step, as will be seen below, enables us to achieve a labeled output stream closer to ground truth in a data-driven way. For a given noise point (p^j), the non-noise gaze points within a temporal range of $\epsilon 2$ to the noise sample are examined; the unique cluster labels of those non-noise points are stored in one set, while their associated 4D centroids (c) are stored in another set, B.

There are studies in the literature that add time as a component dimension, to form a spatial-temporal dimension [69], and works that use or suggest Euclidean distances involving a temporal component such as $d_E = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (t_1 - t_2)^2}$ [70], [71]. Weighted versions of d_E have also been used in clustering tasks [72]. The authors of [73] implement a distance function in the form of a weighted sum of the space and time components for 3D spatial-temporal data $p = (x, y, t)$:

$$3D(p_1, p_2) = w_t \frac{d_t(p_1, p_2)}{MaxT} + (1 - wt)\frac{d_s(p_1, p_2)}{MaxS} \qquad (1)$$

where the authors define the spatial distance component as $d_s(p_1, p_2) = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2}$ and the temporal distance component as $d_t(p_1, p_2) = |t_1 - t_2|$. $MaxS$ and $MaxT$ are used to normalize the spatial and temporal dimensions; these parameters might be set to the maximum obtainable spatial and temporal distances, or can be optimized based on data. Finally, w_t is the weight for temporal distance, while the spatial distances are weighted by $1 - w_t$. We use this distance function in our noise relabeling, where we define $MaxS$ and $MaxT$ as the maximum Euclidean spatial distance and the largest absolute temporal distance from a cluster centroid (stored in B) to the noise sample. The cluster label of the centroid that minimizes this distance function gives the hit object label of the point that was previously labeled as noise.

B. Validation and Testing

To tune our algorithm (Fig. 5) with the hyperparameter set $\Gamma = \{\epsilon 1, \epsilon 2, minPts, w_t\}$, we instructed 12 participants (9 male, 3 female, split between versions A and B), to look at objects (forehead, eyes, mouth, plant, notebook, keyboard, monitor, cabinet, see Fig. 1). An experimenter named objects or facial regions from the list, and the subject promptly shifted their gaze and looked at it steadily. For ground truth, the experimenter recorded the object label and the start/end times of these prompted looks. The order and duration for objects varied across participants and sessions. Each participant took part in 5 sessions (average 287.3s), and in total we collected 4.8 hours of ground-truthed gaze data for tuning. Face region locations and dimensions were slightly different for the two versions because of differences in the interviewers' heads. Hence, for each version, we separately performed subject-wise leave-one-out cross validation (CV), determining the optimal hyperparameter set based on the 25 recordings from 5 individuals, and testing that set on the 5 recordings of the participant left out. The optimal hyperparameter set in each fold was the one that minimized the sum of forehead, eye, and mouth region gaze percentage errors averaged over 25 training recordings, as discussed in the following.

Suppose the subject has been given the i^{th} instruction to look at something (e.g., the eye region) for a ground truth look duration dur^i_{gt}. During this time, the duration of gaze to the eye region measured by our algorithm is dur^i_{alg}, which is the sum of the time spent gazing at the object with the correct object label after applying our gaze processing algorithm. Then, $err^i_{alg} = dur^i_{gt} - dur^i_{alg} \geq 0$. The total time spent looking at a particular region, according to the algorithm, when the subject was not instructed to do so, constitutes the other type of gaze duration error. The participants in this portion of data collection were lab members who were compliant with the instructions to the extent possible, so could be assumed to be looking at the named items. Based on these two duration error types, total errors are computed for each facial region; we divide the error values by the related recording's total duration to produce a percentage error.

For $\epsilon 1$, we evaluated {0.015m, 0.02m, 0.025m, 0.03m, 0.035m}. The candidate set for $\epsilon 2$ was {0.125s, 0.25s, 0.375s, 0.5s, 0.625s, 0.75s, 0.875s, 1s}. For $minPts$, we considered the integers from 1 to 40. Finally, for w_t, we tested 0.25, 0.5, and 0.75. For version A, the optimal hyperparameter set in each fold was {0.03m, 0.5s, 25, 0.5} except for one fold for which the optimal w_t was 0.75. For the 30 recordings, the average total percentage-wise error in unprocessed gaze

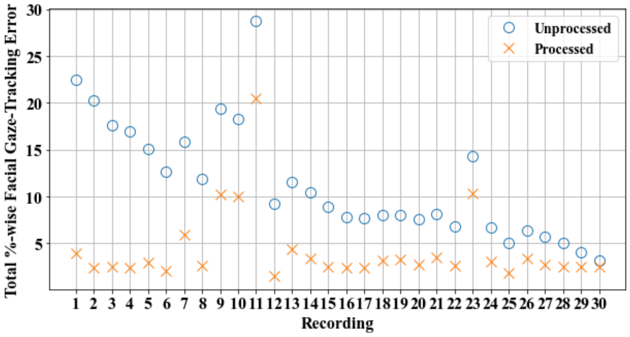

Fig. 6. Average percentage-wise facial gaze-tracking error obtained for each version A recording (the x-axis is ordered by decreasing difference between the error from the unprocessed data and the error our algorithm achieves).

targeting the interviewer's face was 11.46% (33.31s). The duration-wise errors for forehead, eye, and mouth regions were 8.32s (2.86%), 16.62s (5.72%), and 8.37s (2.88%), respectively. Since the experimenter tells the subjects when and where to look, the errors arise from calibration inaccuracy, imprecision in Unity's object collision method, and reaction time between the experimenter giving an instruction and the subject responding with a gaze shift. In young adults, this saccadic reaction time was estimated to be 250ms [74].

This overall gaze error of 11.46% using the unprocessed gaze data was reduced to 4.19% when the gaze data is processed in each fold by our algorithm (see Fig. 5). The average total percentage-wise facial gaze-tracking error of 4.19% amounts to 12.13s, of which 2.84s (0.98%), 6.27s (2.16%), 3.03s (1.05%) are forehead, eye, and mouth region errors, respectively. Fig. 6 exhibits the average total percentage-wise facial gaze-tracking error obtained for each recording using the unprocessed data and the processed gaze data produced by our algorithm tuned in the related fold. The algorithm outperforms the accuracy of unprocessed gaze data for all recordings, so all differences are positive. The level of improvement in gaze-tracking error tends to vary for several reasons. Participants were instructed to gaze at objects and facial regions for different durations in changing orders, and because these regions differ in size and proximity to each other, some are easier to detect. In addition, calibration precision could have been different, and subject behaviors may differ among different recordings.

