The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality

ALEXANDER PLOPSKI, Department of Information Science, University of Otago
TERESA HIRZLE*, Institute of Media Informatics, Ulm University
NAHAL NOROUZI*, College of Engineering and Computer Science, University of Central Florida
LONG QIAN*, Department of Computer Science, Johns Hopkins University
GERD BRUDER, Institute for Simulation and Training, University of Central Florida
TOBIAS LANGLOTZ, Department of Information Science, University of Otago

With innovations in the field of gaze and eye tracking, a new concentration of research in the area of gaze-tracked systems and user interfaces has formed in the field of Extended Reality (XR). Eye trackers are being used to explore novel forms of spatial human-computer interaction, to understand human attention and behavior, and to test expectations and human responses. In this paper, we review gaze interaction and eye tracking research related to XR that has been published since 1985, which includes a total of 215 publications. We outline efforts to apply eye gaze for direct interaction with virtual content, design of attentive interfaces that adapt the presented content based on eye gaze behavior, and discuss how eye gaze has been utilized to improve collaboration in XR. We outline trends and novel directions, and discuss representative high-impact papers in detail.

CCS Concepts: • **Human-centered computing → Interaction devices**; **Interaction paradigms**; **Interaction techniques**.

Additional Key Words and Phrases: eye tracking, gaze, mixed reality, augmented reality, virtual reality, extended reality, interaction, collaboration, selection, interface, survey, literature review, head-mounted, head-worn.

ACM Reference Format:
Alexander Plopski, Teresa Hirzle, Nahal Norouzi, Long Qian, Gerd Bruder, and Tobias Langlotz. 2021. The Eye in Extended Reality: A Survey on Gaze Interaction and Eye Tracking in Head-Worn Extended Reality. *ACM Comput. Surv.* 37, 4, Article 111 (September 2021), 37 pages. https://doi.org/10.1145/1122445.1122456

1 INTRODUCTION

Head-Mounted Displays (HMDs) have been first introduced in the groundbreaking essay on "The Ultimate Display" [241] and the first practical implementation by Sutherland et al. [242]. Over the

*The authors contributed equally to this research and are listed in alphabetical order.

Authors' addresses: Alexander Plopski, alexander.plopski@otago.ac.nz, Department of Information Science, University of Otago, PO Box 56, Dunedin, New Zealand, 9054; Teresa Hirzle, teresa.hirzle@uni-ulm.de, Institute of Media Informatics, Ulm University, James-Franck-Ring 1, Ulm, Germany, 89069; Nahal Norouzi, nahal.norouzi@knights.ucf.edu, College of Engineering and Computer Science, University of Central Florida, 4328 Scorpius St, Orlando, FL, United States, 32826; Long Qian, long.qian@jhu.edu, Department of Computer Science, Johns Hopkins University, 3400 N Charles St, Baltimore, MD, United States, 21218; Gerd Bruder, bruder@ucf.edu, Institute for Simulation and Training, University of Central Florida, 3100 Technology Pkwy, Orlando, FL, United States, 32826; Tobias Langlotz, tobias.langlotz@otago.ac.nz, Department of Information Science, University of Otago, PO Box 56, Dunedin, New Zealand, 9054.

0360-0300/2021/9-ART111 $15.00
https://doi.org/10.1145/1122445.1122456

past 50 years, the design of HMDs has undergone many changes with significant improvements in the design, optical composition, and tracking and rendering capabilities. More than 5.5 million units are expected to have been delivered in 2020, and this number is expected to rise to more than 40 million units in 2025 [231]. This development in HMDs will introduce *Extended Reality* (XR) experiences into our everyday lives ranging from medicine, industry to entertainment and games or even into casually worn eyewear. Hereby, we define XR to include elements of the *Mixed Reality* (MR) continuum defined by Milgram et al. [164] and *Virtual Reality* (VR), more precisely XR contains *Augmented Reality* (AR), VR, and MR.

With the introduction of handheld devices, we shifted from using keyboard and mouse to touch-based interaction with digital information. With the shift towards head-worn devices, such HMDs will likely bring a similar paradigm shift in our interactions with the digital content. Thus far, the interaction method for HMDs has not been universally established or standardized with commercial devices utilizing a variety of interaction methods. Most commercial HMDs utilize handheld controllers (HTC Vive, MagicLeap One, Oculus Rift), touch surfaces on the device itself (Google Glass), voice (Microsoft HoloLens), and gesture interfaces (Microsoft HoloLens, Meta2, Oculus Quest 2). Some researchers have also explored the use of handheld devices as a means of replacing dedicated controllers and to give users a touch surface to interact with [168]. While controllers made of handheld devices can help bridge the gap from the familiar interfaces on handheld devices or game controllers towards wearable interfaces, this requires users to carry around a dedicated device for interaction. Gesture and voice interaction could be a means of overcoming this limitation in some scenarios, but they cannot be used in more crowded environments because of noise or social issues, such as on crowded streets or in public transportation [77]. An ideal interaction method should be readily available, be intuitive and fast, and be inconspicuous without attracting attention from bystanders.

Eye-gaze has the potential to be a key part of this interaction method and has long been envisioned as a natural interaction modality [98]. Our eyes can showcase the users' interest and can provide empirical information about how users perceive a scene, what they notice, pay attention to, or what they are primarily interested in [123, 125].

This wealth of information facilitated cognitive researchers in understanding human cognitive processes [53], and enabled human-computer interfaces with a new modality where user's intent can be inferred from their eye gaze. Such gaze-based interaction techniques resulted in the development of applications for a wide range of purposes from increasing user accessibility [80, 89, 118, 134, 138, 154–156, 256, 260, 270] to entertainment [81, 126].

To classify gaze-based interactions, Majaranta and Bulling introduced an eye tracking application continuum revolving around users' level of intent during an interaction, i.e., intentional vs. unintentional interaction, and the required responsiveness of the system one is interacting with, i.e., online vs. offline [149]. For online/active interfaces, one's gaze can be used to explicitly select or manipulate targets on a computer screen [227], or implicitly adjust the resolution of an image on a display so higher resolution is in line with user's center of attention allocating lower resolution to the periphery [51]. Offline systems are utilized to either create a model of the user's attention and cognitive processes based on their gaze behavior or for diagnostics purposes [149].

Separately, our interactions with computers in one form or another are constantly increasing and are becoming more personalized. Although eye tracking has been conceptualized and investigated as a natural means to streamline such interactions, e.g., by implicitly modifying the behavior of virtual avatars according to user's gaze [173] or triggering interactions with virtual content when the system detects the user's interest in it [218], thus far its applications in XR have been limited, primarily due to the lack of accessible hardware. This limitation is slowly disappearing in newer HMD iterations that integrate eye tracking capabilities, e.g., FOVE, HTC Vive, HoloLens2, and

MagicLeap One. The advances in XR technology and its increased popularity have led to research opportunities for new and interactive ways of utilizing one's eye/gaze information to facilitate various interactions.

Most authors of this review discussed the potential of eye-gaze tracking in XR during the *NII Shonan Seminar on Augmented Reality in Human-Computer Interaction* [177] with participants from HCI and XR. Discussions revealed that in recent years there was a significantly increased interest in eye tracking in XR and HCI, and this review should help stimulate new research opportunities in this area by identifying and structuring existing works and revealing key questions for future work in gaze interaction and eye tracking in XR.

In summary, this review investigates research in the use of eye/gaze tracking in XR environments to provide answers for the following questions:

- **Q1:** What are the main categories of gaze interaction and eye tracking research for XR interfaces?
- **Q2:** What sub-categories within each research category have garnered more attention?
- **Q3:** What are some of the emerging and future research directions for gaze interaction and eye tracking in XR?

In this work, we contribute to the research community by providing summaries of the research efforts in the aforementioned areas from 1985 to 2020 and identifying underrepresented directions, novel solutions, and promising future applications. We hope that our efforts can provide both a historical view and spark innovative ideas for researchers in the fields of eye/gaze tracking, XR, and HCI. In the remainder of this paper, we discuss the methodology for our review and introduce our review topics in Section 2 and provide a high-level analysis of the research contributions. In Section 3, we expand upon research efforts specific to our review topics. We then provide insights on past research trends and future directions in Section 4 and conclude the paper in Section 5.

2 METHODOLOGY

We adopted a two-step procedure for our review of eye and gaze tracking in XR. The first step involved data collection and identification of relevant publications. In the second step, we defined our review topics, further identified the papers that made contributions to these topics and provided summaries on their research findings.

The focus of our review are eye tracking applications in XR and here specifically for HMDs. As such, we identified the related papers on SCOPUS by searching for papers that included XR-related index terms "*Augmented Reality*" OR "*Virtual Reality*" OR "*Mixed Reality*" OR "*Head Mounted Display*" OR "*Head Worn Display*" OR "*Eye Wear*" and eye tracking related terms "*Eye Tracking*" OR "*Eye Gaze*" in the title, keywords, and the abstract fields (see Figure 1). This search resulted in 1278 papers published between 1985 and May 20, 2020 (see Figure 2). We opted to include recent papers in the review to cover recent trends in our review. Furthermore, we did not exclude papers based on the number of citations to ensure that we do not exclude ideas that may be novel but overlooked for many years or are recent publications. After compiling the paper list, we discussed the classification criteria related to eye tracking research and applications for the papers and distributed the papers among the members of our team for an initial review, summary, and classification. We first classified 50 papers to determine any other classification criteria that were predominant in the collected papers and classified all papers according to this expanded classification criteria. The complete set of the 15 classification criteria is shown in Figure 1. After this review cycle, we removed 331 papers that mentioned the keywords but did not utilize XR or eye tracking, e.g., mentioning importance of eye contact in communication or developing an eye tracking algorithm with a mention of XR as a potential application area, 1 paper due to plagiarism, and 90 papers that we could not access. This

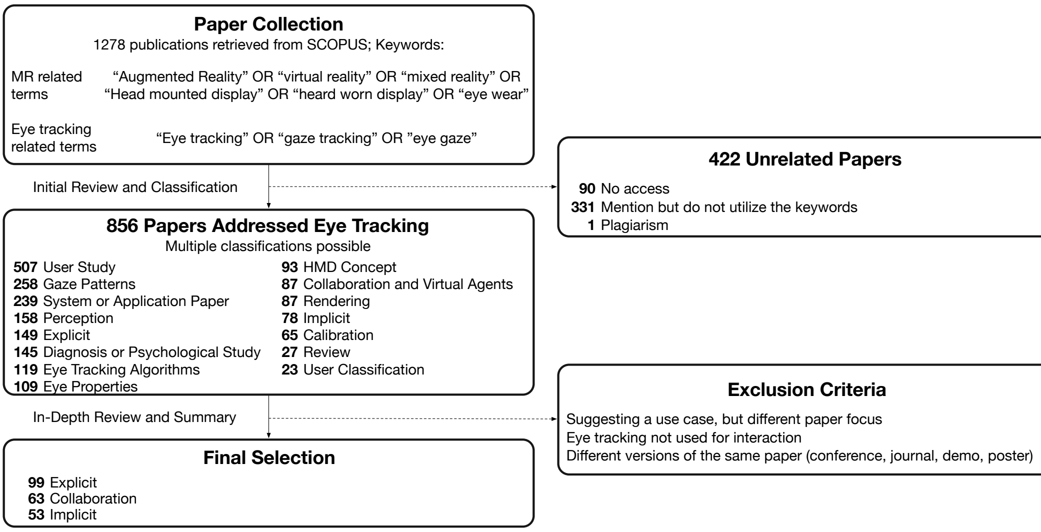


Fig. 1. Selection process of the papers reviewed in this work.

procedure resulted in a list of 856 papers that addressed eye tracking in XR and adhered to our classification criteria.

After the initial classification, the authors discussed and identified areas of relevance for the review that also encompass all the identified papers. Inspired by the continuum of eye tracking applications by Majaranta and Bulling [149], we selected the two areas *explicit gaze interaction* and *implicit gaze interaction*. Furthermore, we selected *collaborative gaze interaction* as the third area. The authors agreed on these topics, as they believed them to be directly relevant to the design of interfaces for a detailed summary.

We, again, distributed the remaining 856 papers and conducted an in-depth review of the papers in each category and removed papers that did not match the focus of each category, e.g., by mentioning eye tracking for selecting targets, but focusing on the development of an eye tracking algorithm; or did not utilize eye tracking for interaction, e.g., assuming that the center of the user's view corresponds to the gaze point or collecting eye gaze information but utilizing other means to facilitate interaction. We also excluded duplicate papers and papers that were different variations of the same paper, e.g., a demo or a poster of a conference paper. This resulted in a final set of 215 papers that were included in the final review. Of these papers, 99 utilized eye tracking for *explicit eye input*, 53 papers presented *implicit user interfaces*, and 63 papers focused on *collaborative gaze interaction* (see the overall process in Figure 1). These papers were again distributed among the members who reviewed the papers in detail, organized them into subcategories, and identified recent and future directions. It is important to note that some works that consider a specific property of the eye are missing from our review even though they may address the identified research categories [35, 124, 240]. Due to their focus on a specific property of the eye they utilize keywords like "saccades" or "pupil dilation" rather than "eye tracking" or "gaze tracking," meaning they did not match our selection criteria. The structure of our paper discussion is shown in Figure 3.

3 RESEARCH TOPICS AND DIRECTIONS

We identified three main categories of eye tracking for interaction in XR. The first group utilizes eye tracking similar to a mouse on the computer, the user can target different objects and select them

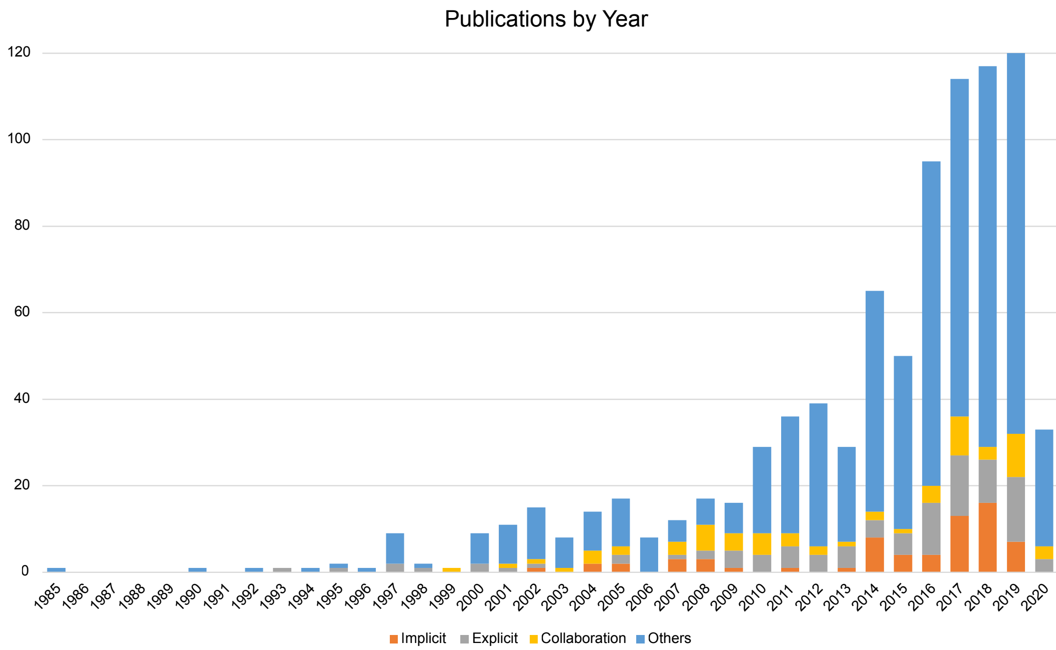


Fig. 2. Published papers by year.

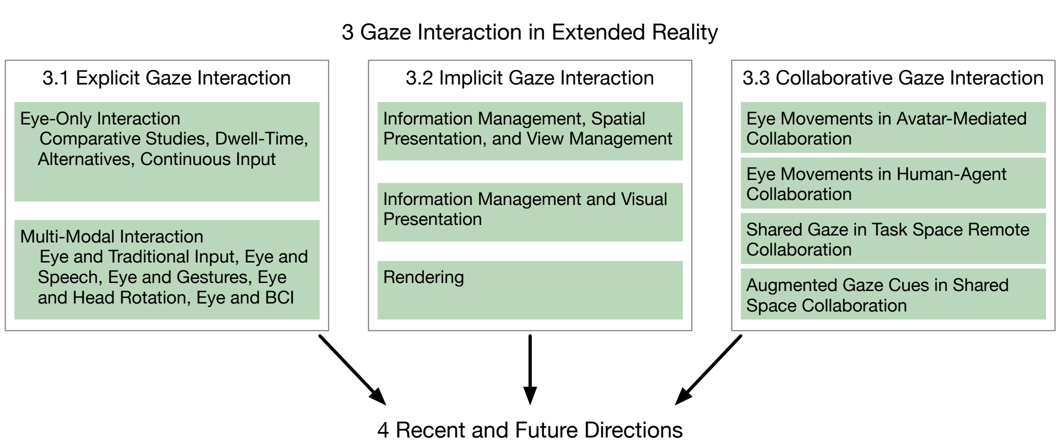


Fig. 3. Overview of the parts that are covered by this survey. Papers were classified into the three categories *explicit gaze interaction*, *implicit gaze interaction*, and *collaborative gaze interaction*. Based on the analysis of the reviewed papers in these categories, we identified recent trends and future directions.

for interaction. The second group analyzes the user's gaze to adapt system parameters without users actively triggering these changes. Finally, eye gaze plays an important role in collaborative XR environments. In this section we summarize research outcomes in each of these categories.

3.1 Explicit Eye Input

Gaze has been identified as a natural means of interaction in the HCI domain, as humans gaze at what they are attending or planning to attend to [125, 149]. Several properties of gaze, such as its fast and direct availability, have been leveraged and applied for intentionally performing an interaction event with the eyes (explicit gaze input). At this, gaze was mainly used for two scenarios: as sole input signal or in combination with other modalities. Figure 4 represents a few of these examples.

One of the most prominent problems that have occurred when using the eyes to control user interfaces is related to the fact that the eyes have evolved to observe and not change the environment. Therefore, the *Midas Touch* problem has often been observed, which describes unintentional gaze behaviors that affect the interaction result (e.g., selection of the wrong menu button because of only glancing at it) [99]. A lot of the previous work in eye-based HCI research has concentrated on developing interaction strategies that prevent the Midas Touch problem. Dwell time is a common approach to solve this problem and researchers tested different dwell times based on the purpose of the interaction, such as target selection and manipulation [218, 226], replicating a single-button mouse for pointing, selecting, and dragging [43], and scrolling while reading [119]. As the dwell time approach can cause fatigue and slow down the interaction in certain applications [92], other approaches were developed to ease the interaction, which were either relying on gaze behavior, such as gaze gestures [93] and fixation patterns [133] or taking advantage of the other modalities in multi-modal platforms [14, 26, 213, 218, 228, 258].

