



Review

Technologies for Multimodal Interaction in Extended Reality—A Scoping Review

Ismo Rakkolainen ^{*}, Ahmed Farooq , Jari Kangas , Jaakko Hakulinen , Jussi Rantala , Markku Turunen
and Roope Raisamo

Faculty of Information Technology and Communication Sciences, Tampere University, FI-33014 Tampere, Finland; Ahmed.Farooq@tuni.fi (A.F.); Jari.a.Kangas@tuni.fi (J.K.); Jaakko.Hakulinen@tuni.fi (J.H.); Jussi.Rantala@tuni.fi (J.R.); Markku.Turunen@tuni.fi (M.T.); Roope.Raisamo@tuni.fi (R.R.)

* Correspondence: Ismo.Rakkolainen@tuni.fi

Abstract: When designing extended reality (XR) applications, it is important to consider multimodal interaction techniques, which employ several human senses simultaneously. Multimodal interaction can transform how people communicate remotely, practice for tasks, entertain themselves, process information visualizations, and make decisions based on the provided information. This scoping review summarized recent advances in multimodal interaction technologies for head-mounted display-based (HMD) XR systems. Our purpose was to provide a succinct, yet clear, insightful, and structured overview of emerging, underused multimodal technologies beyond standard video and audio for XR interaction, and to find research gaps. The review aimed to help XR practitioners to apply multimodal interaction techniques and interaction researchers to direct future efforts towards relevant issues on multimodal XR. We conclude with our perspective on promising research avenues for multimodal interaction technologies.

Keywords: multimodality; human-computer interaction; extended reality; augmented reality; virtual reality; mixed reality



Citation: Rakkolainen, I.; Farooq, A.; Kangas, J.; Hakulinen, J.; Rantala, J.; Turunen, M.; Raisamo, R. Technologies for Multimodal Interaction in Extended Reality—A Scoping Review. *Multimodal Technol. Interact.* **2021**, *5*, 81. <https://doi.org/10.3390/mti5120081>

Academic Editors: Gerd Bruder and Cristina Portalés Ricart

Received: 28 October 2021

Accepted: 29 November 2021

Published: 10 December 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Extended reality (XR) covers an extensive field of research and applications, and it has advanced significantly in recent years. XR augments or replaces the user's view with synthetic objects, typically with head-mounted displays (HMD). XR can be used as an umbrella or unification term (e.g., [1,2]) encompassing virtual reality (VR), augmented reality (AR), and mixed reality (MR). There are many ways to view XR scenes [3], such as various kinds of 3D, light fields, holographic displays, CAVE virtual rooms [4], fog screens [5], and spatial augmented reality [6], which uses projectors and does not require any head-mounted or wearable gear. In this scoping review, we focused on HMD-based XR. XR systems have become capable of generating very realistic synthetic experiences in visual and auditory domains. Current low-cost HMD-based XR systems also increasingly stimulate other senses. Hand-held controllers employing simple vibrotactile haptics are currently the most usual multimodal method beyond vision and audio, but most human perceptual capabilities are underused.

At the time of writing this article, Google Scholar database yields 1,330,000 hits for the search term “virtual reality” and 1,140,000 hits for “multimodal interaction”. The abundant volume of research on XR and multimodality implies that knowledge syntheses and research result consolidations can advance their use and research. The field is changing constantly and rapidly, so up-to-date reviews are useful for the research community.

Our broad research question and purpose was to scope the body of literature on multimodal interaction beyond visuals and audio and to identify research gaps on HMD-based interaction technologies. With such a sea of material and broad scope, we conducted

a scoping review instead of a systematic literature review. Scoping reviews are useful for identifying research gaps and summarizing a field [7–9].

We constructed an overview of modalities, technologies, and trends that can be used for additional synthetic sensations or multimodal interaction for HMD-based XR and assessed their current state. We searched for and selected relevant studies, extracted and charted the data, and collated, summarized, and reported the results. We discussed recent multimodal trends and cutting-edge research results and hardware, which may become relevant in the future. As far as we know, the body of literature on multimodal HMD-based XR has not yet been comprehensively reviewed. This review summarized recent advances in multimodal interaction techniques for HMD-based XR systems.

In Section 2 we present related work and in Section 3 we present the review methodology. Section 4 discusses the multimodal interaction methods beyond standard vision, audio, and simple vibrotactile haptics that are often used with contemporary XR systems. We also highlight some more exotic modalities and methods that may become more important for XR in the future. In Section 5 we discuss the results and further aspects of multimodal interaction for XR, and in Section 6 we provide our conclusions.

2. Background

Multimodal interaction makes use of several simultaneous input and output modalities and human senses in interacting with technology [10]. Human perceptual (input) modalities include visual, auditory, haptic, olfactory, and gustatory modalities. Human output modalities include gestures, speech, gaze, bio-electric measurements, etc. They are essential ingredients for more realistic XR.

Perception is inherently multisensory and cross-modal integration takes place between many senses [11]. A large amount of information is processed and only a small fraction reaches our consciousness. Historically, humans had to adapt to computers through punch cards, line printers, command lines, and machine language programming. Engelbart's system [12], Sutherland's AR display [13], and Bolt's "Put That There" [14] were very visionary demonstrations of multimodal interfaces at their times. Rekimoto and Nagao [15] and Feiner et al. [16] presented early computer augmented interaction in real environments.

Recently computers have started to adapt to humans with the help of cameras, microphones, and other sensors as well as artificial intelligence (AI) methods, and they can recognize human activities and intentions. For example, haptic feedback enables a user to touch virtual objects as if they were real. Human-Computer Interaction (HCI) has become much more multisensory, even though the keyboard and mouse are still the prevalent forms of HCI in many contexts.

User interfaces (UI) change along with changing use contexts and emerging display, sensor, actuator, and user tracking hardware innovations. Research has tried to find more human-friendly, seamless, and intuitive UIs [17], given the contemporary technology available at the time. Perceptual UIs [18] emphasize the multitude of human modalities and their sensing and expression power. They also combine human communication, motor, and cognitive skills. Multimodality and XR match well together. XR and various kinds of 3D UIs [19,20] take advantage of the user's spatial memory, position, and orientation. Many textbooks (e.g., [21–24]) and reviews (e.g., [25]) cover various aspects of multimodal XR.

General HCI results and guidelines cannot always be directly applied to XR. Immersion in VR is one major distinction from most other HCI contexts. VR encloses the user into a synthetically generated world and enables the user to enter "into the image". As users operate with 3D content in XR environments, input and output devices may need to be different, interaction techniques must support spatial interaction, and often embodied cognition has a bigger role. Interactions between humans and virtual environments rely on timely and consistent sensory feedback and spatial information. Effective feedback helps users to get information, notifications, and warnings. Many emerging technologies are enabling e.g., tracking of hands, facial expression, and gaze on HMDs.

There is a large number of all kinds of reviews, surveys, and books on multimodal interaction. As Augstein and Neumayr [26] noted, many of them focus only on the most usual modalities, sometimes only on vision, audition, and haptics. There are also many papers that review selected narrow topics on multimodal interaction for XR, for example, a review on VR-based ball sports performance training and advances in communication, interaction, and simulation [25].

Taxonomies for Multimodal Interaction

Many HCI taxonomies for multimodal input, output, and interaction (e.g., [18,19,26–28]) focus only on those modalities which were feasible and usual for HCI in their time. Augstein and Neumayr [26] discussed in depth the history and types of these taxonomies and the emergence of various enabling technologies. Their taxonomy is targeted for HCI, and it is based both on the input and output capabilities of basic human senses and on the other hand on sensors and actuators employed by computers, i.e., it describes how humans and computers can perceive each other. Their modality classes employ either direct processing (neural oscillation, galvanism), or indirect processing (vision, audition, kinesthetics, touch, olfaction, and gustation). Direct processing (e.g., BCI or EMS) works directly between a computer and the brain or muscles. Indirect processing refers to the multi-stage process where an output stimulus is perceived by a human receptor and then the information is delivered via electrical signals for further processing to the brain. The flow is similar for input stimulus from a human via sensors to the computer.

The taxonomy focuses on the modalities which human senses can perceive and which can actively and consciously be utilized for input or output. It excludes modalities which humans cannot control for interaction purposes, e.g., electrodermal activity. However, a UI could use also those to better interpret the status and intentions of the user.

Even though the Augstein and Neumayr taxonomy is human-centered, it is also partly device-centered, as many modalities such as gaze, gestures, or facial expression can be placed to several classes, depending on the used measurement and sensor technologies. However, the taxonomy fosters and guides the design and research of multimodal XR.

A modified version of Augstein and Neumayr's taxonomy (see Figure 1) forms the base for our review. As the kinesthetic and tactile feedback are closely intertwined and difficult to differentiate, we combined them under Haptics and have subclasses for body-related (kinesthetic) and skin-related (tactile) senses. We included only the non-invasive (without surgical implants) methods of interaction. Similarly, segregating input and output events, as carried out by Augstein and Neumayr, may also be counterintuitive as localized haptic interaction can seldom be isolated to separate input and output events. For that reason, the review considered input and output to be co-localized events and discussed touch interaction accordingly.

One alternative or complementary classification could be contact vs. non-contact interaction [29,30], which enables touchless interaction with digital content through body movements, hand gestures, speech, or other means. For example, ultrasound haptics can create a sensation of touch on the plain hand.

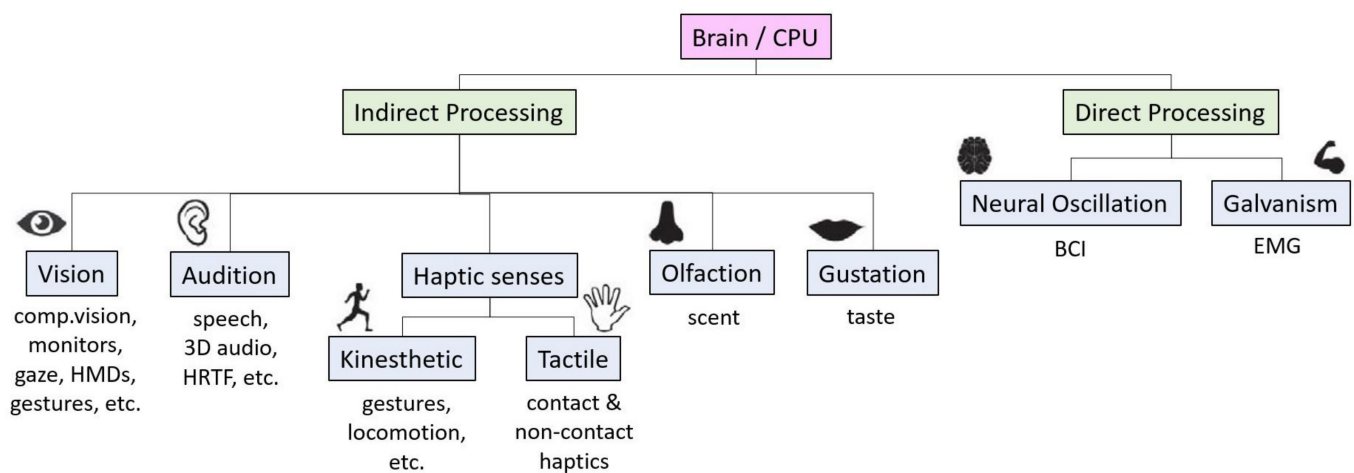


Figure 1. Taxonomy of interaction modalities for XR (adapted from Augstein and Neumayr [26]). The taxonomy is based on human senses and classifies both input and output devices and technologies for multimodal interaction.

3. Review Methodology

We wanted to collect studies on XR which used HMDs, and which employed multimodal interaction beyond standard video or audio. We were particularly interested in primary research on emerging multimodal technologies and not so much on theoretical, perceptual, or application studies of them.

We first identified potentially interesting and relevant peer-reviewed publications in English. We included peer-reviewed journal and conference papers, and excluded patents, books, theses, presentations, and reports. The field was too wide for a systematic literature review or a formal database query.

We carried out informal searches to identify relevant papers on ACM Digital Library, IEEE Xplore, MDPI, and Elsevier databases, which contain most of the relevant peer-reviewed papers related to computer science and various fields of engineering. We also made similar searches on the Google Scholar database, which contains additional scientific publishers and other fields of science. We also used the reference lists of all selected papers and articles to find additional relevant studies. Furthermore, we also searched and read product pages that were relevant for the topics to find out the state of the art and availability of the cutting-edge hardware.

We used our background expert knowledge on the topic as a starting point. After an analysis of text words contained in the title, abstract, or index terms, we manually selected potentially interesting papers for further review. The papers were retained only if they employed an HMD, and if they explicitly explained a multimodal feedback solution for XR interaction, or an emerging technology applicable to multimodal XR.

We made no limitations to the publication date, because if any technology, idea, or approach was not very useful in its time, it may still be useful today with cutting-edge technology or a new use context. The time-variant aspect also shows the emergence of specific fields and trends. We discovered that the number of published multimodal XR studies gradually increased over time, as also noted by Kim et al. [31]. This is at least partly due to improving technology, sensors, and actuators. Our rather wide sample is presumably indicative of the general trends in recent applications and research on multimodal HMD-based XR.

4. Multimodal Interaction Technologies

In this section, we only briefly mention some aspects of visual and audio modalities, as they are covered in other surveys and books (e.g., [2,21,32]). We focus on the other, additional modes of XR interaction. Many experimental and emerging technologies are intriguing for interaction, but not yet deployable for XR. Some of them may become widely used in the future, and others possibly not. Only time will tell.

4.1. Vision

Vision is perhaps the most important human sense, and other senses usually only support vision. However, our real-life experiences and sensations are predominantly multisensory and so XR should also be (depending on the specific application and context).

Visual output interfaces have improved dramatically in recent decades. Advances in hardware (displays, GPU, CPU, memory sizes, high-speed networking, tracking systems) and software (rendering algorithms, computer vision, artificial intelligence (AI), etc.) enable near-realistic XR imagery and XR can be used for many applications.

Improving displays, foveated rendering, and tracking will naturally affect XR interaction. Furthermore, some impactful technologies for XR are improving GPUs and/or practically infinite processing power through supercomputers or quantum computers together with very fast networking and 5G mobile networks, as rendering can then happen on the cloud. This could enable immersive, fully photorealistic remote work or teleconferencing. Being together remotely with other people or visiting remote places could make the experience almost indistinguishable from reality. Low-latency networks and cloud rendering can also make the HMDs relatively simple, low-cost, and lightweight. They might even become contact lens displays (e.g., Mojo Lens (<https://www.mojo.vision/>) accessed on 1 December 2021), which would enable viewing information or watching movies even with shut eyes.

One important recent development in multimodal interaction is the role of artificial intelligence (AI), which can perform many tasks, especially in the visual domain. AI-based systems can, for example, render realistic synthetic scenes and humans, or recognize scenes, people, text, products, or emotions. For example, Microsoft has developed a Seeing AI (<https://www.microsoft.com/en-us/ai/seeing-ai>) accessed on 1 December 2021, which can describe the world around the user. It is helpful, e.g., for blind and visually impaired users.

Visual input interfaces such as eye tracking and gestures have made a lot of progress. Computers or other computing devices can now see the user and improve the UI based on that. In the following, we discuss some interaction methods which are mainly vision-based.

4.1.1. Gestural Interaction in XR

Gestures are used for human-human or human-computer interaction in many contexts (e.g., deaf sign languages). Pointing with hands or a finger is learned in early childhood. The finger is an intuitive, convenient, and universal pointing tool, which is always available, is used across all cultures, and does not require any literacy skills. The meaning of some gestures varies across cultures, e.g., waving “goodbye” in Europe means “come here” in India. The gestures may also have emotional, social, and other meanings and levels.

Gesture recognition of the hands, head, and other body parts is a well-known interaction method for HCI and XR. A tracking system or motion-sensing input device is needed to recognize the moving gestures or static postures and the position and orientation of the HMD. Typically, gesture tracking is vision-based, so gestural interaction can be seen as a part of vision, even though it can also be seen as a part of proprioception. Modern HMDs and XR systems embed many sensors for position, orientation, motion, and gesture tracking.

An early work was Krueger’s installations in the 1970s [33], which enabled interaction with visual art using body movements. Bolt’s “Put-That-There” [14] was another early gestural interface combined with speech input. It enabled pointing at an object on a screen with a hand and giving commands with speech. This is an example of deictic gestures, i.e., pointing indicating objects and locations. They are natural for people, and they are actively studied in VR (see, e.g., [34,35]).

Manual gestures can be split into mid-air gestures, gestures performed with hand-held devices, and touch-based gestures (touchpads on controllers or HMDs). Current hand-held controllers utilize mostly wrist rotations, but pen-type devices can support more precise and faster movements [36].

In VR, two common ways to implement pointing are cursor-based relative pointing (where hand movements move the cursor) and ray casting (where a ray is drawn out relative to the user's body). In AR systems, gestures are also common but rarely utilize controllers. In an early work, Mann [37] and Starner et al. [38] presented early AR systems employing a finger mouse for a head-mounted camera and display. One early work on a gestural UI with computer vision-based hand tracking for HMD was carried out by Kölsch et al. [39].

Direct touching and pointing are natural ways to interact in immersive VR. However, the interaction does not need to be a replica of reality, but it can use more powerful and flexible methods. Li et al. [40] carried out a review on gesture interaction in virtual reality. Chen et al. [41] compared gestures and speech as input modalities for AR. They found that speech is more accurate, but gestures can be faster.

