Evaluating Gaze Interactions within AR for Nonspeaking Autistic Users

Ahmadreza Nazari
ahmadreza.nazari@ucalgary.ca
University of Calgary
Calgary, Alberta, Canada

Dr. Lorans Alabood
lorans.alabood@ucalgary.ca
University of Calgary
Calgary, Alberta, Canada

Molly K. Rathbun
zkr7pa@virginia.edu
University of Virginia
Charlottesville, Virginia, U.S.A.

Dr. Vikram K. Jaswal
jaswal@virginia.edu
University of Virginia
Charlottesville, Virginia, U.S.A.

Dr. Diwakar Krishnamurthy
dkrishna@ucalgary.ca
University of Calgary
Calgary, Alberta, Canada

ABSTRACT

Nonspeaking autistic individuals often face significant inclusion barriers in various aspects of life, mainly due to a lack of effective communication means. Specialized computer software, particularly delivered via Augmented Reality (AR), offers a promising and accessible way to improve their ability to engage with the world. While research has explored near-hand interactions within AR for this population, gaze-based interactions remain unexamined. Given the fine motor skill requirements and potential for fatigue associated with near-hand interactions, there is a pressing need to investigate the potential of gaze interactions as a more accessible option. This paper presents a study investigating the feasibility of eye gaze interactions within an AR environment for nonspeaking autistic individuals. We utilized the HoloLens 2 to create an eye gaze-based interactive system, enabling users to select targets either by fixating their gaze for a fixed period or by gazing at a target and triggering selection with a physical button (referred to as a 'clicker'). We developed a system called HoloGaze that allows a caregiver to join an AR session to train an autistic individual in gaze-based interactions as appropriate. Using HoloGaze, we conducted a study involving 14 nonspeaking autistic participants. The study had several phases, including tolerance testing, calibration, gaze training, and interacting with a complex interface: a virtual letterboard. All but one participant were able to wear the device and complete the system's default eye calibration; 10 participants completed all training phases that required them to select targets using gaze only or gaze-click. Interestingly, the 7 users who chose to continue to the testing phase with gaze-click were much more successful than those who chose to continue with gaze alone. We also report on challenges and improvements needed for future gaze-based interactive AR systems for this population. Our findings pave the way for new opportunities for specialized AR solutions tailored to the needs of this under-served and under-researched population.

VRST '24, October 9–11, 2024, Trier, Germany

ACM ISBN 979-8-4007-0535-9/24/10
https://doi.org/10.1145/3641825.3687743

CCS CONCEPTS

• **Human-centered computing → Mixed / augmented reality**; **Accessibility technologies**.

KEYWORDS

nonspeaking autistic people, augmented reality, eye tracking, assistive technology

ACM Reference Format:
Ahmadreza Nazari, Dr. Lorans Alabood, Molly K. Rathbun, Dr. Vikram K. Jaswal, and Dr. Diwakar Krishnamurthy. 2024. Evaluating Gaze Interactions within AR for Nonspeaking Autistic Users. In *30th ACM Symposium on Virtual Reality Software and Technology (VRST '24), October 9–11, 2024, Trier, Germany*. ACM, New York, NY, USA, 11 pages. https://doi.org/10.1145/3641825.3687743

1 INTRODUCTION

About 30% of autistic people cannot communicate effectively using speech [10] and most do not have access to an effective alternative [7], which significantly limits their ability to participate in educational, social, and employment opportunities. Some nonspeaking autistic people are provided access to picture- or icon-based alternative communication systems that can allow them to request things such as food and activities (e.g., [16]). While these systems can be helpful for basic communication (e.g., requests), they fall short due to obvious limitations of constructing full and meaningful sentences, the basic blocks for effective two-way communication.

Many nonspeaking autistic people have acquired foundational literacy skills [22], which suggests that writing could be a viable alternative to speech. But nonspeaking autistic people have significant attentional, sensory, and motor challenges [15, 21], making it difficult for them to write by hand or type in conventional ways. For example, many nonspeaking autistic people have difficulty with the fine motor control needed to use a pen or pencil [3]. They may also be in constant motion (which seems to serve a regulatory function; [21]), making training to use a keyboard while remaining seated difficult. Some nonspeaking autistic people have learned to communicate by typing, which has allowed them to graduate from college and to write award-winning poetry [8, 38]. But the process by which they learned to type was lengthy and expensive, and often requires the ongoing support of another person [35].

One promising technology that may provide a solution to some of the challenges faced by nonspeaking autistic people in learning to communicate by typing is wearable Augmented Reality (AR) [24]. The flexibility to size and place objects with which the user interacts can to some extent account for limited fine motor control. Furthermore, tracking user movement can accommodate constant motion by ensuring virtual content remains in the field of view. Finally, AR can automate some aspects of support aiming to reduce reliance on another person (e.g., [12, 31]).

Research in wearable AR has investigated the feasibility of near-hand interactions for nonspeaking autistic people [2, 32, 39]. These studies showed that most participants could tolerate wearing the device and learned to use their hands to activate virtual buttons. Taking it one step further, a recently published study investigated the feasibility of virtual typing after just a few minutes of practice and reported positive results [1].

This research, while promising, has an important limitation. Specifically, users interacted with the virtual objects and the virtual letterboard[1] using their hands–activating a button or typing a letter with a single finger. Other research with non-autistic users has shown that typing in virtual environments can be a frustrating experience [13]. Specifically due to the fatigue associated with performing mid-air gestures (including tapping) for a prolonged period, often referred to as 'gorilla arm' [19]. In fact, modern headsets such as the Meta Quest and the Apple Vision Pro encourage other modes of interactions (e.g., gaze and pinch, voice input, or the use of controllers) to address this. [2] Of course, voice input would not be applicable to our target population of nonspeaking autistic people. Using gaze, however, may be a more accessible alternative as it does not require fine motor control and does not require the user to move their hands.

There are two primary interaction modes using eyes: dwell-based option (where a user looks at an object for a certain period of time to activate it) and dwell-free option (where a user looks at an object and triggers a selection via another input such as pinch). We aim to investigate both these interaction types. The device we used is the Microsoft HoloLens 2. As this device's cameras are at the top of the visor facing forward, it has a limited view and can only register gestures such as pinch within a limited range. Therefore, instead of gaze+pinch, we use on a more straightforward solution which is gaze+click (pressing a button on a clicker instead of pinching fingers). To our knowledge, there has not yet been any research investigating nonspeaking autistic people's ability to interact in AR using gaze or gaze+click methods.

There are a number of communication systems that make use of eye tracking and which have been studied with people with other disabilities. For example, some people with ALS or cerebral palsy rely on eye gaze-based communication systems, which interpret eye movements to facilitate interaction with digital interfaces [40]. In autism, however, eye-tracking has typically been limited to basic research in visual attention rather than practical applications in communication (e.g., [30]). Thus, although nonspeaking autistic people could potentially benefit from some of the same kinds of gaze-based interactions that individuals with cerebral palsy or ALS use, their ability to interact with these systems has not yet been studied.

[1]A virtual board with letters of alphabet on it in a grid layout.

[2]e.g., https://developer.apple.com/design/human-interface-guidelines/gestures.