Whereas most traditional interaction devices (e.g. laptop, smartphone) rely on two-dimensional displays, this is different for XR devices. Those augment the real (3D) world with digital information or create even an entirely new three-dimensional world. Still, understanding the eye-only or multi-modal approaches developed for 2D spaces [10, 14–16, 26, 28, 61, 94, 95, 172, 182, 213, 218, 226, 228, 258] can be beneficial for the interactions in 3D space. Therefore, it has to be investigated to which extent existing eye-based interaction techniques can be transferred to 3D space or if new approaches that are specifically tailored to XR requirements have to be developed.

In this section, we present research that investigates eye-based interaction techniques in the XR domain. Hereby, we describe the modalities utilized to facilitate user interactions, identify solutions to the Midas Touch problem, and describe the different application areas benefiting from gaze-based interactions.

3.1.1 Eye-Only Interaction. Many research efforts utilized the eye as the sole input for their XR interaction space. One of the common research questions in this domain was understanding how eye-only interaction compares to other input modalities, such as pointing or head-based interaction. Other researchers focused on developing and investigating solutions aimed at resolving issues specific to eye-only interaction, such as utilizing dwell time for the Midas touch problem. In the following, we provide a detailed description of these works and their findings.

Comparative Studies. In the real world, humans use their body to attend to their environment or communicate their attention to others, such as pointing or directing their head or eyes towards an object of interest. Therefore, these nonverbal cues are natural contenders for further investigation for target selection and manipulation tasks in XR. As XR technology advances and becomes more ubiquitous, the need for understanding the performance, social, and cultural requirements and implications of using different interaction modalities increases. For instance, for a specific task, eye-based interaction might turn out to be faster and more discreet than pointing. Looking at past works, understanding the performance capabilities of eye-based interactions was an important

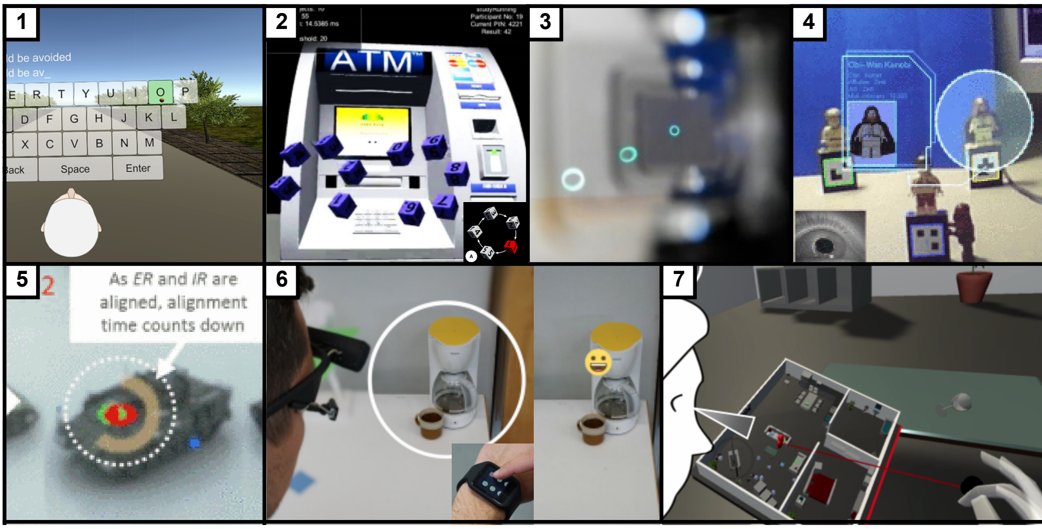


Fig. 4. Examples of explicit eye gaze use for interaction. (1) Gaze typing in VR. Due to the whole keyboard being in the field of view, minimal head movements are needed for selecting the keys. Selection is done via dwell or click [205]. (2) Selection via smooth pursuit eye movements in virtual reality. The person has to follow the rotating cubes with their eyes in order to select a target [111]. (3) Focal depth as the interaction method in augmented reality. Targets are presented at different focal depths to allow for interaction with objects in one line of sight at several depths [248]. (4) Gaze-adapted annotations in augmented reality [129]. (5) Combination of eye gaze and head rotation to select objects. Users had to gaze and nod in order to perform a selection [197]. (6) Gaze interaction in combination with smartwatch input as an annotation system in AR. Users gaze at an object to indicate the object of interest and add an annotation with selecting a smartwatch item [12]. (7) Gaze in combination with freehand gestures for selecting and manipulating items in VR. Here, gaze selects the object of interest, which is then manipulated using freehand gestures [193].

aspect of developing the interaction techniques, usually achieved through comparative studies with other input modalities.

In one of the earlier comparative studies in VR, Tanriverdi et al. compared an eye-based interaction technique with hand pointing for a search and selection task [243]. They found that participants' interactions were faster using the eye tracking based method. Very commonly gaze input was compared with head rotation as input in augmented and virtual reality HMDs [21, 82, 122, 167, 201]. Kyto et al. found that while the proposed head-only interaction technique was more accurate compared to eye tracking, the eye-only technique was faster [122]. However, Qian et al. found faster selection times and higher accuracy values for head-based interaction compared to eye-only interaction [201]. They also found that the head-only technique was overall more fatiguing than the eye-only technique, except neck fatigue, which was also observed by Blattgerste et al., who reported that users found eye-only interaction less exhausting than head-interaction [21]. They further found that mostly less errors were observed using the eye-only method than the head-based method. Minakata et al. [167] found that eye gaze was slower for pointing than head and foot-based controls. Choi et al. compared eye-gaze selection with head-rotation based selection in a VR environment, and found that users preferred eye-gaze selection in terms of convenience and satisfaction, and they preferred head-rotation for ergonomics [38]. Jalaliniya et al. compared target pointing on a head-mounted display using gaze, head and mouse, finding that eye-based pointing is significantly faster, while the users felt that head pointing is more accurate and convenient [100].

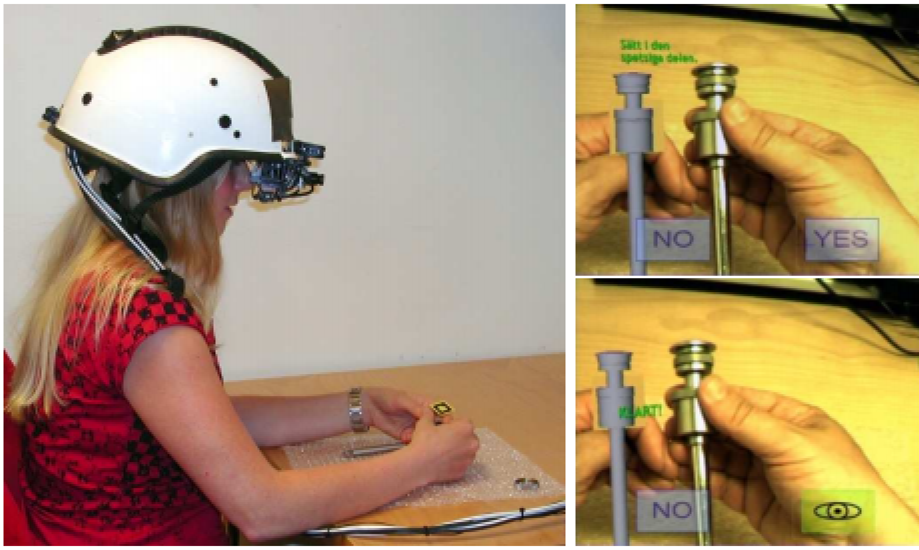


Fig. 5. Application for instructional purposes, where a user (left figure) can progress through steps by fixating on the virtual buttons in their field of view (right figures) [179].

Esteve et al. [58] compared head rotation and eye pursuit in tracking of virtual targets. The results suggested that head-based input can more accurately track moving targets than using the eyes. Zhang et al. [272] compared eye-gaze based and controller based controls in robot teleoperation. In their work, the use of eye-gaze resulted in slower operations with more errors and had a negative impact on the user's situational awareness and recall of the environment. Luro and Sundstedt [145] compared eye gaze and controller based aiming in VR and found that both performed similarly.

Dwell-Time. One of the most common eye-only interaction methods is the *dwell-time* approach. At this, the eyes have to be held on a target for a predefined set of time in order to trigger an input event. To provide ALS patients with more interactive capabilities, Lin et al. developed an HMD eye tracker, calibration, and data processing method to accurately detect user's gaze and activate a speech system linked to different menu items and select those items [138]. Graupner et al. evaluated the usability of a see-through HMD with gaze-based interaction capabilities and measured reaction time and hit rate in point selection tasks and investigated the influence of factors such as noise, sampling rate and target size [74]. Nilsson et al. developed a gaze attentive AR video see through prototype for instructional purposes, illustrated in Figure 5, where users following sequential steps in the task could activate each step using interactive virtual buttons by fixating on them [178, 179]. Rajanna and Hansen [205] compared a dwell-time approach with clicking on a controller for typing on a virtual keyboard. They found that clicking on a controller was faster and produced less errors than the dwell-time approach. Voros et al. [256] developed an interface to allow people with severe speech and physical impairments (SSPI) to select words from the world using gaze, and therefore communicating with others. Giannopoulos [68] used dwell-time based selection in a virtual retail environment. Cottin et al. integrated an optical see through (OST) HMD with an eye tracker, to allow users to select virtual objects on the HMD screen with the dwell time approach in a SmartHome application [45]. Liu et al. [142] designed a gaze-only interface for adjusting the position of an object in 3D by adjusting its position on pre-defined planes.

Dwell-Time Alternatives. The dwell-time approach is rather prone to the Midas Touch problem. To resolve this problem, several other approaches were proposed. To overcome some of the difficulties brought on by dwell-time based gaze interaction methods, Lee et al. developed a novel approach by utilizing half blinks and gaze information to facilitate users with tasks such as target selection which was tested through interacting with augmented annotations in AR [129]. Khamis et al. presented an approach that used smooth pursuit eye movements for selection of 3D targets in a virtual environment [111]. They found that the movement is robust against target size, and detection improves with an increasing movement radius. Gao et al. developed an eye gesture interface, where combinations of eye movements are measured by an amplified AC-coupled electrooculograph [63]. The proposed interface achieved a success rate of 97% in recognizing eye movement. Xiong et al. combined eye fixation and blink in a typing user interface [266]. Toyama et al. combined sequence of eye fixations instead of fixations for each frame with object recognition algorithms to build an AR Museum guidance application [247]. Hirata et al. [85] designed an interface based on conscious change of eye vergence to select objects in 2D and 3D.

Continuous Input. Gaze has been used as a continuous input signal for navigation and control tasks in virtual environments and teleoperation [11, 118, 154, 233, 271]. Gaze has also been explored in narrative and tourism applications that provide users with information about different objects of interest by either detecting the gaze in a highlighted area of interest or during free exploration [121, 267].

Overall, past works indicate a variety of applications where eye-only interaction was used. Although comparisons between eye-only interaction methods with other modalities have not always resulted in consistent findings, differences in type of HMDs and eye trackers utilized, and the interaction tasks can explain some of these inconsistencies. Also, we observed that the ease of using the dwell-time approach has allowed for its adoption in a wide range of research topics from usability assessment [74]to increasing users' accessibility [138, 256]. However, due to certain limitations that this approach can introduce, such as interaction time delays and user fatigue [92, 205], we observed an increasing attention to alternative approaches [63, 85, 111, 129, 247, 266]. The variety of these alternative approaches suggests the potential for eye-based interactions as a flexible interaction mechanism to allow for different user capabilities and interaction contexts. However, open questions exists regarding the usability and performance of these approaches compared to each other and different interaction modalities. Additionally, advances in eye tracking and HMD technologies and artificial intelligence algorithms hold promise for more streamlined interactions in the future.

3.1.2 Multi-Modal Interaction. Understanding the capabilities of eye-only interaction is highly valuable, especially for circumstances concerning specific disabilities, where eyes are the only interaction input. However, combining eye-based interactions with other modalities (e.g., head-based and gesture-based interactions) can create a richer and more expressive experience for the user and also better facilitate certain complex tasks. In the following, we describe previous works that focused on combining eye input with different modalities.

Eye and Traditional Input. Continuous usage of devices that interface through mechanical inputs (e.g., button presses) has become ubiquitous, which includes cell phones and smart watches, making them ideal modality pairs for eye-based interactions for a wider audience. Sidorakis et al. presented a VR user interface combining gaze and an additional mechanical input to signify a selection [224]. The multi-modal interaction scheme is evaluated to be more accurate than traditional mouse/keyboard interaction in an immersive virtual environment. Similar interaction technique is employed in a mobile-based AR game [126] and wheelchair navigation [89]. Sunggeun and Geehyuk [3] explored the benefits of eye-gaze and a control pad attached to a head-mounted display for typing and found this combination to outperform both exclusive eye gaze and control pad input. Bace et al. developed ubiGaze, where the interaction is based on gaze tracking and a smartwatch [12]. The gaze provides selection of real-world objects, and the smartwatch can receive various commands to be executed with regard to the objects. Mardanbegi et al. [151] combined gaze with a control tool attached to the controller to achieve combined selection of an object to interact with and a function.

Eye and Speech. Speech-based interaction is one of the primary modalities used to communicate with intelligent characters in futuristic movies. However, due to limitations in technology, less has been done in pairing speech and eye-based interactions. Beach et al. developed one of the earlier multi-modal prototypes to provide hands-free interaction for users by utilizing speech and discussed the possible use of other modalities such as blinking or fixating on a desired target in case of inaccessibility of speech input [17].

Eye and Gestures. In many eye-based interactions, using eye input for target selection leaves other modalities such as hands free to be utilized as input for other interactions such as object manipulations. Heo et al. developed a multi-modal interaction interface for gaming purposes -which includes eye, hand gesture, and bio-signal inputs [81]. In their setup, pointing towards targets of interest was controlled using gaze, the gestures were used for selection and manipulations and the bio signals controlled the difficulty of the game. Pai et al. [190] combined eye gaze and contractions of arm muscles measured by an EMG for subtle selection and interaction. Novak et al. integrated dwell time and intentional movement for VR-based patient rehabilitation [181]. The system finds the focus of the patient in a VR environment via fixation, and if the patient's intention to move is detected by the rehabilitation robot, the robot will provide sufficient support for the patient.

Other multi-modal interaction approaches include the combination of eye tracking with free-hand 3D gestures [48, 122, 193, 219]. Deng et al. defined the spatial misperception problem that occurs during continuous indirect manipulation with a direct manipulation device [48], and as such is observed when combining gaze and gesture input that leads to manipulation errors and user frustration. The authors introduce three methods, all of which improve the manipulation performance of virtual objects. Pfeuffer et al. introduced the *Gaze + Pinch* interaction technique for virtual reality. Here a user's gaze point is used to indicate the desired object of interaction, whereas pinch gestures are used for its manipulation, as such enabling interaction and manipulation with near and far objects. This technique simultaneously addresses the problem of the virtual hand metaphor that only allows for near interaction, and compared to controller-based methods the user is not required to constantly hold a device.

Eye and Head Rotation. A common approach is to combine eye tracking with head rotation. Techniques have been proposed to allow for hands-free navigation of virtual environments [187, 192, 202, 222, 260]. Findings indicate that navigation techniques benefit from combining eye tracking with head rotation, since its able to correct for common problems related to eye tracking, such as calibration drifts [202]. Sidenmark and Gellersen [222] explored different combinations of eye and head gaze that leverage synergetic movement of eye and gaze for selection and exploration of an environment. It was also found that the combination techniques perform better than the eye only techniques [122, 202]. Piumsomboon et al. proposed three eye-based interaction techniques for navigation and selection in virtual reality [197]. At this, they leveraged specific properties of various eye movements. The Vestibulo-Ocular Reflex (VOR) was for example used for a navigation task, whereas an eye only technique was proposed for selecting targets. These results suggest that different eye-based interaction possibilities should not be used competitively, but that there should be specific interaction possibilities for specific tasks in augmented and virtual environments. Mardanbengi et al. also proposed to use the VOR for improving selection. However, in their work

VOR was explored in the context of 3D gaze estimation in particular in comparison to vergence where their approach using VOR depth estimation showed similar performance in several scenarios despite requiring only one tracked eye [150].

Eye and BCI. Utilizing brain-computer interfaces (BCIs) in a hybrid form (e.g., Eye + BCI) can increase the performance of the whole system [194]. Ma et al. combined a brain-computer interface with eye tracking for typing in virtual reality [147, 269]. A similar setup has been applied in 3D object manipulation [40] and horizontal scrolling and selection interface [156]. Putze et al. [200] combined eye tracking and steady state visually evoked potential to improve the robustness of target selection.

Overall, we observed a wide range of modalities paired with eye-based input spread over various applications, such as increasing accessibility, health care, and entertainment. Some modalities, including traditional input [3, 12, 89, 126, 151, 224], head rotation [150, 187, 192, 197, 202, 222, 260], and gestures [48, 81, 122, 181, 190, 193, 219], were more commonly investigated. This can be explained by the fact that some of these modalities are more well-established (i.e., traditional input), and in some cases, others are already paired in one device or have dedicated resources for pairing, for instance, HMDs with eye trackers like FOVE[1] and HP Omnicept[2] or eye tracking and hand tracking add-ons like Pupil Labs[3] or Leap Motion[4]. Separately, advances in natural language processing, and ubiquity of the speech modality evident from the popularity of digital home assistants, such as Amazon Alexa and Google Home, holds promise for more research on the combination of speech and eye input as we only identified one example in our review [17]. When considering eye gaze for interaction in VR we should not forget the impact of head and torso movement in particular as VR is increasingly moving towards a fully tracked free movement. Sidenmark and Gellersen [221] recently explored the coordination between eye, head, and the torso when looking at targets in VR. Their findings gave insights into the coordination of these body parts and highlighted that when designing gaze-based interfaces these modalities should be considered as a whole and not separately.

3.2 Implicit or Adaptive and Attentive User Interfaces

Apart from XR interfaces that are using eye tracking data for explicit input and selection we identified a second category of XR interfaces that utilizes real-time eye-gaze information. We can summarize this category as adaptive and attentive user interfaces. Adaptive user interfaces are often defined as “an interface that remains well designed even as its world changes” [27]. While initially often used to describe user interfaces that can be adapted explicitly by the user (adaptability) we focus here more on approaches where the user interface is implicitly adapted through the system (adaptivity) [77]. More specifically, in the context of this work the eye and gaze information are used as a context source to control the adaption of the system. Recent works by Grubert et al. [77] highlighted the importance of adaptivity and context-awareness in particular for future AR applications once AR starts to transition from an interface that is sporadically used (e.g., such as an AR app on a mobile phone) to an interface that is continuously used in various contexts (“Pervasive Augmented Reality”). An example of the latter would be HMDs such as the MS HoloLens that are completely designed around the usage of AR as an interface, can serve multiple purposes, and thus can be envisioned to be worn over extended periods and in different contexts. The concept of adaptive user interfaces is related to the concept of attentive user interfaces

[1]https://www.getfove.com/
[2]https://www.hp.com/us-en/vr/reverb-g2-vr-headset-omnicept-edition.html
[3]https://pupil-labs.com/products/vr-ar/
[4]https://www.ultraleap.com/

which could be seen as a subcategory. Attentive user interfaces are defined as interfaces that "are sensitive to the user's attention" [252]. The difference to adaptive and context-aware interfaces is the focus on attention to minimize disruption from the main task and maximize peripheral support. A common example of how eye data can be used here is to adjust the behavior of the interface by processing the user's real-time eye data and predicting the user's focus and interest. There are other definitions of attentive user interfaces [149] with a focus more on implicit user interaction such as non-command user interfaces [176] but we would similarly argue that they are a sub-genre of adaptive user interfaces. We show examples of gaze-adaptive interfaces in Figure 6.