An important field of gestural interaction in XR is head gestures [42–44]. The head can rotate on many axes or move in essentially all directions. Interaction with the head can take many forms such as pointing to objects based on the posture of the head (and thus HMD), or it can signal selections (like nodding or shaking one's head for yes/no) [45].

Superwide field-of-view (FOV) on HMDs can improve immersion, situational awareness, and performance [46] and is generally preferred by audiences. Gestures with them could be much wider than usual, but this is a relatively little-researched field.

Several technologies can be used for gesture tracking for XR [47] and often sensor fusion is used. In recent years, tracking software and hardware have improved tremendously, and environmental tracking and hand tracking are possible for a stand-alone HMD. Often, computer vision (CV) methods are used for tracking the arms, hands, or fingers. It is convenient, as it often requires no user-mounted artifacts.

Hand gesture recognition and hand pose estimation are very challenging due to the complex structure and dexterous movement of the human hand, which has 27 degrees of freedom (DOF). It can also perform very fast and delicate movements. Deep-learning-based methods are very promising in hand-pose estimation.

One widely used tracking method is by a tiny, built-in inertial measurement unit (IMU) for orientation tracking which contains accelerometers, gyroscopes, and magnetometers. Optical trackers use light (often IR light) for tracking. Magnetic (e.g., Polhemus, Razer Hydra) systems use magnetic fields for tracking and are thus not limited to line-of-sight to any device. Acoustic tracking can also be used to locate an object's position. Typically, it uses three ultrasonic sensors and three ultrasonic transmitters on devices. For hand tracking, there are also various kinds of bend and stretch sensors.

An RGB stereo camera and CV algorithms can discern depth information (e.g., Stereolabs ZED 2) and gestures. RGB-D cameras contain a depth sensor, which outputs a per-pixel depth map. The depth sensor can be based on many technologies. One way is to project IR light patterns onto the environment (Kinect 1.0) and calculate the depth based on the distorted patterns. Some cameras emit light pulses and measure the time it arrives back (time-of-flight cameras such as Kinect 2.0, Intel RealSense D400, or Microsoft HoloLens 2). Ultraleap (Controller and Stereo IR 170) uses a stereo IR-camera pair and IR illumination for accurate and latency-free finger, hand, and gesture tracking, and it is often used on HMDs. Solid state LiDAR cameras use MEMS mirror scanning and a laser for high-resolution scanning and they can nowadays be very small (e.g., Intel RealSense L515 with 61 mm diameter and 100 g of weight). The emerging, small-size Google Soli and KaiKuTek Inc.'s 3D gesture sensors both use a 60 GHz frequency-modulated radar signal.

There is a massive pool of literature on gesture tracking methods employing CV-based methods. Rautaray and Agrawal [48] surveyed CV-based hand gesture recognition for HCI. Cheng et al. [49] surveyed hand gesture recognition using 3D depth sensors, 3D hand gesture recognition approaches, the related applications, and typical systems. They also discussed deep-learning-based methods. Vuletic et al. [50] carried out a review of hand gestures used in HCI. Chen et al. [51] carried out a comprehensive and timely review of real-time sensing and modeling of the human hands with wearable sensors

or CV-based methods. Alam et al. [52] provided a comprehensive survey on intelligent speech and vision applications using deep neural networks. Beddiar et al. [53] reviewed and summarized the progress of human activity recognition systems from the computer vision perspective.

Hand-held controllers, data gloves, or full-body VR suits can track the user's movements and possibly also provide some tactile feedback. Advanced hand-held controllers such as Valve Knuckles have a large set of built-in sensors, including grip force sensor and finger tracking. Data gloves or full-body suits can be more precise than cameras and they do not require a line-of-sight to cameras, but a user must put them on, wear, and possibly calibrate them before use. They may have hygiene problems for multiple users, especially in times of pandemics. They may also be tethered and have a limited operational range.

Motion capture (Mo-cap) is a form of gesture recognition. Mo-cap trackers record the position and orientation of human bodies, usually in real-time. Typically, Mo-cap uses optical (e.g., Vicon), magnetic (e.g., Polhemus), or full-body suit tracking systems. Mo-cap is used in medical or sports applications, film making, TV studios, etc.

4.1.2. Facial Expression and Emotion Recognition Interfaces

Facial expressions and emotion recognition are (usually unconscious) elements of human-human communication. They are not widely used as explicit inputs, as they are relatively new concepts in interaction. Because HMDs partly block the user's face, the XR context has specific challenges. However, sensors can be placed inside the HMD, and on the other hand, the HMDs are becoming smaller and may ultimately become ultralight smart glasses. In the last few years, researchers and the industry have integrated facial recognition technology into HMDs to track and relay user expression to virtual avatars. Devices such as Decagear [54] utilize facial tracking and mapping in real-time.

4.1.3. Gaze

Gaze is an important element of human-human communication, providing insight into the attention of other people. Humans use gaze to study their environments, to look at objects, and to communicate with others. The gaze can be used as a control in HCI or as an indicator of the person's mental state or intentions. Eye tracking is a form of gestural interface where only movements of eyes are tracked. Typically, vision-based tracking methods are used, but other sensor technologies are also available. Recent advances in eye tracking technology have lowered the prices and made it more generally available.

Gaze can be utilized in HCI in various ways, e.g., to infer the user's interests based on gaze patterns, to improve foveated rendering, or to provide specific commands. Multimodal gaze and gesture have been utilized in collaboration, e.g., by Bai et al. [55].

As the eye is primarily used for sensing and observing the environment, using gaze as an input method can be problematic since the same modality is then used for both perception and control. The tracking system needs to be able to distinguish casual viewing from the act of intentional selection to prevent the "Midas touch" problem wherein all viewed items are selected. A common method is to introduce a brief delay, "dwell time", in which the user needs to stare at the object for the duration of the dwell time to activate it [56]. Blink-and-wink detection can also be used as a control tool in gaze interaction [57]. In multimodal settings, other control methods, such as body gestures [58], audio commands [59], or a separate physical trigger [60], can also be used to activate a gaze-based selection.

There are three common methods to utilize gaze as an explicit command: dwell-select (described above), gaze gestures [61], and smooth-pursuit-based interactions. Gaze gesture interaction recognizes a sequence of rapid eye movements (saccades) that follow a specified pattern to activate a command [62,63]. Finally, the smooth pursuit interaction is based on recognizing a continuous movement of gaze while tracking a specific moving target [64,65]. The system deduces that if the gaze follows a similar trajectory to that target,

there is a match, and a selection or a (target-specific) command is triggered. An example of pursuit-based selection in VR is a study by Sidenmark et al. [66].

Several techniques have been developed for tracking eye movements [67] and defining gaze position, or gaze vector [68]. The most common method is analyzing video images of the eye (video-oculography, VOG). For each video frame captured by a camera located somewhere close to the user's eye, tracking software detects several visual features, such as pupil size, pupil center, and so on. VOG-based trackers typically require calibration before the gaze point can be estimated. The VOG system fits naturally to HMD as the cameras can be installed close to the display elements facing the user's eye. The cameras for VOG can be installed freely if the eyes are visible, and Khamis et al. [69] used a flying drone as a platform. Eye movements can also be detected by electro-oculography (EOG) based on the cornea-retinal potential difference [70]. EOG systems require sensors to touch the skin close to the eyes, which is easy to arrange in HMDs. EOG is most useful in detecting relative eye movements (e.g., gaze gesture recognition).

Gaze control as an input for HCI has a long history (e.g., [71]). At first, the gaze interaction method was mostly used for special purposes, like typing tools for the disabled [72], but through active research and affordable new trackers, gaze-based interfaces can be included in many new devices. It has been demonstrated how eye tracking could enhance the interaction e.g., with mobile phones [73], tablets [74], smart watches [75], smart glasses [76], and public displays [77]. Gaze can also be used to directly control moving devices, like drones (e.g., [78,79]).

The recent development of gaze tracking for VR has made it easy to study gaze behavior in a virtual environment [80]. There are already several commercial HMDs with integrated eye trackers (e.g., HTC Vive Pro Eye, Fove, Magic Leap 1, Varjo, Pico Neo2 Eye), and eye tracking is expected to become a standard feature. For mixed reality, the gaze-based input has been used in studies with HMDs (e.g., [60,81,82]). Additionally, Meissner et al. [83] used gaze tracking in VR to study shopper behavior. Tobii [84] and Varjo [85] provided integration tools for gaze data collection and analysis for VR-based use cases. Burova et al. [86] utilized gaze tracking in the development of AR solutions using VR technology. By analyzing gaze, it is possible to understand how AR content is perceived and to check also real-world related aspects like the safety of AR use in risky environments. Such safety issues can be found early with a VR prototype. Additionally, Gardony et al. [87] discussed how gaze tracking can be used to evaluate the cognitive capacities and intentions of users to tune the UI for an improved experience in mixed-reality environments.

While the focus of gaze interaction studies has been in intentional control of HCI using gaze, the research interest for other use cases is growing. The systems can also utilize the gaze data to "know" of the user's attention or interests, and optionally to adapt to it. Other examples could be that the system might notice user confusion by following gaze behavior [88,89] and offer help, recognize the cognitive state of a person [90], make early diagnoses of some neurological conditions [91], or analyze the efficiency of advertisement [92]. Human gaze tracking constitutes an important part of the vision of human augmentation [93]. In a way, humans also analyze (often unconsciously) other people's interests, stresses, phobias, or state-of-mind by their gaze behavior.

Research on gaze UIs is often divided into research on gaze-only UIs where the gaze is the sole form used for input, gaze-based UIs where the gaze information is the main input modality, and gaze-added UIs where gaze data is used to add some functionality to a UI. For example, in some assistive systems, gaze interaction is the only method of communication and control [72]. Alternatively, information from the user's gaze behavior can be exploited subtly in the background in attentive interfaces in a wide variety of application areas [94,95]. The research on HMD-integrated gaze tracking naturally falls into the latter category as all the other input modalities are also available.

Burova et al. [86] used gaze data to identify where industrial personnel are looking at when they are performing operations (see Figure 2). This can be used to analyze what users have seen and in which order, and what they have missed. This can be crucial information in

safety-critical environments since it would be possible to detect, e.g., hazardous situations and unsafe working conditions even when other measures (e.g., task completion and error rates) show that the tasks have been completed successfully.

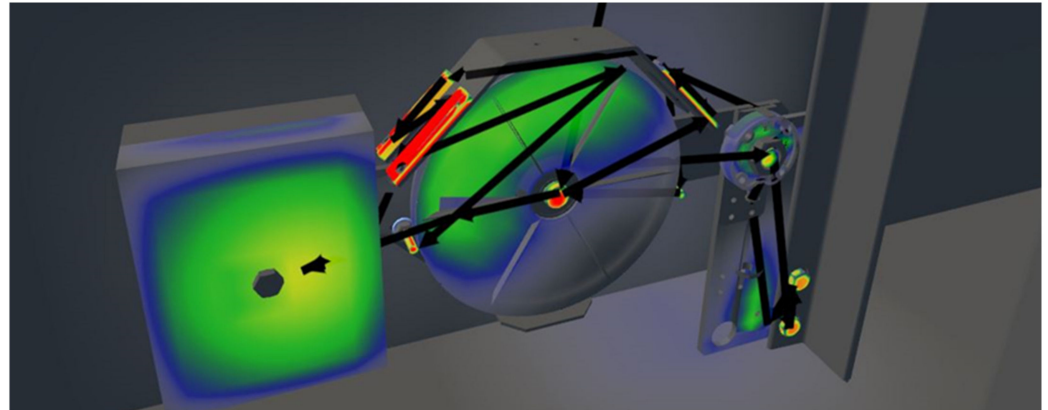


Figure 2. Gaze-data analysis visualized as a heatmap and applied for an industrial operation.

4.2. Audition

Most VR and AR headsets include audio input, and all have audio output support. Auditory interfaces can be utilized in XR, and they come in many forms [96]. Serafin et al. [32] carried out a recent survey of sonic interactions in VR.

The audio can be ambient, directional, musical, speech, or noise. Different forms of audio do not match every context or usage situation since audio can annoy, leak private information, or be hard to hear in a noisy environment. Audio can convey information about the environment in different ways. Spatial information can communicate about shapes and materials, and recognizable sounds help to better understand the environment. Human voices and speech can provide a lot of information, including emotional information.

Audio is public communication, i.e., everybody in the shared space can hear the sounds. However, most XR solutions utilize headphones. Audio can capture users' attention efficiently, even if their visual attention is somewhere else. Audio is thus often used for warnings. It is also temporal; once it is played out, the user cannot go back to it unless through an explicit interaction solution. This is different from typical visual information.

The main audio parameters are frequency, pitch, loudness, timbre, temporal structure, and spatial location. Continuous audio can support the awareness of users as people can spot subtle changes in repeating sounds. Audio can also help in fine control of systems (e.g., while driving a car, people monitor unconsciously the changes in speed via sound).

Auditory icons (c.f., real-world sounds) and earcons (the meaning must be learned) can be used in interfaces. Music can evoke emotions or communicate information. In addition, augmentation can be carried out in form of sounds. Finally, audio created by a user's actions is a significant part of multimodal interaction in some cases. By hearing things like button clicks, interaction with devices can be more efficient than without.

4.2.1. D Audio

Sound has a spatial aspect and humans can detect the direction of sound. Sound arrives at the two ears with slightly varying timing and intensity, and the human brain can estimate the direction of the sound source from them. Depending on the direction, the typical accuracy can be a few degrees (in front of the listener) or dozens of degrees (on the back). Rotating the head can help this process, and blind people typically develop better accuracy. Echoes and reverberation tell us about the size, shape, and materials of an environment. A stone cathedral sounds very different from a room full of pillows. For many uses, simple stereo mixing is enough.

VR simulates real or fictional environments and 3D audio can simulate sounds in the virtual space and act as an element of interaction. Suitable processing can provide realistic 3D sound. As users' heads are usually tracked, directional hearing is possible when the volume and timing of audio sent to left and right ears is adjusted. In AR, similar use of sound is possible, but the presence of real-world sounds must be considered in all designs. Head-related transfer function (HRTF) includes the effects of outer ears and head and improves the realism of 3D audio further. HRTF is slightly different for each person, but recent work on machine learning helps with its personalization [97] and on approaches providing results good enough for most uses [98].

4.2.2. Speech

Speech technologies include speech recognition (speech to text), speech synthesis (text to speech), speaker recognition and identification, and emotion recognition from speech. Speech technology has advanced greatly in the last decades and several companies provide speech and language technologies for numerous languages (e.g., Apple Siri, Google Assistant, Microsoft Cortana, Amazon Alexa, etc.). Deep neural networks play a significant role in it [52]. AI can make it difficult to discern if a remote discussion partner is a computer or a real person. Nonverbal aspects of speech such as psychophysiological states, emotions, intonations, pronunciation, accents, etc. can also be used. Speech-based interaction is now possible in most domains and situations.

There are several motives to use speech in XR environments. First, voice input provides both hands-free and eyes-free usage. This is particularly important in professional settings where users are typically focused on the task at hand, and the benefits of XR are most obvious in tasks where hands-on activities are performed, and user's hands and eyes are occupied. Typical examples include industrial installation and maintenance tasks [86]. Second, voice input is efficient and expressive. It helps to select from large sets of possible values in categories, and people have names for the things they need to talk about, and it can communicate abstract concepts and relations.

Speech is also a natural way for people to communicate and it is often preferred over other modalities. Especially communication with virtual characters is expected to take place in a spoken, natural manner. This requires, however, not only robust speech recognition, but also sophisticated dialogue modeling techniques. Luckily, modern (spoken) natural-language dialogue systems can be applied to XR. The resulting human-like conversational embodied characters can be efficient guides in many applications.

Speech is a relatively slow-output method if large amounts of content are played out. Most people can read faster than they can talk or listen. However, both listening and reading in XR environments differ greatly from a desktop or mobile environment. The rendering of text is more challenging in XR environments than on screen. Spoken output is often private as users wear headphones, and their eyes are typically focused on other tasks. Noisy environments can make both speech input and output challenging, but noise-canceling can reduce the effects of noise even completely.

Error management is critical in speech-based interaction. The user must be kept aware of how the system recognizes their speech and there must be ways to correct the situation. Error management may take significant time, reducing the interaction efficiency. Combining voice input with gestures, gaze, or other modalities can be efficient.

Natural language is weak when one needs to communicate about direction, distances, and detailed spatial relations. For navigation, using place and object names can be efficient via speech. Speech can also be used as an element of multimodal communication, where modalities together can overcome each other's weaknesses. The combination of speech and pointing gestures is a natural match (e.g., the "Put That There" system [14]).

Lately, Google Project Guideline (<https://blog.google/outreach-initiatives/accessibility/project-guideline/>) accessed on 1 December 2021, has helped a blind man to run unassisted. AI approaches are becoming commonplace also for other senses, and this may have seminal implications on the development of multimodal user interfaces.

4.2.3. Exhalation Interfaces

An exhalation interface is a specialized method of gestural interaction, albeit rarely used. It can provide a limitedly controlled hands-free interaction and it is almost always available. Blowing is useful, discreet, and quick when the user's hands are preoccupied with another task. It is typically based on microphones, thus being a part of auditory interaction. Breathing or blowing as an interaction method has been used for VR art, play, and entertainment (e.g., [99,100]). Numerous tiny microphones can be fitted onto a VR headset near the mouth. It has also been proposed for computer, mobile phone, or smartwatch UIs (e.g., [101]).

