Analysis of Gaze, Head Orientation, and Joint Attention in Autism With Triadic VR Interviews

Saygin Artiran, Poorva S. Bedmutha, and Pamela Cosman, *Fellow, IEEE*

***Abstract*— Effective use of gaze and head orientation can strengthen the sense of inclusion in multi-party interactions, including job interviews. Not making significant eye contact with the interlocutors, or not turning towards them, may be interpreted as disinterest, which could worsen job interview outcomes. This study aims to support the situational solo practice of gaze behavior and head orientation using a triadic (three-way) virtual reality (VR) job interview simulation. The system lets users encounter common interview questions and see how they share attention among the interviewers based on their conversational role (speaking or listening). Given the yaw and position readings of the VR headset, we use a machine learning-based approach to analyze head orientations relative to the interviewers in the virtual environment, and achieve low angular error in a low complexity way. We examine the degree to which interviewer backchannels trigger attention shifts or behavioral mirroring and investigate the social modulation of gaze and head orientation for autistic and non-autistic individuals. In both speaking and listening roles, the autistic participants gazed at, and oriented towards the two virtual interviewers less often, and they displayed less behavioral mirroring (mirroring the head turn of one avatar towards another) compared to the non-autistic participants.**

***Index Terms*— Autism, job interview practice, machine learning, social modulation of gaze and head orientation, virtual reality.**

I. Introduction

VIRTUAL reality (VR) has rapidly gained traction in recent years. It is employed in numerous fields, including games, interior design, and healthcare. The immersive and interactive nature of VR lets users activate their senses to blend with the environment. The sensation of becoming physically present in a non-physical world may offer a distinct opportunity for effective experiential learning. For example, users can practice social communicative skills in a supervised setting without fear of real-world repercussions.

Gaze can be used to perceive information from others or to signal a variety of meanings, e.g., wishing to communicate [1]. Conversational roles may alter gaze characteristics, which is called the social modulation of gaze; when listening, individuals make more eye contact than when speaking [2], [3], [4]. Head rotation and gaze direction tend to align about 70% of the time [5]. Head orientation alone can also allude to the visual center of attention in a conversation [6]. Like gaze and head rotation, backchannels can also be used to show interest in the conversation. Conversational backchannels can be verbal, e.g., "uh-huh", "hmm", "yes", "wow", or nonverbal such as head nods, shakes, and tilts.

Observations of individuals' gaze patterns and head orientations, along with their responses to backchannels can yield valuable insights into their propensity for engagement and joint attention. Joint attention refers to the deliberate alignment of an individual's focus of attention with that of another person, resulting in both parties looking at the same subject matter. This capacity holds special significance during early developmental stages as it facilitates children in acquiring object names and object usage guidelines, contributing to later developmental milestones [7]. An individual's focus of attention can be inferred from gaze patterns when their face is visible; head or body orientation can also provide valuable clues [8]. Both gaze and head orientation serve as critical mechanisms for either enacting or evaluating joint attention. In situations involving multiple listeners, speakers can prevent excluding participants by briefly directing their gaze or orienting their bodies towards them [9]. This includes acknowledging listener input. Such actions are pivotal for ensuring that other parties feel acknowledged and integrated in the conversation.

Effective use of gaze and head orientation can boost an individual's chances of securing and maintaining employment. During job interviews, applicants who maintained direct eye contact and kept their head up were generally perceived as more confident and reliable, increasing their likelihood of being offered the job [10]. Other early research supported these findings, revealing a strong connection between higher levels of perceived self-confidence and competence in interviewees who effectively used nonverbal cues such as eye contact and head orientation [11], [12], [13]. In a more recent study, interviewees who maintained optimal eye contact with the interviewer, without excessive gaze wandering or staring, received higher interview scores [14]. Candidates who demonstrated effective head orientation towards conversational partners were selected more frequently for employment opportunities.

Autism is a multifaceted developmental condition that influences an individual's social interactions and information

Manuscript received 3 November 2023; revised 17 January 2024; accepted 4 February 2024. Date of publication 8 February 2024; date of current version 15 February 2024. 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.

The authors are with the Department of Electrical and Computer Engineering, University of California San Diego, La Jolla, CA 92093 USA (e-mail: sartiran@ucsd.edu; pbedmutha@ucsd.edu; pcosman@ucsd.edu).

Digital Object Identifier 10.1109/TNSRE.2024.3363728

processing [15]. One out of every 44 children is autistic [16], and autism is linked to a high unemployment rate, as 69% of individuals with autism express a desire to work, yet only 20% of this demographic finds gainful employment [17], [18]. Researchers have delved into contrasting patterns of gaze behavior and head movements in non-autistic (NA) and autistic individuals. Variations in head movements were identified as potential early indicators of autism [19], [20]. Autistic individuals exhibited a reduced tendency to initiate and sustain eye contact, with less focus on facial features [21], [22], [23], [24], [25], [26]. A study suggested that autistic individuals might fixate on a single person while neglecting others [27], or may face challenges when attempting joint attention [28]. Differences in social communication, coupled with conventional workplace communication norms and hiring procedures, could be hindering the employment prospects of autistic individuals [29], [30], [31], [32].

In this study, our contributions are: (1) A novel triadic job interview system in VR that enables fully automatic analysis of joint attention and of social modulation of gaze and head orientation, (2) A multilayer perceptron (MLP) regressor to adjust head rotation values based on head movements to accurately detect engagement with virtual targets, (3) A novel study of joint attention and social modulation of head orientation for both autistic and NA participants using immersive VR.

The rest of this paper is organized as follows. Section II summarizes related work and Section III describes the triadic VR mock job interview application. In Section IV, we explain the virtual interviewers' backchannel behaviors during the mock job interviews, while Section V presents the system pipeline and data processing procedures. Section VI presents comparative social behavior analyses of our participants, and we conclude with Section VII.

II. Related Work

In this section, we review studies where VR was employed to analyze social behaviors such as gaze, head movements, and mirroring in a variety of environments, including triadic (three-person) conversations. We also report on previous work where the technology was used as a tool to enable the practice of joint attention and social skills in job interviews.

A. Gaze Behavior Analysis

Precise surveillance of eye movements can allow VR systems to understand user gaze patterns, identify objects that draw attention, and investigate the impact of stimuli on gaze behavior. VR eye tracking has found applications across a range of fields, such as gaming [33] and psychology [34]. Fathy et al. [35] built an ML architecture to forecast visual focus in VR, showing the promise of using large-scale gaze data to analyze gaze patterns in VR. Wang et al. [36] investigated visual attention in VR driving simulators.

A few studies investigated gaze behavior in the context of job interviews. The authors of [37] developed a tool to practice maintaining eye contact. During the interviews, the virtual character's level of interest was influenced by the user's gaze direction. If the user gazed towards the virtual character, it appeared interested, otherwise, it behaved as though it were not paying attention. In [38], participants acted as interviewers listening to job applicants in both computer-mediated and face-to-face mock interviews; eye tracking was used to examine how scar-like facial features influenced gaze patterns.

Contemporary eye trackers, particularly those in head-mounted displays (HMDs), encounter difficulties with calibration and data accuracy [39], [40]. Glasses, contact lenses, mascara, and physiological aspects such as eye color can impede gaze tracking. This may necessitate repeated calibration and could reduce measurement trustworthiness. Researchers have devised algorithms to mitigate issues due to calibration drift, headset movement, and blinking [4], [41], [42].

B. Head Orientation Analysis

Head movements can communicate information on emotions, intentions, and conversational engagement. Researchers have used VR headsets to understand the role of head movements in engaging with virtual environments [43], [44]. Xiao et al. [45] aimed to discern the head movements executed by participants in dyadic (two-person) interactions to acknowledge or criticize the other party. In [46], an ML model gauged conversational engagement levels based on head gestures detected using readings from an augmented reality headset.