Our study investigates three primary questions. First, we explore whether nonspeaking autistic people could tolerate the HoloLens 2 and complete its built-in eye calibration routine. Second, we developed a series of modules to investigate whether they could learn to select virtual objects using the gaze and gaze+click options. Third, we examine the feasibility of typing on a virtual letterboard (which requires isolating targets on a complex interface in addition to being able to interact). Throughout, we gather insights to inform the development of more effective and user-friendly AR eye gaze-based systems.

With these aims, we developed a system called HoloGaze. As support is crucial for our target population, we made HoloGaze a multiplayer system that can be used by a caregiver to train autistic students. Designing multiplayer AR applications has challenges; therefore, we created an easy to use framework for multiplayer AR apps that uses OpenXR, and which we make publicly available.

HoloGaze starts by introducing gaze and gaze+click modes of interaction. Participants who can tolerate the device first undergo the HoloLens 2 built-in eye calibration routine. Upon successful completion of the calibration, they engage with practice phases designed to both train participants to use gaze and gaze+click and to investigate the feasibility of these interactions. Finally, participants choose their preferred mode of interaction and have the opportunity to interact with a partial letterboard (one that only shows some letters) and which gradually turns into a full letterboard. This testing phase aims to assess the feasibility of virtual typing using HoloGaze.

We tested HoloGaze with 14 nonspeaking autistic participants. Thirteen tolerated the device and completed calibration. We did not expect this degree of success given that the calibration process requires users to keep their head still, which can be quite challenging for this population. Ten completed all practice phases; the remaining 3 completed half of the practice phases. Seven participants selected the gaze+click interface as their preferred mode of operation, and 6 selected gaze alone. Participants who selected gaze+click were overall more successful in the testing phases than those who chose gaze. Of the 7 participants who chose gaze+click, 6 completed all testing phases while 1 completed half of the testing phases.

Our finding show that most nonspeaking autistic participants could tolerate the AR device, confirming what was previously reported [1, 2, 5, 32, 39], successfully calibrate it to their eyes, and engage in gaze-based interactions supported by these devices. Specifically, our study suggests that gaze interactions can be a viable alternative to near-hand interactions; opening up new possibilities with more advanced devices such as the Apple Vision Pro. This presents a gateway to accessing information, games, and communication using wearable AR, thereby enhancing the quality of life of this under-served and under-researched population.

2 BACKGROUND AND RELATED WORK

For people who cannot communicate using speech and who have fine motor challenges (e.g., some individuals with cerebral palsy or ALS), written communication systems that make use of eye gaze have long been considered a possible alternative. Low-tech solutions can be slow and labor-intensive: They rely on human assistants to judge what letter of the alphabet a user is looking at on a physical letterboard, verify that letter with the user, transcribe manually each letter, and then read aloud the user's comment (e.g., Vocal Eyes [6], E-Tran Frames [29]).

Recent efforts on assistive AR for nonspeaking autistic individuals have aimed at examining the utilization of this technology for digital spelling. One approach involved developing a system called HoloBoard that replicates the familiar environment of physical letterboard training. Researchers reported a study with 23 participants. Sixteen completed a brief training module on the virtual letterboard, with many achieving unexpected levels of independence in spelling and related tasks. Notably, 14 participants engaged in complex tasks, such as spelling full sentences, and 5 used the system solo without a practitioner's support [1, 2]. These findings indicate that AR-based training can significantly enhance communication outcomes for nonspeaking autistic individuals. As noted previously, however, near-hand interaction in AR are not without setbacks and might not be best suited for all users. This motivates the exploration of alternative interactions based on eye gaze.

With advances in eye tracking, high-tech solutions such as the Tobii Dynavox systems (e.g., PCEye [41]) have automated this process. As noted earlier, these devices primarily use two modes of interaction: a) dwell-based interaction, where a user selects a target by gazing at it for a predefined, fixed period of time (e.g., 1 second), and b) dwell-free interactions, where an external input is used to trigger selection while gazing at a target, such as a gesture (e.g., pinch or blink) or a physical input (e.g., a clicker).

Research on eye gaze-based interactive systems has largely focused on improving throughput. Within the context of communication, adults without disabilities can communicate at 190 words per minute (wpm) using speech [9]. Non-disabled adults can reach 20 wpm using an eye typing system with dwell-based selection [27]. Given this large difference in throughput, many studies have focused on creating novel systems to achieve higher throughput. For example, Microsoft researchers created a system that allows communication partners to engage with the user via a mobile app, suggesting words while the user is typing [14]. Other research has focused on dwell-free eye typing. For example, Kristenssson et al. reported that under perfect conditions, a swipe-style keyboard (swiping with eyes rather than fingers) could allow non-disabled users to reach 46 wpm [25]. However, other studies indicate that achieving this rate is highly unlikely even for individuals without disabilities [26, 34].

Eye-tracking research within the context of autism has primarily addressed basic questions in visual attention (e.g., whether autistic people spend more time looking at social or nonsocial aspects of a scene; [30]) and has often failed to evaluate their systems with actual autistic users [17]. Little attention has been paid to how eye tracking systems might be used to support autistic individuals, particularly the 30% of autistic people who cannot communicate effectively using speech.

An eye-tracking study by Jaswal et al. [23], involving 9 nonspeaking autistic participants who use low-tech physical letterboards, found that participants looked at and pointed to letters on the physical board quickly and accurately. It follows from their study that incorporating eye gaze technology into a high-tech Augmentative and Alternative Communication (AAC) system could provide a viable, digital solution for enhancing communication in nonspeaking autistic individuals.

Commercial AR and Virtual Reality (VR) devices with built-in eye tracking offer many unique additional benefits for the nonspeaking population (as opposed to traditional eye trackers). For example, these devices can be used in a wider context, e.g., not just at a special education classroom but also for personal use cases (e.g., [43]). Another example is exploiting the mobility of these devices. One study, for instance, utilized machine learning to offer a personalized and accessible virtual content placement within AR that respond to user's movements [31]. Finally, having a 3-dimensional environment shared between educators and students can facilitate the training process for those who require extensive training (as opposed to collaborating and practising on a computer screen) [36].

Despite these benefits, however, there are no published reports that investigate the ability of nonspeaking autistic users to engage in gaze or gaze+click interactions supported by a commercial wearable AR device. This study investigates the feasibility of gaze and gaze+click interactions for nonspeaking autistic individuals. Through our study, we gather insights to drive the design of future AR eye gaze-based applications for this population.

3 SYSTEM DESIGN

The software used in this study was developed using MRTK3 [28], Netcode for GameObjects [37], and Azure Spatial Anchors [33]. Designing multiplayer AR applications presents several challenges, including ensuring accurate synchronization of virtual objects across multiple devices and maintaining consistent spatial mapping in varied physical environments. There is no straightforward commercial-grade solution for developing multiplayer AR applications using OpenXR where players share the real physical space and the virtual content within that space. The necessity for multiplayer feature in assistive technology is amplified as the support provided by caregivers is often crucial for practice and training. Furthermore, the solution needs to be simple and quick to deploy to utilize the limited time and energy of our autistic participants effectively.