Sra et al. [99] proposed four breathing actions as a directly controlled input interaction for VR games. Their user study showed that breathing UI was found to provide a higher sense of presence and be more fun. They also proposed several design strategies for blowing with games. Chen et al. [101] used a headset microphone as a blowing sensor and classified the input to improve the measuring accuracy. Interaction is limited, as people cannot skillfully control many forms of blowing. Their user tests indicated that blowing improves users' interest and experience.

4.3. Haptics

The word "*haptics*" (Greek "*haptesthai*") relates to the sense of touch. Psychology and neuroscience study human touch sensing, specifically via kinesthetic (force/position) and cutaneous (tactile) receptors, associated with perception and manipulation. Kinesthetics is related to human movements, balance, and acceleration. Kinesthetic and tactile senses are closely intertwined. In HCI and XR interaction, haptics are generally specified as natural or simulated touch feedback between components, devices, humans and real, remote, or simulated environments, in various combinations [102]. This chapter looks at current technologies to create and deliver meaningful haptic feedback for XR interaction.

The sense of touch provides a wealth of information about our environment. Touch is delicately and marvelously built, consisting of a complex interconnected system and pervading the entire body. It comprises cutaneous inputs from various types of mechanoreceptors in the skin and kinesthetic inputs from the muscles, tendons, and joints that are closely integrated. This fine balance helps to provide a variety of information, such as shapes and textures of objects, the position of the limbs for balance, and proprioception of muscle [103] to manage the position and movement of the body.

Classical methods of looking at touch have segregated the various aspect into independent classifications. Earlier work by Asaga et al. [104], Kandel et al. [105], and Proske and Gandevia [106] predominantly categorize touch to consist of kinesthetics and tactility. According to Augstein and Neumayr [26], kinesthetics can be sub-divided into "proprioception", "equilibrioception", and "kinematics", whereas tactility encompasses artificial and natural stimulation (temperature, pressure, vibration, etc.) of the various mechanoreceptors in the skin. These taxonomies can be very useful in understanding clinical elements of the sensing system or its interaction parameters with various parts of the body, but as discussed by Oakley [107] and Barnett [108], usually more than one element of stimulation is utilized, so such segregation may not always be necessary.

We take a more holistic approach towards identifying interaction devices to stimulate various aspects of touch interaction. This chapter explores the role of both kinesthetics and tactility for developing tools and techniques suitable to provide haptic feedback for XR interaction and compares some existing devices that have been employed for it.

4.3.1. Utilizing Kinesthetics and Tactility to Create Meaningful Interaction

Robots, drones, vehicles, and other machinery can be controlled remotely or on-site using VR or AR (e.g., [109–115]). They enable new ways of work and can enhance safety, as dangerous places can be approached remotely. Most such systems provide visual and auditory feedback. Often, haptics are a part of these systems in order to provide better feedback and "feeling" of the operation.

As discussed, cutaneous senses (touch) are responsible for sensations based on the stimulation of receptors in the skin that are activated by, e.g., a touch on the forearm. Proprioception refers to the sense of the position of the limbs, while kinesthesia is related to the sense of movement. For example, during a handshake, information of grip-strength and up-and-down movement of the hand is received through proprioception and kinesthesia, while skin texture and subtle variation of the stretched skin are relayed through tactile sensation collected by the mechanoreceptors. Therefore, haptic devices artificially recreate the various touch sensations for a communication interface between humans and computers (e.g., in interactive computing, virtual worlds, and robot teleoperation). Specific mechanoreceptors in the skin need to be stimulated to produce expedient sensations of touch. To enhance realism and human performance in XR, artificial stimulation of various receptors in the body needs to be calibrated for specific application environments.

Tactile feedback is usually provided in direct contact with skin, which seems intuitive for the sensation of touch. There are several surveys on haptics in general (e.g., [103,116–118]), and recent surveys on haptics for VR [119]. Although some haptic systems highlighted in these surveys can provide both kinesthetic and tactile feedback, most systems focus on one or the other. This is because kinesthetic and tactile receptors may overlap or supersede each other during various interactions; therefore, the perception of complex signals may not be as intended. Such systems need to be dynamically adjusted to ensure that natural haptic feedback can be relayed to the user.

Another issue to consider is the fact that the fidelity of current tactile display technologies is very rudimentary compared with audiovisual displays or the capabilities and complexity of human tactile sensing [103]. The shortcomings amount to several orders of magnitude [120]. Many shortcuts and approximations for features such as device DoF, response time, workspace, input/output position and resolution, continuous and maximum force and stiffness, system latency, dexterity, and isotropy must be used to mass-produce haptic displays for general use. As haptics constitute a personalized method of interaction, any approach can create inconsistent outputs. Moreover, tactile interaction devices generate encoded signals only to the skin which may contribute towards lower information transfer rate and higher cognitive load as compared with visual and auditory modalities. Having said this, when other modalities are restricted, even low-resolution haptic feedback can improve user experience substantially. In any case, end-to-end communication needs to happen with minimum latency to ensure that the multimodal experience is natural and immersive across all the available modalities.

Additionally, an approach to develop meaningful touch interaction is to separate discriminative touch and emotional touch [121]. Humans rely on discriminative touch when manipulating objects or exploring their surroundings. Emotional touch becomes activated via a range of tactile social interactions such as grooming and nurturing. In this section, we focus on discriminative touch and discuss the most relevant techniques for it.

Haptic feedback has the potential to greatly enhance the immersion, performance, and quality of the XR interaction experience [103,119], but the current technology is still very limited. The lack of realistic and natural haptic feedback prevents deep immersion during object contact and manipulation. It can be disappointing and confusing to reach toward a visually accurate virtual object and then feel rudimentary (or no) tactile signals. Most conventional implementations of haptic devices provide only global vibrotactile feedback. In some cases, such devices are easy to use; however, they lack the resolution and functionality necessary for immersive XR interaction. In recent years, research and commercial haptic devices have been specifically developed for XR interaction. These include gloves, full-body suits, wearables with skin-integrated adhesive bandages or patches [122,123], dedicated hand-helds with dynamic tactile and kinesthetic surfaces, and customized HMDs with onboard tactile feedback (e.g., [82,124,125]). This section explores the currently available technologies and devices suitable for multimodal XR interaction and highlights possible future implementation paths.

4.3.2. Tactile Feedback Non-Contact Interaction

Mid-air gestures may partly replace hand-held controllers and create a more seamless interaction space. However, such non-contact interaction lacks tactile feedback, which can feel unnatural and can lead to uncertainty. The ability to ‘feel’ content in mid-air can address usability challenges with gestural interfaces [96].

Using calibrated ultrasound [126–128] or pneumatic transducer arrays [129], researchers have been successful at bringing complex 3D virtual objects to the physical space. Mid-air tactile feedback is unobtrusive and maintains the freedom of movement. It can improve user engagement and create a more immersive interaction experience.

Ultrasound haptics create acoustic radiation force, which produces small skin deformations and thus elicit the sensation of touch. They have been combined with some 3D displays (e.g., [130]) and VR systems [131–134]. The array can be placed, for example, on a table in front of the HMD user [131,133], or directly on the HMD [134], as depicted in Figure 3. The user can see objects through an HMD and feel them in mid-air.

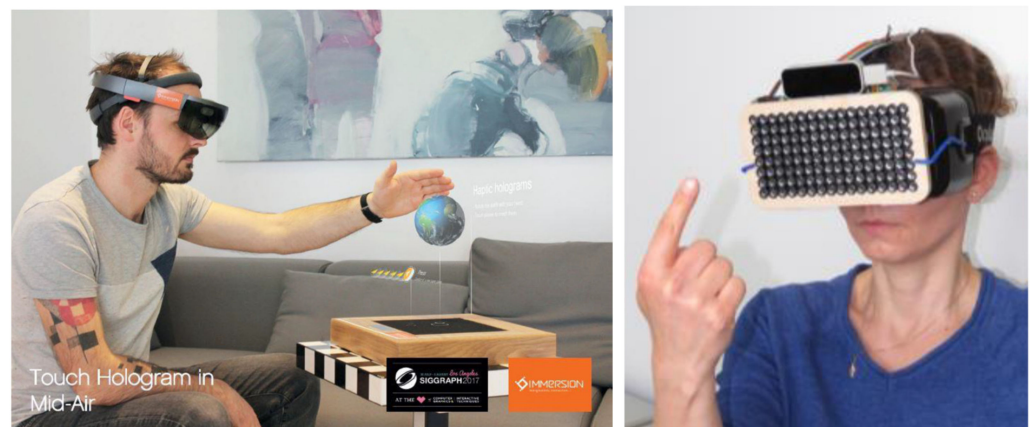


Figure 3. (Left) Mixed reality ultrasound haptics with the array in a fixed position [133]. Credit: Immersion. (Right) Mid-air ultrasound haptics with the array in front of a VR HMD [134]. Credit: I.R.

Research on the perception of mid-air ultrasound haptics suggests that the technique can provide similar properties as vibrotactile feedback on the perceptibility of frequencies. A good form of feedback for a button click is a single 0.2 s burst of 200-Hz modulated ultrasound [135]. The average localization error of a static point is 8.5 mm [136]. Linear shapes are easier to recognize than circular shapes.

Recent ultrasound actuators are only 1 mm thin. In the future, flexible printed circuit technology could enable transparent ultrasonic emitters [137], which could be pasted onto a visual display, and it could also bring down the cost significantly.

Surface-Based or Hand-Held Interaction

Most haptic systems rely on surface-based interaction to track and deliver vibrotactile feedback. Touchscreens and smart surfaces augment virtual objects and environments. Systems such as the Haptic Stylus [138] use a discrete point on touchscreen to locate and track interaction and to deliver tactile and kinesthetic feedback. It regulates a physical manipulandum linked to the virtual environment.

Smartphones have been utilized as either tools [139] or the core platform [140] for XR interaction. Initiatives such as Google’s Cardboard VR, Samsung’s Gear VR, and Apple’s AR kit extend a conventional smartphone to become a rudimentary XR platform. The Portal-ble system [141] utilizes the visual interface of a mobile device but uses rear-mounted sensors to track the user’s hands in real-time.

Other techniques such as “inFORCE” [142] use projection displays to overlay virtual objects to the user’s physical workspace, whereas selected tangible objects provide tactility

needed to complete the immersion. Similar techniques have been used in various mixed interaction environments from creating augmented eating experiences [143,144] to complex training procedures [145–147].

However, the techniques either require custom-designed interaction surfaces or novel hand-held devices that need to be mapped to the virtual environment in real-time to provide meaningful tactile or kinesthetic experiences in combination with other modalities. For example, the PlayStation 5 dual sense controller allows the user to experience various kinesthetic (through adapted trigger buttons) and tactile signals (through individually actuated wings). However, it needs a dedicated encoded haptics layer on top of the visual and auditory layers. Encoding the relevant information across the different modalities needs to be done in such a way that no one overpowers the rest and is fast enough.

4.3.3. Kinesthetic Feedback

Force feedback devices sense the position and movements of body parts and generate forces to alter the position of body parts (e.g., exoskeletons or moving platforms). Equilibrioception interfaces can sense or alter user's balance (e.g., Wii balance board, 4D cinema chairs). Kinematics interfaces can sense or alter user's acceleration (e.g., 4D cinema chairs, Voyager[®] VR chairs).

Audi Holoride [148] is a gaming/VR kinematics platform for car backseat passengers. Instead of causing motion sickness on top of carsickness, it takes advantage of the car stops, accelerations, bends, etc. and transports users to a virtual environment. The motions of the virtual world and the car are in synchrony, and hence the car becomes a locomotion platform. It even seems to reduce symptoms of motion sickness and nausea.

Wearable Device Interaction

Wearable devices using Bluetooth connections and integrated actuation components can relay basic tactile and kinesthetic information to the user. Gloves, rings, wristbands, and watches can serve as an always-on interface between events and triggers within the virtual environment and the physical space. In some cases (fitness tracker and rings), the onboard sensor can provide tracking or movement information which can be relayed to the XR device to improve the overall experience.

However, in most cases, their haptic output is very basic. As they and their batteries are small, they cannot reliably generate sensible feedback for extended sessions. Lastly, they utilize wireless connections that prioritize efficiency over low latency, which translates to unreliable haptic output.

Gloves that track the user's movements and deliver tactile/kinesthetic feedback to a very sensitive part of the body in real-time are an ideal tool for XR interaction. However, existing haptic gloves can restrict the user's natural motion, have limited output force, or be too bulky and heavy to be worn for extended interaction sessions [149].

Until recently, the only reliable force-feedback gloves were the CyberGrasp[™] of Immersion Corp (now CyberGlove Systems) [150] or the Master II of Rutgers University [151]. These and other similar prototype devices used DC motors, artificial muscles, shape memory alloys, or dielectric elastomers to create finger movements and tactile stimulation on the hand. However, as VR/AR interaction has become more mainstream, private companies have started to develop haptic gloves. Companies like HaptX, VRgluv, and Tesla have complex exoskeleton-based force feedback devices that are reliable and manageable for extended sessions.

These devices boost five finger interaction and wireless solutions with enhanced degrees of freedom (5–9 DoF) and can concurrently provide actuation and tracking. Exoskeletons ensure that sufficient force feedback can be generated for an immersive experience, but they have also some limitations. Firstly, most of the devices are work-in-progress and lack reliable driving. Custom haptic encoding and tactile layers need to be added to virtual environments to control each device. Moreover, most of the devices cannot sense environmental and user-specific forces during the interaction, which can make them

susceptible to overdriving the force feedback motor mechanism. Furthermore, even though some of these devices use lightweight alloys, the exoskeleton gloves still weigh 500 g (each) or more. Lastly, as most of these commercial products have not been extensively tested, results on user perception and long-term user experience are limited.

Wearable haptic devices have the potential to create more immersive feedback for XR interaction than previously. There is a growing academic and commercial interest in the field. Table 1 shows current and upcoming haptic gloves and their technical specifications. We excluded mocap gloves, which only track hand movements but do not give haptic feedback. A comprehensive list of mocap gloves can be found in a recent review [152].

Table 1. Currently available haptic gloves and their specifications (revised and updated from [153]).

Device	Type	Fingers	Wireless	Actuator	Force Feedback	Tactile Feedback	Hand Tracking	Active DoFs	Weight (g)	Price
Gloveone ¹	Glove	5	yes	Electromagnetic	no	yes	yes	10	na	499 €
AvatarVR ²	Glove	5	yes	Electromagnetic	no	yes	yes	10	na	1100 €
Senso Glove ³	Glove	5	yes	Electromagnetic	no	yes	yes	5	na	\$599
Cynteract ⁴	Glove	5	yes	Electromagnetic	yes	no	yes	5	na	na
Maestro ⁵	Glove	1	yes	Electromagnetic	no	yes	yes	5	590	na
GoTouchVR ⁶	Thimble	1	yes	Electromagnetic	yes	yes	yes	2	20	na
Exo-Glove Poly ⁷	Thimble	1–3	no	Soft Polymer (20–40N)	yes	yes	no	na	194	na
Prime X Series ⁸	Glove	5	yes	Electromagnetic	no	yes	yes	9	na	3999 €
CyberGrasp ⁹	Exosk.	5	no	Electromagnetic	yes	yes	yes	5	450	\$50,000
Dextarobotics ¹⁰	Exo-Gloves	5 + 5	yes	Electromagnetic	yes	yes	yes	11	320	\$12,000
HaptX ¹¹	Exosk.	5	no	Pneumatic	yes	yes	yes	na	na	na
VRgluv ¹²	Exosk.	5	yes	Electromagnetic	yes	yes	yes	5	na	\$579
Sense Glove 1 ¹³	Exosk.	5	yes	Electromagnetic	no	yes	yes	5	300	999 €
Sense Glove N ¹⁴	Exosk.	5	yes	Electromagnetic	yes	yes	yes	24	na	4500 €
HGlove ¹⁵	Exosk.	3	no	Electromagnetic	yes	yes	no	9	750	30,000 €
Noitom Hi5 ¹⁶	Glove	5	yes	Electromagnetic	no	yes	yes	9	105	\$999
Sensoryx VR Fr ¹⁷	Glove	5 + 5	yes	Electromagnetic	no	yes	yes	10	na	600 €
Tesla Glove ¹⁸	Exosk.	5	yes	Electromagnetic 3 × 3 display/finger	yes	yes	yes	9	300	\$5000
ThermoReal Plus ¹⁹	HMD, glove, sleeve	1 pt	yes	Thermal feedback (hot & cold stimulation)	no	yes	no	na	na	na
WEART Thimble ²⁰	Thimbles	3	yes	Electromagnetic and Thermal actuation	yes	yes	yes	na	na	\$3999

¹ Gloveone—<https://sensorialxr.com/>. ² AvatarVR haptic glove—<https://sensorialxr.com/>. ³ Senso DK3 glove—<https://senso.me/>. ⁴ Cynteract Rehabilitation Glove—<https://cynteract.com/>. ⁵ Maestro Gesture control and Drone operation Glove—<http://maestroglove.com/>. ⁶ GotouchVR—VRtouch glove—<https://www.gotouchvr.com/>. ⁷ Exo-Glove—<https://www.wevolver.com/wevolver.staff/exo.glove>. ⁸ Manus Prime X glove—<https://www.manus-vr.com/haptic-gloves>. ⁹ CyberGrasp's CyberForce CyberTouch 1 & 2, haptic gloves—<http://www.cyberglovesystems.com/cybergasp#specs>. ¹⁰ Dexmo Glove—<https://www.dextarobotics.com/>. ¹¹ HaptX DK2 Glove—<https://haptx.com/>. ¹² VRgluv—<https://www.vrgluv.com/enterprise>. ¹³ Sense glove DK1—<https://www.senseglove.com/>. ¹⁴ Sense Glove Nova VR Training glove—<https://www.senseglove.com/nova/>. ¹⁵ Haptions HGlove—<https://www.haption.com/fr/products-fr/hglove-fr.html>. ¹⁶ Noitom's Hi5 Haptic Glove—<https://www.noitom.com/hi5-vr-glove>. ¹⁷ SensoryX VRfree Tactile Gloves—<https://www.sensoryx.com/products/vrfreehaptics/>. ¹⁸ Tesla Exo Glove—<https://teslasuit.io/blog/teslasuit-introduces-its-brand-new-vr-gloves/>. ¹⁹ Thermoreal Glove, HMD, and sleeve—http://tegway.co/bbs/board.php?bo_table=product&sca=Game&lang=en. ²⁰ <https://www.weart.it/touchdiver/> (accessed on 1 December 2021).