VR technology for analyzing head orientation has shown utility across diverse fields such as spatial cognition and training simulations [47], [48]. Researchers have studied how people organically move their heads during social engagements within immersive virtual settings [49]. These studies revealed that effective head rotations can enhance the sense of inclusion among interlocutors. Beyond gaze, comprehending the patterns of head rotation in social contexts can improve the realism of simulated conversational systems [50].

C. Social Behavior Analysis in Triadic Conversations

Although they can inform about some social communicative behaviors, dyadic interactions have limitations in capturing other behaviors like social exclusion. There are some studies of triadic interactions in real-world scenarios. In [51], unequal distribution of gaze by a salesperson in triadic sales encounters was perceived as favoritism, leading to reduced trust from customers. Zima et al. [52] explored how speakers use gaze to assert or relinquish their conversational turn; gaze aversion from co-starting speakers was an effective strategy in securing speaking opportunities. In [53], head and eye movements of listeners were found to accurately indicate speaker location, regardless of background noise level, although head movements alone slightly undershot the speaker's position.

Unlike the real-world studies, VR use has been limited in studying triadic interactions. Hladek and Seeber [54] expanded upon [53], observing a similar undershooting behavior in a VR-based triadic conversation scenario. Using immersive VR, and head and hand tracking, Miller et al. [55] explored synchrony within triads. Synchrony, which represents the

natural time-dependence of behaviors in human interactions, was influenced by the virtual environment, the dynamics of turn-taking within the triad, and gaze. Tarr et al. [56] designed an experiment in which participants, represented by virtual agents, engaged in a collaborative movement activity with two other participants. Participants in the synchrony condition reported significantly higher social closeness to their virtual co-participants compared to those in the non-synchrony condition.

D. Behavioral Mirroring and Joint Attention

Unintentional behavioral mirroring describes the phenomenon of passively and involuntarily mimicking the postures, expressions, and mannerisms of one's counterparts in social settings. In an earlier non-VR study, Chartrand and Bargh [57] examined how this unconscious mimicry can enhance the fluidity of interactions and foster a greater sense of liking between individuals. Their findings pointed at a direct correlation between the degree of behavioral mirroring and participants' self-perceived rapport. Novotny et al. [58] investigated mirroring in interviews. Following the initial interviews, participants who were paired with interviewers who engaged in mirroring showed a greater willingness to share additional information when compared to a control group in which interviewers intentionally refrained from mirroring.

In [59], researchers examined how head gestures exhibited by virtual interviewers shaped trust and liking towards them. Head gestures that appeared to follow naturalistic patterns and that were realistically mimicked led to enhanced synchronization and a stronger perception of mutual understanding. Hence, virtual interviewers capable of providing lifelike conversational backchannels might enhance the immersiveness and authenticity of VR experiences.

VR technologies have provided opportunities for groups with social communication differences to practice joint attention. In [60], a VR-based joint attention practice module helped autistic school-aged children improve their joint attention abilities, leading to more normative eye contact patterns and increased initiation of interactions. Mei et al. [61] reported that customizable virtual humans can remind participants to focus on task-relevant areas within a VR game, improving their task performance.

E. Job Interview Practice

Various tools have been developed to aid individuals in practicing skills for job interviews. Strickland et al. [62] designed a job interview practice suite that comprised multiple approaches, including VR. Other studies have demonstrated the effectiveness of VR-based job interview practice, with participants gaining familiarity with common interview questions, performing better after the practice, and reporting reduced anxiety and increased self-confidence [63], [64], [65].

In summary, previous research has demonstrated the potential of VR for studying human behavior and social interactions. VR simulations of job interviews can let individuals tailor their social communication skills to increase their employment chances while reducing the potential costs of having that same practice delivered by a human coach [66].

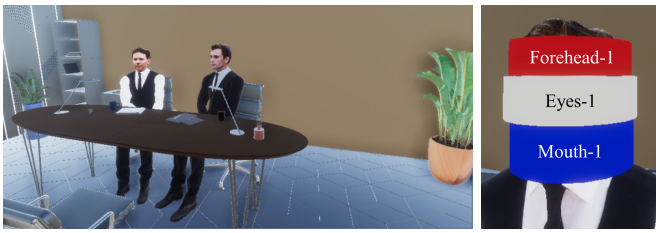

Fig. 1. (a) The virtual office space. (b) Interviewer face regions: Forehead, Eyes, Mouth.

III. VR Interview Application

A. Main Design

In our simulation, the job interview was for a video game company as video gaming is a common interest of young adults, and because neurodivergent individuals are strongly represented in the video game industry [67]. Users signed a consent form before the simulations. They wore an HTC Vive Pro Eye VR headset to visualize the virtual office space. It contained two virtual interviewers along with common office objects (e.g., desk, notebooks, pens, plant, closet; Fig. 1a). The interview consists of 43 questions, such as "What are some skills you would like to gain while working with us?" and "Have you previously worked with any game engines?". To signal that they fully answered a question, the users use the controller's trigger button. If needed, a question could be repeated by pressing the trackpad. This study was approved by the UC San Diego Institutional Review Board under IRB Protocol 210775 (Date of approval 7/1/2021).

The virtual interviewers ask questions one by one, not referring back to previous questions. The users respond as they wish and can pause to think. Each interviewer asks roughly half of the questions, with assignments fixed across users. For increased realism, an interviewer could ask consecutive questions or remain silent for longer durations.

We used the Live Link Face app to record the facial animations and voices of two individuals while they read 43 predefined job interview lines to a camera. The recorded data was processed in Blender and added on head models as key points. Facial textures were created using photos of the individuals. The interviewers looked engaged by means of head nodding, shaking, and tilting animations, performed when the users talk. In addition, the interviewers provide verbal backchannels such as "uh-huh" and "hmm." An interviewer turns their head towards the other interviewer when the other asks a question. They can also turn to the other interviewer when he gives a backchannel during user response. Other than backchannels and head turns, the interviewers do not engage in nonverbal communication such as arm movements. The interviewers' faces were divided into forehead, eyes, and mouth (Fig. 1b).

3D coordinates and angular velocities (pitch, yaw, roll) of the headset are tracked using Unity. Eye tracking uses the built-in tracker of the headset, accessible through Unity using the Vive SRanipal SDK. At each time step, the system collects gaze origin and direction and finds the virtual object that collides with the related gaze ray. The system records

the object label, intersection location, and time spent in that time step which is not fixed across time steps. Detected speech levels are also recorded at each time step, which is also accessible through Unity. During each interview session, in addition to all of the previously mentioned data, information regarding whether the user blinked at a time step, question IDs, and the total time and percentage spent gazing at each face region are recorded and collected in a CSV file. These data are not publicly accessible due to our IRB protocol.

B. Participatory Design

Autism researchers have recently started favoring participatory design (PD) techniques; autistic individuals coordinating with researchers can co-develop practical technologies that reflect the insights of end users [68], [69]. To improve our application, we conducted a PD session with two autistic adults (one college educated, one not) to discuss the acceptability, ethics, and design of the VR job interview simulation. Our initial design was the office environment we introduced in [70]. Although our design partners were not distracted by their surroundings and thought the space was realistic, they suggested populating the desk with more objects which would make it look more cluttered and natural. Both design partners thought the interviewer speech was realistic and immersive, and reported that audio was in sync with mouth movements. They suggested adding subtitles as an option to better accommodate the needs of individuals with impaired hearing. They were generally pleased with the execution and timing of the head turns but reported that although not common, some of them were slightly too fast. One design partner felt interrupted by some of the verbal backchannels as they were sometimes performed while the participant was talking. Per our design partners' suggestions, we put more objects on the virtual desk and modified our design to have the interviewers only perform physical backchannels (e.g., head nods) if the user is speaking, and both backchannel types, otherwise.

IV. Voice Activity and Backchannel Timing

Our system tracks user speech to control interviewer backchannel behavior. The starting point of a user's response is important as we do not want an interviewer to give a backchannel before the user starts speaking. Likewise, voice activity detection (VAD) is important to prevent the interviewers from interrupting the users by giving out of place verbal backchannels as pointed out during the PD session.