We also evaluated our tuned algorithm using the Intersection over Union (IoU) metric. The average IoU score using the unprocessed data was 85.05%, whereas 94.37% using our algorithm. For this metric as well, the algorithm outperforms the accuracy of unprocessed gaze data for all recordings. Performing the same tuning procedure for version B, the hyperparameter set {0.025m, 0.625s, 20, 0.25} was unanimously selected as optimal. For version B, our algorithm improved the overall percentage error from 8.93% (24.69s) to 4.34% (12.08s), and the average IoU score increased from 86.20% to 93.34%. The processing yields improvement for each recording for both metrics.

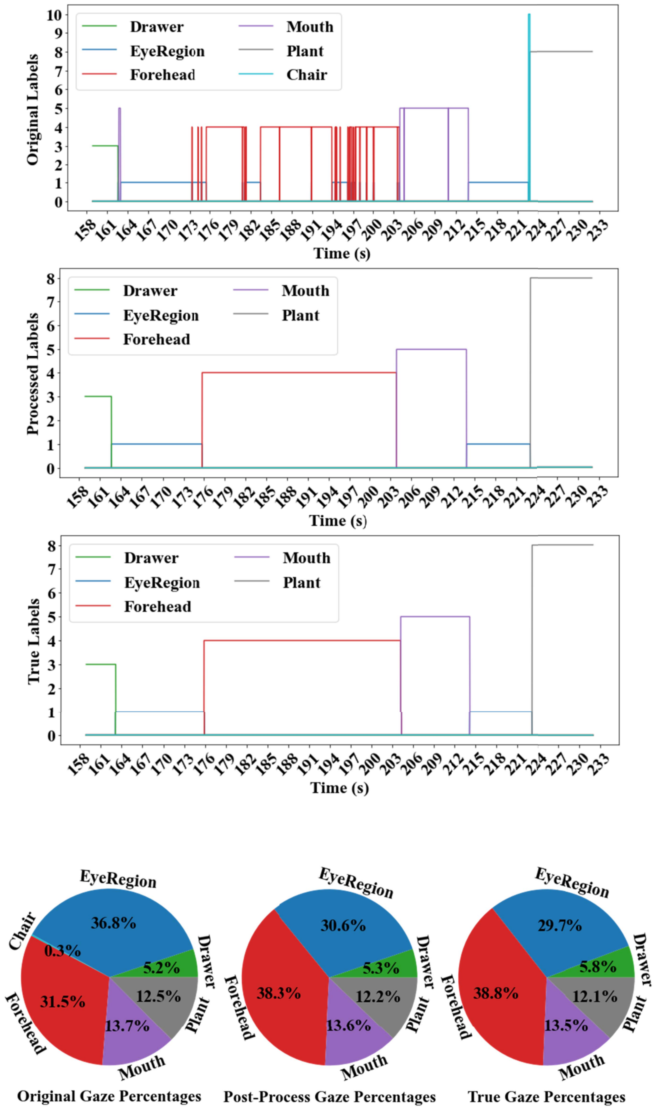

Fig. 7. Section of an original hit object label stream (top), labels after processing using our tuned algorithm (2nd plot), true labels (3rd plot). The first two pie charts visualize the gaze percentages before and after processing, and the third pie chart shows the true gaze percentages.

For the remainder of this paper, we use the hyperparameter set $\hat{\Gamma} = \{0.03\text{m}, 0.5\text{s}, 25, 0.5\}$ for version A, and {0.025m, 0.625s, 20, 0.25} for version B. We tested the pipeline on 2.5 hours of additional unseen data (3 test subjects who completed 5 sessions for each version). For version A, unprocessed gaze data yielded an average facial gaze-tracking error of 7.80% (23.05s) which improved to 3.83% (11.31s) with processing, and the average IoU improved from 88.70% to 94.49%. The average facial region errors for version B improved from 8.44% (24.63s) to 4.73% (13.81s) while average IoU score increased from 87.83% to 93.36%.

Fig. 7 shows an example section of the original and processed virtual hit object label streams of the recording with the highest improvement from the additional set, together with the true labels in that section. The figure also shows gaze percentages as computed originally and as adjusted by our algorithm, as well as true gaze percentages. Some of the errors which get corrected are due to blinking, and some come from calibration inaccuracy. For a relatively small region such as the eyes, a person gazing near the boundary of the region may, with a small calibration inaccuracy, get recorded in the unprocessed gaze stream as having the gaze fall repeatedly inside and outside of the region. We can see in the figure that spikes and discontinuities in correct label flows are eliminated. For larger and relatively isolated objects such as the drawer or the plant, we do not observe as many inaccuracies, hence the algorithm makes fewer corrections around those objects.

C. Specific Problems Mitigated by Processing

We conducted an experiment to investigate our algorithm's ability to mitigate specific eye-tracking inaccuracies due to blinking and headset slippage. During 10 sessions (5 with each version) of about 1 minute each, without interaction with the virtual interviewer, timestamps were recorded at points when the subject blinked or nudged the headset to resemble slippage. A total of 48 blinks and 40 simulated headset slippage examples were recorded. Each occurred while the subject gazed fixedly at some chosen object, so ground truth around each occurrence involves no gaze shift. Examining the unprocessed label streams within 1 second after a blink or headset slippage showed that the shift in the gaze ray was not large enough to move it to a different object for 12 instances of blinking, and 3 instances of slippage. Of the 36 blinks associated with apparent object shifts, our tuned algorithm corrected all but one, and it corrected all 37 occurrences of shifts due to slippage. Fig. 8 shows the original and processed virtual hit object label streams in one session, alongside the true labels. Solid black (cyan) vertical lines mark the time points where a blink (slippage) occurred. The gap between a solid and a successive broken vertical line of same color is 1 second. The average blink lasts about one third of a second [75], but since gaze ray shifts do not consistently emerge right after an event, we look over a 1 second interval.

We also designed a procedure to quantify calibration drift and its impact on eye-tracking accuracy. Step 1: Calibrate the headset's eye-tracker; calibration accuracy was deemed sufficient if the gaze ray fell on the spheres when a participant was told to look at 5 spheres (of radius 3.25cm) placed at the corners and center of a square of edge length 50cm. Step 2: After achieving an acceptable calibration visually, to measure the initial calibration accuracy, participants looked at the front-most point on each sphere for 5 seconds. Step 3: The experimenter prompted the participants for 3 minutes to look steadily at different face regions and objects around the room in a fixed order; timestamps of prompts were recorded. Step 4: The participants interacted with the virtual interviewer for 12 minutes, allowing potential factors of calibration drift to arise such as headset slippage or the lenses fogging up. Step 5: Repeat step 3 to measure post-interview eye-tracking accuracy. Step 6: Repeat step 2 to measure post-interview calibration drift. Three participants each completed the procedure twice, once for each version of the mock job interview application.