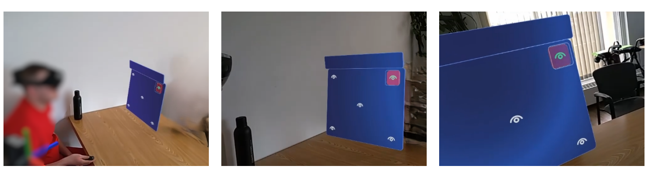
To address these challenges, we developed an easy-to-use Unity framework to facilitate the creation of multiplayer applications, which we have made publicly available[3]. Our solution takes insight from a Unity package published in a recent study [1]. It removes unnecessary computations and integrates Netcode, which is commonly used in Unity game development. With this framework, a host device first creates a spatial anchor at its origin using Azure Spatial Anchors. The anchor ID associated with this anchor is then shared with all other client devices. Once all clients have located the anchor, a single Unity GameObject (called 'Shared Content') is positioned and rotated to match the anchor's position and rotation for each client to facilitate synchronization of shared objects' poses. For a detailed documentation of this framework, please refer to the GitHub repository.

We built a system called HoloGaze on top of the above-mentioned framework. HoloGaze enables multiple clients to join an AR session. In our specific case for example, we have three types of clients: the user, the assistant (i.e., educator), and the researcher. The user is

[3]https://github.com/ETHEREAL-Research-Group/SSA-OpenXR

able to interact with interactable objects in the space using two modes: gaze and gaze+click. Thus, the interactables in the scene only respond to the user's actions. For example, when the user looks at a virtual button, the icon and/or text of that button turn green and a bounding box around it appears, indicating successful gaze engagement. The user then can trigger inputs by either gazing for a fixed period of time, or gazing and pressing the clicker[4] which is connected to the device via Bluetooth. The educator, on the other hand, can observe the animations and state changes, but interactables in the scene do not respond to their gaze. However, they can move the virtual content in the scene (e.g., position an object in a particular location). The researcher can observe all interactions and has access to a hand menu enabling them to navigate to different phases of the study (these are detailed in Section 4). The hand menu also enables toggling between the two modes of interaction (gaze and gaze+click) and setting the dwell time if the interaction mode is set to gaze. However, if this software is to be used with only two users (envisioning the future actual use case), the hand menu will be available to the educator. Figure 1 illustrates the three different views from a snapshot of one the user sessions.

We developed HoloGaze to run on the HoloLens 2 for two primary reasons. First, it is, to our knowledge, the only AR device that has been tested with nonspeaking autistic participants in multiple recent studies [1, 2, 32, 39]. These studies have demonstrated a high level of acceptability and tolerability of the HoloLens 2 among nonspeaking autistic individuals, likely due to features such as its transparent visor, which make it less intrusive. Second, at the time of this study, the HoloLens 2 was one of the few mixed reality devices with integrated eye tracking. While the Quest Pro has eye tracking, it suffers from poor pass-through quality. The Vision Pro, on the other hand, was not available in Canada at the time of the study.



(a) Researcher's view (b) Educator's view (c) Participant's view

Figure 1: Three views of a session. The left panel shows the researcher's view, the middle is the educator's view, and the right is the participant's view. An interactable icon is turned green when the participant gazes at it. The person sitting in front of the participant has been removed using object removal.

4 METHODOLOGY

In this section, we first report on participants. We then provide a comprehensive description of the study protocol. Finally, we describe the collected data and the metrics used.

[4]We used a presentation clicker: https://www.logitech.com/en-us/products/presenters/r500s-laser-presentation-remote.910-006518.html

4.1 Participants

This study was approved by the research ethics boards of our institutions. It was conducted in two provinces in Canada. Potential participants and their families were provided an advertisement poster and a short video explaining the study procedure. Interested families were asked to fill out a form and book a 1-hour session. There were three inclusion criteria: a formal autism diagnosis, an inability to communicate effectively using speech, and experience in communicating using a physical letterboard. If these criteria were met, their session was confirmed, and they received additional information, including consent forms and study questionnaires. We limited our participants to those who can communicate using letterboards since we wanted them to answer follow-up questions on their experience.

We recruited 14 adolescents and young adults (Mean age = 18.79 years; range = 12 − 28 years; all male) with help from speech-language pathologists, occupational therapists, and educators whose primary caseloads included nonspeaking autistic clients. All participants had a clinical diagnosis of autism (of the 11 participants providing the age of diagnosis information, mean age of diagnosis = 3.16 years; range = 1.58 − 5 years), and none were able to communicate effectively using speech. We also asked for an optional developmental history questionnaire. Of the 9 who provided this information (some partially), 3 reported a diagnosis of Obsessive-Compulsive Disorder (OCD), 2 reported Generalized Anxiety Disorder (GAD), 2 reported Sensory Processing Disorder and Apraxia, and 1 reported Non-Verbal Learning Disorder.

The physical letterboard experience levels varied across participants, and we measured this from their first encounter with a letterboard to the day of our data collection (mean length of experience using a letterboard = 68.07 months or 5.67 years; range 11 − 119 months). We saw these 14 participants one at a time for data collection in March and April 2024. They were seen in one of two locations in Canada (6 at one location and 8 at the other). They were paid 20 CAD via electronic gift card for their participation.

All sessions, except one, were conducted either in a private educator's office or a university lab setting. At the request of one family, one participant's session was conducted at the participant's home. Some participants had prior exposure to AR/VR (6 have previously participated in different AR research studies). All sessions were conducted by the first author, with some sessions attended by the second author. Sessions took place in the presence of a trusted other (e.g., parent or known educator). All participants' educators were also asked to sign the consent form as they would be involved in the session too.

4.2 Study Protocol

We collaborated with two professionals who support nonspeaking autistic people and a nonspeaking autistic consultant (who was not part of the study) to design the study protocol. As per their suggestion, we followed the principle of presuming competence of our autistic subjects [11]. Thus, we spoke directly to participants rather than to their caregivers, using language appropriate for adolescents and young adults. We provided multiple ways for them to become familiar with the device even before agreeing to participate (e.g., in written form and in the form of a video). Furthermore, it was emphasized in all sessions that this was not a test of the participant's ability to perform certain tasks but rather intended to identify challenges and problems of the system for future improvement. Feedback from users was collected using the physical letterboard. The different phases of the study are detailed below.

4.2.1 Tolerance and Calibration. Participants initially underwent a tolerance testing phase, where they were asked to wear the HoloLens 2 for 30 consecutive seconds. Participants were given the chance to practice wearing the device for increasing lengths of time if needed (e.g., starting with 10 seconds and advancing to 30). Those who could tolerate then engaged in the calibration process for the HoloLens 2, which typically takes about 1 minute. This process requires the user to hold their head still and follow a gem that appears at different positions of the field of view with their eyes.

We anticipated that the calibration could be challenging for many members of our target population. Therefore, if the calibration process was not successful, participants were encouraged to retry. For this phase, the calibration was loaded, and then the device was handed to the participant. This phase required pressing two buttons with hands (pressing two "next" buttons). By observing the live stream from a participant's HoloLens 2, a researcher assisted the participant (if needed) to press these two buttons.

4.2.2 Practice Phases. Participants engaged in 4 practice phases as detailed below. These phases are also depicted in Figure 2.