Multi-Device and Full-Body Interaction

Another method of providing tactile and force feedback is to use multiple wearable devices that either interact with each other or communicate with the system to provide a comprehensive haptic experience. Some adaptations of this technology can create large-area stimulation through smart clothing [154] or through small puck-like devices (AHD, Autonomous Haptic Devices) that can be attached to any part of the body [155]. However, these techniques do not provide full-body tracking and interaction, which can be very useful in complex XR environments.

Full-body motion reconstruction and haptic output for XR enable better interaction and a higher level of immersion [156]. Such XR applications need to track user's full-body movements and provide real-time feedback throughout the body [157].

Slater and Wilbur [158] illustrated that XR immersion requires the entire virtual body, whereas presence requires the user to also identify with that virtual body (virtual self-image). In other words, for a high sense of presence, the users must recognize the movements of their virtual body as their own movements and be able to sense the interaction in real-time to achieve virtual immersion [159].

Research into full-body haptic stimulation is being carried out focusing on wearable clothing. Although there are several wearable tracking solutions (PrioVR, Perception Neuron 2.0, Smartsuit Pro, Xsens, etc.), most of them only focus on tracking full body or joint movements. Some startups and research labs are developing full-body vibrotactile and kinesthetic feedback using electromagnetic and microfluid actuation technologies. Some of them are still work in progress, whereas many are extensions of wearable devices (i.e., gloves).

As XR is becoming more mainstream, and as realistic haptic feedback is largely missing from XR, commercial interest has evolved, and there are many companies offering data suits. Table 2 shows a selected list of some of the currently available or upcoming full-body haptic suits and their technical specifications. Their technical level and prices vary significantly. The prices of haptic data suits are not yet suitable for average consumers, but they are useful for many professional, industrial, and training applications.

Locomotion Interfaces

Locomotion interfaces enable users to move in a virtual space and make them feel as if they indeed are moving from one place to another [160,161] when, in reality, they are just moving on a pneumatic or some other motion platform. VR can also trick the user visually to feel that they are walking straight when they are actually walking in circles. In a review by Boletsis [162], real walking, walking-in-place, controllers, and redirected walking were the most usual locomotion techniques.

Applications that require users to traverse virtual space often rely on UI tools such as teleportation, avatars, blinking, tunneling, etc. However, each technique may have its issues, especially concerning motion sickness. This is because in a typical implementation, the system is using head or hand-based locomotion. Recently, wearable devices and sensors coupled with HMDs provide hip-based tracking (e.g., DecaMove sensor [54]). A hip tracker makes inverse kinematics and body tracking easier and reduces motion sickness by using hip-based instead of hand or head-based locomotion.

Locomotion by walking supports the user's spatial understanding and helps wayfinding. Redirected walking is useful when the space is limited. It distorts the perceived environment so that the user feels to walk normally even though movements are truncated. Locomotion by walking can use treadmills, where the user can walk naturally while the floor below them is moving—even in any direction. Other ways are curved, slippery walking platforms (keeping the user at the same spot), where the user stepping in place, or sitting on a chair and wearing movement-sensing slippers (e.g., Cybershoes for Oculus Quest (<https://www.kickstarter.com/projects/cybershoes/cybershoes-for-oculus-quest>) accessed on 1 December 2021). Another approach is a "VR hamster ball" VirtuSphere, which fully surrounds the user and enables them to walk in any direction. Large robotic arms carrying a seated user is one way to create wild sensations of motion. Various locomotion platform products include Kat walk (https://www.kat-vr.com/pages/kat_walk_c) accessed on 1 December 2021), Infinadeck (<https://infinadeck.com/>) accessed on 1 December 2021), Omnideck (<https://www.omnifinity.se/>) accessed on 1 December 2021), Stewart platform (https://en.wikipedia.org/wiki/Stewart_platform) accessed on 1 December 2021), and 3dRudder (<https://www.3drudder.com/>) accessed on 1 December 2021).

Table 2. A list of currently available full-body haptic suits and their specification for VR/XR interaction.

Device	Coverage	Tracking	Feedback Type	Haptic Library	Dev. API	HMD Platform	Price (\$)
Nullspace VR ¹	32 independent zones (chest, abdomen, shoulders, arms & hands)	Yes	Vibrotactile (via Bluetooth & wired)	Fixed 117 effects with customization effects editor	Unity 3D	Multiplatform with 3rd party tracking	299 (vest only)
Tesla Suit ²	Full body suit, 80 embedded electrostatic channels	Yes, biometrics, 10 MC sensors	EM Vibrotactile (via Bluetooth)	Customizable: from 1–300 Hz, 1–260 ms and 0.15 A/chan.	Unity 5 and Unreal Engine 4	Multiplatform	13,000 (suit and gloves)
Axon VR/HaptX ³	Full body suit, gloves, and pads	Yes, magnetic motion tracking	Vibrotactile, micro-fluidic, force feedback exoskeleton	Temperature, vibration, motion, shape, and texture	Unity 5, Unreal Engine, Steam VR HaptX SDK	Multiplatform	N/A enterprise solution
Tactsuit x16 Tactsuit x40 (bHaptics) ⁴	Full body suit, gloves, pads, and feet guard, 70 zones (x16, x40)	Yes	Vibrotactile Bluetooth actuator versions: x16, x40	Customizable	Unity 5, Unreal Engine	Multiplatform	x16 = 299 x40 = 499 (pre-order price)
Rapture VR ⁵	HMD and Vest by uploadVR	Yes	Vibrotactile	N/A	Unity 5, Unreal Engine	Only for VOID Experience	N/A
Synesthesia Suit ⁶	Vest, gloves, and pads, 26 active zones	Yes	Vibrotactile	Customizable (triggered by audio feedback)	PS VR, Unity 5, Unreal Engine	Multiplatform including PlayStation	Under development
Hands Omni ⁷	Gloves, vest, pads, treadmill	Yes	Vibrotactile	Customizable	Unity 5, Unreal Engine	Multiplatform	Enterprise solution
HoloSuit ⁸	Glove, jacket & pants, 40 sensors, nine actuation elem.	Yes	Vibrotactile	Customizable	Unity, UE 4 and Motion-Builder	Multiplatform	Under development
Woojer ⁹	Vest (two actuators at sides, back, front) and Waist straps (one actuator)	No	Vibrotactile, audio (1–200 Hz) & TI wireless control	Customizable (triggered along-side audio feedback)	Used over any audio-based interface	Multiplatform/open	Vest 349 Strap 129
NeoSensory exoskin VR suit ¹⁰	Haptic jacket, vest & wrist, 32 actuation motors	No	Vibrotactile, adjustable freq. signals	Customizable	Custom SDK, Unity, and UE	Multiplatform	400 SDK and developer package
Shockwave (Kickstarter) ¹¹	Vest, and wearable straps (on the legs) with eight zones	Fullbody, eight IMUs (wireless)	64 vibrotactile feedback (HD haptics)	Customizable	Unity, Unreal Engine 4	Most VR headsets (requires dev. support)	300 for Kickstarter

¹ VRFocus—NullSpace VR Suit—<https://www.vrfocus.com/tag/nullspace-vr/>. ² Modular Tesla Suit—<https://teslasuit.io/>. ³ HaptX (previously known as AxonVR) Gloves, pads, and vest—<https://haptx.com/virtual-reality/>. ⁴ bHaptics’s fullbody suit—<https://www.bhaptics.com/>. ⁵ Rapture VR’s “The Void” headset and vest—<https://uploadvr.com/void-rapture-ted/>. ⁶ RezInfinite’s Synesthesia Suit—<https://rezinfinite.com/synesthesia-suit/>. ⁷ Hands Omni & Omni Addons—<https://www.virtuix.com/>. ⁸ HoloSuit: Indian Military Robotic Drone Suit—<https://www.holosuit.com/>. ⁹ Woojer Vest and Strap for VR interaction—<https://www.woojer.com/>. ¹⁰ Neosensory ExoSkin Vest and Strap—<https://neosensory.com/>. ¹¹ Shockwave Multipurpose suit—<https://www.kickstarter.com/projects/1110033869/shockwave> (accessed on 1 December 2021).

Tongue Interfaces

Tongue gestures can be used for discreet, hands-free control, and they match well with XR systems. The tongue is a fast and precise organ, comparable to the head, hands, or eyes for the purposes of user interaction. Tongue UI is currently used mostly as an experimental or assistive technology. Tongue movements, including sticking it out of the mouth or onto the cheek, can be tracked with a camera. Other approaches are, e.g., an array of pressure sensors on the user’s cheek or on a mask [163], a wearable device around ears to read tongue muscle signals, EMG signals detected at the underside of the jaw, or intraoral electrode or capacitive touch sensors on a thin mouthguard [164].

Dynamic Physical Environments and Shape-Changing Objects

Dynamic physical environments can be tethered to virtual environments to enhance immersion. The physical elements relay physical forces to virtual actions. Various adaptations of the Haptic Floor [165,166] are prime examples of this. Individual segments of the floor pivot or vibrate to support visual and auditory feedback. Each segment of the floor acts as an individual pixel within the interaction scheme, and it provides meaningful tactile and kinesthetic information enhancing the experience.

Other dynamic environments track the user's physical and virtual movements and supplement auxiliary support or cues with artificial forces. The ZoomWalls [167] creates dynamically adjustable walls that simulate a haptic infrastructure for room-scale VR. Multiple movable wall segments track and follow the user, and, if needed, orient the user to their artificial surroundings by simulating virtual walls, doors, and walkways.

Another such environmental haptics feedback is the CoVR system [168], which utilizes a robotic interface to provide strong kinesthetic feedback (100 N) in a room-scale VR arena. It consists of a physical column mounted on a 2D Cartesian ceiling robot (XY displacements) with the capacity of resisting body-scaled users' actions such as pushing or leaning and acting on the users by pulling or transporting them. The system is also able to carry multiple potentially heavy objects (up to 80 kg) which users can freely manipulate within a joint interaction environment. However, in both cases, virtual and physical tracking plays a crucial role. Several elements are needed to follow the interacting user in real-time. This can be an issue if the HMD has limited or no visual passthrough capabilities, as users may bump into these objects or each other.

However, dynamically adjustable environmental interaction is a new research area and novel solutions may enhance the usability and efficiency of similar approaches. Researchers from Microsoft have created the PIVOT system [169], a wrist-worn haptic device that renders virtual objects into the user's hand on demand. PIVOT uses actuated joints that pivot a haptic handle into and out of the user's hand, rendering the haptic sensations of grasping, catching, or throwing an object anywhere in space. Unlike existing hand-held haptic devices and haptic gloves, PIVOT leaves the user's palm free when not in use. PIVOT also enables rendering forces acting on the held virtual objects, such as gravity, inertia, or air-drag, by actively driving its motor while the user is firmly holding the handle. Authors suggest that wearing a PIVOT device on both hands can add haptic feedback to bimanual interaction, such as lifting larger objects.

4.4. Olfaction

The sense of smell is known as a chemical sense because it relies on chemical transduction. The sense of smell and scents in HCI are more difficult to digitize compared with sounds and light (Obrist et al., 2016). Scents have been underrepresented in VR [19]. However, the technology for enabling scents in XR is advancing rapidly. An increasing body of research shows that scents affect the user in numerous ways. For example, scents can enrich the user experience [19,170], increase immersion [171], sense of reality [172] and presence [173], affect emotion, learning, memory, and task performance [174], and enhance training experience, e.g., in shopping, entertainment, and simulators [175–177].

The easiest way to deliver scents to an XR user is to utilize ambient scent [178] which is lingering in the environment [179]. It can be created with various scent-emitting devices, but it is difficult to rapidly change one scent to another or change its intensity unless the space is small (e.g., the sensory reality pods by Sensiks Inc.). Active directing of scented air with a fan or air cannon enables a little more control.

Scented air can be directed from a remote scent display to an HMD with tubes [180]. Alternatively, it is possible to produce more compact scent displays that are attached to a VR controller [181], worn on the user's body [182], or connected directly to the HMD [173,183–185]. The advantages of wearable scent display typically include better spatial, temporal, and quantitative control because the scents can be delivered near or in the nostrils. Figure 4 illustrates these more precise approaches to deliver scents and shows a recent scent display prototype on an HMD.

Before scents can be delivered to a user, they must be vaporized from a stocked form of odor material. Typical solutions are natural vaporization, accelerated vaporization by airflow, heating, and atomization [186]. A limiting factor in all scent displays is the number of possible scents that can be created, typically 1–15 scents. Recent research indicates that it could be possible to synthesize scents on demand by creating a mixture of odorants that humans perceive as the original scent [187]. This is a major step towards technology that

digitizes and reproduces scents, similarly to what is possible by recording sounds and taking photographs.

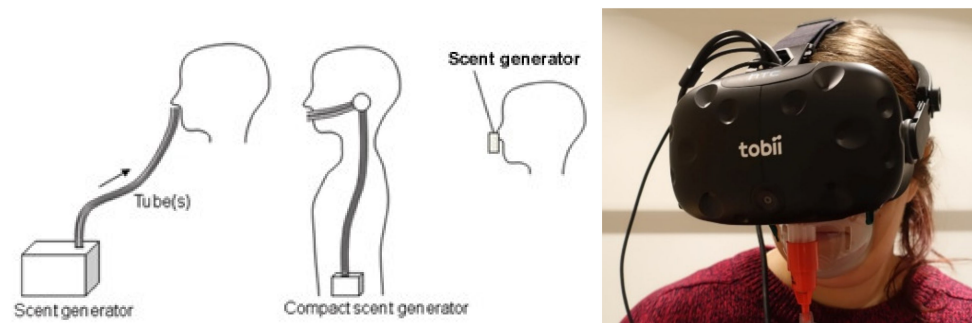


Figure 4. Alternative scent delivery methods suitable for XR [186]. An example of a scent display suitable for XR applications. Credit: J.R.

4.5. Gustation

Taste (gustation) is also a chemical sense and is even less often used than olfaction, especially in XR [177]. Taste perception is often a multimodal sensation composed of chemical substance, sound, smell, and haptic sensations [188]. Taste perception largely originates from the sense of smell [189] because scents travel through orthonasal (sniff) and retronasal (mouth) airways while eating. Many XR applications such as those aimed for augmenting flavor perception have therefore used scents [185,190] instead of attempting to stimulate the sense of taste directly. It is also possible to develop technology for stimulating specifically the sense of taste, targeting one or more of the five basic taste sensations that taste buds can sense: salty, sweet, bitter, sour, and umami.

The three main approaches to creating taste sensations are ingesting chemicals, sensing electrical stimulation on the tongue, and using thermal stimulation [191]. TasteScreen [192] used a questionable method of users licking a computer screen with a layer of flavoring chemical. Vocktail [193] used a glass with embedded electronics for creating electrical stimulation at the tip of the tongue. There is significant interpersonal variation in the robustness of taste perception resulting from electrical stimulation [194,195]. Therefore, it often uses simultaneous stimulation of other senses. The third approach, thermal stimulation, was used in the Affecting Tumbler [196], a cup designed for changing the flavor perception of a drink by heating the skin around the user's nose.

Even though initial empirical findings have suggested that the prototypes can alter taste perceptions (e.g., [196]), more research is needed. Taste stimulation typically requires other supporting modalities to create applications that are meaningful, function robustly, and are pleasant to use. Compared with other modalities, we are still in the early stages of development for taste [176]. However, HMDs and other wearable devices for XR offer a good technological platform for further development.

4.6. Brain-Computer Interface (BCI)

Many potential future UIs are based on measured bioelectrical signals near or in brains (BCI), or in tissues, organs or the nervous system, or feedback given through bioelectrical signals. These include electroencephalogram (EEG), electrocardiogram (ECG), electromyogram (EMG), electrooculography (EOG), mechanomyogram (MMG), magnetoencephalogram (MEG) and galvanic skin response (GSR).