Table I displays the average calibration accuracy and face gaze-tracking error before and after the 12-minute interview sessions. As expected, the average calibration accuracy is worse after the interview session, and, without using the algorithm, the error in face gaze-tracking is also worse. However, using our algorithm to process the gaze stream, gaze-tracking errors before and after the interview session are comparable, and both are significantly lower than the errors for the unprocessed gaze stream.

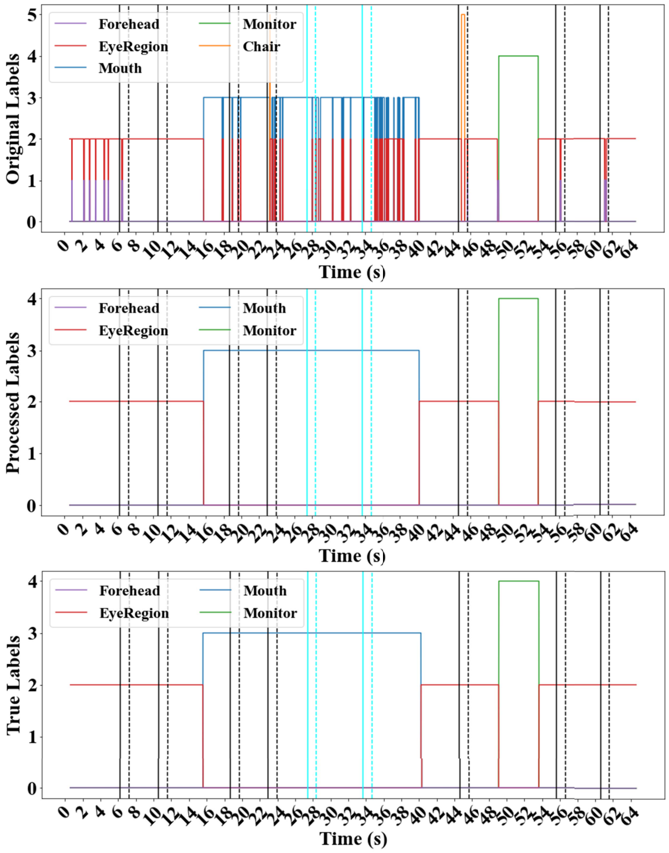

Fig. 8. Original hit object label stream (top), labels after using tuned algorithm (middle), true labels (bottom). Solid black/cyan vertical lines mark blinks/headset slippages; broken lines are 1s after the preceding solid lines.

TABLE I
AVERAGE CALIBRATION ACCURACY AND FACE REGION GAZE TRACKING ERROR

Stage Statistic	Pre-interview	Post-interview
Calibration Accuracy (cm)	3.48	4.5
Error Before Algorithm (s)	9.47	11.95
Error After Algorithm (s)	5.8	4.66

In summary, this subsection provides evidence that the improvement in object label accuracy from processing the gaze stream arises from the algorithm's ability to correct for blinking, headset slippage, and calibration drift.

V. Gaze Data Collection in the Virtual Interview

The gaze data collection described previously involved instructions to look at objects and facial regions, to provide a ground truth for developing the gaze processing algorithm.

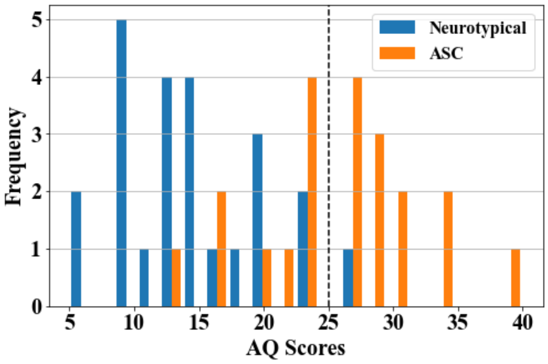

Fig. 9. Autism-Spectrum Quotient (AQ) scores reported by participants. The broken black line marks the threshold introduced in [11]; scores above this point to the probability of having an ASC using this self-report questionnaire.

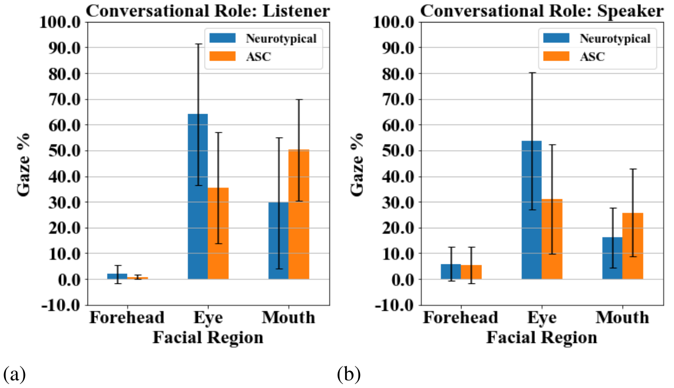

Fig. 10. Plot on left displays the means of the percentages displayed in Fig. 11a, 11b, and 11c, while the plot on right shows the means from Fig. 12. Black error bars mark one standard deviation (s.d.) interval.

Following this, we turned to the main project focus, gaze in VR job interviews. Here, subjects are not instructed to look at anything in particular (other than for the initial calibration) and are not aware that gaze is the focus of study.

Twenty four individuals with ASC (21 male, 3 female) and 24 NT subjects (16 male, 8 female) participated in the virtual interviews. Participants were assigned to the ASC group if they had received a community ASC diagnosis. We recruited 18 subjects who were previous participants in a neurodiversity summer internship at UC San Diego, five current or former trainees at the National Foundation for Autism Research technical training program, and one person through a San Diego-based group Autism Masterminds. Subject ages ranged from 20 to 35 years old. At the start of a session, the built-in eye-tracker was calibrated for each subject using the standard HTC Vive calibration procedure, and the calibration was verified by checking that the detected gaze coincided with the three facial regions when the subject was prompted to look at them. After calibration, participants began interacting with the virtual interviewer. Half of the participants from each population completed version A (average session length 22min 4s, s.d. = 5min 28s), and half did version B (average length 25min 54s, s.d. = 10min 25s).