The system's audio sampling rate is 48kHz. Our application runs at around 45 time points per second, and the root mean squared (RMS) value of the audio samples over 1 second are computed at each time point. A Gaussian weighted window of RMS values centered at a time point is thresholded to decide whether there is voice activity. This approach is defined by the threshold (t_{RMS}), window size (w), and standard deviation of the Gaussian distribution (σ). To validate and tune our VAD algorithm, we had 12 individuals wear the headset and read aloud 3 scripts displayed in VR with and without additional background noise (office sounds on YouTube). Each recording took about 3 minutes; we annotated the time intervals where the reader was silent for longer than 0.5s. In total 1,582 silence segments were marked (average length = 1.3s). The optimal parameter set ($t_{\hat{RMS}}$, $\hat{w}$, $\hat{\sigma}$) = (0.07, 29, 21) maximized the sum of true positive rate, true negative rate, and intersection over union value (92.8%, 89%, and 89.8%).

TABLE I
Time to First Backchannel, and Backchannel Spacing

Type	Avg. Initial Wait Time		Avg. Time In Between	
	$IntQ$	$IntO$	$IntQ$	$IntO$
Verbal	9.3s (4.6s)	10.3s (3.4s)	9.7s (3.5s)	13.7s (8.6s)
Physical	8.1s (0.9s)	9s (3.6s)	8.3s (1.3s)	12.1s (2.3s)

For our application, we had to determine when the virtual interviewers ought to provide a backchannel. We based this decision on timings from 6 in-person 10-minute professional conversations involving an interviewee and two individuals who acted as interviewers. The interviewer wait times for verbal backchannels (VB) and physical backchannels (PB) were similar. From here on, we call the interviewer who asked the most recent question $IntQ$, and the other interviewer $IntO$. On average, $IntQ$ gives the first backchannel at t+8.5s (s.d.=2.8s), and $IntO$ at t+9.5s (s.d.=3.5s) where t marks the start of the interviewee's response. $IntQ$ and $IntO$ spend about 9s (s.d.=2.4s) and 13s (s.d.=5.5s) between consecutive backchannels. The shortest interviewee response that received a backchannel was 2.5s.

The virtual interviewers randomly perform backchannels based on these numbers from real conversations, while never giving a VB in the presence of user speech. If 2.5s of voice activity is detected starting at time t, $IntQ$ gives an initial backchannel at a time drawn uniformly randomly from the interval $t + 8.5\pm2.8$s, while this interval is $t + 9.5\pm3.5$s for $IntO$, as long as the user does not push the trigger button to indicate the end of their answer prior to the backchannel occurrence. For $IntQ$, the duration between consecutive backchannels is drawn from 9 ± 2.4s, and from 13 ± 5.5s for $IntO$ (Table I).

Among the possibilities of interviewer backchannels and head turns towards each other, we consider four cases: $QbOl$: $IntQ$ may give backchannels, $IntO$ merely listens, $QlOb$: $IntQ$ merely listens, $IntO$ may give backchannels, $QbOb$: both interviewers may give backchannels, and $QtOb$: $IntQ$ turns his head towards $IntO$ when $IntO$ gives a backchannel, but does not give backchannels to the user. For those 5 questions that were known from [4] to consistently have user responses shorter than the average initial backchannel times in Table I, the virtual interviewers do not give backchannels. The remaining 38 questions were assigned with 10 each to the first two cases, and 9 each to the other two, where questions with answers of different lengths were split evenly among the cases. These assignments were fixed across participants.

V. System Pipeline

The system pipeline is visualized in Fig. 2. The process begins with a VR session in which the user's voice activity is tracked in real-time to determine the timing of conversational backchannels by the avatars, and all gaze data and head motion data is recorded. After each mock job interview session,

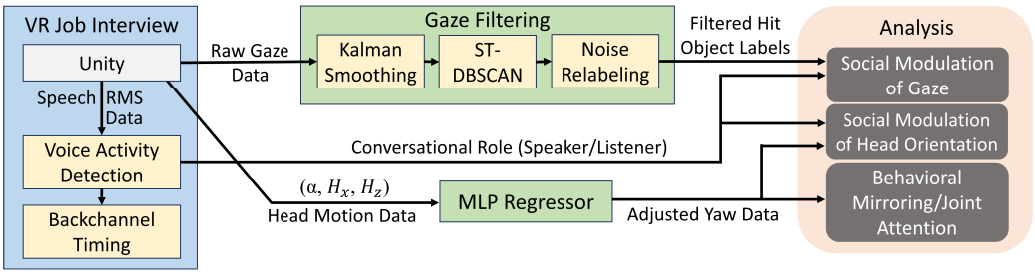

Fig. 2. System pipeline, showing the VR session on the left, the data processing in the middle section, and the behavioral analysis on the right. In the figure, α is the yaw angle measured by the headset, and H_X and H_Z are x and z-positions of the headset. Note that the analysis portion also takes as input (not shown) from the VR application the identity of *IntQ*/*IntO* and whether *IntQ* turned his head towards *IntO* at a time point or not.

the gaze data and head motion data from the headset are processed separately. The middle portion of the diagram (non real-time) shows the data processing, which consists of gaze filtering (top branch, described in Section V-A) and head motion processing (bottom branch, described in Section V-B). Finally, the analysis portion on the right uses the sequence of gaze object labels and head orientation angles to tabulate the user's engagement under different conditions (results shown in Section VI).

A. Gaze Processing

The gaze data is processed using the algorithm from [4] which reduces tracking inaccuracies due to factors like blinking, headset slippage, and abrupt head movements. As shown on the upper branch of Fig. 2, the gaze filtering algorithm begins with Kalman smoothing to remove blinks and jitter, then clusters the gaze locations in time and space using ST-DBSCAN [71]. The algorithm has a final step in which points that were initially labeled as noise (no cluster assignment) by ST-DBSCAN get relabeled based on their spatial-temporal distance from the clusters.

The gaze filtering algorithm is defined by four parameters: maximum spatial ($\epsilon 1$) and temporal ($\epsilon 2$) distance to form/share a cluster, minimum number of gaze points to form a cluster ($minPts$), and the weight assigned to the temporal distance (w_t) in the spatial-temporal distance metric used to relabel initial noise gaze points. Detailed parameter explanations can be found in [4].

We tuned this algorithm for this new triadic conversation version of our simulation. Following the protocol from [4], an experimenter instructed 10 participants to look at an object (e.g., forehead, eyes, mouth, plant, notebook, mug; Fig. 1a); subjects promptly shifted their gaze and fixated on it. Each participant completed 5 sessions (average 185.6s) which adds up to 2.6 hours of annotated gaze data for tuning. Subject-wise leave-one-out cross validation (CV) was used, so 45 recordings from 9 individuals determined the optimal hyperparameter set which was tested on the remaining 5 recordings. The optimal hyperparameter set in each fold was the one that minimized the sum of forehead, eye, and mouth region gaze percentage errors [4] averaged over all 45 recordings.

For $\epsilon 1$, we evaluated {0.015m, 0.02m, 0.025m, 0.03m, 0.035m, 0.04m}. The candidate set for $\epsilon 2$ was {0.5s, 0.75s, 1s, 1.25s, 1.5s, 1.75s, 2s}. For $minPts$, we considered the integers from 1 to 40. Finally, for w_t, we tested the numbers from 0.1 to 0.9 in increments of 0.05. The unanimous optimal hyperparameter set in each fold was {0.03m, 1s, 30, 0.2}. The average total percentage-wise error in unprocessed gaze targeting the interviewers' faces was 6.83% (12.68s). The duration-wise errors for forehead, eye, and mouth regions were 3.66s (1.97%), 6.01s (3.24%), and 3.01s (1.62%), respectively. Using the tuned algorithm, the overall percentage-wise error dropped to 1.4% (2.6s) with 0.47% (0.87s), 0.58% (1.08s), and 0.35% (0.65s) as the individual face region errors.