The ultimate interface would be a “mind-reading” direct link between a user's thoughts and a computer. BCI is two-way (input and output) communication between the brain and a device, unlike one-way neuromodulation. BCI is not a sense per se, but it bypasses all human sensors and nerves and stimulates the brain directly and non-invasively with various signals to create synthetic sensations (output) or to interpret electric brain signals (input). Feeding visual, auditory, haptic, taste, smell, or other sensations directly

to the brain could open up entirely new avenues for XR, but this is presumably still far in the future. To illustrate the potential, there is an intriguing sci-fi movie, *Brainstorm* (<https://www.youtube.com/watch?v=cOGAEA4xJE&t=1385>) accessed on 1 December 2021 (1983), describing the neuromodulation and BCI sensory feeding. Human senses can also be bypassed in several points of action [197]. Figure 5 depicts some possible access points to create a visual sensation.

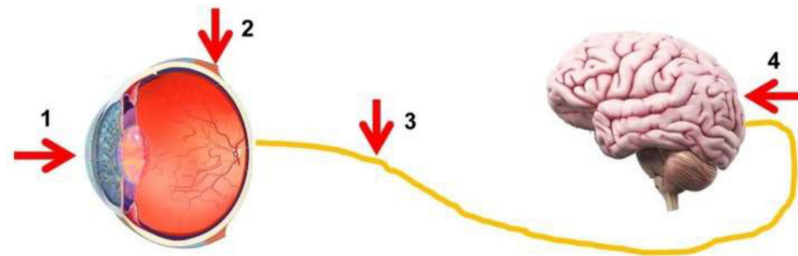


Figure 5. Light is normally delivered through the lens and pupil (1) to the retina. Alternative non-invasive ways are to bring light directly through the tissue to the retina (2), or to elicit phosphenes in the optic nerve (3) or in the visual cortex of brain (4). [197].

BCI input can be based on surgically implanted prostheses (which is usually more effective), or external, non-invasive devices such as EEG sensors [198]. There are several non-invasive neuroimaging methods, such as electroencephalography (EEG), functional magnetic resonance imaging (fMRI), and functional near-infrared spectroscopy (fNIRS). EEG is currently the most widely used for VR. Several non-invasive commercial devices can read human brain activities (e.g., Emotiv, NextMind, MindWave, Neuroware, Open BCI, Brain Co, Neurosity, iDun, Paradromics, Looxidlabs, or NeuroSky). They use it as an input to perform actions with computers or other devices. At least Looxid Labs already sells EEG headsets that can be retrofitted to HMDs. However, their capabilities are extremely limited, and they suit only very simple tasks.

HMDs can have various physiological sensors close to the skin, eyes, and skull. The PhysioHMD system [199] (see Figure 6) merges several types of biometric sensors to an HMD and collects sEMG, EEG, EDA, ECG, and eye-tracking data. EEGlass [200] is a prototype of an HMD employing EEG for BCI. Luong et al. [201] estimated the mental workload of VR applications in real-time with the aid of physiological sensors embedded in the HMD. Barde et al. [202] carried out a review on recording and employing the user's neural activity in virtual environments.

Neuralink Inc. (<http://www.neuralink.com/>) accessed on 1 December 2021 has demonstrated Gertrude, a pig with a coin-sized computer chip implant, and recently also a monkey playing video games using its mind. Human experiments are due soon, and they intend to achieve a variety of things, e.g., to solve ailments such as memory or hearing loss, depression, insomnia, or restoring some movement to people with quadriplegia. Ultimately, they hope to fuse humankind with artificial intelligence.

The feedback (output) can be administered through brain stimulation using various methods. Transcranial magnetic stimulation (TMS) and transcranial-focused ultrasound stimulation (tFUS, TUS) are some possible non-invasive feedback methods. TMS has been used, for example, for helping a blindfolded user to navigate a 2D computer game only with direct brain stimulation [203]. tFUS has superior spatial resolution and the ability to reach deep brain areas. TMS, tFUS [204], and other brain stimulation methods can also elicit synthetic visual perception (phosphenes) when applied onto the visual cortex, even though phosphenes are very coarse with current methods.



Figure 6. PhysioHMD records physiological data through the contact with the skin (credit: Guillermo Bernal, <https://www.media.mit.edu/projects/physiohmd/overview/>, accessed on 1 December 2021, image cropped).

BCI-XR research is high-risk, high-reward work. Potentially, it is a very disruptive technology, and closely related to human augmentation [93,205]. However, the few conducted experiments on creating realistic synthetic sensations have very underwhelming results so far. On the other hand, input through EEG or EMG has achieved better, but very limited results.

BCI is still very rudimentary in its capabilities, and it is used mostly for special purposes such as implanted aid for paralyzed people or to prevent tremors caused by Parkinson's disease. One practical line of research is to create a synthetic vision for blind people. Recently, machine learning has helped, for example, to classify mental or emotional states. BCI will have tremendous challenges with ethical and privacy issues.

4.7. Galvanism

Electromyography (EMG) and other biometric signal interfaces read the electrical activity from muscles and EOG reads specifically the electrical activity of the muscles near the eyes. This can be used for input in XR systems. EMG sensors can be attached to the user's muscles (e.g., Thalmic Labs' discontinued Myo gesture control armband), wearables, or datasuits. Bioelectrical signals can also be used for feedback. For example, Sra et al. [206] added proprioceptive feedback to VR experiences using galvanic vestibular stimulation.

Some HMDs have embedded EMG or EOG physiological sensors. The beforementioned PhysioHMD system [199] merges several types of biometric sensors to an HMD, including EMG. Barde et al. [202] carried out a review on recording the neural activity for their use in virtual environments.

5. Discussion

Can XR systems in the year 2040 offer full immersion for all the user's senses? Near-perfect visuals and audio will be straightforward to implement, and effective kinesthetic, haptic, and scent technologies are maturing, but there will be grand challenges and probably insurmountable obstacles to produce, e.g., unencumbered locomotion or gustation systems. Neuromodulation or BCI may allow a simpler way to produce different sensations. Yet, again, this is task-, context-, and cost-dependent.

Perfectly seamless, unencumbered, and fluid interaction in HCI is not always required or possible for all purposes. People use all kinds of instruments, devices, and tools for various tasks in real life, so this style of interaction should also work well in XR.

Furthermore, the camera-based Kinect gesture sensor was unencumbered, but it never became a long-lasting success story.

Even though the human perceptual system is delicately and meticulously designed, it has some perceptual shortcomings which can be taken advantage of. Many shortcuts, tricks, and approximations can be used to create satisfactory multimodal interaction that makes the audiences believe they are seeing magic.

There are many emerging or disruptive technologies which are not directly interaction methods or modalities, per se, but which may have immense implications for XR interaction and HMD technology. Some of these include XR chatbots (agents), battery technology, nanotechnology, miniaturization of sensors and actuators, IoT, and robots. New materials, flexible sensor and/or actuator patches on the skin or earbuds, or rollable displays might also be useful for unobtrusive HMDs. Distributed ledgers, social media, social gaming in virtual environments (c.f., Facebook, Second Life), neuromodulation, etc. also have potential for XR. Some of these challenges and opportunities were discussed at greater length by Spicer et al. [207].

Human-computer integration and [205,208] and human augmentation [93] are paradigms to extend human abilities and senses seamlessly through technology. In a way, human augmentation involves extreme use of multimodal interaction, not just using a handful of modalities and interfacing with a device, but using a vast number of advanced sensor and actuator technologies that generate a large volume of data which are presented concisely and coherently, integrating them with the user (e.g., [209]). Its processing requires, e.g., machine learning, signal processing, computer vision, and statistics.

Augmentation technologies provide new, smart, and stunning experiences in unobtrusive ways. Lightweight, comfortable, and yet efficient augmentation can be useful and have a significant impact on various human activities. For a person with deteriorated vision, smart glasses can enhance visual information or turn it into speech. The glasses can also augment cognition and support memory. Special clothes can provide augmented skin that senses the touches and movements assisted by a therapist and integrates them with a training program stored in the smart glasses. Embedded sensors in the clothes can notice an imbalance in movements and save information of physical reactions so that the therapy instructions can be adapted accordingly [210]. Physical augmentation is also possible, e.g., with lightweight exoskeletons or robotic prostheses, which amplify the user's physical strength or endurance.

XR has also many social [211], societal [212], health [213], ergonomic [214], educational [215], and other issues to be solved [216,217]. In the current pandemic, people have had to consider hygiene issues [218] of shared XR hardware. Like any new technology, XR is also potentially widening the digital divide between countries, cultures, and people [219]. Additionally, XR may become a new addiction and a form of escapism [220].

Biometric methods such as iris scanning would be easy to embed to an HMD and fingerprint reading to data gloves, and thus personalize and authorize selected content. On the other hand, biometric technologies can be very intimate and thus they have grave ethical and privacy concerns. For example, unscrupulous companies, or criminals such as authoritarian governments or mafia organizations could spy on people and exploit them in multiple ways and, if they can, they will. As Mark Pesce [221] put it: *"The concern here is obvious: When it [Facebook Aria] comes to market in a few years, these glasses will transform their users into data-gathering minions for Facebook. Tens, then hundreds of millions of these AR spectacles will be mapping the contours of the world, along with all of its people, pets, possessions, and peccadilloes. The prospect of such intensive surveillance at planetary scale poses some tough questions about who will be doing all this watching and why"*.

Some multimodal technologies and 3D UIs could become mainstream features and applications of XR in a few years. They have a vast number of applications in industry, health care, entertainment, design, architecture, and beyond. They have the potential to increase the use of XR in many areas (e.g., [34,193]).

IBM's director of automation research, Charles DeCarlo, presented a stunning vision of immersive telepresence in 1968 [222]. It described a home wherein a photorealistic live remote scene was projected onto curved screens and realistic audio completed the immersion. The "immersive home theatre" was used for telepresence and teleconferencing. The author foresaw VR replacing reality, but current XR technology is still struggling to fulfill that vision. Telepresence and remote collaborations have great potential for XR technology, and one of the early demos for Microsoft HoloLens demonstrated this in the form of remote expert guiding in maintenance tasks.

Other industrial tasks, such as assembly and installation [6], have already been applied in everyday work life. The current focus is on education (e.g., learning of professional skills) in VR environments [219]. In future industrial settings, one of the most promising ways to utilize XR technology may be to develop a complete VR—AR continuum, in which people first learn skills in VR and then utilize AR in the field operations when those learned skills are applied in practice [86]. The industry also needs seamless co-operation between people from different locations to solve common problems [55]. Here, the key focus for research is on collaborative XR, which can be seen as one of the future drivers for XR technology.

Learning and training can be more efficient with multimodal feedback, as the richer experiences are more memorable, making the training more realistic and thus faster to learn, and can even enable new features [174]. Training [223,224] can become more immersive and transfer knowledge from simulations to real life so that trainees can recall the correct procedures in real situations. With haptic feedback and better tracking of users' movements, benefits of embodied learning and training can be realized (e.g., [129,176]) and combined with other benefits of XR in training [225], like safe learning of hazardous tasks in VR and context-specific learning with AR [22,30].

Rehabilitation can also benefit from multimodal feedback for many of the same reasons [31,226]. Exercises to regain motor capabilities have utilized haptics in many research prototypes (e.g., [227]) and with developing solutions this could become more efficient and more widely available.

New interaction solutions can also support visually challenged people in new ways. Automatic processing of a camera-based image stream from the environment can further improve this. Examples include user-controlled or semiautomatic enlarging of the relevant part of the scenery to support those with limited visual acuity and using other modalities to present color information for people with color blindness.

For people with motor impairments, haptic feedback may be used to ease physical actions in a virtual environment, enabling XR interaction via BCI and replacing some motor actions with audio-based solutions. Cross-modal presentation of information can be used to support people with limitations with one or more senses [162].

The more fluent use of XR can make tasks more efficient. This can be accomplished with more natural interactions and by enabling alternative ways to do and especially perceive things [223]. This rich HCI can improve also XR-mediated collaboration and human-human communication and can make XR use safer when information for different senses can help users to maintain awareness.

Finally, multimodal solutions can increase the level of immersion [228,229] and sometimes also task performance [229,230]. This enables more immersive entertainment, and the abovementioned other uses can provide richer experiences leading to many of the mentioned benefits.

The technologies must also be fit for human perception and cognition. As stated by Gardony et al. [87]: *"When done poorly, MR experiences frustrate users both mentally and physically leading to cognitive load, divided attention, ocular fatigue, and even nausea. It is easy to blame poor or immature technology but often a general lack of understanding of human perception, human-computer interaction, design principles, and the needs of real users underlie poor MR experiences. ..., if the perceptual and cognitive capacities of the human user are not fully considered then technological advancement alone will not ensure MR technologies proliferate across society"*.

In addition to various multimodal technologies, other issues will also have an impact on future XR technology, usage, and applications—directly or indirectly. Social trends, cultural issues, economy, business, politics, geopolitics, demographics, pandemics, etc. will alter sentiment, prosperity, innovation, and many other things, and these will have an indirect impact on the development and usage of XR. As with any technology, market penetration of XR depends on a wide range of issues such as revenue, marketing, consumer needs and acceptance, IPR, backward compatibility, price, manufacturability, timing, luck, etc.

6. Conclusions

Multimodal interaction can revolutionize the usability of XR, but interaction methods developed for the PC desktop context are not usually effective in XR. New concepts, paradigms, and metaphors are needed for advanced interaction.

Most of the multimodal interaction technologies are still immature for universal use. All the current approaches for multimodal interaction have their strengths and weaknesses. Additionally, no single approach is likely to dominate. The applied technology on a specific use case will largely depend on the context and application. Therefore, XR devices should support a wide range of interaction modalities from which to choose when developing interfaces for different contexts of use.

What a typical XR system will look like 10 or 20 years from now is an open question, but it will likely change substantially. It may not even make any sense to talk about XR systems then anymore, in the same way as multimedia PC is an obsolete term nowadays, as all PCs support multimedia. We may be wearing a 6G XR-capable Flex-Communicator in our pocket, on our hand, or near our eyes, depending on the context of use. The added value of XR lies in augmenting and assisting us in our daily routines and special moments.

There are some especially promising but underused emerging technologies and research avenues for multimodal XR interaction technologies. They have significant potential for wide use and to become standard technologies in HMDs. These include various forms of haptic interfaces, as the sense of touch is an important element of interaction in the real world, but it is currently underused in interaction with technology. Gaze is useful and important for HCI, and it is already becoming a standard element in HMDs. BCI is still far out from most practical uses, but it has tremendous potential if more effective technologies for it are developed.

In conclusion, future XR technologies and proposed “Metaverses” may impact our work and daily life significantly. They might also make our world more accessible (e.g., through telepresence), but they could also create new accessibility barriers and inequality. Well-considered multimodal experiences, emerging technologies, and improved interaction methods may become elements of success for next-generation XR systems.

Author Contributions: Conceptualization, I.R.; methodology, I.R. and R.R.; investigation, all authors; writing—original draft preparation, all authors; writing—review and editing, all authors; supervision, I.R.; project administration, I.R.; funding acquisition, R.R. and M.T. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by Business Finland, grant numbers 3913/31/2019 and 1316/31/2021, and the Academy of Finland, grant number 316804.

Institutional Review Board Statement: Not applicable. Ethical review and approval were waived for this study, due to the study being based on literature review where no human participants were involved.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable; Public data review article.

Conflicts of Interest: The authors declare no conflict of interest.