After each session, we asked the participant to take the Autism-Spectrum Quotient (AQ) test, a tool designed to quantitatively evaluate the expression of autism spectrum traits in an individual, based on their subjective self-assessment [11]. Fig. 9 shows the reported AQ scores (three participants with ASC did not disclose their scores). The mean AQ scores reported by the NT and ASC groups are 14.21 (s.d. = 5.35) and 25.76 (s.d. = 6.31), respectively. While one participant with no prior diagnosis of autism scored above 25, considered in [11] as the lower bound indicating an ASC, this subject was retained in the NT group as they did not receive a community diagnosis of autism and did not self-identify as autistic. Similarly, 9 of our neurodivergent participants scored below 25, and 3 were far from the threshold. Potential reasons for such scores include that they may have taken the publicly-accessible AQ survey multiple times and know how to answer the questions, or that while they once received an autism spectrum diagnosis, they might be unlikely to receive one again based on their current communicative and interaction behaviors [76]. Regardless, participants who had previously received a community diagnosis of autism or autism spectrum were maintained in that subset. The AQ score data distributions were verified to be normal for each group and a Mann-Whitney U test indicated a significant difference between the scores achieved by the two groups, $U = 38$, $p < 0.001$. (Here, 38 represents the sum of the number of autistic participants with a lower AQ score compared to each NT participant separately. Ties contribute 0.5.)

VI. Results

In this section, we compare the gaze behavior, as a function of conversational role, of the autistic and NT participants in our virtual job interview setting. For the tests of significance in this section, we use one of two non-parametric tests, the Mann-Whitney U test, which assumes that two groups are sampled from the same type of distribution (e.g., normal, left-skewed) and which is valid for both normally and non-normally distributed data [77], and the two-sample Kolmogorov-Smirnov (K-S) test, which compares the cumulative distributions of two data sets without assumptions about the distributions. For the K-S test, the statistic D represents the maximum distance between two cumulative distributions. We use the Shapiro-Wilk test of normality and skewness tests to determine which significance test to use.

The average gaze percentages across forehead, eyes, and mouth regions while subjects listen to the virtual interviewer speak are in Fig. 10a, which shows that the average percentage of eye contact while listening was 35.50% (s.d. = 21.58%) for neurodivergent participants compared to 64.04% (s.d. = 27.50%) for NT participants, whereas gaze at the mouth region averaged 50.15% (s.d. = 19.79%) for the ASC group and 29.59% (s.d. = 25.57%) for the NT group. Fig. 11 shows the distribution of those percentages. Although there are some outliers, the majority of NT participants mostly made eye contact when listening. In contrast, autistic participants had a higher tendency to look at the interviewer's mouth compared to his eyes. The authors of [78] suggested that autistic individuals might fixate more on the mouth rather than eyes since the mouth is the source of speech and they aim to maximize their

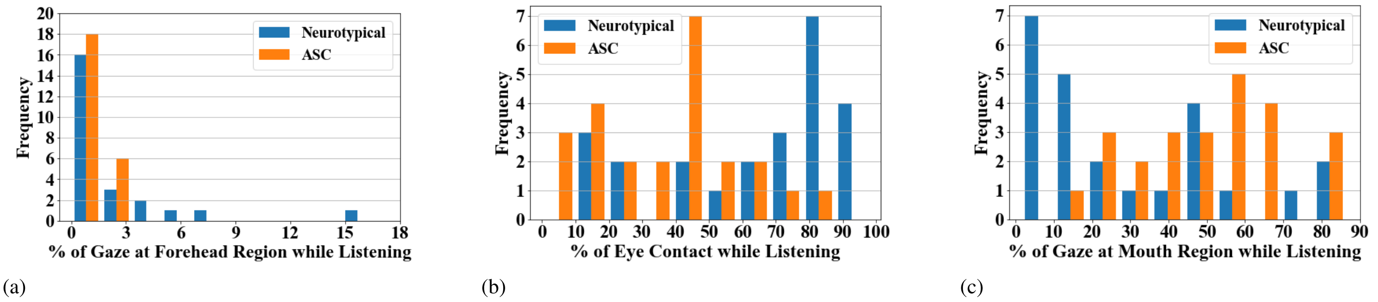

Fig. 11. Distributions of percentages of gaze at the three main regions of the virtual interviewer's face (Fig. 3a) while listening.

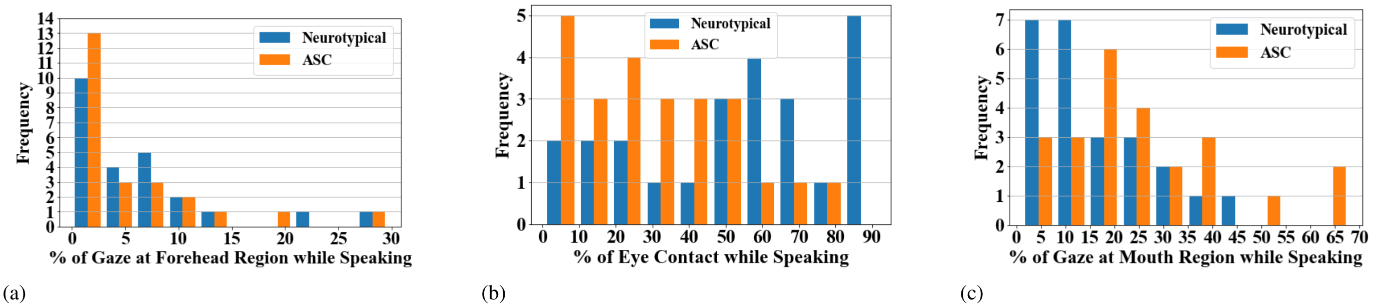

Fig. 12. Distributions of percentages of gaze at the three main regions of the virtual interviewer's face (Fig. 3a) while speaking.

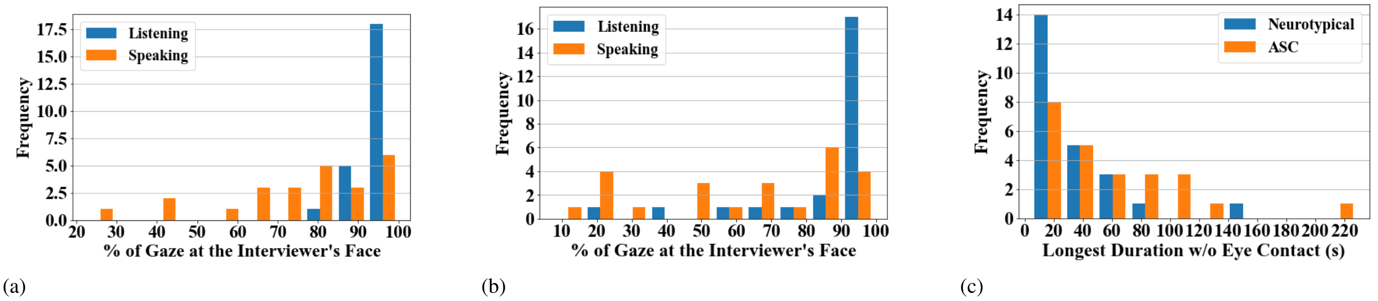

Fig. 13. Distribution of (a) % of gaze at the interviewer's face for NT subjects based on conversational role (b) % of gaze at the face for autistic subjects based on conversational role (c) longest uninterrupted time intervals with no eye contact between the interviewees and virtual interviewer.

understanding of social situations by concentrating on a feature they can comprehend. These observations are in alignment with the average numbers presented in Fig. 10a. These results that NT participants, compared with autistic participants, made more eye contact with an interviewer are consistent with previous results for in-person interviews, android robots, and video calls shown in [47], [48], and [52]. So our results extend to immersive VR the previous research that was not on VR, while being qualitatively consistent with those earlier results.