In the following we discuss the main directions of works in this category. We group the identified works by focusing on the context targets (what is adapted) as proposed by Grubert et al. [77] and try to reuse the original categories for different context targets when applicable.

3.2.1 Information Management, Spatial Presentation, and View Management. View management is a term commonly used to describe the issue of where to show user interface elements or digital overlays within an AR interface [73]. In general, view-management techniques that adapt to the context can be classified as techniques that were initially designed for desktop and handheld interfaces systems but could be applied within VR or AR as well as techniques that were designed specifically for head-mounted displays implementing a VR or AR interface. While some prior works used saliency information to estimate the user's gaze and important scene features worth preserving [73], tracking the human gaze in real-time can also help identify areas where to show or not show digital overlays. A simple example is the work by Scholte et al. [217] who modified the location where important information appears to improve view management for car heads-up displays. In particular, they showed warning information within the direction of the user's gaze to reduce reaction times.

Many concepts of view management have been previously explored on desktop and mobile devices. With the increasing interest in creating virtual [78] or augmented desktop environments [206], similar modification techniques could find application in XR as well. An early concept of an attentive interface for desktop machines is EyeWindows [61]. EyeWindows enlarged windows the user focused at to address clutter when users had multiple windows open at the same time. Enlarging the window currently in the user's focus and shrinking other windows accordingly helped users to more quickly acquire and transcribe information from them. Identifying user activities in the targeted window can be applied for automatic content management. Kumar et al. [119] introduced an automatic scrolling interface for users reading a website or an email. Whenever the user's gaze goes beyond a predefined threshold the content is scrolled automatically with an ever increasing speed as the user's gaze comes closer to the edge of the screen, thus keeping the gaze close to the center of the screen and eliminating the need for continuous scrolling via gestures or peripheral devices. Toyama et al. [249] applied this concept to view management on an OST-HMD. Whenever the system detected the user's gaze on the virtual text it would either highlight where the user stopped reading, or automatically scroll the text if the user is reading. The system could also detect when the user did not check important information for a long time and highlight it through a time-dependent urgency indicated, e.g., an outline [183].

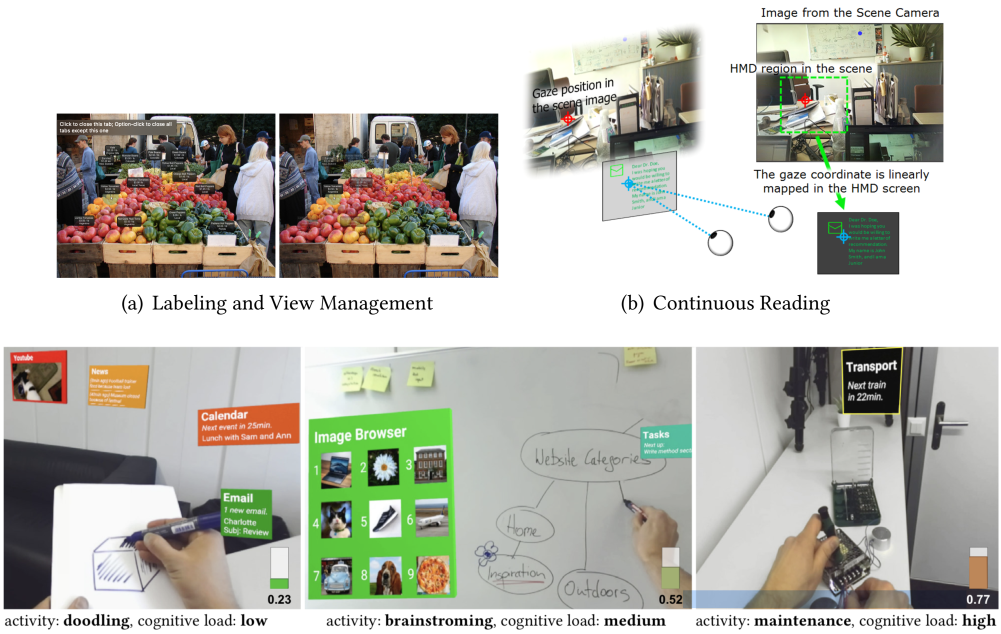
Contrary to traditional displays that present users only with 2D information, virtual content in HMDs is commonly viewed in 3D. Especially when multiple layers of virtual content are shown to the user or overlaid over the scene, it is important to provide a natural interface for switching between the different content planes. The user's focus distance within the 3D space has been envisioned as a natural cue to distinguish what content the user is currently focused on. A common approach is to blend out content that is not in focus [131, 191, 248, 249]. As estimating the user's focus depth is prone to errors [150], common user behavior such as squinting when focusing on an object far away [191], the VOR [249], and other aspects of the scene can help disambiguate the focused object. Saraiji et al. [214] analyzed the saliency of multiple overlapping views shown in VR to determine the most likely layer in focus at the user's gaze location. They blurred out other layers creating an artificial depth-of-field effect.

Another difference to traditional 2D interfaces is that the overlaid content overlays the scene. This means that users may be presented with too much information (information overload) or lose the context due to information being overlaid onto it. Nakao et al. [174] investigated different text visualization techniques for AR HMDs that considered the environment. They initially measured the required attention for a given set of predefined environments and tasks (e.g., for walking stairs) and showed that it is hard to keep the attention on the HMD when doing certain tasks. They then proposed different visualization methods that required less attention but are only briefly evaluated. McNamara et al. [159–161] adjusted the visibility of labels dependent on their proximity to the user's gaze. They suggested a distance-based dimming function that dims labels that are too far from the user's view as well as a time-based dimming approach where labels disappear shortly after the user's gaze moved away. However, they evaluated this approach only in a preliminary study on a desktop and a tablet device. Gebhardt et al. [64] suggested that instead of presenting all additional information in a scene it should be added only when a user's gaze pattern indicates their interest in said object. Although this reduces clutter in the scene, it does not prevent virtual content, e.g., labels, from occluding relevant real content. Tönis and Klinker [245, 246] addressed this by attaching the virtual content to the user's gaze, so it is presented close to but does not overlap with the user's focus. When the user's gaze moves towards the attached information the system registers that the user intends to interact with the virtual information. If, however the user moves the gaze quickly somewhere else the virtual information detaches and moves back to its original location. They found that participants preferred more stabilized virtual content that exhibited less movement.

Finally, very recent works created a model for interactively placing virtual information based on the users cognitive load (measured using eye data), their task, and their environment [139]. The model interactively controls what type of information is shown, the placement of the information, and the amount of information displayed. As such it emphasizes the content aware view and information management also requested by the initial concept of Pervasive Augmented Reality [77]. In summary, adapting the XR view is an important topic when considering long-term usage of XR interfaces and here in particular wearable AR interfaces. If AR glasses become omnipresent and the next mobile phone, then they have to adjust based on the users context. Despite this observation it is obvious that we are only at the beginning as the models for computing the context information using gaze are still often basic and the actual effectiveness of adapting the interface are yet to be explored.

3.2.2 Information Management and Visual Presentation. Eye-tracking is often associated with the user's attention and comprehension. Detecting objects users are interested in, can be applied to present additional relevant information. Toyama et al. [248] applied this principle to present related information about content users are reading on an OST-HMD by analyzing where the user's gaze is in the text.

Presenting additional information does not have to be limited to 2D information, but can also be applied to different objects in the scene. Ajanki et al. developed an augmented reality platform for accessing abstract information in real-world pervasive computing environments by inferring user's focus of attention through signals such as gaze patterns and speech, for applications such as user guides or meetings [5]. Ivaschenko et al. [97] identified objects in user's focus through eye-tracking to optimize what information to show in an AR supported manufacturing application. Moniri et



(a) Labeling and View Management

(b) Continuous Reading

(c) Activity Recognition

Fig. 6. Common examples of adaptive user interfaces: (a) Managing label placement based on the user's gaze [160]. (b) Adjusting visibility of virtual content based on the user's focus distance [249]. (c) Predicting activities and workload from egocentric views [139].

al. [169] considered the amount of presented information about an object to the user. They suggest utilizing the object position relative to the user's gaze and its distance from the user to determine its visibility. Objects that have low visibility could blink to attract the user's attention. When an object has medium visibility, users see few large words, and when an object is in high visibility (looked at) a lot of information is shown. A similar idea was presented by Gras and Yang [72] who adjusted the visualization within a surgery context based on the user's gaze and the state of surgery instruments to either show no overlay, a partial overlay, or a full overlay to the surgeon. Although they tested their system only on a desktop it can be directly transferred to an HMD.

A similar concept was applied by Giannopoulos et al. [66] for navigating users, whenever they came to a cross-road and were unsure what direction to turn to. Their system tracked the user's gaze direction and vibrated the mobile phone when the user looked in the correct direction to move to. They found that participants preferred using their system compared to a map-based navigation. The user's confidence in navigating an environment [8], performing a medical procedure [72], or training could also be derived from their gaze patterns. The system could then provide assistance only when the user requires it, potentially reducing the mental demand and clutter.

Sometimes, instead of presenting additional information about the scene, it is more important to guide the user's gaze towards an important location. Eaddy et al. [52] aimed to guide user's attention to important locations when viewing a map. By detecting the user's gaze, they provided directions towards locations of interest. While Eaddy et al. [52] actively directed the user's gaze towards a target, in some situations, e.g., art exhibitions, it may be preferable to unobtrusively guide the user's attention towards areas of interest. McNamara et al. [157] investigated the effects of subtle modulation of content brightness on gaze attraction in a search task. They utilized eye-tracking to activate modulation when the user's gaze moved away from the target and to deactivate it when it was in the user's focus. They found that this modulation significantly improved the user's answers. Furthermore, increasing the size of the modulation to be more obvious did not significantly improve the results compared to subtle modulation. They extended their work [158] to also investigate the effects of distractors (modulation of other areas in an image) on the search performance. Their results show that despite the additional distractors, reported results were better than when no modulation was presented. While in their work McNamara et al. focused only on brightness modulation, other modulation, such as blur, zooming, or content movement, can be considered as well. Although the effectiveness of content modulation is an effective guidance on displays and in environments where only a small portion of the user's view is augmented, their effectiveness cannot be guaranteed in more natural environments. Instead of modifying the brightness of the target area for both eyes, Grogorick et al. [75] suggested to increase the brightness for one eye, while reducing it for the other eye. They found that although this method can attract the user's gaze, its effectiveness may depend on the complexity of the environment. Grogorick et al. [76] investigated the effectiveness of different gaze guidance techniques within a $160 \times 90°$ FOV immersive scenario. However, they did not find any of the techniques to be outperforming or to achieve attraction rates of more than 50% within 1 second of the stimulus onset. After modifying some methods to repeatedly activate the stimulus the attraction rates rose to 70%. Furthermore, although 42 out of the 102 participants did not detect any of the modifications, they concluded that no technique was truly imperceptible. We can conclude the review of this research direction stating that real-time gaze analysis has been used for guiding the user or providing additional information. However, similar to view management most of the current approaches are in a very early stage and in particular their effectiveness when used outside the lab is not very well understood.

3.2.3 Rendering. As most XR applications primarily target our visual sense, it is natural to exploit the perceptual limitations of our visual system by adapting the graphics according to the location, orientation of the user's eye and the user's focus. In this section, while not providing a detailed review, we briefly introduce some research directions of eye gaze applications in rendering and HMD design but refer to details to the work by Itoh et al. exploring latest trends and challenges in AR HMD design [96].

Since the beginning of computer graphics computational speed constraints and pixel density enforced limitation on the quality of presentable computer graphics (CG), which led to the concept of foveated rendering [132]. As humans see only a small portion of the scene in focus, about 5 degrees around the center of the gaze, it is sufficient to render only a portion of the CG in full resolution. Foveated rendering is often regarded as a means of achieving wide FOV HMDs without sacrificing the perceived rendering quality [239, 262, 263]. The amount of acceptable foveation hereby depends not only on the selected technique but also the latency of the processing pipeline (eye tracking, rendering, displaying the result). Some results suggest that an overall latency of 50-70ms may be tolerable [7, 143]. Although the rendering problems have been resolved for desktop systems with more and more powerful GPUs and CPUs, it is still a big problem for HMDs that require a high framerate with high resolution and low latency. With the move towards untethered devices that allow users to explore the virtual environment foveated rendering is getting attention as a way to reduce the amount of data that needs to be streamed from a processing computer to the HMD [144].

Further efforts to reduce the computational demand focus on what users can see in an HMD. Due to the design of current HMDs, users will usually not see portions of the display that theoretically can be left black thus reducing the overall computational demand. The invisible areas vary as the user looks at different areas on the screen, thus a gaze aware restriction of the rendering area can significantly reduce the computational demand [199]. Liang et al. adjust the undistortion parameters of 360° views to reduce the amount of distortion around the user's gaze point [135]. However, in their chosen scenarios they could not show a benefit of utilizing their undistortion approach for users.

Another application of eye gaze in virtual reality is the replication of visual cues, such as the depth of field (DoF) [209]. The generation of gaze-based DoF effects has been shown to improve the realism, the fun factor, and the overall user experience of virtual environments [84]. This concept can also be applied to generate CG that replicate other features of our vision, such as achromatic aberrations [39], resulting in more realistic depth and appearance of CG. While most research focused on virtual environments, replicating the DoF in AR is important to create the illusion that the CG are really placed in the real world [210]. Estimating the focus depth can also be applied to correct unintended out of focus rendering of CG due to a fixed focal plane of most OST-HMDs [44, 184].

When users explore the virtual environment in constrained surroundings, redirected walking can direct them away from the edges creating the feeling that users are in a larger room than they actually are. While only slight rotations of the scene can be done while the user observes the scene, saccade contingent updating exploits our blindness to system changes during saccades. This should allow larger scene modifications without users becoming alert to the change in the environment. While this idea has been conceptualized more than 10 years ago [250], recently it received renewed attention [22, 110]. Bolte and Lappe [22] investigated the noticeability of scene transformations during saccade suppression. They tested different transformations with ten participants and found that during saccades rotations of up to 5° and 0.5m were not noticeable, compared to a threshold of only 0.23° and 0.02m during fixations. Marwecki et al. [152] showed that a similar concept can be applied for scene management by modifying elements of the scene whenever the user is focused on a different portion of the environment.

One important consideration of XR experiences is the risk of cybersickness [47], simulator sickness [113], and motion sickness [261]. While sometimes used interchangeably it is important to note that although these share some symptoms, their severity and origin is different [108, 229]. Whilst there are different hypotheses on the origins of cybersickness, such as postural instability theory and sensory conflict theory [127], it is still unclear how to fully mitigate its occurrence. Some works have shown that eye gaze can be used to predict the onset of cybersickness [268]. This information could then be used to adjust the rendered content to reduce the severity of cybersickness [152, 175].

Recently, Liu et al. [141] suggested to identify a comfortable brightness value that balances the visibility of the virtual content and the background by learning user preferences and the corresponding pupil size. They then recover the optimal brightness of the virtual content by measuring the brightness of the scene and the size of the user's pupil.

Finally, eye-gaze has been considered imperative for a variety of recent HMD prototypes and commercial devices [101, 148]. Hereby, the application range of eye tracking can be very vast, ranging from determining what image plane the content should be rendered on to present an improved user experience (MagicLeap One), determining what area user's see to reduce computations and ensure a consistent image ([148]), to physically shifting a high resolution inset based on the user's gaze to reduce computational cost while presenting high resolution graphics in the user's focus (Varjo VR-2 pro).

In summary, we can see that gaze information is increasingly relevant for complex rendering. If we know where the user is looking at we can increase the realism of the rendering by approximating visual cues or adjusting the rendering quality to deliver the highest visual fidelity were human vision

(a) Simulating Believable Eye Gaze Behaviors of Virtual Avatars

(b) Correcting Eye Gaze in Immersive Video Conferencing

(c) Enhancing Gaze Cues with Pointers and Cursors

Fig. 7. Examples of eye gaze research in collaborative environments: (a) Development of eye behavior models for realistic eye gaze during avatar-mediated communication in VR [220]. (b) Correcting gaze directions using eye trackers in immersive video conferencing environments [207]. (c) Enhancing shared gaze cues in collaborative environments with pointers and cursors [180].

requires it. Research is already investigating future rendering algorithms and perceptual display technologies for XR that aim to achieve an experience that is visually almost indistinguishable from the reality [96, 264]. However, achieving this requires often computational expensive algorithms and is easy to see the important role of utilising gaze data to reduce some of this additional computational requirements, in particular for less powerful mobile and wearable devices.

3.3 Collaboration

In this section we discuss research on collaborative real and virtual environments that focused on real-time eye-tracking information. In these types of interfaces, users can communicate with other humans or their computer-graphics representations (“avatars”) or computer-controlled entities (“agents”), while the shared spaces and interlocutors can either be co-located or remote. These environments have in common that they rely on shared social cues for the coordination of human actions with respect to themselves and the environment [41]. The *eye-mind hypothesis* states that the location of one’s gaze directly corresponds to the most immediate thought in one’s mind [70, 105]. Human gaze thus provides important social cues for establishing *common ground* in conversations or spatial interaction [25, 41, 65], and establishing *situational awareness* with respect to the interlocutors and the environment [55, 65], e.g., by creating eye contact, aligning one’s gaze with another’s, or coordinating gaze patterns in multi-party conversations.

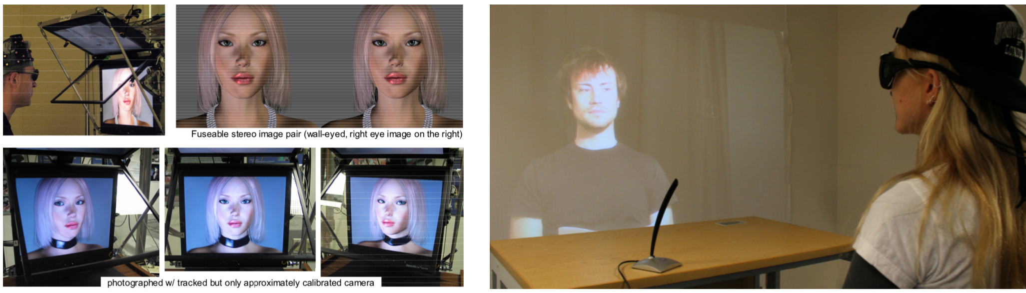
The most impactful previous research on eye tracking in this field focused on four general directions. First, researchers utilized eye trackers to address the inherent challenge in shared VR spaces to communicate a user’s eye movements and attention when embodied in the form of a virtual avatar. Second, researchers in XR leveraged eye trackers to make virtual agents’ gaze react and adapt to the user’s gaze and thus appear more realistic and natural in collaborative environments. Third, researchers worked on sharing tracked eye gaze among a distance between workers and helpers in AR remote collaboration setups. Fourth, researchers introduced augmented gaze cues such as gaze pointers or rays to enhance gaze awareness in shared-space collaboration tasks. In the following, we discuss publications in these four research directions in this category.