References

- Milgram, P.; Kishino, F. Taxonomy of mixed reality visual displays. *Inst. Electron. Inf. Commun. Eng. Trans. Inf. Syst.* **1994**, E77-D, 1321–1329.
- LaValle, S. *Virtual Reality*; National Programme on Technology Enhanced Learning: Bombay, India, 2016.
- Benzie, P.; Watson, J.; Surman, P.; Rakkolainen, I.; Hopf, K.; Urey, H.; Sainov, V.; Von Kopylow, C. A survey of 3DTV displays: Techniques and technologies. *Inst. Electr. Electron. Eng. Trans. Circuits Syst. Video Technol.* **2007**, *17*, 1647–1657. [[CrossRef](#)]
- Cruz-Neira, C.; Sandin, D.J.; DeFanti, T.A. Surround-screen projection-based virtual reality: The design and implementation of the CAVE. In Proceedings of the 20th Annual Conference on Computer Graphics and Interactive Techniques, SIGGRAPH 1993, Anaheim, CA, USA, 2–6 August 1993; pp. 135–142.
- Rakkolainen, I.; Sand, A.; Palovuori, K. Midair User Interfaces Employing Particle Screens. *Inst. Electr. Electron. Eng. Comput. Graph. Appl.* **2015**, *35*, 96–102. [[CrossRef](#)] [[PubMed](#)]
- Bimber, O.; Raskar, R. *Spatial Augmented Reality: Merging Real and Virtual Worlds*; AK Peters: Wellesley, MA, USA, 2005; ISBN 9781439864944.
- Arksey, H.; O'Malley, L. Scoping studies: Towards a methodological framework. *Int. J. Soc. Res. Methodol.* **2005**, *8*, 19–32. [[CrossRef](#)]
- Colquhoun, H.L.; Levac, D.; O'Brien, K.K.; Straus, S.; Tricco, A.C.; Perrier, L.; Kastner, M.; Moher, D. Scoping reviews: Time for clarity in definition, methods, and reporting. *J. Clin. Epidemiol.* **2014**, *67*, 1291–1294. [[CrossRef](#)] [[PubMed](#)]
- Peters, M.D.J.; Godfrey, C.M.; Khalil, H.; McInerney, P.; Parker, D.; Soares, C.B. Guidance for conducting systematic scoping reviews. *Int. J. Evid. Based Healthc.* **2015**, *13*, 141–146. [[CrossRef](#)] [[PubMed](#)]
- Raisamo, R. *Multimodal Human-Computer Interaction: A Constructive and Empirical Study*; University of Tampere: Tampere, Finland, 1999.
- Spence, C. Multisensory contributions to affective touch. *Curr. Opin. Behav. Sci.* **2022**, *43*, 40–45. [[CrossRef](#)]
- Engelbart, D. A demonstration at AFIPS. In Proceedings of the Fall Joint Computer Conference, San Francisco, CA, USA, 9–11 December 1968.
- Sutherland, I.E. A head-mounted three dimensional display. In Proceedings of the Fall Joint Computer Conference, San Francisco, CA, USA, 9–11 December 1968; Association for Computing Machinery Press: New York, NY, USA, 1968; Volume 3, p. 757.
- Bolt, R.A. "Put-that-there": Voice and gesture at the graphics interface. In Proceedings of the 7th Annual Conference on Computer Graphics and Interactive Techniques, Seattle, WA, USA, 14–18 July 1980; pp. 262–270. [[CrossRef](#)]
- Rekimoto, J.; Nagao, K. The world through the computer. In Proceedings of the 8th Annual Association for Computing Machinery Symposium on User Interface and Software Technology—UIST'95, Pittsburgh, PA, USA, 15–17 November 1995; Association for Computing Machinery Press: New York, NY, USA, 1995; pp. 29–36.
- Feiner, S.; MacIntyre, B.; Höllerer, T.; Webster, A. A touring machine: Prototyping 3D mobile augmented reality systems for exploring the urban environment. *Pers. Ubiquitous Comput.* **1997**, *1*, 208–217. [[CrossRef](#)]
- Van Dam, A. Post-WIMP user interfaces. *Commun. Assoc. Comput. Mach.* **1997**, *40*, 63–67. [[CrossRef](#)]
- Turk, M. Multimodal interaction: A review. *Pattern Recognit. Lett.* **2014**, *36*, 189–195. [[CrossRef](#)]
- LaViola, J.J., Jr.; Kruijff, E.; Bowman, D.; Poupyrev, I.P.; McMahan, R.P. *3D User Interfaces: Theory and Practice*, 2nd ed.; Addison-Wesley: Boston, MA, USA, 2017.
- Steed, A.; Takala, T.M.; Archer, D.; Lages, W.; Lindeman, R.W. Directions for 3D User Interface Research from Consumer VR Games. *Inst. Electr. Electron. Eng. Trans. Vis. Comput. Graph.* **2021**, *27*, 4171–4182. [[CrossRef](#)]
- Jerald, J. *The VR Book: Human-Centered Design for Virtual Reality*; Morgan & Claypool: San Rafael, CA, USA, 2015; ISBN 9781970001129.
- Rash, C.; Russo, M.; Letowski, T.; Schmeisser, E. *Helmet-Mounted Displays: Sensation, Perception and Cognition Issues*; Army Aeromedical Research Laboratory: Fort Rucker, AL, USA, 2009.
- Schmalstieg, D.; Höllerer, T. *Augmented Reality: Principles and Practice*; Addison-Wesley Professional: Boston, MA, USA, 2016.
- Billinghurst, M.; Clark, A.; Lee, G. A Survey of Augmented Reality. *Found. Trends[®] Hum.-Comput. Interact.* **2015**, *8*, 73–272. [[CrossRef](#)]
- Rubio-Tamayo, J.L.; Barrio, M.G.; García, F.G. Immersive environments and virtual reality: Systematic review and advances in communication, interaction and simulation. *Multimodal Technol. Interact.* **2017**, *1*, 21. [[CrossRef](#)]
- Augstein, M.; Neumayr, T. A Human-Centered Taxonomy of Interaction Modalities and Devices. *Interact. Comput.* **2019**, *31*, 27–58. [[CrossRef](#)]
- Blattner, M.M.; Glinert, E.P. Multimodal integration. *Inst. Electr. Electron. Eng. Multimed.* **1996**, *3*, 14–24. [[CrossRef](#)]
- Benoit, C.; Martin, J.; Pelachaud, C.; Schomaker, L.; Suhm, B. Audio-visual and Multimodal Speech Systems. In *Handbook of Standards and Resources for Spoken Language Systems-Supplement*; Kluwer: Dordrecht, The Netherlands, 2000; Volume 500, pp. 1–95.
- Koutsabasis, P.; Vogiatzidakis, P. Empirical Research in Mid-Air Interaction: A Systematic Review. *Int. J. Hum. Comput. Interact.* **2019**, *35*, 1747–1768. [[CrossRef](#)]
- Mewes, A.; Hensen, B.; Wacker, F.; Hansen, C. Touchless interaction with software in interventional radiology and surgery: A systematic literature review. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 291–305. [[CrossRef](#)]
- Kim, J.; Laine, T.; Åhlund, C. Multimodal Interaction Systems Based on Internet of Things and Augmented Reality: A Systematic Literature Review. *Appl. Sci.* **2021**, *11*, 1738. [[CrossRef](#)]

32. Serafin, S.; Geronazzo, M.; Erkut, C.; Nilsson, N.C.; Nordahl, R. Sonic Interactions in Virtual Reality: State of the Art, Current Challenges, and Future Directions. *Inst. Electr. Electron. Eng. Comput. Graph. Appl.* **2018**, *38*, 31–43. [[CrossRef](#)]
33. Krueger, M.W.; Gionfriddo, T.; Hinrichsen, K. VIDEOPLACE—An artificial reality. In Proceedings of the 8th Annual Association for Computing Machinery Symposium on User Interface and Software Technology, San Francisco, CA, USA, 1 April 1985; pp. 35–40. [[CrossRef](#)]
34. Mayer, S.; Reinhardt, J.; Schweigert, R.; Jelke, B.; Schwind, V.; Wolf, K.; Henze, N. Improving Humans' Ability to Interpret Deictic Gestures in Virtual Reality. In Proceedings of the Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Volume 20, pp. 1–14.
35. Henrikson, R.; Grossman, T.; Trowbridge, S.; Wigdor, D.; Benko, H. Head-Coupled Kinematic Template Matching: A Prediction Model for Ray Pointing in VR. In Proceedings of the Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; pp. 1–14.
36. Li, N.; Han, T.; Tian, F.; Huang, J.; Sun, M.; Irani, P.; Alexander, J. Get a Grip: Evaluating Grip Gestures for VR Input using a Lightweight Pen. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13.
37. Mann, S. Wearable computing: A first step toward personal imaging. *Computer* **1997**, *30*, 25–32. [[CrossRef](#)]
38. Starner, T.; Mann, S.; Rhodes, B.; Levine, J.; Healey, J.; Kirsch, D.; Picard, R.W.; Pentland, A. Augmented reality through wearable computing. *Presence Teleoperators Virtual Environ.* **1997**, *6*, 386–398. [[CrossRef](#)]
39. Kölsch, M.; Bane, R.; Höllerer, T.; Turk, M. Multimodal interaction with a wearable augmented reality system. *Inst. Electr. Electron. Eng. Comput. Graph. Appl.* **2006**, *26*, 62–71. [[CrossRef](#)] [[PubMed](#)]
40. Li, Y.; Huang, J.; Tian, F.; Wang, H.A.; Dai, G.Z. Gesture interaction in virtual reality. *Virtual Real. Intell. Hardw.* **2019**, *1*, 84–112. [[CrossRef](#)]
41. Chen, Z.; Li, J.; Hua, Y.; Shen, R.; Basu, A. Multimodal interaction in augmented reality. In Proceedings of the 2017 Institution of Electrical Engineers International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; Volume 2017, pp. 206–209.
42. Yi, S.; Qin, Z.; Novak, E.; Yin, Y.; Li, Q. GlassGesture: Exploring head gesture interface of smart glasses. In Proceedings of the 2016 Institution of Electrical Engineers Conference on Computer Communications Workshops (INFOCOM WKSHPS), San Francisco, CA, USA, 10–14 April 2016; Volume 2016, pp. 1017–1018.
43. Zhao, J.; Allison, R.S. Real-time head gesture recognition on head-mounted displays using cascaded hidden Markov models. In Proceedings of the 2017 Institution of Electrical Engineers International International Conference on Systems, Man, and Cybernetics (SMC), Banff, AB, Canada, 5–8 October 2017; Volume 2017, pp. 2361–2366.
44. Yan, Y.; Yu, C.; Yi, X.; Shi, Y. HeadGesture: Hands-Free Input Approach Leveraging Head Movements for HMD Devices. *Proc. Assoc. Comput. Mach. Interact. Mob. Wearable Ubiquitous Technol.* **2018**, *2*, 1–23. [[CrossRef](#)]
45. Zhao, J.; Allison, R.S. Comparing head gesture, hand gesture and gamepad interfaces for answering Yes/No questions in virtual environments. *Virtual Real.* **2020**, *24*, 515–524. [[CrossRef](#)]
46. Ren, D.; Goldschwendt, T.; Chang, Y.; Hollerer, T. Evaluating wide-field-of-view augmented reality with mixed reality simulation. In Proceedings of the 2016 Institution of Electrical Engineers Virtual Reality (VR), Greenville, SC, USA, 19–23 March 2016; Volume 2016, pp. 93–102.
47. Cardoso, J.C.S. *A Review of Technologies for Gestural Interaction in Virtual Reality*; Cambridge Scholars Publishing: Newcastle upon Tyne, UK, 2019; ISBN 9781527535367.
48. Rautaray, S.S.; Agrawal, A. Vision based hand gesture recognition for human computer interaction: A survey. *Artif. Intell. Rev.* **2015**, *43*, 1–54. [[CrossRef](#)]
49. Cheng, H.; Yang, L.; Liu, Z. Survey on 3D Hand Gesture Recognition. *Inst. Electr. Electron. Eng. Trans. Circuits Syst. Video Technol.* **2016**, *26*, 1659–1673. [[CrossRef](#)]
50. Vuletic, T.; Duffy, A.; Hay, L.; McTeague, C.; Campbell, G.; Grealy, M. Systematic literature review of hand gestures used in human computer interaction interfaces. *Int. J. Hum. Comput. Stud.* **2019**, *129*, 74–94. [[CrossRef](#)]
51. Chen, W.; Yu, C.; Tu, C.; Lyu, Z.; Tang, J.; Ou, S.; Fu, Y.; Xue, Z. A survey on hand pose estimation with wearable sensors and computer-vision-based methods. *Sensors* **2020**, *20*, 1074. [[CrossRef](#)]
52. Alam, M.; Samad, M.D.; Vidyaratne, L.; Glandon, A.; Iftekharruddin, K.M. Survey on Deep Neural Networks in Speech and Vision Systems. *Neurocomputing* **2020**, *417*, 302–321. [[CrossRef](#)]
53. Beddiar, D.R.; Nini, B.; Sabokrou, M.; Hadid, A. Vision-based human activity recognition: A survey. *Multimed. Tools Appl.* **2020**, *79*, 30509–30555. [[CrossRef](#)]
54. DecaGear. Available online: <https://www.deca.net/decagear/> (accessed on 3 December 2021).
55. Bai, H.; Sasikumar, P.; Yang, J.; Billinghamurst, M. A User Study on Mixed Reality Remote Collaboration with Eye Gaze and Hand Gesture Sharing. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13.
56. Majaranta, P.; Ahola, U.-K.; Špakov, O. Fast gaze typing with an adjustable dwell time. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2009; pp. 357–360.

57. Kowalczyk, P.; Sawicki, D. Blink and wink detection as a control tool in multimodal interaction. *Multimed. Tools Appl.* **2019**, *78*, 13749–13765. [[CrossRef](#)]
58. Schweigert, R.; Schwind, V.; Mayer, S. EyePointing: A Gaze-Based Selection Technique. In Proceedings of the Mensch und Computer 2019 (MuC'19), Hamburg, Germany, 8–11 September 2019; pp. 719–723.
59. Parisay, M.; Poullis, C.; Kersten-Oertel, M. EyeTAP: Introducing a multimodal gaze-based technique using voice inputs with a comparative analysis of selection techniques. *Int. J. Hum. Comput. Stud.* **2021**, *154*, 102676. [[CrossRef](#)]
60. Nukarinen, T.; Kangas, J.; Rantala, J.; Koskinen, O.; Raisamo, R. Evaluating ray casting and two gaze-based pointing techniques for object selection in virtual reality. In Proceedings of the 24th Association for Computing Machinery Symposium on Virtual Reality Software and Technology, Tokio, Japan, 18 November–1 December 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–2.
61. Hyrskykari, A.; Istance, H.; Vickers, S. Gaze gestures or dwell-based interaction? In Proceedings of the Symposium on Eye Tracking Research and Applications—ETRA'12, Santa Barbara, CA, USA, 28–30 March 2012; Association for Computing Machinery Press: New York, NY, USA, 2012; p. 229.
62. Drewes, H.; Schmidt, A. Interacting with the Computer Using Gaze Gestures. In *Lecture Notes in Computer Science*; Springer: Berlin/Heidelberg, Germany, 2007; Volume 4663, pp. 475–488; ISBN 9783540747994.
63. Istance, H.; Hyrskykari, A.; Immonen, L.; Mansikkamaa, S.; Vickers, S. Designing gaze gestures for gaming: An investigation of performance. In Proceedings of the 2010 Symposium on Eye-Tracking Research & Applications—ETRA'10, Austin, TX, USA, 22–24 March 2010; Association for Computing Machinery Press: New York, NY, USA, 2010; p. 323.
64. Vidal, M.; Bulling, A.; Gellersen, H. Pursuits: Spontaneous interaction with displays based on smooth pursuit eye movement and moving targets. In Proceedings of the 2013 Association for Computing Machinery International Joint Conference on Pervasive and Ubiquitous Computing, Zurich, Switzerland, 8–12 September 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 439–448.
65. Esteves, A.; Velloso, E.; Bulling, A.; Gellersen, H. Orbits. In Proceedings of the 28th Annual Association for Computing Machinery Symposium on User Interface Software & Technology, Charlotte, NC, USA, 8–11 November 2015; Association for Computing Machinery: New York, NY, USA, 2015; pp. 457–466.
66. Sidenmark, L.; Clarke, C.; Zhang, X.; Phu, J.; Gellersen, H. Outline Pursuits: Gaze-assisted Selection of Occluded Objects in Virtual Reality. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13.
67. Duchowski, A. *Eye Tracking Methodology: Theory and Practice*; Springer: London, UK, 2007; ISBN 978-1-84628-608-7.
68. Hansen, D.W.; Qiang, J. In the Eye of the Beholder: A Survey of Models for Eyes and Gaze. *Inst. Electr. Electron. Eng. Trans. Pattern Anal. Mach. Intell.* **2010**, *32*, 478–500. [[CrossRef](#)]
69. Khamis, M.; Kienle, A.; Alt, F.; Bulling, A. GazeDrone. In Proceedings of the 4th Association for Computing Machinery Workshop on Micro Aerial Vehicle Networks, Systems, and Applications, Munich, Germany, 10–15 June 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 66–71.
70. Majaranta, P.; Bulling, A. Eye Tracking and Eye-Based Human–Computer Interaction. In *Advances in Physiological Computing*; Gilleade, S.F.K., Ed.; Springer: Berlin/Heidelberg, Germany, 2014; pp. 39–65.
71. Hutchinson, T.E.; White, K.P.; Martin, W.N.; Reichert, K.C.; Frey, L.A. Human-computer interaction using eye-gaze input. *Inst. Electr. Electron. Eng. Trans. Syst. Man Cybern.* **1989**, *19*, 1527–1534. [[CrossRef](#)]
72. Majaranta, P.; Räihä, K.J. Twenty years of eye typing: Systems and design issues. In Proceedings of the Eye Tracking Research and Applications Symposium (ETRA), New Orleans, LA, USA, 25–27 March 2002; pp. 15–22.
73. Rozado, D.; Moreno, T.; San Agustín, J.; Rodriguez, F.B.; Varona, P. Controlling a smartphone using gaze gestures as the input mechanism. *Hum.-Comput. Interact.* **2015**, *30*, 34–63. [[CrossRef](#)]
74. Holland, C.; Komogortsev, O. Eye tracking on unmodified common tablets: Challenges and solutions. In Proceedings of the Symposium on Eye Tracking Research and Applications—ETRA'12, Santa Barbara, CA, USA, 28–30 March 2012; Association for Computing Machinery Press: New York, NY, USA, 2012; p. 277.
75. Akkil, D.; Kangas, J.; Rantala, J.; Isokoski, P.; Spakov, O.; Raisamo, R. Glance Awareness and Gaze Interaction in Smartwatches. In Proceedings of the 33rd Annual Association for Computing Machinery Conference Extended Abstracts on Human Factors in Computing Systems, Seoul, Korea, 18–23 April 2015; Association for Computing Machinery: New York, NY, USA, 2015; Volume 18, pp. 1271–1276.
76. Zhang, L.; Li, X.Y.; Huang, W.; Liu, K.; Zong, S.; Jian, X.; Feng, P.; Jung, T.; Liu, Y. It starts with iGaze: Visual attention driven networking with smart glasses. In Proceedings of the Annual International Conference on Mobile Computing and Networking, MOBICOM, Maui, HI, USA, 7–11 September 2014; pp. 91–102.
77. Zhang, Y.; Bulling, A.; Gellersen, H. SideWays: A gaze interface for spontaneous interaction with situated displays. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Paris, France, 27 April–2 May 2013; Association for Computing Machinery: New York, NY, USA, 2013; pp. 851–860.
78. Hansen, J.P.; Alapetite, A.; MacKenzie, I.S.; Møllenbach, E. The use of gaze to control drones. In Proceedings of the Symposium on Eye Tracking Research and Applications, Safety Harbor, FL, USA, 26–28 March 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 27–34.