For NT participants as listeners, we compared the percentage of gaze at eyes and gaze at mouth using the K-S test as the skewness values pointed to moderate skewness in opposite directions [79]. The difference between the two groups of gaze percentages was deemed significant ($D = 0.54$, $p = 0.001$). For autistic participants as listeners, the difference between percentages of gaze at eyes and gaze at mouth was in the opposite direction but also significant ($U = 78$, $p = 0.02$, where the Mann-Whitney U test was used because both sets of data satisfied normality). A final two-sample K-S test on the difference between the eye contact percentages of the two populations was also significant ($D = 0.54$, $p = 0.001$).

The distributions of percentages of gaze to the forehead, eye and mouth regions while speaking are displayed in Fig. 12, with means in Fig. 10b. While speaking, NT participants made eye contact with the virtual interviewer 53.69% (s.d. = 26.61%) of the time on average, whereas this number was 30.99% (s.d. = 21.28%) for autistic participants. Both groups came from a normal distribution, and the Mann-Whitney U test deemed this inter-population difference significant, $U = 429$, $p = 0.003$. The authors of [47], [48], and [52] reported the same type of result within their respective contexts, so our results are qualitatively consistent with them while extending those results to immersive VR.

Fig. 13a and Fig. 13b display the distributions of gaze at the interviewer's face overall, when listening and speaking. While listening, the NT group gazed at the interviewer's face on average 95.66% (s.d. = 4.75%) of the time, while this value was 86.51% (s.d. = 19.74%) for the ASC group. These

TABLE II
LONGEST UNINTERRUPTED TIME INTERVALS WITHOUT EYE CONTACT AVERAGED WITHIN POPULATION

Role Population	Listener	Speaker
Neurodivergent	20.19s (12.29s)	32.13s (22.37s)
Neurotypical	13.73s (13.67s)	20.25s (16.84s)

numbers dropped to 75.77% (s.d.= 19.64%) and 62.16% (s.d. = 28.29%), when speaking. These drops based on conversational role were significant according to K-S tests ($D = 0.71$, $p < 0.001$ for the NT group and $D = 0.58$, $p < 0.001$ for the ASC group). For both groups, individuals tend to more frequently avert their gaze while speaking, a general finding that comports with [46], [47], [51], and [52]. Furthermore, regardless of conversational role, NT participants gazed at the interviewer's face more often than autistic participants. This result is in agreement with [45], [49], and [50] which involved different contexts.

We also investigated the longest uninterrupted time interval during which interviewees did not make eye contact with the interviewer, an aspect of gaze behavior previously unexamined. Such time intervals were longer for the ASC group (Fig. 13c), with average longest time of 58.49s (s.d. = 50.01s) versus 32.04s (s.d.= 32.97s) for the NT group. The maximum of such durations is 153.63s for the NT group, whereas it reaches 228.39s among the ASC group. A Mann-Whitney U test considered the difference to be significant, $U = 172$, $p = 0.02$ (both distributions were highly right skewed).

The impact of conversational role on this measurement is in Table II. The table presents average values for each population, with standard deviations in parentheses. Both groups tended to gaze away from the interviewer's eyes more while speaking. When gazing away from the eyes, as listeners, both populations mostly fixated on the interviewer's mouth, whereas they mainly scanned the surrounding objects as speakers. In agreement with the trend presented so far, autistic participants had longer maximum times with no eye contact for both conversational roles. The values in Table II which examine lack of eye contact within a conversational role are substantially shorter than the no-contact periods mentioned above which ignore conversational role, because it is a common pattern for no-eye-contact periods to cross role boundaries, for example where a subject listens to a question with eyes averted then keeps eyes averted while they formulate and begin an answer.

VII. CONCLUSION

Companies around the world are recognizing the benefits that diverse teams bring to productivity and creative problem-solving [80], [81]. Autistic people are able to focus on tasks well, they can accomplish tasks more efficiently, and work with honesty and dedication [82]. However, individuals with ASC can find it challenging to obtain employment because of social appropriateness and professionalism in an interview setting [83]. Although there have been attempts to make job hiring processes more inclusive and approachable [83], these are slow to spread.

We have designed a VR-interview practice application which provides an immersive experience for users to interact with a human-like virtual interviewer and to answer his questions. While it provides an opportunity to prepare for job interviews, and to get exposed to questions that are commonly asked in real job interviews, our system simultaneously tracks eye movements in the background. The system accurately measures gaze to different face regions during different conversational roles (speaking and listening). This allows us to identify differences in gaze behavior of autistic and neurotypical individuals. Some of these types of measures exist in the literature for non-VR or non-immersive-VR systems, and our findings are consistent with those results, validating the viability of our approach, and extending previous results to this immersive VR setup. Other measures presented are novel.

We make three main contributions in this paper. First, we introduce the novel portable VR system that is capable of analyzing distinct gaze tendencies, as well as changes in gaze behavior based on conversational role. Secondly, we design a signal processing-based algorithm that can reduce the problem of calibration drift of the HMD eye-trackers and potential inaccuracies due to external factors like headset slippage or blinking which were commonly reported in [44], [45], [46], and [47]. Our algorithm mitigates such errors, providing significantly more accurate gaze tracking in this conversational context. Finally, researchers have previously used in-person interviews or video calls to investigate social modulation of gaze for neurotypical/diverse populations [46], [47], [51], [52], and our work is the first to study these behaviors in the context of immersive VR.

For future work, we aim to construct an interview practice tool that provides automated feedback to warn users in real time if, for example, interviewees fail to make eye contact for a long time, and also that will provide a comparative analysis with respect to NT gaze behavior as this is still the expectation in most workplace settings. This will allow users to access solo practice for job interviews by answering professional questions. They will also get the chance to adjust their gaze behavior, if needed, to suit society's mismatched expectations about what constitutes engaged social conversation appropriate for an interview setting. The application may be useful for neurotypical participants as well; anecdotally some participants reported that using the system made them think about questions they had not faced previously, and that it was useful to practice answering these questions in a real-time "think on your feet" situation.