3.3.1 Eye Movements in Avatar-Mediated Collaboration. Collaborative virtual environments connect remote or co-located users within a shared virtual space to create a spatial and social context for interpersonal interaction. Users' body is generally tracked and represented as a three-dimensional avatar, allowing them to turn their head and interact with their body, thus providing different non-verbal social cues additionally to speech. However, users' eye gaze was traditionally not captured or represented in the form of avatars' eye movements in such environments.

Vertegaal et al. [253, 254] evaluated the importance of eye gaze and correlations between gaze and attention in multiple highly impactful studies involving virtual avatars and agents. For instance, they showed that gaze is a strong predictor of conversational attention, with a high probability that the person looked at is the person listened to (88%) or spoken to (77%) [254]. They further showed that participants were 22% more likely to speak when an avatar's gaze was synchronized with conversational attention compared to random gaze, but that the amount of gaze is more important than its synchronization [253]. These results highlight the importance of eye gaze in avatar-mediated communication.

In a highly impactful collaborative effort among three universities with their own CAVE systems, Wolff et al. [265] and Steptoe et al. [238] presented one of the first systems in 2008 called *EyeCVE*, which used mobile eye-trackers in three separate CAVEs to map users' gaze to their virtual avatar, thus supporting mutual eye contact and awareness of others' gaze in a shared virtual workspace. Their system was based on head-worn eye trackers mounted on shutter glasses. Informal user trials suggested that such gaze cues support multiparty conversational scenarios [238], even though the system latency was comparatively high [265].

The researchers later investigated different factors within this and extended versions of this system. For instance, they evaluated the importance of realistic deformations of avatars' eyelids, eyebrows, and surrounding areas during eye gaze, showing that the added realism significantly improved users' perceived authenticity but also that the realism made it harder to identify what avatars were looking at, suggesting a trade-off and potential benefits of more abstract representations depending on the task [235, 236]. They showed for a collaborative puzzle-solving task that tracked eye gaze leads to superior performance compared to gaze models that simulate eye movements based on the user's head orientation and the environment [189, 234]. They further compared their system to video conferencing and physical co-location as a baseline and confirmed that its advantages compared to video conferencing mainly lie in the ability to walk around naturally and not be limited by a single camera viewpoint, while also pointing out limitations of the head-worn eye tracker system [207] (see Figure 7b). They showed in an experiment that tracked eye gaze is essential for users to correctly identify what object a user is looking at in an environment [171]. Steptoe et al. [237] later integrated pupil size and blink rate tracking and showed that such cues in avatar-mediated communication resulted in higher lie detection rates than video conferencing (see Figure 8b). Later systems investigated real-time 3D reconstruction of users' body and gaze from multiple live video streams, highlighting the difficulties in reproducing viable eye movements [203, 211], in particular when wearing shutter glasses for stereoscopic displays [59]. Moreover, researchers investigated related effects, such as Borland et al. [23], who showed that accurate eye movements are important to improve self-identification with one's virtual avatar, e.g., when one sees it in a mirror, and related body-ownership illusions. Recently, security and privacy of eye tracking information has gained a lot of attention [24, 34, 87, 103]. John et al. [102] evaluated how blurring of the captured eye images to improve the security of the iris biometrics affects the perception of the avatar's gaze direction. They found that applying a blur of up to $\sigma = 3.5$ did not noticeably affect the perceived movement of the avatar's gaze while improving the security aspect. In summary, while a significant effort has been undertaken to support eye gaze in avatar-mediated



(a) Eye Contact for Intelligent Virtual Agents (b) Eye Pupil Size and Blink Rate

Fig. 8. Examples of eye gaze research focusing on believable gaze behaviors for virtual agents and avatars: (a) Simulating believable eye contact of interactive virtual characters with real users [230]. (b) Integrating pupil size and blink rate tracking, e.g., showing that such gaze cues can result in higher lie detection rates than video conferencing [237].

communication with a wide range of display technologies, more research is needed to advance these solutions beyond prototypical states.

3.3.2 Eye Behavior in Human-Agent Collaboration. A large body of literature focused on the development of algorithmic gaze behavior models for intelligent virtual agents to make them appear more realistic and elicit more natural responses in human users during human-agent collaboration [4, 30, 42, 46, 69, 109, 130, 188, 220, 255, 273] (see Figure 7). While traditional models were limited in the sense that they did not react to users' gaze, newer models can incorporate eye trackers to create more natural *bidirectional* gaze behavior for agents taking into account the user's gaze.

For example, Bee et al. [18] developed a model for natural eye behavior for virtual agents during face-to-face conversations with a real user. They instrumented the user with an eye tracker and used a dynamic behavioral model to improve the agent's reactions, e.g., by making the agent avert their gaze when the user stared at them. They further designed an eye behavior model for an interactive storytelling application in which they used an eye tracker to characterize when the user looked into the eyes of a female virtual agent, impersonating her lover [19]. State [230] presented a behavioral model for believable eye contact between humans and virtual agents, e.g., determining whether the agent's eyes should converge on the user's left or right pupil (see Figure 8a). Vertegaal et al. [254] presented the *FRED* system, which uses a behavioral gaze model to react to users or agents looking at them, making them listen or talk to the person in line with the conversational flow. Morency et al. [170] presented an approach using eye trackers to generate realistic conversational behaviors for agents with *backchannel feedback* based on nodding when the user is talking. Andrist et al. [9] introduced a sophisticated bidirectional gaze model, in which an agent provided gaze cues in a sandwich-making task but also elicited and responded to the user's tracked eye gaze, e.g., by creating eye contact. Kim et al. [114] further looked at eye behaviors that indicate whether users or agents initiate or respond to joint attention cues. Eichner et al. [54] described a system in which users were equipped with an eye tracker to determine their attention and interests when watching a virtual presentation given by an agent. They found that agents were judged as more realistic and responsive if they tuned the presentation to the user's gaze. Keh et al. [107] developed a behavioral gaze model to improve the effectiveness of sports training with virtual opponents, using gaze to present controlled cues about their intentions. Khokhar et al. [112] conceptualized that a teaching avatar could determine if a student follows the lesson from their gaze and adjust their behavior accordingly.

Caruana et al. [31] investigated the intention monitoring processes involved in differentiating communicative and non-communicative gaze shifts during a search task and found that communicative gaze shifts have an important measurable influence on subsequent joint attention behavior between humans and virtual agents. Krum et al. [117] further applied the approaches to a system involving head-mounted projectors to effectively reduce the "Mona Lisa Effect" that arises when a projected virtual agent appears to simultaneously gaze at all observers in the room regardless of their location.

Similar related research focused on collaboration between humans and robotic agents. For instance, Sidner et al. [223] proposed a behavioral model for a social robot agents that could track a user's face and adjust its gaze accordingly, and a human-subject study showed that users established mutual gaze with the robot. Chadalavada et al. [33] investigated how users react to different navigation cues projected by a robot and what their gaze can tell about their intended movement direction. Other work focused on robots with gaze behavior models for establishing joint attention, regulating turn-taking, and disambiguating speakers [162, 225].

In summary, gaze behavioral models for intelligent agents have advanced considerably over the last two decades, resulting in a range of sophisticated solutions for selected collaborative contexts.

3.3.3 Shared Gaze in Task Space Remote Collaboration. While most research on teleconferencing focuses on face-to-face collaboration, a distinct research direction aims to develop systems that help a user perform tasks in the real world with the aid of one or multiple remote collaborators, also called *asymmetric* collaboration [196, 198, 216, 259]. One of the first systems in this field was *SharedView*, in which a camera was mounted on the worker's head, which was then shared with a remote helper who viewed it on a computer screen [90, 120]. A helper can then in turn provide cues back to the worker, e.g., verbally or visually via a head-mounted display (HMD), helping them complete the task. Such remote collaboration systems have different limitations, in particular related to the shared view, which alone is not sufficient to inform the remote helper and/or worker about what the other is attending to or looking at.

To address this limitation, different systems and techniques have been presented [14]. For instance, Fussell et al. [62] introduced an early system in which they used a head-worn eye tracker such that the worker's eye gaze was shared in the form of a pointer in the camera view provided to the remote helper. A study showed mixed results without a clear benefit of eye tracking, which might be because they did not use an HMD for visual stimulus presentation to the worker in their early system. In later work, Ou et al. [185] showed that the worker's focus of attention can be inferred from the shared gaze points, suggesting advantages of eye tracking for such setups over speech-only communication. In 2016, Gutpa et al. [20, 79] and Masai et al. [153] presented one of the first fully-integrated systems in which a user was equipped with a head-mounted display, camera, and eye tracker while a remote helper could see the user's view and gaze points on a computer screen. Using this system, they showed for a 3D LEGO construction task that the eye tracker significantly improved the users' sense of co-presence and performance [79]. In their work, the remote helper used a mouse cursor to annotate the shared view for the worker. Chetwood et al. [37] turned this around and shared the remote helper's gaze with the worker in a DaVinci surgery system, which significantly reduced errors. Wang et al [259] compared a head and gaze pointer for remote assistance in an assembly task and found that head gaze was more stable resulting in better performance. Later work by [128, 196] realized a *bidirectional* shared gaze interface where both the worker and remote helper could see each other's gaze points on the shared view. They showed that this mutually shared gaze significantly improved collaboration and communication. In summary, the design space of asymmetric collaboration interfaces continues to be mapped out by different research groups, focusing in particular on different shared gaze cues and their directionality.

3.3.4 Augmented Gaze Cues in Shared Space Collaboration. Humans are generally capable of inferring visual attention from the direction in which another human's eyes are pointing, which is important for collaborative tasks when establishing common ground or situational awareness [25, 41, 55, 65]. However, different factors can reduce the effectiveness of such natural human gaze cues such as wearing glasses, turning away from the observer, occlusion with scene objects, or the presence of distractors like other humans. Researchers thus tried to augment the natural human gaze cues using artificial visual gaze information [67].

For instance, Vertegaal introduced a *gaze pointer* in the *GAZE Groupware System* by drawing a circle around the target where users in a shared virtual environment were looking at, and discussed its benefits in establishing who is talking about what in cooperative work [251]. This target circle indicated the *point of regard* similar to a laser pointer used in presentations. In a related approach, Duchowski et al. [50] introduced a colored "lightspot" as a *visual deictic reference* in collaborative spaces, indicating the point the user is looking at. They compared eye-slaved and head-slaved lightspots that illuminate the target in the direction their eyes or head are facing, respectively, and found that eye-slaved lightspots help disambiguate the deictic point of reference. Similar findings have also been made by Špakov et al. [257]. Luxenburger et al. [146] further communicated the person's visual field via colored elliptic shapes. Piumsomboon et al. [195] presented the user's total visual field as a frustum as well as the gaze direction as a ray. Sadasivan et al. [212] combined gaze rays with a colored target dot in a collaborative training environment. In later research [163], they extended the system with a *decaying trace* stimulus, which provided a brief positional history of the sequence of target dots that faded out over 200ms. They further introduced a semi-transparent *cone-shaped ray*, which extended the gaze ray by communicating the direction of the ray from the user's head to the target. They compared the stimuli and found that the decaying trace performed best for a collaborative inspection and search task, compared to a single target dot or ray. Rahman et al. [204] suggested different cues, such as trails, arrows, and highlights, to communicate a learner's gaze to a supervisor.

While previous research mainly focused on virtual environments, Norouzi and Erickson et al. [56, 57, 180] evaluated the effectiveness of sharing *gaze rays* between two interlocutors in an AR environment (see Figure 7c). Their task consisted of identifying a target among a crowd of people based on another person's gaze rays. They simulated different limitations of AR shared gaze setups including factors related to the eye tracker (accuracy and precision) and the network (latency and frame drops), and they identified subjective and objective thresholds for acceptable performance.

Hosobori and Kakehi [88] investigated non-visual gaze cues to augment shared space collaboration. They introduced a technique called *Eyefeel*, which converts and delivers the gaze of another person as tactile information, and *EyeChime*, which converts events such as gazing at another person or eye contact to sound.

In summary, augmenting shared gaze cues has shown promise for enhancing collaboration in different application contexts, but more research is needed to explore and evaluate the approaches.

4 RECENT AND FUTURE DIRECTIONS

In this section we extrapolate the insights about previous research trends and directions in the XR field to the future. Recent advances in gaze input and user interfaces are largely fueled by continuing improvements of the base technologies related to eye trackers, gaze estimation algorithms, and their display integration. We expect these improvements to continue over the next decade, resulting in eye tracking becoming available to the broader research community and ubiquitous in the head-worn display market. We are already seeing some new hardware approaches that have a lot of potential to mitigate the issues that held gaze input back in the past, such as the low angular accuracy. Examples are the infrared mirrors employed in the VIVE Pro Eye, and sensor fusion with camera-based eye tracking and electrooculography (EOG) [22, 49] as well as other non-infrared based eye tracking technologies [215]. In the following, we discuss some of the more prominent trends and directions for gaze input in the field of XR.

4.1 Explicit Eye Input

The majority of the publications in the area of explicit eye input focused on studying various approaches to facilitate *general interaction* in XR, more specifically selection, manipulation, and navigation tasks. Other fields of applications that were studied are accessibility, daily tasks and entertainment, healthcare, telepresence, and military fields. This trend is understandable, considering that the general interaction techniques can be re-purposed to support various specific applications. We assume that this trend will be further integrated into XR experiences in the future, expanding the interaction space, for example by reaching out to distant objects that are not accessible using traditional interaction methods. Gaze input was shown to provide a suitable interaction technique especially for the people with disabilities, where it can serve as a substitute for traditional hand-based interaction techniques [80, 138, 147, 154, 266]. This community can especially benefit from systems and eye trackers getting more easily available. Application areas are diverse and include navigation, control of extra limbs, or simply enabling access to general interfaces by providing gaze-based interactions at a larger scale.

A limitation that pervasively exists in the reviewed literature is the lack of a baseline for evaluation. As introduced in Sect. 3.1, various gaze-based targeting techniques have been developed, but the evaluation of them is conducted in different settings, including the evaluation tasks, subjective and objective metrics. The lack of common ground in the evaluation leads to diversified understanding within the community. For instance, Blattgerste et al. found that gaze-based interaction is more accurate than head-based interaction [21], but Kyto et al. came up with the opposite conclusion [122]. Overall, a similar pattern is observed with many of the eye-only interaction comparative studies on a number of factors, such as interaction speed and accuracy [38, 58, 100, 167, 201]. It is a definite future need to develop a set of common tasks and evaluation metrics, to compare the performance of different interaction methods, and to ensure the repeatability of evaluation results. Such efforts can help clarify questions such as, in what cases do we really need eye tracking, and when can we substitute/approximate it with head direction? Separately, more standardized evaluation methods can shed light on the contribution of each modality to user's performance and comfort when multi-modal approaches are utilized. Therefore, researchers and developers can pick from a menu of modalities based on the needs of their application and their target population.

As discussed in Section 3.1, dwell time has been a popular approach for target selection and manipulation in many eye tracking applications both for 2D and 3D interaction spaces [68, 74, 138, 178, 205, 256]. Although popular, this approach cannot entirely resolve the Midas Touch problem and is not the most efficient. We noticed the development of novel approaches such as half-blink detection and gaze gestures, aimed at resolving the slow interaction times and potential incorrect selections [63, 85, 111, 129, 247, 266]. Still, further research is required to understand the performance benefits of these novel approaches in comparison with each other and the type of tasks that are better facilitated by these methods. Also, we identified opportunities for further research in understanding the performance and usability benefits of these novel approaches compared with current multi-modal techniques and the impact of user profile and task type for utilizing either multi-modal approaches or dwell time alternative methods.

Last, we could not identify any longitudinal investigations on the usability of eye-based interactions and their long-term effects on users' behaviors and preferences based on the papers reviewed under *explicit eye input* in Section 3.1. Due to the limited availability of mixed reality systems equipped with eye trackers in the past, long-term studies were very difficult to conduct. One challenge that we foresee for future research is the scalability of gaze-based interaction techniques. For now, studies are usually conducted in very limited, mostly laboratory, settings. It is unsolved how gaze interaction techniques perform in less-restricted circumstances. Also, similar to traditional interaction methods, gaze-based interaction methods produce fatigue, which is minimally considered in the literature we reviewed (e.g., see [21, 38, 201]). Another challenge that has to be addressed is visual discomfort that is produced by mixed reality glasses. It has to be discussed how the use of these devices influences visual comfort and well-being. One prominent example is the vergence-accommodation conflict. Future projects should investigate how the decoupling of vergence and accommodation responses influences our visual system and what solutions there are on the interaction side, besides technical ones.

4.2 Implicit or Adaptive and Attentive User Interfaces

We support the idea that for a continuous use of an XR interface it has to adapt to the user's context and we think that human gaze can play an important role. However, from the literature we see that we are only at the beginning.

As such our review identified a lack of research on adaptive and implicit interfaces compared to interfaces that utilize gaze for explicit interaction. One can also argue that current approaches are relatively simple demonstrating the early stage of this research direction. This is because 1) the models used for computing the context based on gaze are simple and 2) because the chosen context targets are only a subset of possible targets for adaption. E.g., the work by Lindlbauer et al. [139] is important as it makes first steps but only considers cognitive load as context source and information placement as context target. It is easy to see that a continuously used XR interface might consider other sources and targets.

Going further we expect to see more works targeting XR interfaces by using more complex models for context recognition. There are different approaches that explore activity recognition (e.g. reading) based on gaze data and electrooculography (e.g., [29, 91]). Similarly some works explored the identification of the onset of cybersickness [268] from changes in the user's gaze patterns and approximating the mental state of the user has been suggested as the ultimate goal of several works that focus on explorations of eye gaze behavior [2, 32, 36, 83, 83, 86, 274]. Although those ideas are appealing so far they have not been explored for adapting an interface. It remains to be seen how this can be accomplished for XR and how well it is received by end-users but we definitely see this as a trend in adaptive XR interfaces.

We also realised that so far many works adapt existing solutions from 2D interfaces to XR [119, 183, 249] but do not fully reflect on the 3D nature of most XR interfaces. We expect that these adaptations will continue but need to consider how the 2D modes can be expanded to the 3D environment in XR. Thus far we have seen few methods that go beyond blending in and out of different layers based on the user's current focus distance [131, 191, 248, 249]. We have also identified different approaches to manage the presented information in consideration of the background environment, however, these were only focused on 2D labels rather than more complex 3D objects commonly found in XR [159–161, 174]. With virtually unlimited space to place content in the user's surroundings content management information overload and clutter become a significant concern [60, 64, 244]. We expect that techniques that modify the arrangement, placement, and visibility of virtual content will gain importance. We have also observed increased interest in gaze guidance in XR environments [52, 75, 76, 157, 158], which is especially of interest for the emerging

XR entertainment industry. We expect interest in this area to continue in the near term, potentially with expansions into environments that adapt to the user's gaze. Here, we expect the research to incorporate findings from the collaborative work and interactions with virtual avatars covered in the next section.