79. Yuan, L.; Reardon, C.; Warnell, G.; Loianno, G. Human gaze-driven spatial tasking of an autonomous MAV. *Inst. Electr. Electron. Eng. Robot. Autom. Lett.* **2019**, *4*, 1343–1350. [CrossRef]
80. Clay, V.; König, P.; König, S.U. Eye tracking in virtual reality. *J. Eye Mov. Res.* **2019**, *12*. [CrossRef]
81. Piumsomboon, T.; Lee, G.; Lindeman, R.W.; Billingham, M. Exploring natural eye-gaze-based interaction for immersive virtual reality. In Proceedings of the 2017 Institute of Electrical and Electronics Engineers Symposium on 3D User Interfaces (3DUI), Los Angeles, CA, USA, 18–19 March 2017; Volume 3DUI, pp. 36–39.
82. Nukarinen, T.; Kangas, J.; Rantala, J.; Pakkanen, T.; Raisamo, R. Hands-free vibrotactile feedback for object selection tasks in virtual reality. In Proceedings of the 24th Association for Computing Machinery Symposium on Virtual Reality Software and Technology, Tokyo, Japan, 18 November–1 December 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 1–2.
83. Meißner, M.; Pfeiffer, J.; Pfeiffer, T.; Oppewal, H. Combining virtual reality and mobile eye tracking to provide a naturalistic experimental environment for shopper research. *J. Bus. Res.* **2019**, *100*, 445–458. [CrossRef]
84. Tobii VR. Available online: <https://vr.tobii.com/> (accessed on 3 December 2021).
85. Varjo Eye. Tracking in VR. Available online: <https://varjo.com/blog/how-to-do-eye-tracking-studies-in-virtual-reality/> (accessed on 3 December 2021).
86. Burova, A.; Mäkelä, J.; Hakulinen, J.; Keskinen, T.; Heinonen, H.; Siltanen, S.; Turunen, M. Utilizing VR and Gaze Tracking to Develop AR Solutions for Industrial Maintenance. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–13.
87. Gardony, A.L.; Lindeman, R.W.; Brunyé, T.T. Eye-tracking for human-centered mixed reality: Promises and challenges. In *Optical Architectures for Displays and Sensing in Augmented, Virtual, and Mixed Reality (AR, VR, MR)*; Kress, B.C., Peroz, C., Eds.; International Society for Optics and Photonics: Bellingham, WA, USA, 2020; Volume 11310, p. 27.
88. Sims, S.D.; Conati, C. A Neural Architecture for Detecting User Confusion in Eye-tracking Data. In Proceedings of the 2020 International Conference on Multimodal Interaction, Online, 25–29 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; Volume ICMI'20, pp. 15–23.
89. DeLucia, P.R.; Preddy, D.; Derby, P.; Tharanathan, A.; Putrevu, S. Eye Movement Behavior During Confusion. *Proc. Hum. Factors Ergon. Soc. Annu. Meet.* **2014**, *58*, 1300–1304. [CrossRef]
90. Marshall, S.P. Identifying cognitive state from eye metrics. *Aviat. Sp. Environ. Med.* **2007**, *78*, B165–B175.
91. Boraston, Z.; Blakemore, S.J. The application of eye-tracking technology in the study of autism. *J. Physiol.* **2007**, *581*, 893–898. [CrossRef]
92. Drèze, X.; Hussherr, F.-X. Internet advertising: Is anybody watching? *J. Interact. Mark.* **2003**, *17*, 8–23. [CrossRef]
93. Raisamo, R.; Rakkolainen, I.; Majaranta, P.; Salminen, K.; Rantala, J.; Farooq, A. Human augmentation: Past, present and future. *Int. J. Hum. Comput. Stud.* **2019**, *131*, 131–143. [CrossRef]
94. Hyrskykari, A.; Majaranta, P.; Räihä, K.J. From Gaze Control to Attentive Interfaces. In Proceedings of the 11th International Conference on Human-Computer Interaction, Las Vegas, NV, USA, 22–27 July 2005.
95. Hansen, J.; Hansen, D.; Johansen, A.; Elvesjö, J. Mainstreaming gaze interaction towards a mass market for the benefit of all. In *Universal Access in HCI: Exploring New Interaction Environments*; Stephanidis, C., Ed.; Lawrence Erlbaum Associates, Inc.: Mahwah, NJ, USA, 2005; Volume 7.
96. Freeman, E.; Wilson, G.; Vo, D.-B.; Ng, A.; Politis, I.; Brewster, S. Multimodal feedback in HCI: Haptics, non-speech audio, and their applications. In *The Handbook of Multimodal-Multisensor Interfaces: Foundations, User Modeling, and Common Modality Combinations*; ACM Books: New York, NY, USA, 2017; Volume 1, pp. 277–317.
97. Miccini, R.; Spagnol, S. HRTF Individualization using Deep Learning. In Proceedings of the 2020 Institute of Electrical and Electronics Engineers Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; pp. 390–395.
98. Wolf, M.; Trentsios, P.; Kubatzki, N.; Urbanietz, C.; Enzner, G. Implementing Continuous-Azimuth Binaural Sound in Unity 3D. In Proceedings of the 2020 Institute of Electrical and Electronics Engineers Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Atlanta, GA, USA, 22–26 March 2020; pp. 384–389.
99. Sra, M.; Xu, X.; Maes, P. BreathVR: Leveraging breathing as a directly controlled interface for virtual reality games. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; Volume 2018, pp. 1–12.
100. Kusabuka, T.; Indo, T. IBUKI: Gesture Input Method Based on Breathing. In Proceedings of the 33rd Annual Association for Computing Machinery Symposium on User Interface Software and Technology, Online, 20–23 October 2020; pp. 102–104.
101. Chen, Y.; Bian, Y.; Yang, C.; Bao, X.; Wang, Y.; De Melo, G.; Liu, J.; Gai, W.; Wang, L.; Meng, X. Leveraging Blowing as a Directly Controlled Interface. In Proceedings of the 2019 Institute of Electrical and Electronics Engineers SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCOM/IOP/SCI), Los Alamitos, CA, USA, 19–23 August 2019; pp. 419–424.
102. Goldstein, E.B. *Sensation & Perception*, 5th ed.; Brooks/Cole Publishing Company: Pacific Grove, CA, USA, 1999.
103. Biswas, S.; Visell, Y. Emerging Material Technologies for Haptics. *Adv. Mater. Technol.* **2019**, *4*, 1900042. [CrossRef]

104. Asaga, E.; Takemura, K.; Maeno, T.; Ban, A.; Toriumi, M. Tactile evaluation based on human tactile perception mechanism. *Sens. Actuators A Phys.* **2013**, *203*, 69–75. [[CrossRef](#)]
105. Kandel, E.; Schwartz, J.; Jesell, T.; Siegelbaum, S. *Hudspeth Principles of Neural Science*; McGraw-Hill: New York, NY, USA, 2013.
106. Proske, U.; Gandevia, S.C. The proprioceptive senses: Their roles in signaling body shape, body position and movement, and muscle force. *Physiol. Rev.* **2012**, *92*, 1651–1697. [[CrossRef](#)] [[PubMed](#)]
107. Oakley, I.; McGee, M.R.; Brewster, S.; Gray, P. Putting the feel in 'look and feel'. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI'00, The Hague, The Netherlands, 1–6 April 2000; Association for Computing Machinery Press: New York, NY, USA, 2000; pp. 415–422.
108. Barnett-Cowan, M. Vestibular Perception is Slow: A Review. *Multisens. Res.* **2013**, *26*, 387–403. [[CrossRef](#)] [[PubMed](#)]
109. Morphew, M.E.; Shively, J.R.; Casey, D. Helmet-mounted displays for unmanned aerial vehicle control. In Proceedings of the Helmet- and Head-Mounted Displays IX: Technologies and Applications, Orlando, FL, USA, 12–13 April 2004; Volume 5442, p. 93.
110. Mollet, N.; Chellali, R. Virtual and Augmented Reality with Head-Tracking for Efficient Teleoperation of Groups of Robots. In Proceedings of the 2008 International Conference on Cyberworlds, Hangzhou, China, 22–24 September 2008; pp. 102–108.
111. Higuchi, K.; Fujii, K.; Rekimoto, J. Flying head: A head-synchronization mechanism for flying telepresence. In Proceedings of the 2013 23rd International Conference on Artificial Reality and Telexistence (ICAT), Tokyo, Japan, 11–13 December 2013; pp. 28–34.
112. Smolyanskiy, N.; Gonzalez-Franco, M. Stereoscopic first person view system for drone navigation. *Front. Robot. AI* **2017**, *4*, 11. [[CrossRef](#)]
113. Pittman, C.; LaViola, J.J. Exploring head tracked head mounted displays for first person robot teleoperation. In Proceedings of the 19th International Conference on Intelligent User Interfaces, Haifa, Israel, 24–27 February 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 323–328.
114. Teixeira, J.M.; Ferreira, R.; Santos, M.; Teichrieb, V. Teleoperation Using Google Glass and AR, Drone for Structural Inspection. In Proceedings of the 2014 XVI Symposium on Virtual and Augmented Reality, Piata Salvador, Brazil, 12–15 May 2014; pp. 28–36.
115. Doisy, G.; Ronen, A.; Edan, Y. Comparison of three different techniques for camera and motion control of a teleoperated robot. *Appl. Ergon.* **2017**, *58*, 527–534. [[CrossRef](#)]
116. Culbertson, H.; Schorr, S.B.; Okamura, A.M. Haptics: The Present and Future of Artificial Touch Sensation. *Annu. Rev. Control. Robot. Auton. Syst.* **2018**, *1*, 385–409. [[CrossRef](#)]
117. Bermejo, C.; Hui, P. A Survey on Haptic Technologies for Mobile Augmented Reality. *Assoc. Comput. Mach. Comput. Surv.* **2022**, *54*, 1–35. [[CrossRef](#)]
118. Choi, S.; Kuchenbecker, K.J. Vibrotactile Display: Perception, Technology, and Applications. *Proc. Inst. Electr. Electron. Eng.* **2013**, *101*, 2093–2104. [[CrossRef](#)]
119. Wang, D.; Ohnishi, K.; Xu, W. Multimodal haptic display for virtual reality: A survey. *Inst. Electr. Electron. Eng. Trans. Ind. Electron.* **2020**, *67*, 610–623. [[CrossRef](#)]
120. Hamza-Lup, F.G.; Bergeron, K.; Newton, D. Haptic Systems in User Interfaces. In Proceedings of the 2019 Association for Computing Machinery Southeast Conference, Kennesaw, GA, USA, 18–20 April 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 141–148.
121. McGlone, F.; Vallbo, A.B.; Olausson, H.; Loken, L.; Wessberg, J. Discriminative touch and emotional touch. *Can. J. Exp. Psychol. Can. Psychol. Expérimentale* **2007**, *61*, 173–183. [[CrossRef](#)]
122. Pacchierotti, C.; Sinclair, S.; Solazzi, M.; Frisoli, A.; Hayward, V.; Prattichizzo, D. Wearable haptic systems for the fingertip and the hand: Taxonomy, review, and perspectives. *Inst. Electr. Electron. Eng. Trans. Haptics* **2017**, *10*, 580–600. [[CrossRef](#)]
123. Yu, X.; Xie, Z.; Yu, Y.; Lee, J.; Vazquez-Guardado, A.; Luan, H.; Ruban, J.; Ning, X.; Akhtar, A.; Li, D.; et al. Skin-integrated wireless haptic interfaces for virtual and augmented reality. *Nature* **2019**, *575*, 473–479. [[CrossRef](#)]
124. De Jesus Oliveira, V.A.; Nedel, L.; Maciel, A.; Brayda, L. Spatial discrimination of vibrotactile stimuli around the head. In Proceedings of the 2016 Institute of Electrical and Electronics Engineers Haptics Symposium (HAPTICS), Philadelphia, PA, USA, 8–11 April 2016; pp. 1–6.
125. Kaul, O.B.; Rohs, M. HapticHead: A spherical vibrotactile grid around the head for 3D guidance in virtual and augmented reality. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; Volume 2017, pp. 3729–3740.
126. Iwamoto, T.; Tatezono, M.; Shinoda, H. Non-contact Method for Producing Tactile Sensation Using Airborne Ultrasound. In *Haptics: Perception, Devices and Scenarios*; Springer: Berlin/Heidelberg, Germany, 2008; Volume 5024, pp. 504–513; ISBN 3540690565.
127. Long, B.; Seah, S.A.; Carter, T.; Subramanian, S. Rendering volumetric haptic shapes in mid-air using ultrasound. *Assoc. Comput. Mach. Trans. Graph.* **2014**, *33*, 1–10. [[CrossRef](#)]
128. Rakkolainen, I.; Freeman, E.; Sand, A.; Raisamo, R.; Brewster, S. A Survey of Mid-Air Ultrasound Haptics and Its Applications. *Inst. Electr. Electron. Eng. Trans. Haptics* **2021**, *14*, 2–19. [[CrossRef](#)]
129. Farooq, A.; Evreinov, G.; Raisamo, R.; Hippula, A. Developing Intelligent Multimodal IVI Systems to Reduce Driver Distraction. In *Intelligent Human Systems Integration 2019. IHSI 2019. Advances in Intelligent Systems and Computing*; Springer: Cham, Switzerland, 2019; Volume 903, pp. 91–97.

130. Hoshi, T.; Abe, D.; Shinoda, H. Adding tactile reaction to hologram. In Proceedings of the RO-MAN 2009—The 18th Institute of Electrical and Electronics Engineers International Symposium on Robot and Human Interactive Communication, Toyama, Japan, 27 September–2 October 2009; pp. 7–11.
131. Martinez, J.; Griffiths, D.; Biscione, V.; Georgiou, O.; Carter, T. Touchless Haptic Feedback for Supernatural VR Experiences. In Proceedings of the 2018 Institute of Electrical and Electronics Engineers International Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, 18–22 March 2018; pp. 629–630.
132. Furumoto, T.; Fujiwara, M.; Makino, Y.; Shinoda, H. BaLuna: Floating Balloon Screen Manipulated Using Ultrasound. In Proceedings of the 2019 Institute of Electrical and Electronics Engineers International Conference on Virtual Reality and 3D User Interfaces (VR), Osaka, Japan, 23–27 March 2019; pp. 937–938.
133. Kervegant, C.; Raymond, F.; Graeff, D.; Castet, J. Touch hologram in mid-air. In Proceedings of the Association for Computing Machinery SIGGRAPH 2017 Emerging Technologies, Los Angeles, CA, USA, 30 July–3 August 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1–2.
134. Sand, A.; Rakkolainen, I.; Isokoski, P.; Kangas, J.; Raisamo, R.; Palovuori, K. Head-mounted display with mid-air tactile feedback. In Proceedings of the 21st Association for Computing Machinery Symposium on Virtual Reality Software and Technology, Beijing, China, 13–15 November 2015; Association for Computing Machinery: New York, NY, USA, 2015; Volume 13, pp. 51–58.
135. Palovuori, K.; Rakkolainen, I.; Sand, A. Bidirectional touch interaction for immaterial displays. In Proceedings of the 18th International Academic MindTrek Conference on Media Business, Management, Content & Services—AcademicMindTrek'14, Tampere, Finland, 4–6 November 2014; Association for Computing Machinery Press: New York, NY, USA, 2014; pp. 74–76.
136. Wilson, G.; Carter, T.; Subramanian, S.; Brewster, S.A. Perception of ultrasonic haptic feedback on the hand. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Toronto, ON, Canada, 26 April–1 May 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 1133–1142.
137. Van Neer, P.; Volker, A.; Berkhoff, A.; Schrama, T.; Akkerman, H.; Van Breemen, A.; Peeters, L.; Van Der Steen, J.L.; Gelinck, G. Development of a flexible large-area array based on printed polymer transducers for mid-air haptic feedback. *Proc. Meet. Acoust.* **2019**, *38*, 45008. [[CrossRef](#)]
138. Farooq, A.; Weitz, P.; Evreinov, G.; Raisamo, R.; Takahata, D. Touchscreen Overlay Augmented with the Stick-Slip Phenomenon to Generate Kinetic Energy. In Proceedings of the 29th Annual Symposium on User Interface Software and Technology, Tokyo, Japan, 16–19 October 2016; Association for Computing Machinery: New York, NY, USA, 2016; pp. 179–180.
139. Desai, A.P.; Pena-Castillo, L.; Meruvia-Pastor, O. A Window to Your Smartphone: Exploring Interaction and Communication in Immersive VR with Augmented Virtuality. In Proceedings of the 2017 14th Conference on Computer and Robot Vision (CRV), Edmonton, AB, Canada, 16–19 May 2017; Volume 2018, pp. 217–224.
140. Chuah, J.H.; Lok, B. Experiences in Using a Smartphone as a Virtual Reality Interaction Device. *Int. J. Virtual Real.* **2012**, *11*, 25–31. [[CrossRef](#)]
141. Qian, J.; Ma, J.; Li, X.; Attal, B.; Lai, H.; Tompkin, J.; Hughes, J.F.; Huang, J. Portal-ble: Intuitive Free-hand Manipulation in Unbounded Smartphone-based Augmented Reality. In Proceedings of the 32nd Annual Association for Computing Machinery Symposium on User Interface Software and Technology, New Orleans, LA, USA, 20–23 October 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 133–145.
142. Nakagaki, K.; Fitzgerald, D.; Ma, Z.J.; Vink, L.; Levine, D.; Ishii, H. InFORCE: Bi-directional “Force” Shape Display For Haptic Interaction. In Proceedings of the Thirteenth International Conference on Tangible, Embedded, and Embodied Interaction, Tempe, AR, USA, 17–20 March 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 615–623.
143. Allman-Farinelli, M.; Ijaz, K.; Tran, H.; Pallotta, H.; Ramos, S.; Liu, J.; Wellard-Cole, L.; Calvo, R.A. A Virtual Reality Food Court to Study Meal Choices in Youth: Design and Assessment of Usability. *JMIR Form. Res.* **2019**, *3*, e12456. [[CrossRef](#)]
144. Stelick, A.; Penano, A.G.; Riak, A.C.; Dando, R. Dynamic Context Sensory Testing—A Proof of Concept Study Bringing Virtual Reality to the Sensory Booth. *J. Food Sci.* **2018**, *83*, 2047–2051. [[CrossRef](#)]
145. Kaluschke, M.; Weller, R.; Zachmann, G.; Pelliccia, L.; Lorenz, M.; Klimant, P.; Knopp, S.; Atze, J.P.G.; Mockel, F. A Virtual Hip Replacement Surgery Simulator with Realistic Haptic Feedback. In Proceedings of the 2018 Institute of Electrical and Electronics Engineers International Conference on Virtual Reality and 3D User Interfaces (VR), Reutlingen, Germany, 18–22 March 2018; pp. 759–760.
146. Brazil, A.L.; Conci, A.; Clua, E.; Bittencourt, L.K.; Baruque, L.B.; da Silva Conci, N. Haptic forces and gamification on epidural anesthesia skill gain. *Entertain. Comput.* **2018**, *25*, 1–13. [[CrossRef](#)]
147. Karafotias, G.; Korres, G.; Sefo, D.; Boomer, P.; Eid, M. Towards a realistic haptic-based dental simulation. In Proceedings of the 2017 Institute of Electrical and Electronics Engineers International Symposium on Haptic, Audio and Visual Environments and Games (HAVE), 21–22 October 2017; Volume 2017, pp. 1–6.
148. Holoride: Virtual Reality Meets the Real World. Available online: <https://www.audi.com/en/experience-audi/mobility-and-trends/digitalization/holoride-virtual-reality-meets-the-real-world.html> (accessed on 3 December 2021).
149. Ma, Z.; Ben-Tzvi, P. Design and optimization of a five-finger haptic glove mechanism. *J. Mech. Robot.* **2015**, *7*, 041008. [[CrossRef](#)]
150. Turner, M.L.; Gomez, D.H.; Tremblay, M.R.; Cutkosky, M.R. Preliminary tests of an arm-grounded haptic feedback device in telemanipulation. In Proceedings of the 2001 ASME International Mechanical Engineering Congress and Exposition, New York, NY, USA, 11–16 November 2001; Volume 64, pp. 145–149.