ACKNOWLEDGMENT

The authors thank Zachary Burns and Trent Simmons for helping them model the interviewers, Khalil Jackson, Thomas Circle, and Corly Huang for contributing to the Unity applications, Sarah Hacker and Ara Jung for managing the IRB approval and subject payments, and the participants in their study.

REFERENCES

[1] R. Cañigueral and A. F. D. C. Hamilton, "The role of eye gaze during natural social interactions in typical and autistic people," *Frontiers Psychol.*, vol. 10, p. 560, Mar. 2019.

[2] L. W. Renninger, P. Verghese, and J. Coughlan, "Where to look next? Eye movements reduce local uncertainty," *J. Vis.*, vol. 7, no. 3, p. 6, Feb. 2007.

[3] M. Hayhoe and D. Ballard, "Eye movements in natural behavior," *Trends Cogn. Sci.*, vol. 9, no. 4, pp. 188–194, 2005.

[4] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz, "Visual memory and motor planning in a natural task," *J. Vis.*, vol. 3, no. 1, p. 6, 2003.

[5] S. Ohl and M. Rolfs, "Saccadic eye movements impose a natural bottleneck on visual short-term memory," *J. Exp. Psychol., Learn., Memory, Cogn.*, vol. 43, no. 5, p. 736, 2017.

[6] J. Kwon and J. Y. Kim, "Meaning of gaze behaviors in individuals' perception and interpretation of commercial interior environments: An experimental phenomenology approach involving eye-tracking," *Frontiers Psychol.*, vol. 12, p. 3290, Aug. 2021.

[7] B. D. Corneil and D. P. Munoz, "The influence of auditory and visual distractors on human orienting gaze shifts," *J. Neurosci.*, vol. 16, no. 24, pp. 8193–8207, Dec. 1996.

[8] J. D. Ryan, D. E. Hannula, and N. J. Cohen, "The obligatory effects of memory on eye movements," *Memory*, vol. 15, no. 5, pp. 508–525, Jul. 2007.

[9] R. E. Parker, "Picture processing during recognition," *J. Exp. Psychol., Hum. Perception Perform.*, vol. 4, no. 2, p. 284, 1978.

[10] K. G. Gruters, D. L. Murphy, C. D. Jenson, D. W. Smith, C. A. Shera, and J. M. Groh, "The eardrums move when the eyes move: A multisensory effect on the mechanics of hearing," *Proc. Nat. Acad. Sci. USA*, vol. 115, no. 6, pp. E1309–E1318, 2018.

[11] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The autism-spectrum quotient (AQ): Evidence from asperger syndrome/high-functioning autism, malesand females, scientists and mathematicians," *J. Autism Develop. Disorders*, vol. 31, no. 1, pp. 5–17, 2001.

[12] C. C. A. H. Bours *et al.*, "Emotional face recognition in male adolescents with autism spectrum disorder or disruptive behavior disorder: An eye-tracking study," *Eur. Child Adolescent Psychiatry*, vol. 27, no. 9, pp. 1143–1157, Sep. 2018.

[13] Q. Guillon, N. Hadjikhani, S. Baduel, and B. Rogé, "Visual social attention in autism spectrum disorder: Insights from eye tracking studies," *Neurosci. Biobehav. Rev.*, vol. 42, pp. 279–297, May 2014.

[14] J. B. Wagner, S. B. Hirsch, V. K. Vogel-Farley, E. Redcay, and C. A. Nelson, "Eye-tracking, autonomic, and electrophysiological correlates of emotional face processing in adolescents with autism spectrum disorder," *J. Autism Develop. Disorders*, vol. 43, no. 1, pp. 188–199, Jan. 2013.

[15] M. Hanley, D. M. Riby, C. Carty, A. M. McAteer, A. Kennedy, and M. McPhillips, "The use of eye-tracking to explore social difficulties in cognitively able students with autism spectrum disorder: A pilot investigation," *Autism*, vol. 19, no. 7, pp. 868–873, Oct. 2015.

[16] S. W. White, B. B. Maddox, and R. K. Panneton, "Fear of negative evaluation influences eye gaze in adolescents with autism spectrum disorder: A pilot study," *J. Autism Develop. Disorders*, vol. 45, no. 11, pp. 3446–3457, Nov. 2015.

[17] V. Yaneva, L. A. Ha, S. Eraslan, Y. Yesilada, and R. Mitkov, "Detecting high-functioning autism in adults using eye tracking and machine learning," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 6, pp. 1254–1261, Jun. 2020.

[18] J. D. Stokes, A. Rizzo, J. J. Geng, and J. B. Schweitzer, "Measuring attentional distraction in children with ADHD using virtual reality technology with eye-tracking," *Frontiers Virtual Reality*, vol. 3, p. 23, Mar. 2022.

[19] G. Nie *et al.*, "An immersive computer-mediated caregiver-child interaction system for young children with autism spectrum disorder," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 884–893, 2021.

[20] P. R. K. Babu and U. Lahiri, "Multiplayer interaction platform with gaze tracking for individuals with autism," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 28, no. 11, pp. 2443–2450, Nov. 2020.

[21] D. C. Strickland, C. D. Coles, and L. B. Southern, "JobTIPS: A transition to employment program for individuals with autism spectrum disorders," *J. Autism Develop. Disorders*, vol. 43, no. 10, pp. 2472–2483, Oct. 2013.

[22] M. J. Maenner *et al.*, "Prevalence of autism spectrum disorder among children aged 8 years—Autism and developmental disabilities monitoring network, 11 sites, United States, 2016," *MMWR Surveill. Summaries*, vol. 69, no. 4, p. 1, 2020.

[23] A. V. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell, "Costs of autism spectrum disorders in the United Kingdom and the United States," *J. Amer. Med. Assoc. Pediatrics*, vol. 168, no. 8, pp. 721–728, 2014.

[24] R. Cameto, C. Marder, M. Wagner, and D. Cardoso, "Youth employment," *NLTS2 Data Brief*, vol. 2, no. 2, pp. 1–6, 2003.

[25] J. L. Chen, G. Leader, C. Sung, and M. Leahy, "Trends in employment for individuals with autism spectrum disorder: A review of the research literature," *Rev. J. Autism Develop. Disorders*, vol. 2, no. 2, pp. 115–127, Jun. 2015.

[26] D. B. Burt, S. P. Fuller, and K. R. Lewis, "Brief report: Competitive employment of adults with autism," *J. Autism Develop. Disorders*, vol. 21, no. 2, pp. 237–242, Jun. 1991.

[27] D. Hagner and B. F. Cooney, "'I do that for everybody': Supervising employees with autism," *Focus Autism Other Develop. Disabilities*, vol. 20, no. 2, pp. 91–97, 2005.