- Gaze practice #1: Gaze Training with Flashing Tiles– Once calibration was completed, participants went through a custom gaze training phase. In this phase, participants saw a virtual rectangle. A flashing tile appeared at the top left corner of the rectangle. When the participant gazed at this tile for 1 second, it stopped flashing, turned green, and a bounding box appeared around it to indicate successful eye gaze engagement. The person assisting the user could also observe the tile's colour (because they were also wearing a device), enabling them to provide verbal prompts to guide the user's attention if necessary. Subsequently, four more tiles appeared at the other three corners and the middle of the board one by one (the order is: top left, top right, middle, bottom left, bottom right). Previously appeared tiles remained in the field of view but were disabled (i.e., did not respond to gaze). The objective remained the same: gaze at each tile one by one until it disappeared. The total number of interactions in this phase was 5.
- Gaze+Click Practice #1: Gaze+Click Training with Flashing Tiles– This phase was similar to the previous one, with the difference being all 5 tiles were present in the field of view (FoV) from the start of the phase. Tiles started flashing one by one, and all other tiles except the flashing one were disabled. The order of flashing was the same as the previous phase. However, participants were now required to press the physical button on the physical clicker while maintaining eye gaze on the tile to select it and for it to stop flashing. The person assisting the user could offer hand-over-hand support for those who found this task challenging. However, the phase repeated until participants could complete all 5 interactions independently.
- Gaze Practice #2: Gaze Training with Flashing Letters– This phase was similar to Gaze Practice #1, but instead of flashing tiles with an eye icon at the center of the tile, there were letters. The reason for having another practice phase involving letters was that, as noted earlier, all participants had experience spelling on a physical letterboard. The top left was 'A', the top right was 'E', the middle was 'M', the bottom left was 'U', and the bottom right was 'Z'.
- Gaze+Click Practice #2: Gaze+Click Training with Flashing Letters– The task was the same as Gaze+Click Practice #1, but instead of flashing tiles with an eye icon at the centre of the tile, the tiles contained letters (similar to Gaze Practice #2).

4.2.3 Choosing Preferred Interaction Mode. At the end of the practice, participants took a short break and were asked to use a physical letterboard to choose their preferred mode of interaction: either gaze or gaze+click. If gaze was picked, the initial dwell time was set to 1 second and was adjusted as per the educator's suggestion and researcher's observation in subsequent phases. Based on in-session observations, if the participant took their gaze away from the correct target onto the next one before a click was registered, the dwell time was reduced in steps of 0.1 seconds.

4.2.4 Testing Phases. Participants engaged in two testing phases that involved spelling simple words. We focus on dictated spelling for evaluation for two primary reasons. First, we aim to leverage their existing experience in spelling on the physical letterboard as suggested by Alabood et al. [1]. Second, it is aligned with the ultimate goal of improving communication outcomes for nonspeaking autistic individuals. We acknowledge that independent spelling using such a new system requires extensive practice and training; thus, we encouraged educators to assist in the process by providing attentional prompts and cues as needed. We discuss this assistance provided later in Section 6. Selected words were simple, and in a way to cover most of the 26-letters of the English alphabet.

- Assisted Spelling: Participants proceeded to spell seven 3-letter words displayed to them one by one at the top of the letteboard. Educators were encouraged to read the word to participants. The words were: JET, DRY, EVE, FAN, GUM, RUG, IVY. For each word, the letter that needed to be selected flashed until the participant selected it using the interface the participant had chosen earlier–either gaze or gaze+click. At the end of each word, participants had to select 'Done' on the virtual board.
  To increase their visual load gradually, participants did not see the full letterboard at the beginning of the testing phase. Instead, only the letters in the first word were presented. After the first word (and after all subsequent words), the additional letters required to spell the next word were added. This design was suggested by our nonspeaking autistic consultant to reduce visual clutter initially as the participant learned the affordances of a new interface.
- Unassisted Spelling: Participants proceeded to spell 7 four-letter words displayed to them. The words were: ARCH,

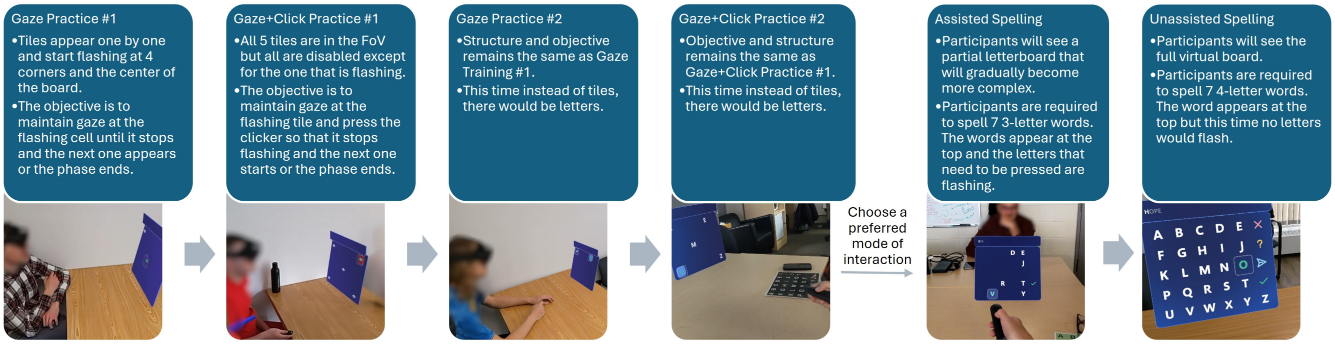


Figure 2: Diagram of Practice and Testing Phases. All snapshots are from user sessions we conducted.

BALL, DUCK, EARL, FALL, GIFT, HOPE. This time, however, the entire letterboard was visible from the first word and the letters in a given word did not flash in sequence (except for the 'Done' button at the end).

Finally, for exploratory purposes, we asked participants to fill out the NASA TLX [20] questionnaire (without pair-wise comparison) using the physical letterboard, and if they had time and interest we asked them to answer five questions with one word answers on the virtual board. This data is not reported here because these tasks were completed by only a subset of participants and because of space limitations. Specifically, the NASA TLX data was initially collected to compare the workload demand between the two types of interactions. However, due to several factors such as the small and imbalanced sample size (3 responses for gaze and 7 for gaze+click), the high standard deviations in responses, the varying number of phases completed by each participant (as explained in Section 5), and technical issues that affected participant responses (as discussed in Section 6) this data was ultimately deemed insufficient for making a meaningful comparison.

4.3 Data and Metrics

We recorded the first-person view from all three HoloLens devices (i.e., the participant, the assistant, and the researcher), along with quantitative data automatically captured by the system. The data captured by the system includes event information such as phase start times, button clicks, gaze hover entries and exits on specific GameObjects, index finger poses of both hands, the pose of the virtual letterboard, the pose of the gaze origin, and gaze hit targets (both the pose of the target and the name of the target GameObject). Although we do not utilize all this data in this study, it will be used in future research to gain a deeper understanding of interactions. Device wear on and wear off events were added in post-processing based on the videos.

As explained previously, the practice phases are divided into four sub-phases. For each sub-phase, we report the interaction throughput, defined as the number of interactions over elapsed time in interactions per minute (ipm). The elapsed time is measured from the first successful interaction (to exclude time spent on giving instructions and initiation) until the last interaction is registered. If a break was taken in between (due to technical issues or participant requests), the break time is deducted from the elapsed time.