All this also requires more studies that are carried out over an extended period. So far we identified this as a common limitation of ideas explored in the reviewed papers. Many works focused on presenting a prototype system to showcase the underlying idea, without thorough evaluation or even missing evaluations with actual users. We also found that very few papers compared the developed prototypes with other interaction methods and scenarios.

4.3 Collaboration

In the field of *avatar-mediated collaboration*, we are seeing an increasing trend that XR developer communities make use of the eye trackers integrated into XR HMDs such as the VIVE Pro Eye, FOVE, or HP Omnicept, and add-ons from Pupil Labs, SMI, and Tobii. Based on the body of literature discussed in Section 3.3 that showed clear benefits of tracked self-avatar eye movements for virtual collaboration (e.g., [171, 189, 234, 253, 254]), we expect this to become standard for social multi-user XR platforms in the near future. We expect the related practical challenges with respect to eye models for rigged avatar characters to be largely resolved over the next years. In the mid-term, once eye tracked self-avatars become more common, we predict that more research will focus on documenting the occurrences of social miscommunication and its causes in collaborative virtual environments due to gaze-related latency. We believe that this will be accompanied by more system/algorithm-oriented research focusing on means to reduce gaze latency in XR, such as eye trackers with higher frame rates and eye motion prediction algorithms to reduce the effects of network latency. Moreover, we see more and more research focusing on the subtle information conveyed by the eyes in conjunction with the surrounding facial muscles, such as discussed by Masai et al. [153] in their "Empathy Glasses" prototype, and recently integrated in the commercial HP Omnicept HMD, which tracks the user's facial muscles together with gaze directions and pupillometry. In the long term, we see some very interesting research becoming possible when macro- and micro-expressions can be tracked in real time, represented and rendered in real time, manipulated in real time, and effectively leveraged and employed during face-to-face conversations in XR in the future.

For *human-agent collaboration*, we predict continued efforts towards realistic eye behaviors of virtual agents in contexts such as education [112], training [107], and entertainment [19]. We expect that one of the major fueling factors will be the increased availability and use of eye trackers throughout our society, which will provide the opportunity to collect larger annotated data sets of natural eye movements that can then be leveraged to develop effective machine learning solutions for this classical challenge [115].

In the direction of *shared gaze and augmented gaze cues*, either in remote collaboration or with co-located users, we predict an increasing amount of research interest. In the near term, we see some natural extensions of current research trends that become possible due to improved AR scene understanding [1], e.g., allowing visual deictic references to be extended with knowledge about automatically classified scene elements and related semantics. Such extended approaches could go beyond communicating a point in space ("Look there!") to a richer and more nuanced non-verbal human gaze expression and communication, including emotional influences and gaze-directed attentional cueing (e.g., "I like that!"). Also, new SLAM-based scene mapping methods in AR could improve the performance of shared gaze cues (such as rays or points) to that previously shown in VR, including natural occlusion from a user's point of view, which so far has not been possible and resulted in lower performance of such cues in AR compared to VR [56]. We also see some interesting extensions in the use of multimodal and non-visual cues in shared gaze environments, with initial work by Hosobori and Kakehi [88]. Last but not least, we also believe that these approaches could be extended to a more general theory of interpersonal attention and emotional processing, with implications for understanding how social referencing is impaired in autism and other disorders of social cognition [71], as well as an improved cross-culture understanding of gaze behavior [116]. Related methods could potentially compensate for such effects using AR enhanced/translated cues. For instance, AR cues could support persons on the autism spectrum to make eye contact or provide visual or non-visual cues about others' social referencing.

5 CONCLUSION

In this paper we report on our review of gaze-based interfaces in XR environments. We reviewed papers from a wide range of journals and conferences indexed by Scopus, resulting in overall 215 papers from 1985-2020 that utilized eye gaze. We identified three emerging areas that utilise gaze in XR, namely explicit eye input, adaptive and attentive interfaces, and collaboration in XR. Our results show that especially in recent years the number of papers that incorporate eye gaze as some sort of input or system parameter has been significantly increasing, with previous concepts being rediscovered with the improved accessibility to hardware that incorporates eye tracking capabilities. However, while we believe in the potential and relevance of the identified areas that emerged, we also showed that each area is probably just in the beginning with explicit gaze input probably best explored. An example is the need for context-aware user interfaces for XR that could utilise gaze information to sense the user context and mental state. While the potential has been recognized, actual works demonstrating the actual use in an XR context are rare. We furthermore found that in many cases eye gaze has been incorporated into prototype systems but identified a significant lack of comparative studies. In some cases, we also found contradicting results without a clear consensus.

As with every work our approach also has some shortcomings. There are the general search terms and database used which still leaves the chance for relevant papers being missed because they do not utilise the wide set of search terms used in our search. Furthermore, considering the large number of papers focusing on eye tracking in XR that appeared in our search, we decided to adjust our review solely to cover gaze-based interactions to allow for a deeper exploration of the topic, which is also aligned with two of the applications of Majaranata and Bulling's eye tracking continuum [149]. However, deeper investigations of other applications identified within the eye tracking continuum (i.e., *Gaze-based user modeling* [104] and *Passive eye monitoring* [149]) and beyond it (e.g., privacy and security [106]) is vital to form a coherent picture of the trajectory of eye tracking research in XR. For instance, the area of privacy and security has attracted a lot of attention with the increased popularity of XR technology in the consumer market that is capable of tracking a wide range of users' behaviors and expressions [13, 186]. One of the implications of collecting and processing this wide range of data, such as gaze data, is the high probability of identifying users without their knowledge and various researchers has been exploring solutions to maintain user's privacy when eye tracking data is involved [6, 24, 34, 87, 103, 136, 137, 140, 165, 166, 208, 232].

Finally, large parts of this work focus on the review and discussion of general directions observed within the field. This is a consequence of the wide utilisation of gaze in XR. We do not focus on a fine grained analysis of trends but rather focused on the overall picture. This however leaves room for future work and here in particular in the field of explicit input using gaze data where we see the potential for a more focused survey or review that also takes a more detailed look at the results from user studies to put them in context.

ACKNOWLEDGMENTS

Dr. Langlotz is partially supported by the Marsden Fund Council from Government funding. Ms. Norouzi and Dr. Bruder are supported in part by the National Science Foundation under Award Number 1564065 (Dr. Ephraim P. Glinert, IIS) and Collaborative Award Numbers 1800961, 1800947, and 1800922 (Dr. Ephraim P. Glinert, IIS), as well as the Office of Naval Research under Award Numbers N00014-17-1-2927 and N00014-21-1-2578 (Dr. Peter Squire, Code 34). Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the supporting institutions.

REFERENCES

[1] S. Aarthi and S. Chitrakala. 2017. Scene understanding - A survey. In *Proceedings of the International Conference on Computer, Communication and Signal Processing*. 1–4.

[2] Hamdi Ben Abdessalem, Maher Chaouachi, Marwa Boukadida, and Claude Frasson. 2019. Toward Real-Time System Adaptation Using Excitement Detection from Eye Tracking. In *Proceedings of the International Conference on Intelligent Tutoring Systems*. 214–223.

[3] Sunggeun Ahn and Geehyuk Lee. 2019. Gaze-Assisted Typing for Smart Glasses. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 857–869.

[4] Tagduda Ait Challal and Ouriel Grynszpan. 2018. What Gaze Tells Us About Personality. In *Proceedings of the International Conference on Human-Agent Interaction*. 129–137.

[5] Antti Ajanki, Mark Billinghurst, Hannes Gamper, Toni Järvenpää, Melih Kandemir, Samuel Kaski, Markus Koskela, Mikko Kurimo, Jorma Laaksonen, Kai Puolamäki, Teemu Ruokolainen, and Timo Tossavainen. 2011. An Augmented Reality Interface to Contextual Information. *Virtual Reality* 15, 2-3 (2011), 161–173.

[6] Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Combining Pairwise Feature Matches from Device Trajectories for Biometric Authentication in Virtual Reality Environments. In *Proceedings of IEEE AIVR*. 9–16.

[7] Rachel Albert, Anjul Patney, David Luebke, and Joohwan Kim. 2017. Latency Requirements for Foveated Rendering in Virtual Reality. *ACM Transactions on Applied Perception* 14, 4, Article 25 (2017), 13 pages.

[8] Rawan Alghofaili, Yasuhito Sawahata, Haikun Huang, Hsueh-Cheng Wang, Takaaki Shiratori, and Lap-Fai Yu. 2019. Lost in Style: Gaze-Driven Adaptive Aid for VR Navigation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 348, 12 pages.

[9] Sean Andrist, Michael Gleicher, and Bilge Mutlu. 2017. Looking Coordinated: Bidirectional Gaze Mechanisms for Collaborative Interaction with Virtual Characters. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2571–2582.

[10] Kikuo Asai, Noritaka Osawa, Hideaki Takahashi, Yuji Y. Sugimoto, Satoshi Yamazaki, Masahiro Samejima, and Taiki Tanimae. 2000. Eye Mark Pointer in Immersive Projection Display. In *Proceedings of the IEEE Virtual Reality*. 125–132.

[11] Pavan Kumar B. N., Adithya Balasubramanyam, Ashok Kumar Patil, Chethana B., and Young Ho Chai. 2020. GazeGuide: An Eye-Gaze-Guided Active Immersive UAV Camera. *Applied Sciences* 10, 5, Article 1668 (2020), 18 pages.

[12] Mihai Bâce, Teemu Leppänen, David Gil De Gomez, and Argenis Ramirez Gomez. 2016. UbiGaze: Ubiquitous Augmented Reality Messaging Using Gaze Gestures. In *Proceedings of the SIGGRAPH Asia Mobile Graphics and Interactive Applications*. Article 11, 5 pages.

[13] Jeremy Bailenson. 2018. Protecting Nonverbal Data Tracked in Virtual Reality. *JAMA Pediatrics* 172, 10 (2018), 905–906.

[14] István Barakonyi, Helmut Prendinger, Dieter Schmalstieg, and Mitsuru Ishizuka. 2007. Cascading Hand and Eye Movement for Augmented Reality Videoconferencing. In *Proceedings of IEEE 3DUI*. 71–78.

[15] Florin Bărbuceanu, Mihai Duguleană, Stoianovici Vlad, and Adrian Nedelcu. 2011. Evaluation of the Average Selection Speed Ratio between an Eye Tracking and a Head Tracking Interaction Interface. In *Proceedings of the Doctoral Conference on Computing, Electrical and Industrial Systems*. 181–186.

[16] Richard Bates and Howell Istance. 2005. Towards Eye Based Virtual Environment Interaction for Users With High-Level Motor Disabilities. *International Journal on Disability and Human Development* 4, 3 (2005), 217–224.

[17] Glenn Beach, Charles J. Cohen, Jeff Braun, and Gary Moody. 1998. Eye Tracker System for Use With Head Mounted displays. In *Proceedings of the IEEE International Conference on Systems, Man, and Cybernetics*, Vol. 5. 4348–4352.

[18] Nikolaus Bee, Johannes Wagner, Elisabeth André, Fred Charles, David Pizzi, and Marc Cavazza. 2010. Interacting with a Gaze-Aware Virtual Character. In *Proceedings of ICMI*. 71–77.

[19] Nikolaus Bee, Johannes Wagner, Elisabeth André, Thurid Vogt, Fred Charles, David Pizzi, and Marc Cavazza. 2010. Discovering Eye Gaze Behavior during Human-Agent Conversation in an Interactive Storytelling Application. In *Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for*

Multimodal Interaction. Article 9, 8 pages.

[20] Mark Billinghurst, Kunal Gupta, Masai Katsutoshi, Youngho Lee, Gun Lee, Kai Kunze, and Maki Sugimoto. 2017. Is It in Your Eyes? Explorations in Using Gaze Cues for Remote Collaboration. In *Proceedings of Collaboration Meets Interactive Spaces*. 177–199.

[21] Jonas Blattgerste, Patrick Renner, and Thies Pfeiffer. 2018. Advantages of Eye-gaze over Head-gaze-based Selection in Virtual and Augmented Reality Under Varying Field of Views. In *Proceedings of the Workshop on Communication by Gaze Interaction*. Article 1, 9 pages.

[22] Benjamin Bolte and Markus Lappe. 2015. Subliminal Reorientation and Repositioning in Immersive Virtual Environments using Saccadic Suppression. *IEEE TVCG* 21, 4 (2015), 545–552.

[23] David Borland, Tabitha Peck, and Mel Slater. 2013. An Evaluation of Self-Avatar Eye Movement for Virtual Embodiment. *IEEE TVCG* 19, 4 (2013), 591–596.

[24] Efe Bozkir, Ali Burak Ünal, Mete Akgün, Enkelejda Kasneci, and Nico Pfeifer. 2020. Privacy Preserving Gaze Estimation Using Synthetic Images via a Randomized Encoding Based Framework. In *Proceedings of ACM ETRA*. Article 21, 5 pages.

[25] Susan E. Brennan, Joy E. Hanna, Gregory J. Zelinsky, and Kelly J. Savietta. 2012. Eye Gaze Cues for Coordination in Collaborative Tasks. In *Proceedings of the Workshop on Dual Eye Tracking in CSCW*. 8.

[26] Jeffrey Breugelmans, Yingzi Lin, Ronald R. Mourant, and Maura Daly Iversen. 2010. Biosensor-Based Video Game Control for Physically Disabled Gamers. 54, 28 (2010), 2383–2387.

[27] Dermot Browne, Peter Totterdell, and Mike Norman (Eds.). 1990. *Adaptive User Interfaces*. Academic Press Ltd., London.

[28] Martin Buckley, Ravi Vaidyanathan, and Walterio Mayol-Cuevas. 2011. Sensor Suites for Assistive Arm Prosthetics. In *Proceedings of the International Symposium on Computer-Based Medical Systems*. 1–6.

[29] Andreas Bulling, Jamie A Ward, Hans Gellersen, and Gerhard Tröster. 2010. Eye Movement Analysis for Activity Recognition Using Electrooculography. *IEEE TPAMI* 33, 4 (2010), 741–753.

[30] George Caridakis, Stylianos Asteriadis, Kostas Karpouzis, and Stefanos Kollias. 2011. Detecting human behavior emotional cues in natural interaction. In *Proceedings of the International Conference on Digital Signal Processing*. 1–6.

[31] Nathan Caruana, Genevieve McArthur, Alexandra Woolgar, and Jon Brock. 2017. Detecting Communicative Intent in a Computerised Test of Joint Attention. *PeerJ* 5, Article e2899 (2017), 16 pages.

[32] Berk Cebeci, Ufuk Celikcan, and Tolga K Capin. 2019. A Comprehensive Study of the Affective and Physiological Responses Induced by Dynamic Virtual Reality Environments. *Computer Animation and Virtual Worlds* 30, 3-4, Article e1893 (2019), 12 pages.

[33] Ravi Teja Chadalavada, Henrik Andreasson, Maike Schindler, Rainer Palm, and Achim J Lilienthal. 2020. Bi-Directional Navigation Intent Communication Using Spatial Augmented Reality and Eye-Tracking Glasses for Improved Safety in Human–Robot Interaction. *Robotics and Computer-Integrated Manufacturing* 61, Article 101830 (2020), 15 pages.

[34] Aayush Kumar Chaudhary and Jeff B. Pelz. 2020. Privacy-Preserving Eye Videos using Rubber Sheet Model. In *Proceedings of ACM ETRA*. Article 22, 5 pages.

[35] Hao Chen, Arindam Dey, Mark Billinghurst, and Rob Lindeman. 2017. Exploring Pupil Dilation in Emotional Virtual Reality Environments. In *Proceedings of theInternational Conference on Artificial Reality and Telexistence and the Eurographics Symposium on Virtual Environments*. 169–176.

[36] Lu Chen, Tom Gedeon, Md Zakir Hossain, and Sabrina Caldwell. 2017. Are You Really Angry? Detecting Emotion Veracity as a Proposed Tool for Interaction. In *Proceedings of the Australian Conference on Computer-Human Interaction*. 412–416.

[37] Andrew S. A. Chetwood, Ka Wai Kwok, Loi Wah Sun, George P. Mylonas, James Clark, Ara Darzi, and Guang Zhong Yang. 2012. Collaborative Eye Tracking: A Potential Training Tool in Laparoscopic Surgery. *Surgical Endoscopy* 26, 7 (2012), 2003–2009.

[38] Seung-Hwan Choi, Hyun-Jin Kim, Sang-Woong Hwang, and Jae-Young Lee. 2017. Natural Interaction for Media Consumption in VR Environment. In *SIGGRAPH Asia Posters*. Article 26, 2 pages.

[39] Steven A. Cholewiak, Gordon D. Love, Pratul P. Srinivasan, Ren Ng, and Martin S. Banks. 2017. Chromablur: Rendering Chromatic Eye Aberration Improves Accommodation and Realism. *ACM TOG* 36, 6, Article 210 (2017), 12 pages.

[40] Jinsung Chun, Byeonguk Bae, and Sungho Jo. 2016. BCI Based Hybrid Interface for 3D Object Control in Virtual Reality. In *Proceedings of the International Winter Conference on Brain-Computer Interface*. 1–4.

[41] Herbert H. Clark. 1996. *Using Language*. Cambridge University Press, Cambridge.

[42] R. Alex Colburn, Michael Cohen, and Steven Drucker. 2000. *The Role of Eye Gaze in Avatar Mediated Conversational Interfaces*. Technical Report. Microsoft Research.

[43] Carlo Colombo and Alberto Del Bimbo. 1997. Interacting Through Eyes. *Robotics and Autonomous Systems* 19, 3-4 (1997), 359–368.

[44] Trey Cook, Nate Phillips, Kristen Massey, Alexander Plopski, Christian Sandor, and J Edward Swan. 2018. User Preference for SharpView-Enhanced Virtual Text during Non-Fixated Viewing. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 1–7.

[45] Tim Cottin, Eugen Nordheimer, Achim Wagner, and Essameddin Badreddin. 2016. Gaze-Based Human-SmartHome-Interaction by Augmented Reality Controls. In *International Conference on Robotics in Alpe-Adria Danube Region*. 378–385.

[46] Matthieu Courgeon, Gilles Rautureau, Jean Claude Martin, and Ouriel Grynszpan. 2014. Joint Attention Simulation Using Eye-Tracking and Virtual Humans. *IEEE Transactions on Affective Computing* 5, 3 (2014), 238–250.

[47] Simon Davis, Keith Nesbitt, and Eugene Nalivaiko. 2014. A Systematic Review of Cybersickness. In *Proceedings of the Conference on Interactive Entertainment*. 1–9.

[48] Shujie Deng, Nan Jiang, Jian Chang, Shihui Guo, and Jian J. Zhang. 2017. Understanding the Impact of Multimodal Interaction Using Gaze Informed Mid-air Gesture Control in 3D Virtual Objects Manipulation. *International Journal of Human-Computer Studies* 105 (2017), 68–80.

[49] Murtaza Dhuliawala, Juyoung Lee, Junichi Shimizu, Andreas Bulling, Kai Kunze, Thad Starner, and Woontack Woo. 2016. Smooth Eye Movement Interaction Using EOG Glasses. In *Proceedings of the ACM International Conference on Multimodal Interaction*. 307–311.