[28] L. A. Sperry and G. B. Mesibov, "Perceptions of social challenges of adults with autism spectrum disorder," *Autism*, vol. 9, no. 4, pp. 362–376, Oct. 2005.

[29] S. Artiran, L. Chukoskie, A. Jung, I. Miller, and P. Cosman, "HMM-based detection of head nods to evaluate conversational engagement from head motion data," in *Proc. 29th EUSIPCO*, Aug. 2021, pp. 1301–1305.

[30] O. N. Tepencelik, W. Wei, L. Chukoskie, P. C. Cosman, and S. Dey, "Body and head orientation estimation with privacy preserving LiDAR sensors," in *Proc. 29th EUSIPCO*, Aug. 2021, pp. 766–770.

[31] X. Pan and A. F. D. C. Hamilton, "Why and how to use virtual reality to study human social interaction: The challenges of exploring a new research landscape," *Brit. J. Psychol.*, vol. 109, no. 3, pp. 395–417, Aug. 2018.

[32] S. Bonaccio, J. O'Reilly, S. L. O'Sullivan, and F. Chiocchio, "Nonverbal behavior and communication in the workplace: A review and an agenda for research," *J. Manag.*, vol. 42, no. 5, pp. 1044–1074, 2016.

[33] R. D. Arvey and J. E. Campion, "The employment interview: A summary and review of recent research," *Personnel Psychol.*, vol. 35, no. 2, pp. 281–322, 1982.

[34] M. Nyström, R. Andersson, K. Holmqvist, and J. Van De Weijer, "The influence of calibration method and eye physiology on eyetracking data quality," *Behav. Res. Methods*, vol. 45, no. 1, pp. 272–288, 2013.

[35] J. D. Morgante, R. Zolfaghari, and S. P. Johnson, "A critical test of temporal and spatial accuracy of the Tobii T60XL eye tracker," *Infancy*, vol. 17, no. 1, pp. 9–32, Jan. 2012.

[36] S. Davis, K. Nesbitt, and E. Nalivaiko, "Comparing the onset of cybersickness using the Oculus Rift and two virtual roller coasters," in *Proc. 11th Austral. Conf. Interact. Entertainment (IE)*, vol. 27. Sydney, NSW, Australia: Australian Computing Society Sydney, 2015, p. 30.

[37] C. Yildirim, "Don't make me sick: Investigating the incidence of cybersickness in commercial virtual reality headsets," *Virtual Reality*, vol. 24, no. 2, pp. 231–239, 2020.

[38] S. Martirosov, M. Bureš, and T. Zítka, "Cyber sickness in low-immersive, semi-immersive, and fully immersive virtual reality," *Virtual Reality*, vol. 26, no. 1, pp. 15–32, Mar. 2022.

[39] L. Bozgeyikli, A. Raij, S. Katkoori, and R. Alqasemi, "A survey on virtual reality for individuals with autism spectrum disorder: Design considerations," *IEEE Trans. Learn. Technol.*, vol. 11, no. 2, pp. 133–151, Apr./Jun. 2018.

[40] N. Newbutt, C. Sung, H.-J. Kuo, M. J. Leahy, C.-C. Lin, and B. Tong, "Brief report: A pilot study of the use of a virtual reality headset in autism populations," *J. Autism Develop. Disorders*, vol. 46, no. 9, pp. 3166–3176, Sep. 2016.

[41] M. Schmidt, C. Schmidt, N. Glaser, D. Beck, M. Lim, and H. Palmer, "Evaluation of a spherical video-based virtual reality intervention designed to teach adaptive skills for adults with autism: A preliminary report," *Interact. Learn. Environ.*, vol. 29, no. 3, pp. 345–364, Apr. 2021.

[42] T. Starner, *Wearable Computing*, vol. 318. Cambridge, MA, USA: MIT Media Lab., 2015.

[43] J. Van Waveren, "The asynchronous time warp for virtual reality on consumer hardware," in *Proc. 22nd ACM Conf. Virtual Reality Softw. Technol.*, Nov. 2016, pp. 37–46.

[44] H. Grillon and D. Thalmann, "Eye contact as trigger for modification of virtual character behavior," in *Proc. Virtual Rehabil.*, Aug. 2008, pp. 205–211.

[45] P. R. Krishnappa Babu, P. Oza, and U. Lahiri, "Gaze-sensitive virtual reality based social communication platform for individuals with autism," *IEEE Trans. Affect. Comput.*, vol. 9, no. 4, pp. 450–462, Oct. 2018.

[46] R. Vertegaal, R. Slagter, G. Van Der Veer, and A. Nijholt, "Eye gaze patterns in conversations: There is more to conversational agents than meets the eyes," in *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.*, Mar. 2001, pp. 301–308.

[47] M. Freeth and P. Bugembe, “Social partner gaze direction and conversational phase; factors affecting social attention during face-to-face conversations in autistic adults?” *Autism*, vol. 23, no. 2, pp. 503–513, Feb. 2019.

[48] Y. Yoshikawa, H. Kumazaki, Y. Matsumoto, M. Miyao, M. Kikuchi, and H. Ishiguro, “Relaxing gaze aversion of adolescents with autism spectrum disorder in consecutive conversations with human and Android robot—A preliminary study,” *Frontiers Psychiatry*, vol. 10, p. 370, Jun. 2019.

[49] B. Noris, J. Nadel, M. Barker, N. Hadjikhani, and A. Billard, “Investigating gaze of children with ASD in naturalistic settings,” *PLoS One*, vol. 7, no. 9, Sep. 2012, Art. no. e44144.

[50] J. Reichenberger, M. Pfaller, and A. Mühlberger, “Gaze behavior in social fear conditioning: An eye-tracking study in virtual reality,” *Frontiers Psychol.*, vol. 11, p. 35, Jan. 2020.

[51] A. Kendon, “Some functions of gaze-direction in social interaction,” *Acta Psychol.*, vol. 26, pp. 22–63, Jan. 1967.

[52] H. Mansour and G. Kuhn, “Studying ‘natural’ eye movements in an ‘unnatural’ social environment: The influence of social activity, framing, and sub-clinical traits on gaze aversion,” *Quart. J. Exp. Psychol.*, vol. 72, no. 8, pp. 1913–1925, 2019.

[53] J. R. Aiello, “A further look at equilibrium theory: Visual interaction as a function of interpersonal distance,” *Environ. Psychol. Nonverbal Behav.*, vol. 1, no. 2, pp. 122–140, Mar. 1977.

[54] N. F. Russo, “Eye contact, interpersonal distance, and the equilibrium theory,” *J. Pers. Social Psychol.*, vol. 31, no. 3, p. 497, 1975.