B. Head Motion Processing and Machine Learning-Based Yaw Adjustment

Shown on the lower branch of Fig. 2, we track head movements by tracking the headset object's position and orientation. Our goal is to explore how users engage with the virtual interviewers whose locations are known with respect to the center C of the virtual chair (Fig. 3). Irrespective of headset location, a yaw angle of 0° corresponds to facing forward. However, a target placed 30° left of C will not be 30° left of the headset if the headset moves. Hence, an algorithm is needed to connect headset readings and known target locations.

The distance from C to the interviewers is 1.5m. The one on the right of Fig. 3 is centered at 22.5° (spanning 22.5°±9°), and the other spans −22.5°±10° with respect to the z (forward) axis of the virtual chair. Since the interviewers appear as targets in the horizontal direction for a user in the virtual chair, we are interested in yaw measurements. In [70], we developed a geometric approach to compute the yaw β around C that corresponds to facing the same virtual target the user is facing from a different position (H_x, H_y, H_z) and rotation angle α (Fig. 4a).

To validate, we instructed users to face spheres placed 1.5m from C at angles between −90° and 90°, in increments of 15° (Fig. 4b). An experimenter told 12 subjects to orient towards specific spheres in a random order. This protocol was repeated for 9 different positions: left, center, right, back-left, back, back-right, front-left, front, and front-right. The center position corresponds to sitting at C. The other positions correspond to being seated half a meter away from C in the given direction. All of the spheres were targeted from all positions.

Assuming minimal head position shifts, the center position does not require yaw adjustment. We examine the mean absolute difference (MAD) between the measured yaw angle (α)

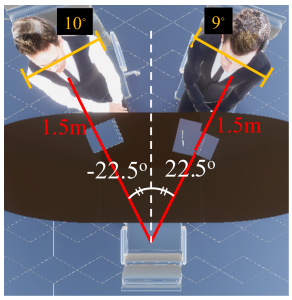

Fig. 3. Position and orientation of the interviewers with respect to the center of the virtual chair.

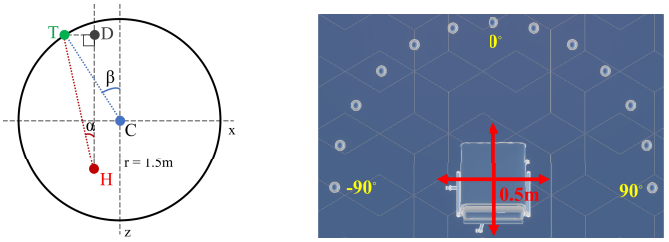

Fig. 4. (a) *C*, *H*, and *T* represent the center of the virtual chair, headset, and target location that the user is facing. *D* is the projection of *T* on the forward *z* axis of the headset. *x*-axis is the horizontal motion axis in Unity. α is the yaw angle measured by the headset, with respect to the forward *z* axis of the headset object. β is the output yaw angle of the algorithm, (b) Head rotation data collection setup. Nine chair positions were used during data collection, over the area spanned by the red arrows.

and the horizontal angle with respect to C at which the designated sphere was located (β^{GT}) as a baseline. For the center position, MAD $= (\sum_{i=1}^{S} |\beta_i^{GT} - \alpha_i|)/S$, where S is the number of spheres that were faced from a position. $S > 13$ because all 13 spheres were oriented towards and because of repeated instructions. The center MAD averaged over all subjects was 3°. This small error is due to the fact that subjects cannot orient exactly towards the spheres. For positions other than center, MAD $= (\sum_{i=1}^{S} |\beta_i^{GT} - \beta_i|)/S$. The average MAD for unprocessed yaw measurements over all positions was 15.1°, whereas the average MAD achieved by our previously developed geometric approach was 3°, equal to the baseline.

Ray casting is commonly used for accurate 3D target acquisition, especially when there is a clear path between the user and the target. Precise object selection in virtual environments is possible through head tracking-based ray casting [72], [73]. Evaluating the head tracking-based ray casting using the spheres yielded an average MAD of 3°, equal to that achieved by the lower complexity geometric approach.

Although the simple three degrees-of-freedom geometric approach works well, it cannot account for some behaviors. For some positions, the participants tended to tilt their heads or rotate their shoulders, not just turn their heads, to accurately orient towards the target sphere. The average absolute pitch angle was 4°, 4°, and 4.8° for the positions in the back, middle, and front rows, and the maximum pitch was 8° (for the sphere at 45° from the front-right position). This means participants tilted their head upwards when facing nearby spheres. The average roll angles were 2°, 2°, and 2.2°

TABLE II
MAD VALUES AFTER PROCESSING THE YAW VALUES WITH GEOMETRIC OR MLP-BASED APPROACH

Case:	Left	Right	Back-Left	Back	Back-Right	Front-Left	Front	Front-Right
Geometric (°):	2.6	2.5	3.3	3.1	3.7	3.1	3.2	2.8
MLP-based (°):	2.4	2.3	3.4	2.6	2.8	2.8	2.4	2.1

for the positions in the back, middle, and front row. The largest roll value was recorded when facing the sphere at −90° from the front position (5.3°). As an ML model can learn that although head positions might change due to head tilts, they still correspond to facing the same target, we developed an MLP regressor-based yaw adjustment module. An MLP is a fully-connected feed-forward artificial neural network with at least three layers (at least one hidden layer) with a nonlinear activation function.

The subjects were instructed to face a sphere for 4.4s on average. For training and validation of the MLP model, N random data points were sampled from each instruction. Our MLP model is defined by 6 potential inputs from the VR headset: yaw, pitch, roll, H_x, H_y, H_z. We used Random Forest Recursive Feature Elimination (RF-RFE) [74] in a subject-wise leave-one-out CV manner, using the data from 11 subjects to train the random forest (including different configurations) and testing it on the outstanding data. On average, the recorded yaw value (α) was the most important feature with 92%, followed by H_x (4%), and H_z (3%). As pitch, roll, and H_y had average importance values less than 1%, we elected to use (α, H_x, H_z), the same set used in the geometric approach.

Next, using the data from 12 subjects, we ran subject-wise leave-one-out CV to train and validate the MLP, while also optimizing N. We used adaptive learning rate. The optimal model had 3 hidden layers with 32 nodes in each layer. Best results were achieved for $\hat{N}$=45 which corresponds to randomly sampling 1s of data from each turn towards a sphere. For this model, the average MAD over all positions was 2.6°, compared to 3° for the geometric approach (see Table II). By exploiting the strong correlations between the pitch/roll rotations and the recorded headset positions (Pearson correlation $|r|$=0.52, p=0.045, on average), the machine learning method is able to more accurately reflect whether the subject is orienting towards one or another interviewer.

VI. RESULTS

In this section, we compare participants' gaze and head orientation tendencies in our triadic VR job interview simulation as a function of conversational role and neurodivergence status, which has not been previously attempted in the context of VR. We make use of two non-parametric significance tests. The Mann-Whitney U test assumes that two groups are sampled from the same distribution (e.g., normal, right-skewed) and is valid for both normally and non-normally distributed data [75]. The two-sample Kolmogorov-Smirnov (K-S) test evaluates the cumulative distributions of two data sets with no assumptions on the distributions; statistic D represents the maximum

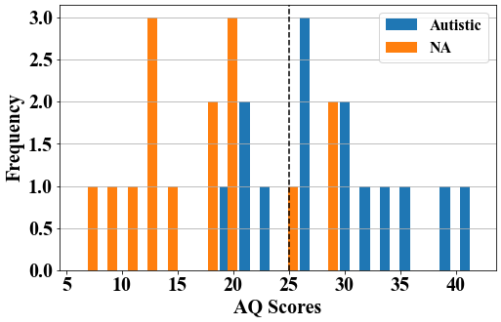

Fig. 5. Autism-Spectrum Quotient (AQ) scores reported by participants. The broken black line marks the threshold introduced in [76]; scores higher than this point to an increased chance of autism based on this self-report questionnaire.

distance between the cumulative distributions of two groups. To decide which test to use, we employ the Shapiro-Wilk test of normality, accompanied by skewness tests.