For the two test phases, we report throughput measured as the number of correct interactions (not counting spelling errors) over elapsed time (in interactions per minute). Similar to the practice phases, elapsed time is measured from the first successful interaction until the last, with break times excluded. We also report the error rate, defined as the total number of errors divided by the sum of the number of errors and correct interactions. For example, an error rate of 0.5 indicates that for each correct interaction, there is, on average, one error.

Some participants had multiple attempts for each phase–some due to researcher's decision either because the participant used their hand (instead of gaze) to select targets, or in general for practice if the participant required a lot of assistance. Therefore, we report the throughput for the first successful completion of each phase (successful meaning they did not use their hands). Finally, we use paired t-tests to determine statistically significant difference in data.

5 RESULTS

Table 1 summarizes the findings. Out of the 14 participants, 13 were able to tolerate the HoloLens 2 device sufficiently to proceed with the study, while one participant was unable to tolerate the device and therefore unable to continue. All 13 who tolerated the device successfully completed the eye calibration. Ten of these participants completed all practice phases, and 3 successfully completed half of the practice phases. In terms of which mode of interaction they preferred, 6 participants picked gaze and 7 picked gaze+click. 12 participants attempted the testing phases; 6 completed both test phases while 3 spelled a subset of words. From the table, those who used gaze+click were significantly more successful in the testing phases. In the following subsections we provide more in-depth results on each phase of the study.

5.1 Tolerance and Calibration

As noted previously, 13 of 14 participants (92.86%) tolerated the device to start the study session, which replicates HoloLens 2 tolerance findings in other studies involving nonspeaking autistic adolescents and adults [1, 32]. One participant (P14) experienced extreme sensory sensitivity to the device. This participant was a 12-year-old male who attempted to wear the device multiple times but did not manage to keep it on. During one of these attempts, he was exposed to one virtual object. When asked to point out the

Table 1: Summary of results for the 13 participants who tolerated the device. The participants are grouped by their preferred mode of interaction.

Participant	Calibration Attempt	Practice Phases	Preferred Mode	Test Phases
P03	1	4/4	gaze+click	2/2
P04	1	4/4	gaze+click	1/2
P06	1	4/4	gaze+click	2/2
P09	1	4/4	gaze+click	2/2
P11	1	4/4	gaze+click	2/2
P12	1	4/4	gaze+click	2/2
P13	1	4/4	gaze+click	2/2
P01	3	4/4	gaze	1 word
P02	1	2/4	gaze	-
P05	3	4/4	gaze	-
P07	1	2/4	gaze	-
P08	1	2/4	gaze	-
P10	1	4/4	gaze	3 Words

problem, he replied on the physical letterboard, "just anxiety and new sensory tripping my flight control." When asked, "do you think practice would help?" he replied, "most likely." Nine participants (64.29%) did not have any issues wearing the device for the full duration of the study while 4 (28.57%) required some breaks in-between to either cool off or get used to the device at first (P02, P04, P08, P10).

All 13 participants who tolerated the device successfully completed the default eye calibration process. Eleven (84.16%) calibrated on the first try, and 2 (P01 and P05) succeeded on their third attempt. This was a surprising result because we had anticipated that keeping the head still and moving only the eyes to follow the moving holographic objects could be challenging for nonspeaking autistic people.

5.2 Practice Phases

Almost all participants who tolerated the device completed all practice phases (10 of 13, 76.92%). P02 completed Gaze Practice #1 and Gaze+Click Practice #1, initially using near-hand interactions and then eye gaze. At this point, he decided to stop, explaining (on the physical letterboard) that he was "at capacity." P07 attempted Gaze+Click Practice #1 5 times and Gaze+Click Practice #2 2 times using near-hand interactions; he was unable to use the clicker. P08 also had difficulty using the clicker and hence could not complete Gaze+Click Practice #1 (and did not attempt Gaze+Click Practice #2 per the researcher's suggestion).

For both gaze and gaze+click, interaction throughput increased from Practice #1 (flashing tiles) to Practice #2 (flashing letters). Figure 3 depicts the throughput and the distribution of the throughput for each of the practice phases from participants who completed both Gaze or both Gaze+Click practice phases. Mean throughput from Gaze Practice #1 to Gaze Practice #2 increased from 14.09 ipm (SD=8.65) to 21.37 ipm (SD=11.30). Similarly mean throughput from Gaze+Click Practice #1 to Gaze+Click Practice #2 increased from 24.32 ipm (SD=16.87) to 34.39 ipm (SD=19.90).

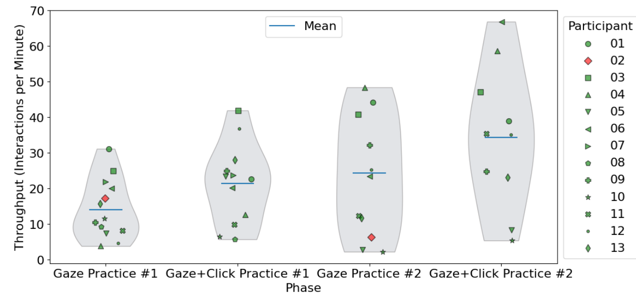


Figure 3: Practice Phases Throughput and Distribution. The distribution shown includes data from participants who completed both Gaze or both Gaze+Click practice phases. P02 is marked in red as he did not complete both phases pertaining to one mode of interaction.

There was a marginally significant increase in throughput across time for both interaction modes. Specifically, paired t-tests were conducted to compare the throughput between Gaze Practice #1 and Gaze Practice #2, and between Gaze+Click Practice #1 and Gaze+Click Practice #2, filtering those who did not complete both practice phases for one interaction mode. The t-statistic for Gaze Practice #2 vs Gaze Practice #1 was $t(11) = 2.16$ ($p = 0.054$), and for Gaze+Click Practice #2 vs Gaze+Click Practice #1 was $t(9) = 2.19$ ($p = 0.056$).

On average, gaze+click practice interactions were marginally faster than gaze alone. For those who completed all practice phases, the average throughput for gaze+click interactions was 29.36 ipm ($SD = 16.96$) and the average throughput for gaze interactions was 18.25 ipm ($SD = 8.19$), $t(9) = 2.18$, $p = 0.057$.

5.3 Preferred Mode of Interaction

As noted previously, our participants were about evenly split as to whether they preferred gaze+click (7 participants) or gaze (6 participants). The researcher suggested to two participants that they use for the testing phase the interaction mode that they had not chosen. P07 selected gaze+click as he thought having the clicker in hand would help him not do near-hand interactions, but the researcher reminded him about the difficulty he had with utilizing the clicker. P12 on the other hand picked gaze but as per the researcher's suggestion, changed his mind to gaze+click.

In general, participants selected the mode where they had better throughput. P01 and P13 were exceptions to this trend. However, we note that P13 did exceptionally well in the testing phases, suggesting that he too made the correct selection.

5.4 Test Phases

Twelve of the 13 participants who tolerated the device attempted the testing phases that involved spelling. (As noted, P02 indicated that he was "at capacity" and declined to continue.) Half of those who tried the testing phases (6 of 12) completed both the phase where the letters flashed in sequence ("assisted") and the phase where the letters did not flash in sequence ("unassisted") (P03, P06, P09, P11, P12, P13). This is a remarkable number given this was their first experience using eye gaze interactions using a head-mounted AR device.