[50] Andrew T. Duchowski, Nathan Cournia, Brian Cumming, Daniel Mccallum, and Richard A. Tyrrell. 2004. Visual Deictic Reference in a Collaborative Virtual Environment. In *Proceedings of ETRA*. 35–40.

[51] Andrew T. Duchowski, Nathan Cournia, and Hunter Murphy. 2004. Gaze-Contingent Displays: A Review. *CyberPsychology & Behavior* 7, 6 (2004), 621–634.

[52] Marc Eaddy, Gabor Blasko, Jason Babcock, and Steven Feiner. 2004. My Own Private Kiosk: Privacy-Preserving Public Displays. In *Proceedings of the International Symposium on Wearable Computers*, Vol. 1. 132–135.

[53] Maria K. Eckstein, Belén Guerra-Carrillo, Alison T. Miller Singley, and Silvia A. Bunge. 2017. Beyond Eye Gaze: What else can Eyetracking Reveal About Cognition and Cognitive Development? *Developmental Cognitive Neuroscience* 25 (2017), 69–91.

[54] T. Eichner, H. Prendinger, E. Andre, and M. Ishizuka. 2007. Attentive Presentation Agents. In *Proceedings of the International Workshop on Intelligent Virtual Agents*. 283–295.

[55] Mica R. Endsley. 1995. Toward a Theory of Situation Awareness in Dynamic Systems. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 37, 1 (1995), 32–64.

[56] Austin Erickson, Nahal Norouzi, Kangsoo Kim, Joseph J. LaViola, Gerd Bruder, and Gregory F. Welch. 2020. Effects of Depth Information on Visual Target Identification Task Performance in Shared Gaze Environments. *IEEE TVCG* 26, 5 (2020), 1934–1944.

[57] Austin Erickson, Nahal Norouzi, Kangsoo Kim, Ryan Schubert, Jonathan Jules, Joseph J. LaViola, Gerd Bruder, and Gregory F. Welch. 2020. Sharing Gaze Rays for Visual Target Identification Tasks in Collaborative Augmented Reality. *Journal on Multimodal User Interfaces: Special Issue on Multimodal Interfaces and Communication Cues for Remote Collaboration* 14, 4 (2020), 353–371.

[58] Augusto Esteves, David Verweij, Liza Suraiya, Rasel Islam, Youryang Lee, and Ian Oakley. 2017. SmoothMoves: Smooth Pursuits Head Movements for Augmented Reality. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 167–178.

[59] Allen J. Fairchild, Simon P. Campion, Arturo S. Garcia, Robin Wolff, Terrence Fernando, and David J. Roberts. 2016. A Mixed Reality Telepresence System for Collaborative Space Operation. *IEEE Transactions on Circuits and Systems for Video Technology* 27, 4 (2016), 814–827.

[60] Vicente Ferrer, Yifan Yang, Alex Perdomo, and John Quarles. 2013. Consider your Clutter: Perception of Virtual Object Motion in AR. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 1–6.

[61] David Fono and Roel Vertegaal. 2005. EyeWindows: Evaluation of Eye-Controlled Zooming Windows for Focus Selection. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 151–160.

[62] Susan R. Fussell, Leslie D. Setlock, and Robert E. Kraut. 2003. Effects of Head-mounted and Scene-oriented Video Systems on Remote Collaboration on Physical Tasks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 513–520.

[63] Dekun Gao, Naoaki Itakura, Tota Mizuno, and Kazuyuki Mito. 2013. Improvement of Eye Gesture Interface. *Journal of Advanced Computational Intelligence and Intelligent Informatics* 17, 6 (2013), 843–850.

[64] Christoph Gebhardt, Brian Hecox, Bas van Opheusden, Daniel Wigdor, James Hillis, Otmar Hilliges, and Hrvoje Benko. 2019. Learning Cooperative Personalized Policies from Gaze Data. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 197–208.

[65] Darren Gergle, Robert E. Kraut, and Susan R. Fussell. 2013. Using Visual Information for Grounding and Awareness in Collaborative Tasks. *Human-Computer Interaction* 28, 1 (2013), 1–39.

[66] Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2015. GazeNav: Gaze-Based Pedestrian Navigation. In *Proceedings of the International Conference on Human-Computer Interaction with Mobile Devices and Services*. 337–346.

[67] Ioannis Giannopoulos, Peter Kiefer, and Martin Raubal. 2015. Watch What I Am Looking At! Eye Gaze and Head-Mounted Displays. In *Proceedings of the CHI 2015 Workshop on Mobile Collocated Interactions: From Smartphones to Wearables*. 1–4.

[68] Ioannis Giannopoulos, Johannes Schöning, Antonio Krüger, and Martin Raubal. 2016. Attention as an Input Modality for Post-WIMP Interfaces Using the viGaze Eye Tracking Framework. *Multimedia tools and applications* 75 (2016), 2913–2929.

[69] Marco Gillies and Daniel Ballin. 2004. Affective Interactions Between Expressive Characters. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 2. 1589–1594.

[70] Sean P. Goggins, Matthew Schmidt, Jesus Guajardo, and Joi Moore. 2010. Assessing multiple perspectives in three dimensional virtual worlds: Eye tracking and All Views Qualitative Analysis (AVQA). In *Proceedings of the Annual Hawaii International Conference on System Sciences*. 1–10.

[71] Reiko Graham and Kevin S. LaBar. 2012. Neurocognitive Mechanisms of Gaze-Expression Interactions in Face Processing and Social Attention. *Neuropsychologia* 50, 5 (2012), 553–566.

[72] Gauthier Gras and Guang-Zhong Yang. 2019. Context-Aware Modeling for Augmented Reality Display Behaviour. *IEEE Robotics and Automation Letters* 4, 2 (2019), 562–569.

[73] Raphaël Grasset, Tobias Langlotz, Denis Kalkofen, Markus Tatzgern, and Dieter Schmalstieg. 2012. Image-Driven View Management for Augmented Reality Browsers. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 177–186.

[74] Sven-Thomas Graupner, Michael Heubner, Sebastian Pannasch, and Boris M Velichkovsky. 2008. Evaluating Requirements for Gaze-Based Interaction in a See-through Head Mounted Display. In *Proceedings of ETRA*. 91–94.

[75] Steve Grogorick, Georgia Albuquerque, Jan-Philipp Tauscher, Marc Kassubeck, and Marcus Magnor. 2019. Towards VR Attention Guidance: Environment-Dependent Perceptual Threshold for Stereo Inverse Brightness Modulation. In *Proceedings of the ACM Symposium on Applied Perception*. Article 22, 5 pages.

[76] Steve Grogorick, Georgia Albuquerque, Jan-Philipp Tauscher, and Marcus Magnor. 2018. Comparison of Unobtrusive Visual Guidance Methods in an Immersive Dome Environment. *ACM Transactions on Applied Perception* 15, 4, Article 27 (2018), 11 pages.

[77] Jens Grubert, Tobias Langlotz, Stefanie Zollmann, and Holger Regenbrecht. 2017. Towards Pervasive Augmented Reality: Context-Awareness in Augmented Reality. *IEEE TVCG* 23, 6 (2017), 1706–1724.

[78] Jens Grubert, Eyal Ofek, Michel Pahud, and Per Ola Kristensson. 2018. The Office of the Future: Virtual, Portable, and Global. *IEEE Computer Graphics and Applications* 38, 6 (2018), 125–133.

[79] Kunal Gupta, Gun A. Lee, and Mark Billinghurst. 2016. Do You See What I See? The Effect of Gaze Tracking on Task Space Remote Collaboration. *IEEE Transactions on Visualization and Computer Graphic* 22, 11 (2016), 2413–2422.

[80] John Paulin Hansen, Alexandre Alapetite, Martin Thomsen, Zhongyu Wang, Katsumi Minakata, and Guangtao Zhang. 2018. Head and Gaze Control of a Telepresence Robot with an HMD. In *Proceedings of ACM ETRA*. Article 82, 3 pages.

[81] Hwan Heo, Eui Chul Lee, Kang Ryoung Park, Chi Jung Kim, and Mincheol Whang. 2010. A Realistic Game System Using Multi-Modal User Interfaces. *IEEE Transactions on Consumer Electronics* 56, 3 (2010), 1364–1372.

[82] Katharina Anna Maria Heydn, Marc Philipp Dietrich, Marcus Barkowsky, Götz Winterfeldt, Sebastian von Mammen, and Andreas Nüchter. 2019. The Golden Bullet: A Comparative Study for Target Acquisition, Pointing and Shooting. In *Proceedings of the International Conference on Virtual Worlds and Games for Serious Applications*. 1–8.

[83] Steven Hickson, Nick Dufour, Avneesh Sud, Vivek Kwatra, and Irfan Essa. 2019. Eyemotion: Classifying Facial Expressions in VR Using Eye-Tracking Cameras. In *Proceedings of the IEEE Winter Conference on Applications of Computer Vision*. 1626–1635.

[84] Sébastien Hillaire, Anatole Lécuyer, Rémi Cozot, and Géry Casiez. 2008. Using an Eye-Tracking System to Improve Camera Motions and Depth-of-Field Blur Effects in Virtual Environments. In *Proceedings of the IEEE Virtual Reality Conference*. 47–50.

[85] Yuki Hirata, Hiroki Soma, Munehiro Takimoto, and Yasushi Kambayashi. 2019. Virtual Space Pointing Based on Vergence. In *Proceedings of the International Conference on Human-Computer Interaction*. 259–269.

[86] Christian Hirt, Marcel Eckard, and Andreas Kunz. 2020. Stress Generation and Non-Intrusive Measurement in Virtual Environments Using Eye Tracking. *Journal of Ambient Intelligence and Humanized Computing* 11, 12 (2020), 5977–5989.

[87] Diane Hosfelt and Nicole Shadowen. 2020. Privacy Implications of Eye Tracking in Mixed Reality. (2020). arXiv:2007.10235 https://arxiv.org/abs/2007.10235

[88] Asako Hosobori and Yasuaki Kakehi. 2014. Eyefeel & EyeChime: A Face to Face Communication Environment by Augmenting Eye Gaze Information. In *Proceedings of AH*. Article 7, 4 pages.

[89] Shigeyuki Ishida, Munehiro Takimoto, and Yasushi Kambayashi. 2017. AR Based User Interface for Driving Electric Wheelchairs. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*. 144–154.

[90] Takemochi Ishii, Michitaka Hirose, Hideaki Kuzuoka, T. Takahara, and Takeshi Myoi. 1990. Collaboration System for Manufacturing System in the 21st Century. In *Proceedings of the International Conference on Manufacturing Systems and Environment—Looking Toward the 21st Century*. 295–300.

[91] Shoya Ishimaru, Kai Kunze, Koichi Kise, Jens Weppner, Andreas Dengel, Paul Lukowicz, and Andreas Bulling. 2014. In the Blink of an Eye: Combining Head Motion and Eye Blink Frequency for Activity Recognition with Google Glass. In *Proceedings of AH*. Article 15, 4 pages.

[92] Howell Istance, Richard Bates, Aulikki Hyrskykari, and Stephen Vickers. 2008. Snap Clutch, a Moded Approach to Solving the Midas Touch Problem. In *Proceedings of ETRA*. 221–228.

[93] Howell Istance, Aulikki Hyrskykari, Lauri Immonen, Santtu Mansikkamaa, and Stephen Vickers. 2010. Designing Gaze Gestures for Gaming: An Investigation of Performance. In *Proceedings of the Symposium on Eye-Tracking Research & Applications*. 323–330.

[94] Howell Istance, Aulikki Hyrskykari, Stephen Vickers, and Thiago Chaves. 2009. For Your Eyes Only: Controlling 3D Online Games by Eye-Gaze. In *Proceedings of the IFIP Conference on Human-Computer Interaction*. 314–327.

[95] Howell Istance, Stephen Vickers, and Aulikki Hyrskykari. 2009. Gaze-Based Interaction with Massively Multiplayer on-Line Games. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*. 4381–4386.

[96] Yuta Itoh, Tobias Langlotz, Jonathan Sutton, and Alexander Plopski. 2021. Towards Indistinguishable Augmented Reality: A Survey on Optical See-through Head-Mounted Displays. *Comput. Surveys* 54, 6, Article 120 (2021), 36 pages.

[97] Anton Ivaschenko, Anastasia Khorina, and Pavel Sitnikov. 2018. Accented Visualization by Augmented Reality for Smart Manufacturing Aplications. In *Proceedings of the IEEE Industrial Cyber-Physical Systems*. 519–522.

[98] Robert J. K. Jacob. 1991. The Use of Eye Movements in Human-Computer Interaction Techniques: What You Look at is What You Get. *ACM Transactions on Information Systems* 9, 2 (1991), 152–169.

[99] Robert J. K. Jacob. 1995. *Eye Tracking in Advanced Interface Design*. Oxford University Press, Inc., 258–288.

[100] Shahram Jalaliniya, Diako Mardanbeigi, Thomas Pederson, and Dan Witzner Hansen. 2014. Head and Eye Movement as Pointing Modalities for Eyewear Computers. In *Proceedings of the International Conference on Wearable and Implantable Body Sensor Networks Workshops*. 50–53.

[101] Changwon Jang, Kiseung Bang, Seokil Moon, Jonghyun Kim, Seungjae Lee, and Byoungho Lee. 2017. Retinal 3D: Augmented Reality Near-Eye Display via Pupil-Tracked Light Field Projection on Retina. *ACM TOG* 36, 6, Article 190 (2017), 13 pages.

[102] Brendan John, Sophie Jörg, Sanjeev Koppal, and Eakta Jain. 2020. The Security-Utility Trade-off for Iris Authentication and Eye Animation for Social Virtual Avatars. *IEEE TVCG* 26, 5 (2020), 1880–1890.

[103] Brendan John, Ao Liu, Lirong Xia, Sanjeev Koppal, and Eakta Jain. 2020. Let It Snow: Adding Pixel Noise to Protect the User's Identity. In *Proceedings of the ACM Symposium on Eye Tracking Research and Applications*. Article 43, 3 pages.

[104] Brendan John, Pallavi Raiturkar, Arunava Banerjee, and Eakta Jain. 2018. An Evaluation of Pupillary Light Response Models for 2D Screens and VR HMDs. In *Proceedings of ACM VRST*. Article 19, 11 pages.

[105] Marcel A. Just and Patricia A. Carptenter. 1980. A Theory of Reading: From Eye Fixations to Comprehension. *Psychology Review* 87, 4 (1980), 329–354.

[106] Christina Katsini, Yasmeen Abdrabou, George E. Raptis, Mohamed Khamis, and Florian Alt. 2020. The Role of Eye Gaze in Security and Privacy Applications: Survey and Future HCI Research Directions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 711, 21 pages.

[107] Huan-Chao Keh and Yang Wang. 2008. Using Detected Physiological Traits to Revive Sports Training. *International Journal of Modelling and Simulation* 28, 4 (2008), 430–439.

[108] Robert S. Kennedy, Norman E. Lane, Kevin S. Berbaum, and Michael G. Lilienthal. 1993. Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness. *The International Journal of Aviation Psychology* 3, 3 (1993), 203–220.

[109] Stevanus Kevin, Yun Suen Pai, and Kai Kunze. 2018. Virtual Gaze: Exploring use of Gaze as Rich Interaction Method with Virtual Agent in Interactive Virtual Reality Content. In *Proceedings of ACM VRST*. 130:1–130:2.

[110] Maryam Keyvanara and Robert S Allison. 2018. Sensitivity to natural 3D image transformations during eye movements. In *Proceedings of the 2018 ACM ETRA*. ACM, 64.

[111] Mohamed Khamis, Carl Oechsner, Florian Alt, and Andreas Bulling. 2018. VRpursuits: Interaction in Virtual Reality Using Smooth Pursuit Eye Movements.

[112] Adil Khokhar, Andrew Yoshimura, and Christoph W. Borst. 2019. Pedagogical Agent Responsive to Eye Tracking in Educational VR. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 1018–1019.

[113] Hyunjeong Kim and Ji Hyung Park. 2019. Effects of Simulator Sickness and Emotional Responses When Inter-pupillary Distance Misalignment Occurs. In *Proceedings of IHSI*. 442–447.

[114] Kwanguk Kim and Peter Mundy. 2012. Joint Attention, Social-Cognition, and Recognition Memory in Adults. *Frontiers in Human Neuroscience* 6, Article 172 (2012), 11 pages.

[115] Ahmad F. Klaib, Nawaf O. Alsrehin, Wasen Y. Melhem, Haneen O. Bashtawi, and Aws A. Magableh. 2021. Eye Tracking Algorithms, Techniques, Tools, and Applications with an Emphasis on Machine Learning and Internet of Things Technologies. *Expert Systems with Applications* 166 (2021).

[116] Tomoko Koda, Taku Hirano, and Takuto Ishioh. 2017. Development and Perception Evaluation of Culture-Specific Gaze Behaviors of Virtual Agents. In *Proceedings of the International Conference on Intelligent Virtual Agents*. Springer International Publishing, 213–222.

[117] David M. Krum, Sin-Hwa Kang, Thai Phan, Lauren Cairco Dukes, and Mark Bolas. 2017. Social Impact of Enhanced Gaze Presentation Using Head Mounted Projection. In *Proceedings of the International Conference on Distributed, Ambient, and Pervasive Interactions*. 61–76.

[118] Sofia Ira Ktena, William Abbott, and A Aldo Faisal. 2015. A Virtual Reality Platform for Safe Evaluation and Training of Natural Gaze-Based Wheelchair Driving. In *Proceedings of the International IEEE/EMBS Conference on Neural Engineering*. 236–239.

[119] Manu Kumar, Terry Winograd, Terry Winograd, and Andreas Paepcke. 2007. Gaze-Enhanced Scrolling Techniques. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*. 2531–2536.

[120] Takeshi Kurata, Nobuchika Sakata, Masakatsu Kourogi, Hideaki Kuzuoka, and Mark Billinghurst. 2004. Remote Collaboration Using a Shoulder-Worn Active Camera/Laser. In *Proceedings of the International Symposium on Wearable Computers*. 62–69.

[121] Tiffany C.K. Kwok, Peter Kiefer, Victor R. Schinazi, Benjamin Adams, and Martin Raubal. 2019. Gaze-Guided Narratives: Adapting Audio Guide Content to Gaze in Virtual and Real Environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 491, 12 pages.

[122] Mikko Kytö, Barrett Ens, Thammathip Piumsomboon, Gun A. Lee, and Mark Billinghurst. 2018. Pinpointing: Precise Head- and Eye-Based Target Selection for Augmented Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 81, 14 pages.

[123] Michael F. Land and Sophie Furneaux. 1997. The Knowledge Base of the Oculomotor System. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* 352, 1358 (1997), 1231–1239.

[124] Eike Langbehn, Frank Steinicke, Markus Lappe, Gregory F. Welch, and Gerd Bruder. 2018. In the Blink of an Eye: Leveraging Blink-Induced Suppression for Imperceptible Position and Orientation Redirection in Virtual Reality. *ACM TOG* 37, 4 (2018), 1–11.

[125] Stephen R. H. Langton, Roger J. Watt, and Vicki Bruce. 2000. Do the Eyes Have It? Cues to the Direction of Social Attention. *Trends in Cognitive Sciences* 4, 2 (2000), 50–59.