[55] K. Asada, Y. Tojo, H. Osanai, A. Saito, T. Hasegawa, and S. Kumagaya, “Reduced personal space in individuals with autism spectrum disorder,” *PLoS One*, vol. 11, no. 1, Jan. 2016, Art. no. e0146306.

[56] A. Perry, E. Levy-Gigi, G. Richter-Levin, and S. G. Shamay-Tsoory, “Interpersonal distance and social anxiety in autistic spectrum disorders: A behavioral and ERP study,” *Social Neurosci.*, vol. 10, no. 4, pp. 354–365, 2015.

[57] J. M. Madera and M. R. Hebl, “Discrimination against facially stigmatized applicants in interviews: An eye-tracking and face-to-face investigation,” *J. Appl. Psychol.*, vol. 97, no. 2, p. 317, 2012.

[58] F. Tian, S. Okada, and K. Nitta, “Analyzing eye movements in interview communication with virtual reality agents,” in *Proc. 7th Int. Conf. Hum.-Agent Interact.*, Sep. 2019, pp. 3–10.

[59] X. Wang *et al.*, “Eye contact conditioning in autistic children using virtual reality technology,” in *Proc. Int. Symp. Pervasive Comput. Paradigms Mental Health*. Cham, Switzerland: Springer, 2014, pp. 79–89.

[60] Y. Cheng and J. Ye, “Exploring the social competence of students with autism spectrum conditions in a collaborative virtual learning environment—The pilot study,” *Comput. Educ.*, vol. 54, no. 4, pp. 1068–1077, 2010.

[61] V. Jyoti and U. Lahiri, “Virtual reality based joint attention task platform for children with autism,” *IEEE Trans. Learn. Technol.*, vol. 13, no. 1, pp. 198–210, Jan. 2020.

[62] A. Z. Amat, H. Zhao, A. Swanson, A. S. Weitlauf, Z. Warren, and N. Sarkar, “Design of an interactive virtual reality system, InViRS, for joint attention practice in autistic children,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 29, pp. 1866–1876, 2021.

[63] U. Lahiri, E. Bekele, E. Dohrmann, Z. Warren, and N. Sarkar, “Design of a virtual reality based adaptive response technology for children with autism,” *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 21, no. 1, pp. 55–64, Jan. 2013.

[64] N. S. Rosenfield, K. Lamkin, J. Re, K. Day, L. Boyd, and E. Linstead, “A virtual reality system for practicing conversation skills for children with autism,” *Multimodal Technol. Interact.*, vol. 3, no. 2, p. 28, 2019.

[65] M. Elgarf, S. Abdennadher, and M. Elshahawy, “I-interact: A virtual reality serious game for eye contact improvement for children with social impairment,” in *Proc. Joint Int. Conf. Serious Games*. Cham, Switzerland: Springer, 2017, pp. 146–157.

[66] *Recording Facial Animation From an iOS Device*. Accessed: Apr. 25, 2021. [Online]. Available: https://docs.unrealengine.com/4.27/en-US/AnimatingObjects/SkeletalMeshAnimation/FacialRecordingiPhone/

[67] D. Birant and A. Kut, “ST-DBSCAN: An algorithm for clustering spatial–temporal data,” *Data Knowl. Eng.*, vol. 60, no. 1, pp. 208–221, Jan. 2007.

[68] M. Ester *et al.*, “A density-based algorithm for discovering clusters in large spatial databases with noise,” in *Proc. 2nd Int. Conf. Knowl. Discovery Data Mining*, 1996, vol. 96, no. 34, pp. 226–231.

[69] Z. Shi and L. Pun-Cheng, “Spatiotemporal data clustering: A survey of methods,” *ISPRS Int. J. Geo-Inf.*, vol. 8, no. 3, p. 112, Feb. 2019.

[70] H. F. Tork, “Spatio-temporal clustering methods classification,” in *Proc. Doctoral Symp. Inform. Eng.*, vol. 1, no. 1. Lisbon, Portugal: Faculdade de Engenharia da Universidade do Porto, 2012, pp. 199–209.

[71] R. Trasarti, “Mastering the spatio-temporal knowledge discover process,” Ph.D. dissertation, Dept. Comput. Sci., Univ. Pisa, Pisa, Italy, 2010.

[72] S. Godara, R. Singh, and S. Kumar, “Proposed density based clustering with weighted Euclidean distance,” *Int. J. Adv. Res. Comput. Sci. Softw. Eng.*, vol. 7, no. 6, pp. 409–412, Jun. 2017.

[73] R. Oliveira, M. Y. Santos, and J. M. Pires, “4D+SNN: A spatio-temporal density-based clustering approach with 4D similarity,” in *Proc. IEEE 13th Int. Conf. Data Mining Workshops*, Dec. 2013, pp. 1045–1052.

[74] B. Kenward *et al.*, “Saccadic reaction times in infants and adults: Spatiotemporal factors, gender, and interlaboratory variation,” *Develop. Psychol.*, vol. 53, no. 9, p. 1750, 2017.

[75] I. Fatt and B. A. Weissman, *Physiology Eye: An Introduction to Vegetative Functions*. Oxford, U.K.: Butterworth, 2013.

[76] D. Fein *et al.*, “Optimal outcome in individuals with a history of autism,” *J. Child Psychol. Psychiatry*, vol. 54, no. 2, pp. 195–205, Feb. 2013.

[77] T. D. V. Swinscow *et al.*, *Statistics at Square One*. London, U.K.: BMJ, 2002.

[78] A. Klin, W. Jones, R. Schultz, F. Volkmar, and D. Cohen, “Visual fixation patterns during viewing of naturalistic social situations as predictors of social competence in individuals with autism,” *Arch. Gen. Psychiatry*, vol. 59, no. 9, pp. 809–816, 2002.

[79] M. Roelofs, *AIMMS Language Reference*. Abu Dhabi, United Arab Emirates: Lulu.com, 2010.

[80] F. G. Stevens, V. C. Plaut, and J. Sanchez-Burks, “Unlocking the benefits of diversity: All-inclusive multiculturalism and positive organizational change,” *J. Appl. Behav. Sci.*, vol. 44, no. 1, pp. 116–133, 2008.

[81] R. D. Austin and G. P. Pisano, “Neurodiversity as a competitive advantage,” *Harvard Bus. Rev.*, vol. 95, no. 3, pp. 96–103, 2017.

[82] R. Cope and A. Remington, “The strengths and abilities of autistic people in the workplace,” *Autism Adulthood*, vol. 4, no. 1, pp. 22–31, Mar. 2022.

[83] A. Krzeminska, R. D. Austin, S. M. Bruyère, and D. Hedley, “The advantages and challenges of neurodiversity employment in organizations,” *J. Manag. Org.*, vol. 25, no. 4, pp. 453–463, Jul. 2019.