Of the remaining 6 participants, one (P04) decided to stop after assisted spelling, and two had to stop due to lack of time (P01 and P10 typed one and three words in the assisted spelling phase, respectively). Technical issues (described in Section 6) prevented P05 and P08 from continuing. These glitches also impacted P07 who had to stop prematurely after typing two words in the assisted phase.

For those who completed at least one of the testing phases, we report the throughput and error rate (see Figures 4 and 5). Figures 4 and 5 also show the distribution for those who completed both testing phases. The rest of this analysis focuses on those who completed both phases.

Considering the increased task complexity, the mean throughput decreased from 13.49 ipm (range: 4.41 − 26.40) for assisted spelling to 10.53 (range: 5.48 − 17.39) for unassisted spelling. The mean error rate, however, surprisingly improved from 0.42 (range: 0.07 − 0.79) to 0.39 (range: 0.08 − 0.63). Paired t-tests were conducted to compare the throughput and error rate across assisted and unassisted spelling. Neither comparisons yielded a significant difference (unassisted vs assisted– throughput: $t(5) = -0.84$ $p = 0.44$; error rate: $t(5) = -0.37$, $p = 0.72$).

From Table 1, participants who chose gaze+click were overall more successful in the test phases. While 6 of the 7 gaze+click participants completely finished the testing phases, none of those who selected gaze were able to fully complete those phases.

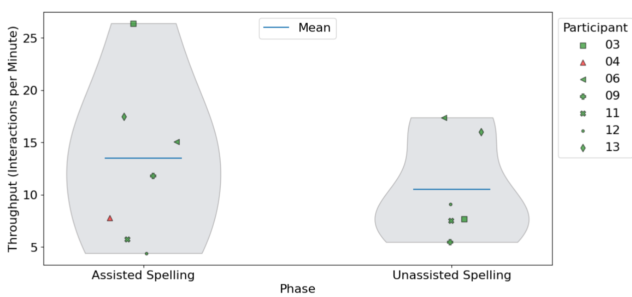


Figure 4: Testing Phases Throughput and Distribution. The distribution shown includes only those who completed both spelling phases. P04, who did not complete both phases, is shown in red.

6 DISCUSSIONS

Our study results indicate that eye gaze-based interactions in AR could be an effective mode of interaction for many nonspeaking autistic individuals. Most of our participants tolerated the HoloLens 2 device, calibrated it to their eyes, and engaged with the system performing various gaze tasks including the familiar task of spelling. Given our results indicate that gaze technology can be effective for this target population, an important questions arises: How can we make this technology more inclusive and accessible?

In this section, we discuss our study results in the context of technology inclusion and accessibility. Our discussions are divided into three main points: process considerations, design considerations and system considerations.

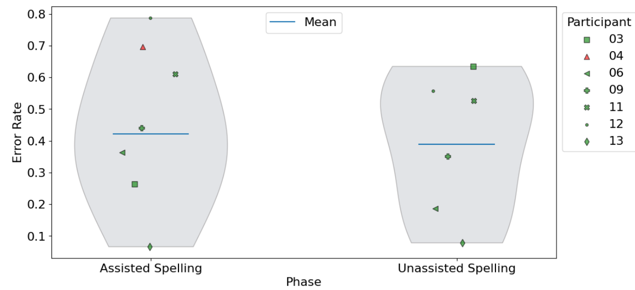


Figure 5: Testing Phases Error Rate and Distribution. The distribution shown includes only those who completed both spelling phases. P04, who did not complete both phases, is shown in red.

6.1 Process Considerations

We received feedback from practitioners on the process we employed for the user sessions. A key overarching request was for a dry run prior to the actual session without the autistic participant involved and with the researcher acting as the participant. Practitioners felt that this would have allowed them to assist participants more effectively. Some pointed out that having multiple buttons in the clicker was distracting and suggested taping the unused ones. These observations suggest the need for more comprehensive community consultations.

6.2 Design Considerations

In our study, we consulted members of the community and some professionals during the initial phases of design and development (user research phase). However, one main aspect not accounted for was the users' technological literacy and the effect of common User Interface (UI) design on the study results.

For instance, it is common practice in UI design to display different audio/visual effects when hovering over an element compared to when pressing and releasing it. In our case, however, many users lacked an understanding of this convention. For example, in the assisted spelling phase, participants were asked to select a sequence of flashing letters to spell a word. However, when they gazed at a non-target letter (i.e., one that was not flashing), the visual highlighting effects this triggered were perceived by some participants as a cue to select that non-target letter (or perhaps triggered an impulse to select it), until we explained otherwise.

For some, this tendency to select letters on hover (or maybe randomly as a self-regulation strategy) was amplified at the start of the unassisted spelling phase. P03, for example, started the unassisted spelling phase by making many errors. In fact, most of his errors occurred before he spelled the first three words (explaining the only case where error rate increased from assisted to unassisted spelling). When the researcher asked about the reason behind these initial errors, he responded (on the physical board), "My OCD took over at first; I needed time to play." Indeed, it is possible that these errors were due to increased impulsivity associated with OCD [18].

As autism and OCD frequently co-occur [4], it will be important to explore potential solutions to address such issues. For example, future research could explore the effectiveness of gaze cursors (instead of colour or scale change visual feedback) as an alternative gaze indicator.

Another example of possible limitations of the design choices we made pertains to the colour scheme and animation options we used. We used the default UI elements of the MRTK3. Although visually more appealing, we note that the subtle animations (e.g., button press animation) and the lack of apparent contrast between shades of green and blue were not the most effective choice. This is further amplified as we used the HoloLens 2, which uses a transparent display (compared to video pass-through). These examples highlight the importance of a more comprehensive user-centric design process when working with our target population.

Finally, UI design should account for the difficulty with isolating targets within a complex interface without making many errors. These errors can be of many kinds. As noted for example, some users impulsively selected incorrect letters. Some participants repeatedly pressed the same letter. Many had issues pressing the 'Done' button (possibly due to the fact that this was new, compared to the rest of the letterboard which presented a familiar layout). Reducing visual clutter, e.g., using text prediction techniques to display only the highly probable letter selections, could be effective in addressing this issue.

6.3 System Considerations

Multiplayer capability is extremely important for assistive AR software, where the involvement of caregivers and practitioners in the AR session is critical for regulation and training. Using our system, other people were able to provide assistance to the participants during the initial training phases by directing their attention through verbal cues, e.g., "look for the flashing letter", and by pointing toward the target that needed to be selected. Thus, we highly recommend future AR applications follow a similar approach, at least in the early stages of technology adoption, and then gradually reduce the need for practitioner involvement.

The tendency of the HoloLens 2 to sometimes trigger a re-calibration in the middle of a session affected user experience. Specifically, this issue prevented two participants (P05 and P08) from starting the testing phases and one (P07) from continuing after he started. These participants cancelled the re-calibration when they were prompted and this went unnoticed by the researchers. Consequently, the device reverted to head gaze, leading researchers to perceive that either the participant could not utilize gaze or the device was not tracking accurately due to a technical glitch. While in hindsight we could have handled this issue by sharing the re-calibration dialog window between devices, hardware with a more stable calibration process is critical from an accessibility perspective.