Fifteen autistic (9 male, 4 female, 2 non-binary) and 15 NA individuals (11 male, 4 female) took part in the simulations. Participant ages ranged from 18 to 28, except for one, who was 43. Individuals were assigned to the former group if they had received a community autism diagnosis in the past. The NA participants were drawn from the university graduate student population with the criteria that they did not identify as autistic and had not participated previously in this VR mock job interview simulation. The autistic participants, also mostly university students, were recruited through a neurodiversity-focused technical summer internship program. We contacted all interns from two summers for participation, and accepted all those who chose to participate; the number of NA participants was selected to match the number of autistic participants. Before each session, the eye-tracker was calibrated following the headset's default calibration procedure. The average session length was 22.3 minutes (s.d.=8.3 minutes).

After a session, participants were asked to complete the Autism-Spectrum Quotient (AQ) survey, a quantitative self-evaluation of autistic traits in an individual [76]. Fig. 5 shows the reported AQ scores (one autistic participant did not disclose their score). The average AQ score for the autistic participants was 29.2 (s.d.=6.6), whereas it was 16.9 (s.d.=6.7) for the NA group. The NA participants with AQ scores higher than 25, which is the lower bound indicating autism according to [76], were kept in the NA group as they had not received a community autism diagnosis, nor self-identified as autistic. Likewise, autistic individuals with scores lower than 25 remained in the autism group. Potential reasons for such scores include familiarity with the test and its normative answers, or a decrease in autistic traits over time after an initial autism diagnosis [77].

A. Gaze and Head Orientation Analysis

Fig. 6 displays the effect of conversational role on gaze behavior. On average, both neurotypes gazed at the interviewers' faces more when they were listening to a question, compared to when they were speaking. Higher overall gaze percentages were recorded for the NA participants. The NA participants looked more at interviewers' eyes than did autistic participants (63.9% (s.d.=18.7%) versus 40.8% (s.d.=31.3%) while listening, and 54.2% (s.d.=26.1%) versus 23.4% (s.d.=19.9%) while speaking). For the listener role, the eye contact percentage distributions for both groups were normal. A Mann-Whitney U test deemed the difference between these distributions significant, U=65, p=0.05. For the speaker role, the eye contact percentage distributions for the autistic and NA participants were left-skewed and normal. A K-S test revealed the difference between these percentages to also be significant, D=0.53, p=0.01. These results comport with the findings of [4].

We also investigated the social modulation of head orientation behavior (Fig. 7). We defined 3 yaw angle regions: *Interviewers:* 22.5°±9° and -22.5°±10°, *Interior:* [−12.5°, 13.5°], which represents the region in between the two interviewers, and *Exterior:* [−180°, −32.5°] and [31.5°, 180°]. Both neurotypes mostly faced the interviewers regardless of conversational role. However, as speakers, participants turned more to the region between the virtual interviewers. The NA participants oriented towards the interviewers more than the autistic group did (86.5% versus 70.6% for listening, which was marginally insignificant, and 76.2% (s.d.=19.8%, right-skewed) versus 58.8% (s.d.=28.6%, left-skewed) for speaking, which was significant, D=0.53, p=0.03), whereas the autistic participants faced the $Interior$ region more (29.1% versus 15.1% for listening, which was also marginally insignificant, and 41% (s.d.=28.7%, right-skewed) versus 23.7% (s.d.=19.7%, left-skewed) for speaking, which turned out to be significant with D=0.53, p=0.03). The participants did not turn to the $Exterior$ region, so we exclude it hereafter.

Our design allows for a more granular analysis based on the identity of $IntQ$, the interviewer who asked the most recent question. We split the face regions in Fig. 6 into smaller ones, producing regions FQ, FO, EQ, EO, MQ, and MO, where F, E, M, Q, O denote forehead, eyes, mouth, $IntQ$, and $IntO$. Fig. 8 shows that the NA participants mainly gazed at $IntQ$'s eyes regardless of their conversational role. They made eye contact with $IntO$ more in the speaker role compared to the listener role (12.4% (s.d.=10.3%, normal) versus 3.7% (s.d.=1.5%, normal) which was significant, U=56, p=0.02). The autistic participants looked at $IntQ$'s mouth the most as listeners. Both parties gazed at the interviewers' faces less when speaking (85.3% (s.d.=8.4%, normal) versus 45.2% (s.d.=25.6%, normal) for the autistic participants which was found to be significant, U=202, p<0.001; 93.2% (s.d.=6.3%, right-skewed) versus 66.2% (s.d.=20.5%, normal) for the NA participants which was significant, D=0.8, p<0.001). Overall, the autistic participants tended to avert their eyes more from the interviewers.

Similarly, the spatial regions in Fig. 7 can be divided to show whether they relate to $IntQ$ or $IntO$. The $Interviewers$ region is divided into $IntQ$ and $IntO$, while the $Interior$ region is divided into $Interior$-Q and $Interior$-O representing the yaw angles from 0° to either −12.5° or 13.5° based on which interviewer asked the most recent question. Fig. 9 shows that when listening to a question,

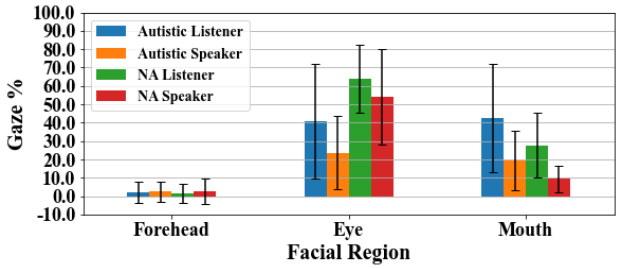

Fig. 6. Social modulation of gaze behavior.

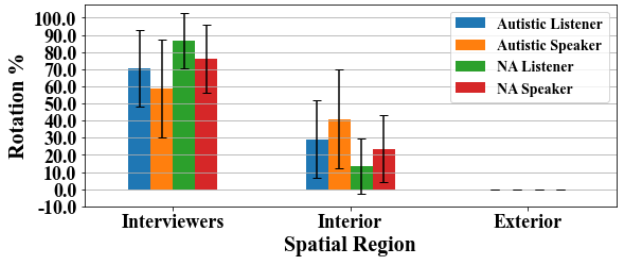

Fig. 7. Social modulation of head orientation behavior.

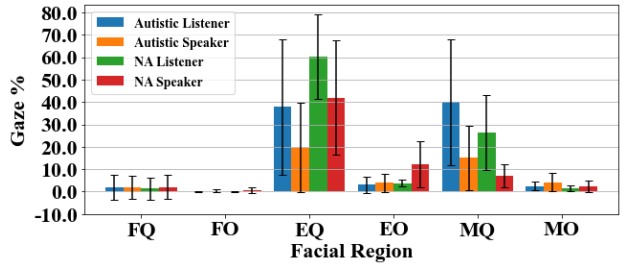

Fig. 8. Social modulation of gaze based on interviewer roles.

the participants mostly dwelt around $IntQ$. As listeners, both groups mainly faced $IntQ$, and they faced the region between him and the 0° line more than $IntO$. As speakers, both groups oriented towards $IntO$ more than they did when they were listening; however, this increase was statistically insignificant for the autism group (5.2% (s.d.=5.6%) versus 3.4% (s.d.=3.3%)) compared to the NA group's significant attention shift (12.2% (s.d.=11.3%, left-skewed) versus 3.7% (s.d.=1.8%, normal), D=0.6, p=0.009).