[126] Michael Lankes and Barbara Stiglbauer. 2016. GazeAR: Mobile Gaze-Based Interaction in the Context of Augmented Reality Games. In *Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. 397–406.

[127] Joseph J. LaViola. 2000. A Discussion of Cybersickness in Virtual Environments. *ACM SIGCHI Bulletin* 32, 1 (2000), 47–56.

[128] Gun Lee, Seungwon Kim, Youngho Lee, Arindam Dey, Thammathip Piumsomboon, Mitchell Norman, and Mark Billinghurst. 2017. Improving Collaboration in Augmented Video Conference Using Mutually Shared Gaze. In *Proceedings of ICAT-EGVE*. 197–204.

[129] Jae-Young Lee, Hyung-Min Park, Seok-Han Lee, Tae-Eun Kim, and Jong-Soo Choi. 2011. Design and Implementation of an Augmented Reality System Using Gaze Interaction. In *Proceedings of the International Conference on Information Science and Applications*. 1–8.

[130] Sooha Park Lee, Jeremy B. Badler, and Norman I. Badler. 2002. Eyes Alive. *ACM TOG* 21, 3 (2002), 637–644.

[131] Youngho Lee, Choonsung Shin, Thammathip Piumsomboon, Gun Lee, and Mark Billinghurst. 2017. Automated Enabling of Head Mounted Display Using Gaze-depth Estimation. In *Proceedings of the SIGGRAPH Asia Mobile Graphics & Interactive Applications*. Article 21, 4 pages.

[132] Marc Levoy and Ross Whitaker. 1990. Gaze-Directed Volume Rendering. In *ACM SIGGRAPH Computer Graphics*, Vol. 24. 217–223.

[133] Michael Li and Ted Selker. 2001. Eye Pattern Analysis in Intelligent Virtual Agents. In *Proceedings of the International Workshop on Intelligent Virtual Agents*. Springer, 23–35.

[134] Songpo Li, Xiaoli Zhang, and Jeremy D Webb. 2017. 3-D-Gaze-Based Robotic Grasping Through Mimicking Human Visuomotor Function for People With Motion Impairments. *IEEE Transactions on Biomedical Engineering* 64, 12 (2017), 2824–2835.

[135] Feng Liang, Stevanus Kevin, Kai Kunze, and Yun Suen Pai. 2019. PanoFlex: Adaptive Panoramic Vision to Accommodate 360° Field-of-View for Humans. In *Proceedings of ACM VRST*. Article 83, 2 pages.

[136] Jonathan Liebers, Mark Abdelaziz, Lukas Mecke, Alia Saad, Jonas Auda, Uwe Grünefeld, Florian Alt, and Stefan Schneegass. 2021. Understanding User Identification in Virtual Reality Through Behavioral Biometrics and the Effect of Body Normalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 517, 11 pages.

[137] Daniel J. Liebling and Sören Preibusch. 2014. Privacy Considerations for a Pervasive Eye Tracking World. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*. 1169–1177.

[138] Chern-Sheng Lin, Kai-Chieh Chang, and Young-Jou Jain. 2002. A New Data Processing and Calibration Method for an Eye-Tracking Device Pronunciation System. *Optics & Laser Technology* 34, 5 (2002), 405–413.

[139] David Lindlbauer, Anna Maria Feit, and Otmar Hilliges. 2019. Context-Aware Online Adaptation of Mixed Reality Interfaces. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 147–160.

[140] Ao Liu, Lirong Xia, Andrew Duchowski, Reynold Bailey, Kenneth Holmqvist, and Eakta Jain. 2019. Differential Privacy for Eye-Tracking Data. In *Proceedings of ACM ETRA*. Article 28, 10 pages.

[141] Chang Liu, Alexander Plopski, Kiyoshi Kiyokawa, Photchara Ratsamee, and Jason Orlosky. 2018. IntelliPupil: Pupillometric Light Modulation for Optical See-Through Head-Mounted Displays. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 98–104.

[142] Chang Liu, Alexander Plopski, and Jason Orlosky. 2020. OrthoGaze: Gaze-based Three-dimensional Object Manipulation using Orthogonal Planes. *Computers & Graphics* 89 (2020), 1–10.

[143] Lester C. Loschky and Gary S. Wolverton. 2007. How Late Can You Update Gaze-Contingent Multiresolutional Displays without Detection? *ACM Transactions on Multimedia Computing, Communications, and Applications* 3, 4, Article 7 (2007), 10 pages.

[144] Pietro Lungaro, Rickard Sjöberg, Alfredo José Fanghella Valero, Ashutosh Mittal, and Konrad Tollmar. 2018. Gaze-Aware Streaming Solutions for the Next Generation of Mobile VR Experiences. *IEEE TVCG* 24, 4 (2018), 1535–1544.

[145] Francisco Lopez Luro and Veronica Sundstedt. 2019. A Comparative Study of Eye Tracking and Hand Controller for Aiming Tasks in Virtual Reality. In *Proceedings of ACM ETRA*. Article 68, 9 pages.

[146] Andreas Luxenburger, Mohammad Mehdi Moniri, Alexander Prange, and Daniel Sonntag. 2016. MedicalVR: Towards Medical Remote Collaboration Using Virtual Reality. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Proceedings*. 321–324.

[147] Xinyao Ma, Zhaolin Yao, Yijun Wang, Weihua Pei, and Hongda Chen. 2018. Combining Brain-Computer Interface and Eye Tracking for High-Speed Text Entry in Virtual Reality. In *Proceedings of the International Conference on Intelligent User Interfaces*. 263–267.

[148] Andrew Maimone, Douglas Lanman, Kishore Rathinavel, Kurtis Keller, David Luebke, and Henry Fuchs. 2014. Pinlight Displays: Wide Field of View Augmented Reality Eyeglasses Using Defocused Point Light Sources. *ACM TOG*, Article 89 (2014), 11 pages.

[149] Päivi Majaranta and Andreas Bulling. 2014. *Eye Tracking and Eye-Based Human–Computer Interaction*. Springer, London, 39–65.

[150] Diako Mardanbegi, Tobias Langlotz, and Hans Gellersen. 2019. Resolving Target Ambiguity in 3D Gaze Interaction Through VOR Depth Estimation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 612, 12 pages.

[151] Diako Mardanbegi, Benedikt Mayer, Ken Pfeuffer, Shahram Jalaliniya, Hans Gellersen, and Alexander Perzl. 2019. EyeSeeThrough: Unifying Tool Selection and Application in Virtual Environments. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 474–483.

[152] Sebastian Marwecki, Andrew D. Wilson, Eyal Ofek, Mar Gonzalez Franco, and Christian Holz. 2019. Mise-Unseen: Using Eye Tracking to Hide Virtual Reality Scene Changes in Plain Sight. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 777–789.

[153] Katsutoshi Masai, Kai Kunze, Maki sugimoto, and Mark Billinghurst. 2016. Empathy Glasses. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*. 1257–1263.

[154] Luca Maule, Alberto Fornaser, Malvina Leuci, Nicola Conci, Mauro Da Lio, and Mariolino De Cecco. 2016. Development of innovative HMI strategies for eye controlled wheelchairs in virtual reality. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. 358–377.

[155] Luca Maule, Alberto Fornaser, Paolo Tomasin, Mattia Tavernini, Gabriele Minotto, Mauro Da Lio, and Mariolino De Cecco. 2017. Augmented Robotics for Electronic Wheelchair to Enhance Mobility in Domestic Environment. In *Proceedings of the International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. 22–32.

[156] Paul McCullagh, Leo Galway, and Gaye Lightbody. 2013. Investigation into a Mixed Hybrid Using SSVEP and Eye Gaze for Optimising User Interaction within a Virtual Environment. In *Proceedings of the International Conference on Universal Access in Human-Computer Interaction*. 530–539.

[157] Ann McNamara, Reynold Bailey, and Cindy Grimm. 2008. Improving Search Task Performance Using Subtle Gaze Direction. In *Proceedings of the Symposium on Applied Perception in Graphics and Visualization*. 51–56.
[158] Ann McNamara, Reynold Bailey, and Cindy Grimm. 2009. Search Task Performance Using Subtle Gaze Direction with the Presence of Distractions. *ACM Transactions on Applied Perception* 6, 3, Article 17 (2009), 19 pages.
[159] Ann McNamara, Katherine Boyd, Joanne George, Weston Jones, Somyung Oh, and Annie Suther. 2019. Information Placement in Virtual Reality. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 1765–1769. https://doi.org/10.1109/VR.2019.8797910
[160] Ann McNamara, Chethna Kabeerdoss, and Conrad Egan. 2015. Mobile User Interfaces based on User Attention. In *Proceedings of the Workshop on Future Mobile User Interfaces*. 1–3.
[161] Ann McNamara, Laura Murphy, and Conrad Egan. 2014. Investigating the Use of Eye-Tracking for View Management. In *Proceedings of the ACM SIGGRAPH Posters*. Article 31, 1 pages.
[162] Gregor Mehlmann, Markus Häring, Kathrin Janowski, Tobias Baur, Patrick Gebhard, and Elisabeth Andre. 2014. Exploring a Model of Gaze for Grounding in Multimodal HRI. In *Proceedings of the ACM International Conference on Multimodal Interaction*. 247–254.
[163] P. Mehta, Sajay Sadasivan, Joel S. Greenstein, Anand K. Gramopadhye, and Andrew T. Duchowski. 2005. Evaluating Different Display Techniques for Communicating Search Strategy Training in a Collaborative Virtual Aircraft Inspection Environment. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*. 2244–2248.
[164] Paul Milgram, Haruo Takemura, Akira Utsumi, and Fumio Kishino. 1995. Augmented Reality: A Class of Displays on the Reality-Virtuality Continuum. In *Proceedings of SPIE, Telemanipulator and Telepresence Technologies*, Vol. 2351. International Society for Optics and Photonics, SPIE, 282 – 292.
[165] Mark Roman Miller, Fernanda Herrera, Hanseul Jun, James A. Landay, and Jeremy N. Bailenson. 2020. Personal Identifiability of User Tracking Data During Observation of 360-degree VR Video. *Scientific Reports* 10, Article 17404 (2020).
[166] Robert Miller, Ashwin Ajit, Natasha Kholgade Banerjee, and Sean Banerjee. 2019. Realtime Behavior-Based Continual Authentication of Users in Virtual Reality Environments. In *Proceedings of IEEE AIVR*. 253–254.
[167] Katsumi Minakata, John Paulin Hansen, I. Scott MacKenzie, Per Bækgaard, and Vijay Rajanna. 2019. Pointing by Gaze, Head, and Foot in a Head-Mounted Display. In *Proceedings of ACM ETRA*. Article 69, 9 pages.
[168] Peter Mohr, Markus Tatzgern, Tobias Langlotz, Andreas Lang, Dieter Schmalstieg, and Denis Kalkofen. 2019. TrackCap: Enabling Smartphones for 3D Interaction on Mobile Head-Mounted Displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. Article 585, 11 pages.
[169] Mohammad Mehdi Moniri, Daniel Sonntag, and Andreas Luxenburger. 2016. Peripheral view calculation in virtual reality applications. In *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 333–336.
[170] Louis-Philippe Morency, Iwan De Kok, and Jonathan Gratch. 2008. Predicting Listener Backchannels: A Probabilistic Multimodal Approach. In *Proceedings of the International Workshop on Intelligent Virtual Agents*. 176–190.
[171] Norman Murray, Dave Roberts, Anthony Steed, Paul Sharkey, Paul Dickerson, John Rae, and Robin Wolff. 2009. Eye Gaze in Virtual Environments: Evaluating the Need and Initial Work on Implementation. *Concurrency Computation Practice and Experience* 21, 11 (2009), 1437–1449.
[172] Lennart E. Nacke, Sophie Stellmach, Dennis Sasse, Jörg Niesenhaus, and Raimund Dachselt. 2011. LAIF: A Logging and Interaction Framework for Gaze-based Interfaces in Virtual Entertainment Environments. *Entertainment Computing* 2, 4 (2011), 265–273.
[173] Yukiko I. Nakano and Ryo Ishii. 2010. Estimating User's Engagement from Eye-Gaze Behaviors in Human-Agent Conversations. In *Proceedings of the International Conference on Intelligent User Interfaces*. 139–148.
[174] Masayuki Nakao, Tsutomu Terada, and Masahiko Tsukamoto. 2014. An Information Presentation Method for Head Mounted Display Considering Surrounding Environments. In *Proceedings of AH*. Article 47, 8 pages.
[175] Guang-Yu Nie, Henry Been-Lirn Duh, Yue Liu, and Yongtian Wang. 2019. Analysis on Mitigation of Visually Induced Motion Sickness by Applying Dynamical Blurring on a User's Retina. *IEEE TVCG* 26, 8 (2019), 2535–2545.
[176] Jakob Nielsen. 1993. Noncommand User Interfaces. *Commun. ACM* 36, 4 (1993), 83–99.
[177] NII. 2018. Augmented Reality in Human-Computer Interaction. Retrieved September 19, 2021 from https://shonan.nii.ac.jp/seminars/135/
[178] Susanna Nilsson, Torbjörn Gustafsson, and Per Carleberg. 2007. Hands Free Interaction with Virtual Information in a Real Environment. *Proceedings of the Workshop on Communication by Gaze Interaction* (2007), 53–57.
[179] Susanna Nilsson, Torbjörn Gustafsson, and Per Carleberg. 2009. Hands Free Interaction with Virtual Information in a Real Environment: Eye Gaze as an Interaction Tool in an Augmented Reality System. *PsychNology Journal* 7, 2 (2009), 175–196.
[180] Nahal Norouzi, Austin Erickson, Kangsoo Kim, Ryan Schubert, Joseph J. LaViola Jr., Gerd Bruder, and Gregory F. Welch. 2019. Effects of Shared Gaze Parameters on Visual Target Identification Task Performance in Augmented

Reality. In *Proceedings of the ACM Symposium on Spatial User Interaction*. 1–11.
[181] Domen Novak and Robert Riener. 2013. Enhancing Patient Freedom in Rehabilitation Robotics Using Gaze-based Intention Detection. In *Proceedings of the IEEE International Conference on Rehabilitation Robotics*. 1–6.
[182] Stephen D. O'Connell, Martin Castor, Jerry Pousette, and Martin Krantz. 2012. Eye Tracking-Based Target Designation in Simulated Close Range Air Combat. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (2012), 46–50.
[183] Jason Orlosky, Chang Liu, Denis Kalkofen, and Kiyoshi Kiyokawa. 2019. Visualization-Guided Attention Direction in Dynamic Control Tasks. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 372–373.
[184] Kohei Oshima, Kenneth R. Moser, Damien Constantine Rompapas, Edward J. Swan II, Sei Ikeda, Goshiro Yamamoto, Takafumi Taketomi, Christian Sandor, and Hirokazu Kato. 2016. SharpView: Improved Clarity of Defocused Content on Optical See-Through Head-Mounted Displays. In *Proceedings of IEEE 3DUI*. 173–181.
[185] Jiazhi Ou, Lui Min Oh, Susan R. Fussell, Tal Blum, and Jie Yang. 2008. Predicting Visual Focus of Attention From Intention in Remote Collaborative Tasks. *IEEE Transactions on Multimedia* 10, 6 (2008), 1034–1045.
[186] Jessica Outlaw and Susan Persky. 2019. Industry Review Boards are Needed to Protect VR User Privacy. In *World Economic Forum*, Vol. 29.
[187] Benjamin I. Outram, Yun Suen Pai, Tanner Person, Kouta Minamizawa, and Kai Kunze. 2018. Anyorbit: Orbital Navigation in Virtual Environments with Eye-tracking. In *Proceedings of ACM ETRA*. Article 45, 5 pages.
[188] Oyewole Oyekoya, Anthony Steed, and Xueni Pan. 2011. Short Paper: Exploring the Object Relevance of a Gaze Animation Model. In *Proceedings of the Eurographics Conference on Virtual Environments & Joint Virtual Reality*. 111–114.
[189] Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2009. A Saliency-Based Method of Simulating Visual Attention in Virtual Scenes. In *Proceedings of ACM VRST*. 199–206.
[190] Yun Suen Pai, Tilman Dingler, and Kai Kunze. 2019. Assessing Hands-Free Interactions for VR Using Eye Gaze and Electromyography. *Virtual Reality* 23, 2 (2019), 119–131.
[191] Yun Suen Pai, Benjamin Outram, Noriyasu Vontin, and Kai Kunze. 2016. Transparent Reality: Using Eye Gaze Focus Depth as Interaction Modality. In *Proceedings of UIST*. 171–172.
[192] Yun Suen Pai, Benjamin I. Outram, Benjamin Tag, Megumi Isogai, Daisuke Ochi, and Kai Kunze. 2017. GazeSphere: Navigating 360-degree-video Environments in VR Using Head Rotation and Eye Gaze. In *ACM SIGGRAPH 2017 Posters (SIGGRAPH '17)*. ACM, New York, NY, USA, Article 23, 2 pages.
[193] Ken Pfeuffer, Benedikt Mayer, Diako Mardanbegi, and Hans Gellersen. 2017. Gaze + Pinch Interaction in Virtual Reality. In *Proceedings of ACM SUI*. 99–108.
[194] Gert Pfurtscheller, Brendan Z. Allison, Günther Bauernfeind, Clemens Brunner, Teodoro Solis Escalante, Reinhold Scherer, Thorsten O. Zander, Gernot Mueller-Putz, Christa Neuper, and Niels Birbaumer. 2010. The Hybrid BCI. *Frontiers in Neuroscience* 4, Article 30 (2010), 11 pages.
[195] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2017. [POSTER] CoVAR: Mixed-Platform Remote Collaborative Augmented and Virtual Realities System with Shared Collaboration Cues. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 218–219.
[196] Thammathip Piumsomboon, Arindam Dey, Barrett Ens, Gun Lee, and Mark Billinghurst. 2019. The Effects of Sharing Awareness Cues in Collaborative Mixed Reality. *Frontiers in Robotics and AI* 6, Article 5 (2019), 18 pages.
[197] Thammathip Piumsomboon, Gun Lee, Robert W. Lindeman, and Mark Billinghurst. 2017. Exploring Natural Eye-Gaze-based Interaction for Immersive Virtual Reality. In *Proceedings of IEEE 3DUI*. 36–39.
[198] Thammathip Piumsomboon, Youngho Lee, Gun A. Lee, Arindam Dey, and Mark Billinghurst. 2017. Empathic Mixed Reality: Sharing What You Feel and Interacting with What You See. In *Proceedings of the International Symposium on Ubiquitous Virtual Reality*. 38–41.
[199] Daniel Pohl, Xucong Zhang, Andreas Bulling, and Oliver Grau. 2016. Concept for Using Eye Tracking in a Head-Mounted Display to Adapt Rendering to the User's Current Visual Field. In *Proceedings of the ACM Conference on Virtual Reality Software and Technology*. 323–324.
[200] Felix Putze, Dennis Weiß, Lisa-Marie Vortmann, and Tanja Schultz. 2019. Augmented Reality Interface for Smart Home Control using SSVEP-BCI and Eye Gaze. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*. 2812–2817.
[201] Yuan Yuan Qian and Robert J. Teather. 2017. The Eyes Don't Have It: An Empirical Comparison of Head-based and Eye-based Selection in Virtual Reality. In *Proceedings of ACM SUI*. 91–98.
[202] Yuan Yuan Qian and Robert J. Teather. 2018. Look to Go: An Empirical Evaluation of Eye-Based Travel in Virtual Reality. In *Proceedings of ACM SUI*. 130–140.
[203] John P. Rae, William Steptoe, and David J. Roberts. 2011. Some Implications of Eye Gaze Behavior and Perception for the Design of Immersive Telecommunication Systems. In *Proceedings of the IEEE International Symposium on*

Distributed Simulation and Real-Time Applications. 108–114.