When participants shook their head vigorously or when they took frequent breaks, the devices lost their synchronization of shared holograms. Once we noticed this problem, we developed a second version of HoloGaze that performed adaptive synchronization. Since autistic users can engage in repetitive movements and since they may require multiple breaks, AR hardware selected for this population should be carefully tested to determine if it meets such needs.

7 LIMITATIONS

Our study has several limitations. First, our participant pool is limited to 14 participants. Further, all our participants were male. This limitation was due to multiple factors. First, autism is four times more likely to be diagnosed in males than females [42]. Second, as the target population of this particular study is relatively narrow (i.e., nonspeaking autistic participants who communicate using a letterboard), we opted for a convenience sample. Even though it is extremely difficult to have a balanced participant pool, it is important to deliberately recruit more diverse participants in any future work. Finally, future studies are required to generalize our findings to other AR devices.

8 CONCLUSIONS AND FUTURE WORK

In this study we introduced HoloGaze, a multiplayer gaze-based interactive system that allows nonspeaking autistic students to be provided training on how to interact with virtual objects using their eyes. HoloGaze runs on the HoloLens 2 and is built on an easy-to-use multiplayer framework for Unity and OpenXR.

Gaze-based interactions will allow nonspeaking autistic users to experience novel AR experiences that are likely to become increasingly prevalent with the proliferation of devices such as Apple's Vision Pro. Yet, prior to the current work, there were no existing studies that characterize the feasibility of gaze-based interactions for this population and that offer design insights to make such interactions effective for them.

Our study involving 14 nonspeaking autistic participants shows that gaze-based interactions could indeed be a viable mode of interaction for nonspeaking autistic users in AR. All but one participant tolerated and successfully completed the HoloLens 2 device's eye calibration. Twelve completed practice phases for at least one mode of interaction and attempted to interact with a gradually more complex interface that required them to isolate and select virtual letters to spell out a dictated word.

Our study opens new opportunities for the use of AR in the nonspeaking world. For example, this technology can expand upon traditional, low-tech communication alternatives, such as physical letterboards. Furthermore, a hands-free, untethered gaze-based virtual letterboard enabled by head-mounted AR can reduce the reliance of nonspeaking autistic individuals on a support from another person for their communication needs thereby increasing their autonomy and privacy. Future work will explore a more modern AR device, recruit more participants, and increase the diversity of participant pool to enhance the generalizability of our findings.

ACKNOWLEDGMENTS

This work was financially supported by SSHRC, NSERC, Alberta Innovates, and University of Calgary. We thank the participants and their families for their enthusiastic participation. We acknowledge the generous assistance of Dr. Morgan Barense, Hridhay Bashyam, Melody Grewal, Madison Imber, Giorgena Sarantopoulos, and Elizabeth Vosseller. Finally, we appreciate the constructive comments from the anonymous reviewers, which helped refine the paper.

REFERENCES

[1] Lorans Alabood, Travis Dow, Kaylyn B Feeley, Vikram K Jaswal, and Diwakar Krishnamurthy. 2024. From Letterboards to Holograms: Advancing Assistive Technology for Nonspeaking Autistic Individuals with the HoloBoard. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[2] Lorans Alabood, Travis Dow, Kate M. Kaufman, Vikram K. Jaswal, and Diwakar Krishnamurthy. 2023. Can Cross-Reality Help Nonspeaking Autistic People Transition to AR Typing?. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 62, 7 pages. https://doi.org/10.1145/3544549.3585859

[3] Jill Ashburner, Jenny Ziviani, and Ana Pennington. 2012. The Introduction of Keyboarding to Children With Autism Spectrum Disorders With Handwriting Difficulties: A Help or a Hindrance? *Australasian Journal of Special Education* 36, 1 (2012), 32–61. https://doi.org/10.1017/jse.2012.6

[4] Claudia Aymerich, Malein Pacho, Ana Catalan, Noorulain Yousaf, Violeta Pérez-Rodríguez, Matthew J Hollocks, Mara Parellada, Georgina Krebs, Bruce Clark, and Gonzalo Salazar de Pablo. 2024. Prevalence and Correlates of the Concurrence of Autism Spectrum Disorder and Obsessive Compulsive Disorder in Children and Adolescents: A Systematic Review and Meta-Analysis. *Brain Sciences* 14, 4 (2024), 379.

[5] Valentin Bauer, Tifanie Bouchara, Olivier Duris, Charlotte Labossière, Marie-Noëlle Clément, and Patrick Bourdot. 2022. Evaluating the acceptability and usability of a head-mounted augmented reality approach for autistic children with high support needs. In *International Conference on Virtual Reality and Mixed Reality*. Springer, 53–72.

[6] Gary Becker. [n. d.]. *Vocal Eyes: Becker Communication System.* Retrieved November 7, 2023 from https://jasonbecker.com/archive/eye_communication.html

[7] Amanda Brignell, Karen V Chenausky, Huan Song, Jianwei Zhu, Chen Suo, and Angela T Morgan. 2018. Communication Interventions for Autism Spectrum Disorder in Minimally Verbal Children. *Cochrane Database of Systematic Reviews* 11 (2018).

[8] Bill Chappell. 2022. A Nonspeaking Valedictorian with Autism Gives her College's Commencement Speech. *National Public Radio NPR* (2022).

[9] Michelle Sunm Kim Adams D. Jeffery Higginbotham, Ann M. Bisantz and Fen Yik. 2009. The effect of context priming and task type on augmentative communication performance. *Augmentative and Alternative Communication* 25, 1 (2009), 19–31. https://doi.org/10.1080/07434610802131869 arXiv:https://doi.org/10.1080/07434610802131869 PMID: 18608144.

[10] Charlotte DiStefano, Wendy Shih, Ann Kaiser, Rebecca Landa, and Connie Kasari. 2016. Communication growth in minimally verbal children with ASD: The importance of interaction. *Autism Research* 9, 10 (2016), 1093–1102.

[11] Anne M Donnellan. 1984. The criterion of the least dangerous assumption. *Behavioral Disorders* 9, 2 (1984), 141–150.

[12] Travis Dow, Pratishtha Pratishtha, Lorans Alabood, Vikram K Jaswal, and Diwakar Krishnamurthy. 2024. Towards an Augmented Reality Agent to Support Communication for Nonspeaking Autistic People. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[13] John Dudley, Hrvoje Benko, Daniel Wigdor, and Per Ola Kristensson. 2019. Performance envelopes of virtual keyboard text input strategies in virtual reality. In *2019 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 289–300.

[14] Alexander Fiannaca, Ann Paradiso, Mira Shah, and Meredith Ringel Morris. 2017. AACrobat: Using Mobile Devices to Lower Communication Barriers and Provide Autonomy with Gaze-Based AAC. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing* (Portland, Oregon, USA) *(CSCW '17)*. Association for Computing Machinery, New York, NY, USA, 683–695. https://doi.org/10.1145/2998181.2998215

[15] Kimberly A Fournier, Chris J Hass, Sagar K Naik, Neha Lodha, and James H Cauraugh. 2010. Motor Coordination in Autism Spectrum Disorders: A Synthesis and Meta-Analysis. *Journal of autism and developmental disorders* 40, 10 (2010), 1227–1240.

[16] Lori Frost. 2002. The picture exchange communication system. *Perspectives on Language Learning and Education* 9, 2 (2002), 13–16.