We also measured individuals' levels of social exclusion while responding to a question. Since it is natural to look at a person ($IntQ$) when they ask one a question, here we examine exclusion based on whether $IntO$ is also attended to during the response. We looked at the longest duration without interacting (gaze or head turn) with $IntO$ while answering a question, averaged over all questions and over all participants. For the autistic participants, this number was 17.4s (s.d.=7.5s), and for the NA participants, it was 13.2s (s.d.=5.2s), and the distributions are in Fig. 10. There is substantial overlap between the two groups. It is apparent that a few participants, both autistic and NA, tend to exclude the interviewer who did not ask the question for 25-30 seconds at a time, while primarily attending to the interviewer that asked the question. This suggests that this VR tool might be useful for solo situational practice by both autistic and NA

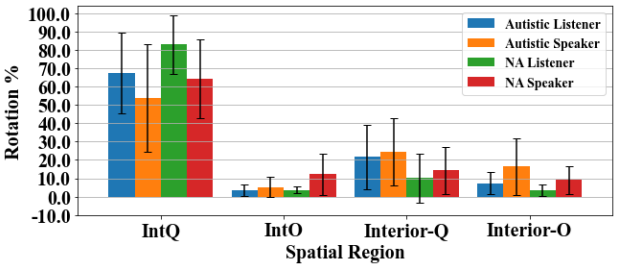

Fig. 9. Social modulation of head orientation based on interviewer roles.

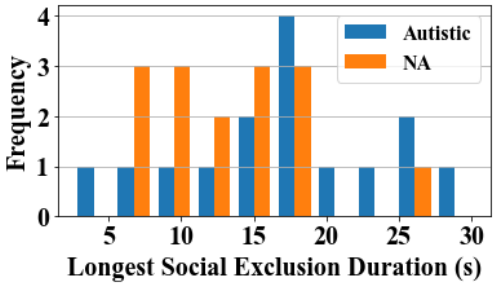

Fig. 10. Distribution of average longest social exclusion durations during a response.

individuals, who could practice answering questions while also bestowing attention on both interviewers.

B. Effects of Interviewer Backchannels and Head Turns

As introduced in Section IV, this VR simulation setup can also be used to explore the influence that backchannel cues and head turns performed by the virtual interviewers have on participant behavior. We investigated 4 cases: $QbOl$, $QlOb$, $QbOb$, and $QtOb$, where Q and O denote $IntQ$ and $IntO$, and b, l, t indicate whether an interviewer can give backchannels, listens only and gives no backchannels, or turns to the other interviewer when the other gives a backchannel. We measured the maximum unsigned yaw angle changes that emerged as a reaction to interviewer backchannels or head turns, averaged over the total number of backchannels or head turns for each case. In total, for $QbOl$, $QlOb$, and $QbOb$, 305, 313, and 184 backchannels were given, and for $QtOb$, 242 head turns were performed. These backchannels and head turns (hereafter, *interviewer cue*) that the interviewers perform are animations with known durations. Including an extra 1 second to give users time to react, the average lengths of the animations for VBs, PBs, and head turns during the mock interviews were 3s, 6s, and 5s, respectively.

We are interested in the largest head turn performed by the participant due to an interviewer cue. This metric can be described as the maximum unsigned difference between the yaw angle (β) recorded at the start of an interviewer cue and during the cue, including the extra reaction time. For the 5 questions where the interviewers did not give any backchannels, the average maximum absolute yaw change was 9.8° (s.d.=8.9°) for the autistic participants, and 10.4° (s.d.=7.4°) for the NA participants. For the other 38 questions, for the autistic/NA participants, the largest head turns on

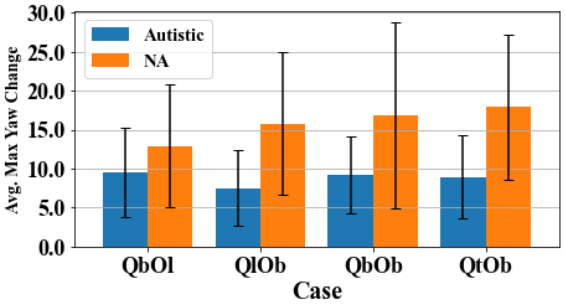

Fig. 11. Maximum unsigned yaw shifts for each interviewer cue type, averaged across all participants.

average for the 4 cases were 9.5°/12.9° (s.d.=5.7°/7.9°), 7.5°/15.8° (s.d.=4.9°/9.1°), 9.2°/16.9° (s.d.=5°/11.9°), and 8.9°/17.9° (s.d.=5.3°/9.3°) (Fig. 11).

The NA participants reacted to $QtOb$ the most; when $IntQ$ turns towards $IntO$ following a backchannel by $IntO$, the NA participants tend to mirror $IntQ$'s behavior. This joint attention pattern was less prominent for the autistic participants which is consistent with a previous study involving in-person conversations [28]. For the NA participants, the maximum unsigned yaw shift values were larger in the presence of interviewer backchannels compared to the short answer cases which had no backchannels, which could be due to the interviewer backchannels, whereas autistic individuals turned their heads slightly less on average for the 4 backchannel cases, compared to the short answer cases without backchannels. Across the 4 cases, autistic individuals behaved similarly regardless of interviewer cues. Mann-Whitney U tests found that the average maximum yaw shifts were significantly different for the two participant groups for $QlOb$ and $QtOb$, with U=55, p=0.01, and U=53, p=0.008, respectively. Therefore, the two neurotypes tended to differ the most when $IntQ$ did not give backchannels but rather stayed idle or acknowledged $IntO$'s backchannels.

C. User Experience

Upon completing the VR simulation, participants were asked about the repeatability, complexity, user-friendliness, and realism of the application. Table III presents the results. They were asked to respond with their level of agreement on a 5-point Likert scale from 1 = "Not realistic at all" to 5 = "Acceptably realistic" for the last 3 items in Table III, and from 1 = "Strongly Disagree" to 5 = "Strongly Agree" for all the other items. High repeatability, realism, and user-friendliness, and low complexity scores verify the suitability of our triadic self-deliverable VR mock job interview simulation which can let users practice for job interviews. Open-ended questions in the survey asked about what the participants liked the most, and what design components could be improved. The participants enjoyed being able to practice for a job interview in a low stakes environment and with human-like interviewers. They found some interviewer head turns a bit fast and abrupt, and suggested having more background noise. Some wished the headset were more comfortable, and suggested the app should facilitate taking a break. Overall, 25 participants out of 30 reported that they would use the application for solo job interview practice.

TABLE III
SUMMARY OF REPEATABILITY, COMPLEXITY, USER-FRIENDLINESS, AND REALISM SCORES

	Avg.	S.D.
Repeatability		
If the system had multiple practice interviews with different questions each time, I think that I would like to use this system multiple time	3.7	1.4
Complexity		
I found the system unnecessarily complex	1.5	0.6
I think that I would need the support of a coach to be able to use this system again	1.9	1.1
I needed a lot of explaining before I could get going with this system	1.6	0.8
User-friendliness		
I thought the system was easy to use	4.4	0.7
The system worked without major glitches and distractions.	3.8	1.1
I would imagine that most people would learn to use this system very quickly	4.4	0.9
I felt very confident using the system	4.4	0.7
Realism		
I think the interview questions were an appropriate mix of easy and difficult questions	4.3	0.8
Considering this is a VR simulation intended for job interview practice, what do you think of the interviewers' appearances?	4.1	1
Considering this is a VR simulation intended for job interview practice, what do you think of the interviewers' behaviors?	3.8	1.2
Considering this is a VR simulation intended for job interview practice, what do you think of the office space?	4.6	0.6

VII. CONCLUSION

In professional settings such as job interviews, applicants who use gaze and head orientation effectively can improve their likelihood of employment [14]. In this study, we presented a triadic VR mock job interview that lets users familiarize themselves with popular interview questions while getting informed about their attention distribution and social exclusion tendencies. Our participants favored the repeatability, user-friendliness, and realism of our design.

Using a machine learning architecture to accurately detect head turns towards the virtual interviewers, and a signal processing-based algorithm that can mitigate the common problems with eye tracking in VR [4], we showed that regardless of their conversational role or their neurotype, our participants primarily interacted with the interviewer who posed the most recent question. As listeners, the autistic participants gazed at the interviewers' mouths more than at their eyes, and more than the NA individuals gazed at the mouths. This was also reported in [4] on dyadic VR conversations, which shows that the autistic participants followed similar gaze trends in triadic cases.