[204] Yitoshee Rahman, Sarker Monojit Asish, Adil Khokhar, Arun K Kulshreshth, and Christoph W Borst. 2019. Gaze Data Visualizations for Educational VR Applications. In *Proceedings of ACM SUI*. Article 23, 2 pages.

[205] Vijay Rajanna and John Paulin Hansen. 2018. Gaze Typing in Virtual Reality: Impact of Keyboard Design, Selection Method, and Motion. In *Proceedings of ACM ETRA*. Article 15, 10 pages.

[206] Ramesh Raskar, Greg Welch, Matt Cutts, Adam Lake, Lev Stesin, and Henry Fuchs. 1998. The Office of the Future: A Unified Approach to Image-based Modeling and Spatially Immersive Displays. In *Proceedings of the Annual Conference on Computer Graphics and Interactive Techniques*. 179–188.

[207] David Roberts, Robin Wolff, John Rae, Anthony Steed, Rob Aspin, Moira McIntyre, Adriana Pena, Oyewole Oyekoya, and William Steptoe. 2009. Communicating Eye-gaze Across a Distance: Comparing an Eye-gaze enabled Immersive Collaborative Virtual Environment, Aligned Video Conferencing, and Being Together. In *Proceedings of the IEEE Virtual Reality Conference*. 135–142.

[208] Cynthia E. Rogers, Alexander W. Witt, Alexander D. Solomon, and Krishna K. Venkatasubramanian. 2015. An Approach for User Identification for Head-Mounted Displays. In *Proceedings of the ACM International Symposium on Wearable Computers*. 143–146.

[209] Przemyslaw Rokita. 1996. Generating Depth of-Field Effects in Virtual Reality Applications. *IEEE Computer Graphics and Applications* 16, 2 (1996), 18–21.

[210] Damien Constantine Rompapas, Aitor Rovira, Alexander Plopski, Christian Sandor, Takafumi Taketomi, Goshiro Yamamoto, Hirokazu Kato, and Sei Ikeda. 2017. EyeAR: Refocusable Augmented Reality Content through Eye Measurements. *Multimodal Technologies and Interaction* 1, 4, Article 22 (2017).

[211] Daniel Roth, Gary Bente, Peter Kullmann, David Mal, Chris Felix Purps, Kai Vogeley, and Marc Erich Latoschik. 2019. Technologies for Social Augmentations in User-Embodied Virtual Reality. In *Proceedings of ACM VRST*. Article 5, 12 pages.

[212] Sajay Sadasivan, P. Mehta, Joel S. Greenstein, Anand K. Gramopadhye, and Andrew T. Duchowski. 2005. Gaze Display in a Collaborative Virtual Aircraft Inspection Training Environment. In *Proceedings of the IIE Annual Conference*. 1–6.

[213] Javier San Agustin, John Paulin Hansen, Dan Witzner Hansen, and Henrik Skovsgaard. 2009. Low-Cost Gaze Pointing and EMG Clicking. In *Proceedings of the SIGCHI Conference Extended Abstracts on Human Factors in Computing Systems*. 3247–3252.

[214] MHD Yamen Saraiji, Shota Sugimoto, Charith Lasantha Fernando, Kouta Minamizawa, and Susumu Tachi. 2016. Layered Telepresence: Simultaneous Multi Presence Experience Using Eye Gaze Based Perceptual Awareness Blending. In *Proceedings of the ACM SIGGRAPH Emerging Technologies*. Article 14, 2 pages.

[215] Niladri Sarkar, Duncan Strathearn, Geoffrey Lee, Mahdi Olfat, Arash Rohani, and Raafat R. Mansour. 2015. A Large Angle, Low Voltage, Small Footprint Micromirror for Eye Tracking and Near-Eye Display Applications. In *Proceedings of the International Conference on Solid-State Sensors, Actuators and Microsystems*. 855–858.

[216] Prasanth Sasikumar, Lei Gao, Huidong Bai, and Mark Billinghurst. 2019. Wearable RemoteFusion: A Mixed Reality Remote Collaboration System with Local Eye Gaze and Remote Hand Gesture Sharing. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality Adjunct*. 393–394.

[217] Maike Scholtes, Philipp Seewald, and Lutz Eckstein. 2018. Implementation and Evaluation of a Gaze-Dependent In-Vehicle Driver Warning System. In *Proceedings of the International Conference on Applied Human Factors and Ergonomics*. 895–905.

[218] William E. Schroeder. 1993. Head-Mounted Computer Interface Based on Eye Tracking. In *Proceedings of SPIE, Visual Communications and Image Processing*, Vol. 2094. 1114–1124.

[219] Robin Schweigert, Valentin Schwind, and Sven Mayer. 2019. EyePointing: A Gaze-Based Selection Technique. In *Proceedings of Mensch Und Computer*. 719–723.

[220] Sven Seele, Sebastian Misztal, Helmut Buhler, Rainer Herpers, and Jonas Schild. 2017. Here's Looking At You Anyway! How Important is Realistic Gaze Behavior in Co-located Social Virtual Reality Games?. In *Proceedings of the Annual Symposium on Computer-Human Interaction in Play*. 531–540.

[221] Ludwig Sidenmark and Hans Gellersen. 2019. Eye, Head and Torso Coordination During Gaze Shifts in Virtual Reality. *ACM Transactions on Computer-Human Interaction* 27, 1, Article 4 (2019), 40 pages.

[222] Ludwig Sidenmark and Hans Gellersen. 2019. Eye&Head: Synergetic Eye and Head Movement for Gaze Pointing and Selection. In *Proceedings of the Annual ACM Symposium on User Interface Software and Technology*. 1161–1174.

[223] Candace L. Sidner, Cory D. Kidd, Christopher Lee, and Neal Lesh. 2004. Where to Look: A Study of Human-Robot Engagement. In *Proceedings of the ACM International Conference on Intelligent User Interfaces*. 78–84.

[224] Nikolaos Sidorakis, George Alex Koulieris, and Katerina Mania. 2015. Binocular Eye-Tracking for the Control of a 3D Immersive Multimedia User Interface. In *Proceedings of the IEEE Workshop on Everyday Virtual Reality*. 15–18.

[225] Gabriel Skantze, Anna Hjalmarsson, and Catharine Oertel. 2014. Turn-Taking, Feedback and Joint Attention in Situated Human-Robot Interaction. *Speech Communication* 65 (2014), 50–66.

[226] Robert Skerjanc and Siegmund Pastoor. 1997. New Generation of 3D Desktop Computer Interfaces. In *Proceedings of SPIE, Stereoscopic Displays and Virtual Reality Systems IV*, Vol. 3012. 439–447.

[227] Henrik Skovsgaard, Kari-Jouko Räihä, and Martin Tall. 2012. Computer Control by Gaze. In *Gaze Interaction and Applications of Eye Tracking: Advances in Assistive Technologies*. IGI Global, 78–102.

[228] Dana Slambekova, Reynold Bailey, and Joe Geigel. 2012. Gaze and Gesture Based Object Manipulation in Virtual Worlds. In *Proceedings of ACM VRST*. 203–204.

[229] Kay M. Stanney, Robert S. Kennedy, and Julie M. Drexler. 1997. Cybersickness is Not Simulator Sickness. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting*, Vol. 41. 1138–1142.

[230] Andrei State. 2007. Exact Eye Contact With Virtual Humans. In *Proceedings of the International Workshop on Human-Computer Interaction*. 138–145.

[231] Statista. 2020. Forecast unit shipments of augmented (AR) and virtual reality (VR) headsets from 2020 to 2025. Retrieved September 19, 2021 from https://www.statista.com/statistics/653390/worldwide-virtual-and-augmented-reality-headset-shipments/

[232] Julian Steil, Inken Hagestedt, Michael Xuelin Huang, and Andreas Bulling. 2019. Privacy-Aware Eye Tracking Using Differential Privacy. In *Proceedings of ACM ETRA*. Article 27, 9 pages.

[233] Sophie Stellmach and Raimund Dachselt. 2012. Designing Gaze-Based User Interfaces for Steering in Virtual Environments. In *Proceedings of ETRA*. 131–138.

[234] William Steptoe, Oyewole Oyekoya, Alessio Murgia, Robin Wolff, John Rae, Estefania Guimaraes, David Roberts, and Anthony Steed. 2009. Eye Tracking for Avatar Eye Gaze Control During Object-Focused Multiparty Interaction in Immersive Collaborative Virtual Environments. In *Proceedings of the IEEE Virtual Reality Conference*. 83–90.

[235] William Steptoe, Oyewole Oyekoya, and Anthony Steed. 2010. Eyelid Kinematics for Virtual Characters. *Computer Animation and Virtual Worlds* 21, 3-4 (2010), 161–171.

[236] William Steptoe and Anthony Steed. 2008. High-Fidelity Avatar Eye-Representation. In *Proceedings of the IEEE Virtual Reality Conference*. 111–114.

[237] William Steptoe, Anthony Steed, Aitor Rovira, and John Rae. 2010. Lie Tracking: Social Presence, Truth and Deception in Avatar-Mediated Telecommunication. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 1039–1048.

[238] William Steptoe, Robin Wolff, Alessio Murgia, Estefania Guimaraes, John Rae, Paul Sharkey, David Roberts, and Anthony Steed. 2008. Eye-Tracking for Avatar Eye-Gaze and Interactional Analysis in Immersive Collaborative Virtual Environments. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 197–200.

[239] Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-guided foveation for light field displays. *ACM TOG* 36, 6, Article 192 (2017), 13 pages.

[240] Qi Sun, Anjul Patney, Li-Yi Wei, Omer Shapira, Jingwan Lu, Paul Asente, Suwen Zhu, Morgan McGuire, David Luebke, and Arie Kaufman. 2018. Towards Virtual Reality Infinite Walking: Dynamic Saccadic Redirection. *ACM TOG* 37, 4, Article 67 (2018), 13 pages.

[241] Ivan E Sutherland. 1965. The Ultimate Display. In *Proceedings of IFIP Congress*. 506–508.

[242] Ivan E Sutherland. 1968. A Head-Mounted Three Dimensional Display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*. ACM, 757–764.

[243] Vildan Tanriverdi and Robert J. K. Jacob. 2000. Interacting with Eye Movements in Virtual Environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 265–272.

[244] Markus Tatzgern, Valeria Orso, Denis Kalkofen, Giulio Jacucci, Luciano Gamberini, and Dieter Schmalstieg. 2016. Adaptive Information Density for Augmented Reality Displays. In *Proceedings of the IEEE Virtual Reality*. 83–92.

[245] Marcus Tönnis and Gudrun Klinker. 2014. Boundary Conditions for Information Visualization with Respect to the User's Gaze. In *Proceedings of AH*. Article 44, 8 pages.

[246] Marcus Tönnis and Gudrun Klinker. 2014. [DEMO] Placing Information near to the Gaze of the User. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality*. 377–378.

[247] Takumi Toyama, Thomas Kieninger, Faisal Shafait, and Andreas Dengel. 2012. Gaze Guided Object Recognition Using a Head-Mounted Eye Tracker. In *Proceedings of the Symposium on Eye Tracking Research and Applications*. 91–98.

[248] Takumi Toyama, Jason Orlosky, Daniel Sonntag, and Kiyoshi Kiyokawa. 2014. A Natural Interface for Multi-Focal Plane Head Mounted Displays Using 3D Gaze. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*. 25–32.

[249] Takumi Toyama, Daniel Sonntag, Jason Orlosky, and Kiyoshi Kiyokawa. 2015. Attention Engagement and Cognitive State Analysis for Augmented Reality Text Display Functions. In *Proceedings of the International Conference on Intelligent User Interfaces*. 322–332.

[250] Jochen Triesch, Brian T. Sullivan, Mary M. Hayhoe, and Dana H. Ballard. 2002. Saccade Contingent Updating in Virtual Reality. In *Proceedings of ETRA*. ACM, 95–102.

[251] Roel Vertegaal. 1999. The GAZE Groupware System: Mediating Joint Attention in Multiparty Communication and Collaboration. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 294–301.
[252] Roel Vertegaal. 2003. Attentive User Interfaces. *Commun. ACM* 46, 3 (2003), 30–33.
[253] Roel Vertegaal and Yaping Ding. 2002. Explaining Effects of Eye Gaze on Mediated Group Conversations: Amount or Synchronization?. In *Proceedings of the ACM Conference on Computer Supported Cooperative Work*. 41–48.
[254] Roel Vertegaal, Robert Slagter, Gerrit van der Veer, and Anton Nijholt. 2001. Eye Gaze Patterns in Conversations: There is More to Conversational Agents Than Meets the Eyes. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 301–308.
[255] Vinoba Vinayagamoorthy, Maia Garau, Anthony Steed, and Mel Slater. 2004. An Eye Gaze Model for Dyadic Interaction in an Immersive Virtual Environment: Practice and Experience. *Comput Graph Forum* 23, 1 (2004), 1–11.
[256] Gyula Vörös, Anita Verő, Balázs Pintér, Brigitta Miksztai-Réthey, Takumi Toyama, András Lőrincz, and Daniel Sonntag. 2014. Towards a Smart Wearable Tool to Enable People with SSPI to Communicate by Sentence Fragments. In *International Symposium on Pervasive Computing Paradigms for Mental Health*. Springer, 90–99.
[257] Oleg Špakov, Howell Istance, Kari-Jouko Räihä, Tiia Viitanen, and Harri Siirtola. 2019. Eye Gaze and Head Gaze in Collaborative Games. In *Proceedings of ACM ETRA*. Article 85, 9 pages.
[258] Jian Wang. 1995. Integration of eye-gaze, voice and manual response in multimodal user interface. In *Proceedings of the IEEE International Conference on Systems, Man and Cybernetics*, Vol. 5. 3938–3942.
[259] Peng Wang, Shusheng Zhang, Xiaoliang Bai, Mark Billinghurst, Weiping He, Shuxia Wang, Xiaokun Zhang, Jiaxiang Du, and Yongxing Chen. 2019. Head Pointer or Eye Gaze: Which Helps More in MR Remote Collaboration?. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces*. 1219–1220.
[260] Ginger S. Watson, Yiannis E. Papelis, and Katheryn C. Hicks. 2016. Simulation-Based Environment for the Eye-Tracking Control of Tele-Operated Mobile Robots. In *Proceedings of the Modeling and Simulation of Complexity in Intelligent, Adaptive and Autonomous Systems and Space Simulation for Planetary Space Exploration*. Society for Computer Simulation International, Article 4, 7 pages.
[261] Nicholas A. Webb and Michael J. Griffin. 2002. Optokinetic Stimuli: Motion Sickness, Visual Acuity and Eye Movements. *Aviation, Space and Environmental Medicine* 73, 4 (2002), 351–358.
[262] Martin Weier, Thorsten Roth, André Hinkenjann, and Philipp Slusallek. 2018. Foveated Depth-of-Field Filtering in Head-Mounted Displays. *ACM Transactions on Applied Perception* 15, 4, Article 26 (2018), 14 pages.
[263] Martin Weier, Thorsten Roth, Ernst Kruijff, André Hinkenjann, Arsène Pérard-Gayot, Philipp Slusallek, and Yongmin Li. 2016. Foveated Real-Time Ray Tracing for Head-Mounted Displays. *Comput Graph Forum* 35, 7 (2016), 289–298.
[264] Gordon Wetzstein, Anjul Patney, and Qi Sun. 2020. *State of the Art in Perceptual VR Displays*. Springer International Publishing, Cham, 221–243.
[265] Robin Wolff, David Roberts, Alessio Murgia, Norman Murray, John Rae, William Steptoe, Anthony Steed, and Paul Sharkey. 2008. Communicating Eye Gaze across a Distance without Rooting Participants to the Spot. In *Proceedings of the IEEE/ACM International Symposium on Distributed Simulation and Real-Time Applications*. 111–118.
[266] Jianbin Xiong, Weichao Xu, Wei Liao, Qinruo Wang, Jianqi Liu, and Qiong Liang. 2013. Eye Control System Base on Ameliorated Hough Transform Algorithm. *IEEE Sensors Journal* 13, 9 (2013), 3421–3429.
[267] Jing Yang and Cheuk Yu Chan. 2019. Audio-Augmented Museum Experiences with Gaze Tracking. In *Proceedings of the International Conference on Mobile and Ubiquitous Multimedia*. Article 46, 5 pages.
[268] Jiawei Yang, Guangtao Zhai, and Huiyu Duan. 2019. Predicting the Visual Saliency of the People with VIMS. In *Proceedings of the IEEE Visual Communications and Image Processing*. 1–4.
[269] Zhaolin Yao, Xinyao Ma, Yijun Wang, Xu Zhang, Ming Liu, Weihua Pei, and Hongda Chen. 2018. High-Speed Spelling in Virtual Reality with Sequential Hybrid BCIs. *IEICE Transactions on Information and Systems* 101, 11 (2018), 2859–2862.
[270] Hong Zeng, Yanxin Wang, Changcheng Wu, Aiguo Song, Jia Liu, Peng Ji, Baoguo Xu, Lifeng Zhu, Huijun Li, and Pengcheng Wen. 2017. Closed-Loop Hybrid Gaze Brain-Machine Interface Based Robotic Arm Control with Augmented Reality Feedback. *Frontiers in Neurorobotics* 11, Article 60 (2017), 13 pages.
[271] Guangtao Zhang and John Paulin Hansen. 2019. A Virtual Reality Simulator for Training Gaze Control of Wheeled Tele-Robots. In *Proceedings of ACM VRST*. Article 49, 2 pages.
[272] Guangtao Zhang, John Paulin Hansen, and Katsumi Minakata. 2019. Hand- and Gaze-Control of Telepresence Robots. In *Proceedings of ACM ETRA*. Article 70, 8 pages.
[273] Hui Zhang, Damian Fricker, Thomas G. Smith, and Chen Yu. 2010. Real-Time Adaptive Behaviors in Multimodal Human-Avatar Interactions. In *Proceedings of the International Conference on Multimodal Interfaces and the Workshop on Machine Learning for Multimodal Interaction*. Article 4, 8 pages.
[274] Lian Zhang, Joshua W. Wade, Dayi Bian, Amy Swanson, Zachary Warren, and Nilanjan Sarkar. 2014. Data Fusion for Difficulty Adjustment in an Adaptive Virtual Reality Game System for Autism Intervention. In *Proceedings of the HCI International - Posters' Extended Abstracts*. 648–652.