[17] Karri Gillespie-Smith and Sue Fletcher-Watson. 2014. Designing AAC Systems for Children with Autism: Evidence from Eye Tracking Research. *Augmentative and Alternative Communication* 30, 2 (2014), 160–171. https://doi.org/10.3109/07434618.2014.905635 arXiv:https://doi.org/10.3109/07434618.2014.905635

[18] Giacomo Grassi, Stefano Pallanti, Lorenzo Righi, Martijn Figee, Mariska Mantione, Damiaan Denys, Daniele Piccagliani, Alessandro Rossi, and Paolo Stratta. 2015. Think twice: Impulsivity and decision making in obsessive–compulsive disorder. *Journal of behavioral addictions* 4, 4 (2015), 263–272.

[19] Jeffrey T Hansberger, Chao Peng, Shannon L Mathis, Vaidyanath Areyur Shanthakumar, Sarah C Meacham, Lizhou Cao, and Victoria R Blakely. 2017. Dispelling the gorilla arm syndrome: the viability of prolonged gesture interactions. In *Virtual, Augmented and Mixed Reality: 9th International Conference, VAMR 2017, Held as Part of HCI International 2017, Vancouver, BC, Canada, July 9-14, 2017, Proceedings 9*. Springer, 505–520.

[20] Sandra G Hart. 1986. NASA task load index (TLX). (1986).

[21] Naoki Higashida. 2013. *The Reason I Jump: The Inner Voice of a Thirteen-Year-Old Boy with Autism*. Knopf Canada.

[22] Vikram K Jaswal, Andrew J Lampi, and Kayden M Stockwell. 2024. Literacy in nonspeaking autistic people. *Autism* (2024), 13623613241230709.

[23] Vikram K Jaswal, Allison Wayne, and Hudson Golino. 2020. Eye-Tracking Reveals Agency in Assisted Autistic Communication. *Scientific reports* 10, 1 (2020), 1–10.

[24] Diwakar Krishnamurthy, Vikram Jaswal, Ahmadreza Nazari, Ali Shahidi, Pranav Subbaraman, and Mea Wang. 2022. HoloType: Lived Experience Based Communication Training for Nonspeaking Autistic People. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, Article 337, 6 pages. https://doi.org/10.1145/3491101.3519869

[25] Per Ola Kristensson and Keith Vertanen. 2012. The potential of dwell-free eye-typing for fast assistive gaze communication. In *Proceedings of the symposium on eye tracking research and applications*. 241–244.

[26] Andrew Kurauchi, Wenxin Feng, Ajjen Joshi, Carlos Morimoto, and Margrit Betke. 2016. EyeSwipe: Dwell-Free Text Entry Using Gaze Paths. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1952–1956. https://doi.org/10.1145/2858036.2858335

[27] Päivi Majaranta, Ulla-Kaija Ahola, and Oleg Špakov. 2009. Fast Gaze Typing with an Adjustable Dwell Time. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Boston, MA, USA) *(CHI '09)*. Association for Computing Machinery, New York, NY, USA, 357–360. https://doi.org/10.1145/1518701.1518758

[28] marlenaklein msft, AMollis, nmanis, cheles, RogPodge, MaxWang-MS, Zee2, thinkuldeep, cre8ivepark, vtieto, Sean-Kerawala, keveleigh, qianw211, and cmcclister. 2023. *Mixed Reality Toolkit 3 Developer Documentation.* Retrieved June 17, 2024 from https://learn.microsoft.com/en-us/windows/mixed-reality/mrtk-unity/mrtk3-overview/

[29] Communication Matters. [n. d.]. *E-Tran Frames.* Retrieved November 7, 2023 from https://www.communicationmatters.org.uk/what-is-aac/types-of-aac/e-tran-frames/

[30] James C McPartland, Sara Jane Webb, Brandon Keehn, and Geraldine Dawson. 2011. Patterns of visual attention to faces and objects in autism spectrum disorder. *Journal of autism and developmental disorders* 41 (2011), 148–157.

[31] Ahmadreza Nazari, Lorans Alabood, Kaylyn B Feeley, Vikram K Jaswal, and Diwakar Krishnamurthy. 2024. Personalizing an AR-based Communication System for Nonspeaking Autistic Users. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 731–741.

[32] Ahmadreza Nazari, Ali Shahidi, Kate M. Kaufman, Julia E Bondi, Lorans Alabood, Vikram K. Jaswal, Diwakar Krishnamurthy, and Mea Wang. 2023. Interactive AR Applications for Nonspeaking Autistic People? - A Usability Study. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 708, 15 pages. https://doi.org/10.1145/3544548.3580721

[33] pamistel, RamonArguelles, msftradford, mattwojo, and SyntaxC4. 2024. *Azure Spatial Anchors overview - Azure Spatial Anchors.* Retrieved June 17, 2024 from https://learn.microsoft.com/en-us/azure/spatial-anchors/overview

[34] Diogo Pedrosa, Maria Da Graça Pimentel, Amy Wright, and Khai N. Truong. 2015. Filteryedping: Design Challenges and User Performance of Dwell-Free Eye Typing. *ACM Trans. Access. Comput.* 6, 1, Article 3 (mar 2015), 37 pages. https://doi.org/10.1145/2724728

[35] Edlyn Vallejo Peña. 2019. *Communication Alternatives in Autism: Perspectives on Typing and Spelling Approaches for the Nonspeaking.* McFarland.

[36] Catlin Pidel and Philipp Ackermann. 2020. Collaboration in virtual and augmented reality: a systematic overview. In *Augmented Reality, Virtual Reality, and Computer Graphics: 7th International Conference, AVR 2020, Lecce, Italy, September 7–10, 2020, Proceedings, Part I 7*. Springer, 141–156.

[37] Amy Reeve. 2024. *About Netcode for GameObjects | Unity Multiplayer Networking.* Retrieved June 17, 2024 from https://docs-multiplayer.unity3d.com/netcode/current/about/

[38] David James Savarese. 2019. Coming to My Senses. *Autism in Adulthood* 1, 2 (2019), 90–92.

[39] Ali Shahidi, Lorans Alabood, Kate M Kaufman, Vikram K Jaswal, Diwakar Krishnamurthy, and Mea Wang. 2023. AR-Based Educational Software for Nonspeaking Autistic People-A Feasibility Study. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 543–552.

[40] Rosella Spataro, Maria Ciriacono, Cecilia Manno, and Vincenzo La Bella. 2014. The eye-tracking computer device for communication in amyotrophic lateral sclerosis. *Acta Neurologica Scandinavica* 130, 1 (2014), 40–45.

[41] Tobii Dynavox US. 2024. *PCEye eye tracker for accessibility.* Retrieved November 7, 2023 from https://us.tobiidynavox.com/pages/pceye

[42] Donna M Werling. 2016. The Role of Sex-Differential Biology in Risk for Autism Spectrum Disorder. *Biology of sex differences* 7, 1 (2016), 1–18.

[43] Liuchuan Yu, Huining Feng, Rawan Alghofaili, Boyoung Byun, Tiffany O'Neal, Swati Rampalli, Yoosun Chung, Vivian Genaro Motti, and Lap-Fai Yu. 2024. HoloAAC: A Mixed Reality AAC Application for People with Expressive Language Difficulties. In *International Conference on Human-Computer Interaction*. Springer, 304–324.