We have findings that are novel in the realm of immersive VR; NA participants engaged with the interviewer who did not pose the most recent question significantly more when speaking compared to listening, whereas the autistic participants did not. We also discovered differences in joint attention tendencies; NA participants tend to mirror an interviewer's behavior in turning to the other interviewer, whereas autistic participants do this significantly less. These behaviors have not been previously addressed using VR. Also, a few participants of both neurotypes tended to exclude the interviewer who did not ask the most recent question, suggesting that this system might be useful in the general population to practice interview and engagement skills.

Although our system can provide a useful solo practice opportunity for job interviews, it has some limitations, including the relatively small number of participants, and that the AQ scores of the two participant groups show significant overlap, likely because the groups were based only on community diagnosis and self identification as autistic or not. We also have some limitations in the design of the VR application, in particular that some avatar head turns were too fast, and users were not able to go back to a previous question (useful if a user accidentally skips a question).

In future work, we intend to expand our participant set. We will design multiple question sets so that the users can encounter a more diverse set of questions, and can use the application more than once. We plan to address the feedback we received. We will add captions to make our design more inclusive for people with hearing disabilities, and will redesign some head turn animations to make them more natural. We also aim to create an interview practice tool that offers automated feedback, such as alerting users in case of social exclusion.

Acknowledgment

The authors would like to thank Zachary Burns and Trent Simmons for helping them model the interviewers, Aaron Li for contributing to the Unity application, and their participants for their involvement.

References

[1] S. Ho, T. Foulsham, and A. Kingstone, "Speaking and listening with the eyes: Gaze signaling during dyadic interactions," *PLoS ONE*, vol. 10, no. 8, Aug. 2015, Art. no. e0136905.

[2] A. Kendon, "Some functions of gaze-direction in social interaction," *Acta Psychologica*, vol. 26, pp. 22–63, Jan. 1967.

[3] 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.

[4] S. Artiran, R. Ravisankar, S. Luo, L. Chukoskie, and P. Cosman, "Measuring social modulation of gaze in autism spectrum condition with virtual reality interviews," *IEEE Trans. Neural Syst. Rehabil. Eng.*, vol. 30, pp. 2373–2384, 2022.

[5] R. Stiefelhagen and J. Zhu, "Head orientation and gaze direction in meetings," in *Proc. CHI Extended Abstr. Hum. Factors Comput. Syst.*, Apr. 2002, pp. 858–859.

[6] S. O. Ba and J.-M. Odobez, "A study on visual focus of attention recognition from head pose in a meeting room," in *Machine Learning for Multimodal Interaction*. Cham, Switzerland: Springer, 2006, pp. 75–87.

[7] P. Venuprasad et al., "Characterizing joint attention behavior during real world interactions using automated object and gaze detection," in *Proc. 11th ACM Symp. Eye Tracking Res. Appl.*, Jun. 2019, pp. 1–8.

[8] D. I. Perrett, J. K. Hietanen, M. W. Oram, and P. J. Benson, "Organization and functions of cells responsive to faces in the temporal cortex," *Philos. Trans. Roy. Soc. London, B, Biol. Sci.*, vol. 335, no. 1273, pp. 23–30, 1273.

[9] W. R. Steele, *Presentation Skills 201: How to Take It to the Next Level as a Confident, Engaging Presenter*. Outskirts Press, 2009. Accessed: Oct. 3, 2023. [Online]. Available: https://books.google.com/books?id=PYJmPgAACAAJ

[10] J. G. Amalfitano and N. C. Kalt, "Effects of eye contact on the evaluation of job applicants," *J. Employment Counseling*, vol. 14, no. 1, pp. 46–48, Mar. 1977.

[11] J. K. Burgoon, V. Manusov, P. Mineo, and J. L. Hale, "Effects of gaze on hiring, credibility, attraction and relational message interpretation," *J. Nonverbal Behav.*, vol. 9, no. 3, pp. 133–146, 1985.

[12] R. J. Forbes and P. R. Jackson, "Non-verbal behaviour and the outcome of selection interviews," *J. Occupational Psychol.*, vol. 53, no. 1, pp. 65–72, Mar. 1980.

[13] C. K. Parsons and R. C. Liden, "Interviewer perceptions of applicant qualifications: A multivariate field study of demographic characteristics and nonverbal cues," *J. Appl. Psychol.*, vol. 69, no. 4, pp. 557–568, 1984.

[14] 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.

[15] *Diagnostic and Statistical Manual of Mental Disorders: DSM-5*, Amer. Psychiatric Assoc., Washington, DC, USA, 2013.

[16] 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, 2016.

[17] J. Rochkind, "Marking World Day, UN calls on businesses to commit to employing people with autism," UN News, Apr. 2, 2015. Accessed: Oct. 14, 2023. [Online]. Available: https://news.un.org/en/story/2015/04/495002-marking-world-day-un-calls-businesses-commit-employing-people-autism

[18] A. V. S. Buescher, Z. Cidav, M. Knapp, and D. S. Mandell, "Costs of autism spectrum disorders in the United Kingdom and the United States," *JAMA Pediatrics*, vol. 168, no. 8, p. 721, Aug. 2014.

[19] H. Gima et al., "Early motor signs of autism spectrum disorder in spontaneous position and movement of the head," *Exp. Brain Res.*, vol. 236, no. 4, pp. 1139–1148, Apr. 2018.

[20] Z. Zhao et al., "Identifying autism with head movement features by implementing machine learning algorithms," *J. Autism Develop. Disorders*, vol. 52, no. 7, pp. 3038–3049, Jul. 2022.

[21] C. Trepagnier, M. M. Sebrechts, and R. Peterson, "Atypical face gaze in autism," *CyberPsychology Behav.*, vol. 5, no. 3, pp. 213–217, Jun. 2002.

[22] M. Sigman, A. Dijamco, M. Gratier, and A. Rozga, "Early detection of core deficits in autism," *Mental Retardation Develop. Disabilities Res. Rev.*, vol. 10, no. 4, pp. 221–233, Nov. 2004.

[23] 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.

[24] 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.

[25] 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.

[26] 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.

[27] A. McParland, S. Gallagher, and M. Keenan, "Investigating gaze behaviour of children diagnosed with autism spectrum disorders in a classroom setting," *J. Autism Develop. Disorders*, vol. 51, no. 12, pp. 4663–4678, Dec. 2021.

[28] C. Kasari, M. Sigman, P. Mundy, and N. Yirmiya, "Affective sharing in the context of joint attention interactions of normal, autistic, and mentally retarded children," *J. Autism Develop. Disorders*, vol. 20, no. 1, pp. 87–100, Mar. 1990.

[29] 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.

[30] 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.

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

[32] 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.

[33] P. M. Corcoran, F. Nanu, S. Petrescu, and P. Bigioi, "Real-time eye gaze tracking for gaming design and consumer electronics systems," *IEEE Trans. Consum. Electron.*, vol. 58, no. 2, pp. 347–355, May 2012.

[34] A. Skulmowski, A. Bunge, K. Kaspar, and G. Pipa, "Forced-choice decision-making in modified trolley dilemma situations: A virtual reality and eye tracking study," *Frontiers Behav. Neurosci.*, vol. 8, p. 426, Dec. 2014.

[35] F. Fathy, Y. Mansour, H. Sabry, M. Refat, and A. Wagdy, "Virtual reality and machine learning for predicting visual attention in a daylit exhibition space: A proof of concept," *Ain Shams Eng. J.*, vol. 14, no. 6, Jun. 2023, Art. no. 102098.

[36] Z. Wang, R. Zheng, T. Kaizuka, and K. Nakano, "Relationship between gaze behavior and steering performance for driver–automation shared control: A driving simulator study," *IEEE Trans. Intell. Vehicles*, vol. 4, no. 1, pp. 154–166, Mar. 2019.

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

[38] 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, pp. 317–330, 2012.

[39] 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, Mar. 2013.

[40] 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.

[41] R. Mantiuk, B. Bazyluk, and R. K. Mantiuk, "Gaze-driven object tracking for real time rendering," *Comput. Graph. Forum*, vol. 32, no. 2, pp. 163–173, May 2013.

[42] M. Weier, T. Roth, A. Hinkenjann, and P. Slusallek, "Predicting the gaze depth in head-mounted displays using multiple feature regression," in *Proc. ACM Symp. Eye Tracking Res. Appl.*, Jun. 2018, pp. 1–9.

[43] J. Brookes, M. Warburton, M. Alghadier, M. Mon-Williams, and F. Mushtaq, "Studying human behavior with virtual reality: The unity experiment framework," *Behav. Res. Methods*, vol. 52, no. 2, pp. 455–463, Apr. 2020.

[44] J. R. J. Neo, A. S. Won, and M. M. Shepley, "Designing immersive virtual environments for human behavior research," *Frontiers Virtual Reality*, vol. 2, Mar. 2021, Art. no. 603750.

[45] B. Xiao, P. Georgiou, B. Baucom, and S. S. Narayanan, "Head motion modeling for human behavior analysis in dyadic interaction," *IEEE Trans. Multimedia*, vol. 17, no. 7, pp. 1107–1119, Jul. 2015.

[46] 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 Eur. Signal Process. Conf. (EUSIPCO)*, Aug. 2021, pp. 1301–1305.

[47] S. Kloiber et al., "Immersive analysis of user motion in VR applications," *Vis. Comput.*, vol. 36, nos. 10–12, pp. 1937–1949, Oct. 2020.

[48] A. S. Won, J. N. Bailenson, S. C. Stathatos, and W. Dai, "Automatically detected nonverbal behavior predicts creativity in collaborating dyads," *J. Nonverbal Behav.*, vol. 38, no. 3, pp. 389–408, Sep. 2014.

[49] 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.

[50] S. Sheikhi and J.-M. Odobez, "Combining dynamic head pose–gaze mapping with the robot conversational state for attention recognition in human–robot interactions," *Pattern Recognit. Lett.*, vol. 66, pp. 81–90, Nov. 2015.

[51] A. D. Arndt, L. Khoshghadam, and K. Evans, "Who do I look at? Mutual gaze in triadic sales encounters," *J. Bus. Res.*, vol. 111, pp. 91–101, Apr. 2020.

[52] E. Zima, C. Weiß, and G. Brône, "Gaze and overlap resolution in triadic interactions," *J. Pragmatics*, vol. 140, pp. 49–69, Jan. 2019.

[53] H. Lu, M. F. McKinney, T. Zhang, and A. J. Oxenham, "Investigating age, hearing loss, and background noise effects on speaker-targeted head and eye movements in three-way conversations," *J. Acoust. Soc. Amer.*, vol. 149, no. 3, pp. 1889–1900, Mar. 2021.

[54] L. Hladek and B. U. Seeber, "Behavior in triadic conversations in conditions with varying positions of noise distractors," in *Proc. Fortschritte Akustik*, 2023, pp. 916–918.

[55] M. R. Miller, N. Sonalkar, A. Mabogunje, L. Leifer, and J. Bailenson, "Synchrony within triads using virtual reality," *Proc. ACM Hum.-Comput. Interact.*, vol. 5, no. CSCW2, pp. 1–27, Oct. 2021.

[56] B. Tarr, M. Slater, and E. Cohen, "Synchrony and social connection in immersive virtual reality," *Sci. Rep.*, vol. 8, no. 1, p. 3693, Feb. 2018.

[57] T. L. Chartrand and J. A. Bargh, "The chameleon effect: The perception-behavior link and social interaction," *J. Personality Social Psychol.*, vol. 76, no. 6, pp. 893–910, 1999.

[58] E. Novotny, M. G. Frank, and M. Grizzard, "A laboratory study comparing the effectiveness of verbal and nonverbal rapport-building techniques in interviews," *Commun. Stud.*, vol. 72, no. 5, pp. 819–833, Sep. 2021.

[59] N. Aburumman, M. Gillies, J. A. Ward, and A. F. D. C. Hamilton, "Nonverbal communication in virtual reality: Nodding as a social signal in virtual interactions," *Int. J. Hum.-Comput. Stud.*, vol. 164, Aug. 2022, Art. no. 102819.

[60] V. Ravindran, M. Osgood, V. Sazawal, R. Solorzano, and S. Turnacioglu, "Virtual reality support for joint attention using the floreo joint attention module: Usability and feasibility pilot study," *JMIR Pediatrics Parenting*, vol. 2, no. 2, Sep. 2019, Art. no. e14429.

[61] C. Mei, B. T. Zahed, L. Mason, and J. Ouarles, "Towards joint attention training for children with ASD—A VR game approach and eye gaze exploration," in *Proc. IEEE Conf. Virtual Reality 3D User Interface (VR)*, Mar. 2018, pp. 289–296.

[62] 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.

[63] R. M. Aysina, Z. A. Maksimenko, and M. V. Nikiforov, "Feasibility and efficacy of job interview simulation training for long-term unemployed individuals," *Psychnol. J.*, vol. 14, no. 1, pp. 41–60, 2016.

[64] B. Chang, J.-T. Lee, Y.-Y. Chen, and F.-Y. Yu, "Applying role reversal strategy to conduct the virtual job interview: A practice in second life immersive environment," in *Proc. IEEE 4th Int. Conf. Digit. Game Intell. Toy Enhanced Learn.*, Mar. 2012, pp. 177–181.

[65] H. Grillon, F. Riquier, B. Herbelin, and D. Thalmann, "Virtual reality as a therapeutic tool in the confines of social anxiety disorder treatment," *Int. J. Disability Hum. Develop.*, vol. 5, no. 3, pp. 243–250, Jan. 2006.

[66] J. Bersin, "The corporate learning factbook 2014: Benchmarks, trends, and analysis of the U.S. training market," Deloitte, Oakland, CA, USA, 2014. Accessed: Oct. 12, 2023. [Online]. Available: http://www.cedma-europe.org/newsletter%20articles/Brandon%20Hall/The%20Corporate%20Learning%20Factbook%202014%20(Jan%2014).pdf

[67] M. Taylor. (2022). *U.K. Games Industry Census 2022*. [Online]. Available: https://ukie.org.uk/resources/uk-games-industry-census-2022

[68] J. A. Martin, "Research with adults with Asperger's syndrome—Participatory or emancipatory research?" *Qualitative Social Work*, vol. 14, no. 2, pp. 209–223, Mar. 2015.

[69] L. Crane, F. Adams, G. Harper, J. Welch, and E. Pellicano, "'Something needs to change': Mental health experiences of young autistic adults in England," *Autism*, vol. 23, no. 2, pp. 477–493, Feb. 2019.

[70] S. Artiran, P. S. Bedmutha, A. Li, and P. Cosman, "Gaze and head rotation analysis in a triadic VR job interview simulation," in *Proc. IEEE Int. Symp. Mixed Augmented Reality Adjunct (ISMAR-Adjunct)*, Oct. 2023, pp. 381–386.

[71] 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.

[72] Y. Y. Qian and R. J. Teather, "The eyes don't have it: An empirical comparison of head-based and eye-based selection in virtual reality," in *Proc. 5th Symp. Spatial User Interact.*, 2017, pp. 91–98.

[73] M. Kytö, B. Ens, T. Piumsomboon, G. A. Lee, and M. Billinghurst, "Pinpointing: Precise head- and eye-based target selection for augmented reality," in *Proc. CHI Conf. Hum. Factors Comput. Syst.*, Apr. 2018, pp. 1–14.

[74] P. M. Granitto, C. Furlanello, F. Biasioli, and F. Gasperi, "Recursive feature elimination with random forest for PTR-MS analysis of agroindustrial products," *Chemometric Intell. Lab. Syst.*, vol. 83, no. 2, pp. 83–90, Sep. 2006.

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

[76] S. Baron-Cohen, S. Wheelwright, R. Skinner, J. Martin, and E. Clubley, "The Autism-Spectrum Quotient (AQ): Evidence from Asperger syndrome/high-functioning autism, males and females, scientists and mathematicians," *J. Autism Develop. Disorders*, vol. 31, pp. 5–17, Feb. 2001.

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