151. Bouzit, M.; Popescu, G.; Burdea, G.; Boian, R. The Rutgers Master II-ND force feedback glove. In Proceedings of the 10th Symposium on Haptic Interfaces for Virtual Environment and Teleoperator Systems, Orlando, FL, USA, 24–25 March 2002; pp. 145–152.
152. Perret, J.; Poorten, E. Vander Touching virtual reality: A review of haptic gloves. In Proceedings of the ACTUATOR 2018—16th International Conference and Exhibition on New Actuators and Drive Systems, Bremen, Germany, 25–27 June 2018; pp. 270–274.
153. Caeiro-Rodríguez, M.; Otero-González, I.; Mikic-Fonte, F.A.; Llamas-Nistal, M. A systematic review of commercial smart gloves: Current status and applications. *Sensors* **2021**, *21*, 2667. [[CrossRef](#)]
154. Lindeman, R.W.; Page, R.; Yanagida, Y.; Sibert, J.L. Towards full-body haptic feedback: The design and deployment of a spatialized vibrotactile feedback system. In Proceedings of the Association for Computing Machinery Symposium on Virtual Reality Software and Technology—VRST'04, Tokyo, Japan, 28 November–1 December 2018; Association for Computing Machinery Press: New York, NY, USA, 2004; p. 146.
155. Farooq, A.; Coe, P.; Evreinov, G.; Raisamo, R. Using Dynamic Real-Time Haptic Mediation in VR and AR Environments. In *Advances in Intelligent Systems and Computing*; Ahram, T., Taiar, R., Colson, S., Choplin, A., Eds.; Springer: Cham, Switzerland, 2020; Volume 1018, pp. 407–413; ISBN 9783030256289.
156. Kasahara, S.; Konno, K.; Owaki, R.; Nishi, T.; Takeshita, A.; Ito, T.; Kasuga, S.; Ushiba, J. Malleable Embodiment: Changing sense of embodiment by spatial-temporal deformation of virtual human body. In Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017; Association for Computing Machinery: New York, NY, USA, 2017; Volume 2017, pp. 6438–6448.
157. Jiang, F.; Yang, X.; Feng, L. Real-time full-body motion reconstruction and recognition for off-the-shelf VR devices. In Proceedings of the 15th Association for Computing Machinery SIGGRAPH Conference on Virtual-Reality Continuum and Its Applications in Industry—Volume 1, Zhuhai, China, 3–4 December 2016; Association for Computing Machinery: New York, NY, USA, 2016; Volume 1, pp. 309–318.
158. Slater, M.; Wilbur, S. A framework for immersive virtual environments (FIVE): Speculations on the role of presence in virtual environments. *Presence Teleoperators Virtual Environ.* **1997**, *6*, 603–616. [[CrossRef](#)]
159. Caserman, P.; Garcia-Agundez, A.; Gobel, S. A Survey of Full-Body Motion Reconstruction in Immersive Virtual Reality Applications. *Inst. Electr. Electron. Eng. Trans. Vis. Comput. Graph.* **2020**, *26*, 3089–3108. [[CrossRef](#)]
160. Olivier, A.H.; Bruneau, J.; Kulpa, R.; Pettre, J. Walking with Virtual People: Evaluation of Locomotion Interfaces in Dynamic Environments. *Inst. Electr. Electron. Eng. Trans. Vis. Comput. Graph.* **2018**, *24*, 2251–2263. [[CrossRef](#)]
161. Nilsson, N.C.; Serafin, S.; Steinicke, F.; Nordahl, R. Natural walking in virtual reality: A review. *Comput. Entertain.* **2018**, *16*, 1–22. [[CrossRef](#)]
162. Boletsis, C. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technol. Interact.* **2017**, *1*, 24. [[CrossRef](#)]
163. Suzuki, Y.; Sekimori, K.; Yamato, Y.; Yamasaki, Y.; Shizuki, B.; Takahashi, S. A Mouth Gesture Interface Featuring a Mutual-Capacitance Sensor Embedded in a Surgical Mask. In *Lecture Notes in Computer Science*; Springer: Cham, Switzerland, 2020; Volume 12182, pp. 154–165.
164. Hashimoto, T.; Low, S.; Fujita, K.; Usumi, R.; Yanagihara, H.; Takahashi, C.; Sugimoto, M.; Sugiura, Y. TongueInput: Input Method by Tongue Gestures Using Optical Sensors Embedded in Mouthpiece. In Proceedings of the 2018 57th Annual Conference of the Society of Instrument and Control Engineers of Japan (SICE), Nara, Japan, 11–14 September 2018; pp. 1219–1224.
165. Visell, Y.; Law, A.; Cooperstock, J.R. Touch is everywhere: Floor surfaces as ambient haptic interfaces. *Inst. Electr. Electron. Eng. Trans. Haptics* **2009**, *2*, 148–159. [[CrossRef](#)] [[PubMed](#)]
166. Bouillot, N.; Seta, M. A Scalable Haptic Floor Dedicated to Large Immersive Spaces. In Proceedings of the 17th Linux Audio Conference (LAC-19), Stanford, CA, USA, 23–26 March 2019.
167. Yixian, Y.; Takashima, K.; Tang, A.; Tanno, T.; Fujita, K.; Kitamura, Y. ZoomWalls: Dynamic walls that simulate haptic infrastructure for room-scale VR world. In Proceedings of the 33rd Annual Association for Computing Machinery Symposium on User Interface Software and Technology, Online, 20–23 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 223–235.
168. Bouzbib, E.; Bailly, G.; Haliyo, S.; Frey, P. CoVR: A Large-Scale Force-Feedback Robotic Interface for Non-Deterministic Scenarios in VR. In Proceedings of the 33rd Annual Association for Computing Machinery Symposium on User Interface Software and Technology, Online, 20–23 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 209–222.
169. Kovacs, R.; Ofek, E.; Gonzalez Franco, M.; Siu, A.F.; Marwecki, S.; Holz, C.; Sinclair, M. Haptic PIVOT: On-demand handhelds in VR. In Proceedings of the 33rd Annual Association for Computing Machinery Symposium on User Interface Software and Technology, Online, 20–23 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1046–1059.
170. Munyan, B.G.; Neer, S.M.; Beidel, D.C.; Jentsch, F. Olfactory Stimuli Increase Presence in Virtual Environments. *PLoS ONE* **2016**, *11*, e0157568. [[CrossRef](#)]
171. Hopf, J.; Scholl, M.; Neuhofer, B.; Egger, R. Exploring the Impact of Multisensory VR on Travel Recommendation: A Presence Perspective. In *Information and Communication Technologies in Tourism 2020*; Springer: Cham, Switzerland, 2020; pp. 169–180.
172. Baus, O.; Bouchard, S.; Nolet, K. Exposure to a pleasant odour may increase the sense of reality, but not the sense of presence or realism. *Behav. Inf. Technol.* **2019**, *38*, 1369–1378. [[CrossRef](#)]

173. Ranasinghe, N.; Jain, P.; Thi Ngoc Tram, N.; Koh, K.C.R.; Tolley, D.; Karwita, S.; Lien-Ya, L.; Liangkun, Y.; Shamaiah, K.; Eason Wai Tung, C.; et al. Season Traveller: Multisensory narration for enhancing the virtual reality experience. In Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems, Montreal, QC, Canada, 21–26 April 2018; Association for Computing Machinery: New York, NY, USA, 2018; Volume 2018, pp. 1–13.
174. Tortell, R.; Luigi, D.-P.; Dozois, A.; Bouchard, S.; Morie, J.F.; Ilan, D. The effects of scent and game play experience on memory of a virtual environment. *Virtual Real.* **2007**, *11*, 61–68. [[CrossRef](#)]
175. Murray, N.; Lee, B.; Qiao, Y.; Muntean, G.M. Olfaction-enhanced multimedia: A survey of application domains, displays, and research challenges. *Assoc. Comput. Mach. Comput. Surv.* **2016**, *48*, 1–34. [[CrossRef](#)]
176. Obrist, M.; Velasco, C.; Vi, C.T.; Ranasinghe, N.; Israr, A.; Cheok, A.D.; Spence, C.; Gopalakrishnakone, P. Touch, Taste, & Smell User Interfaces. In Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems, San Jose, CA, USA, 7–12 May 2016; Association for Computing Machinery: New York, NY, USA, 2016; Volume 7, pp. 3285–3292.
177. Cheok, A.D.; Karunanayaka, K. Virtual Taste and Smell Technologies for Multisensory Internet and Virtual Reality. In *Human-Computer Interaction Series*; Springer International Publishing: Cham, Switzerland, 2018; ISBN 978-3-319-73863-5.
178. Spence, C.; Obrist, M.; Velasco, C.; Ranasinghe, N. Digitizing the chemical senses: Possibilities & pitfalls. *Int. J. Hum. Comput. Stud.* **2017**, *107*, 62–74. [[CrossRef](#)]
179. Spangenberg, E.R.; Crowley, A.E.; Henderson, P.W. Improving the Store Environment: Do Olfactory Cues Affect Evaluations and Behaviors? *J. Mark.* **1996**, *60*, 67–80. [[CrossRef](#)]
180. Salminen, K.; Rantala, J.; Isokoski, P.; Lehtonen, M.; Müller, P.; Karjalainen, M.; Väliaho, J.; Kontunen, A.; Nieminen, V.; Leivo, J.; et al. Olfactory Display Prototype for Presenting and Sensing Authentic and Synthetic Odors. In Proceedings of the 20th Association for Computing Machinery International Conference on Multimodal Interaction, Boulder, CO, USA, 16–20 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 73–77.
181. Niedenthal, S.; Lunden, P.; Ehrndal, M.; Olofsson, J.K. A Handheld Olfactory Display For Smell-Enabled VR Games. In Proceedings of the 2019 Institute of Electrical and Electronics Engineers International Symposium on Olfaction and Electronic Nose (ISOEN), Fukuoka, Japan, 26–29 May 2019; pp. 1–4.
182. Wang, Y.; Amores, J.; Maes, P. On-Face Olfactory Interfaces. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–9.
183. Brooks, J.; Nagels, S.; Lopes, P. Trigeminal-based Temperature Illusions. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–12.
184. Kato, S.; Nakamoto, T. Wearable Olfactory Display with Less Residual Odor. In Proceedings of the 2019 Institute of Electrical and Electronics Engineers International Symposium on Olfaction and Electronic Nose (ISOEN), Fukuoka, Japan, 26–29 May 2019; pp. 1–3.
185. Narumi, T.; Nishizaka, S.; Kajinami, T.; Tanikawa, T.; Hirose, M. Augmented reality flavors. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Vancouver, BC, Canada, 7–12 May 2011; Association for Computing Machinery: New York, NY, USA, 2011; pp. 93–102.
186. Yanagida, Y. A survey of olfactory displays: Making and delivering scents. In Proceedings of the 11th Institute of Electrical and Electronics Engineers Sensors Conference, Taipei, Taiwan, 28–31 October 2012; pp. 1–4. [[CrossRef](#)]
187. Ravia, A.; Snitz, K.; Honigstein, D.; Finkel, M.; Zirlor, R.; Perl, O.; Secundo, L.; Laudamiel, C.; Harel, D.; Sobel, N. A measure of smell enables the creation of olfactory metamers. *Nature* **2020**, *588*, 118–123. [[CrossRef](#)]
188. Iwata, H. Taste interfaces. In *HCI Beyond the GUI: Design for Haptic, Speech, Olfactory, and Other Nontraditional Interfaces*; Kortum, P., Ed.; Elsevier Inc.: Cambridge, MA, USA, 2008.
189. Auvray, M.; Spence, C. The multisensory perception of flavor. *Conscious. Cogn.* **2008**, *17*, 1016–1031. [[CrossRef](#)]
190. Aisala, H.; Rantala, J.; Vanhatalo, S.; Nikinmaa, M.; Pennanen, K.; Raisamo, R.; Sözer, N. Augmentation of Perceived Sweetness in Sugar Reduced Cakes by Local Odor Display. In Proceedings of the 2020 International Conference on Multimodal Interaction, Utrecht, The Netherlands, 25–29 October 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 322–327.
191. Kerruish, E. Arranging sensations: Smell and taste in augmented and virtual reality. *Senses Soc.* **2019**, *14*, 31–45. [[CrossRef](#)]
192. Maynes-Aminzade, D. Edible Bits: Seamless Interfaces between People, Data and Food. In Proceedings of the 2005 Association for Computing Machinery Conference on Human Factors in Computing Systems (CHI'2005), Portland, OR, USA, 2–7 April 2005; pp. 2207–2210.
193. Ranasinghe, N.; Nguyen, T.N.T.; Liangkun, Y.; Lin, L.-Y.; Tolley, D.; Do, E.Y.-L. Vocktail: A virtual cocktail for pairing digital taste, smell, and color sensations. In Proceedings of the 25th Association for Computing Machinery International Conference on Multimedia, Mountain View, CA, USA, 23–27 October 2017; Association for Computing Machinery: New York, NY, USA, 2017; Volume MM'17, pp. 1139–1147.
194. Nakamura, H.; Miyashita, H. Development and evaluation of interactive system for synchronizing electric taste and visual content. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, Austin, TX, USA, 5–10 May 2012; Association for Computing Machinery: New York, NY, USA, 2012; pp. 517–520.

195. Ranasinghe, N.; Cheok, A.; Nakatsu, R.; Do, E.Y.-L. Simulating the sensation of taste for immersive experiences. In Proceedings of the 2013 Association for Computing Machinery International Workshop on Immersive Media Experiences—ImmersiveMe'13, Barcelona, Spain, 22 October 2013; Association for Computing Machinery Press: New York, NY, USA, 2013; pp. 29–34.
196. Suzuki, C.; Narumi, T.; Tanikawa, T.; Hirose, M. Affecting tumbler: Affecting our flavor perception with thermal feedback. In Proceedings of the 11th Conference on Advances in Computer Entertainment Technology, Funchal, Portugal, 11–14 November 2014; Association for Computing Machinery: New York, NY, USA, 2014; pp. 1–10.
197. Koskinen, E.; Rakkolainen, I.; Raisamo, R. Direct retinal signals for virtual environments. In Proceedings of the 23rd Association for Computing Machinery Symposium on Virtual Reality Software and Technology, Gothenburg, Sweden, 8–10 November 2017; Association for Computing Machinery: New York, NY, USA, 2017; Volume F1319, pp. 1–2.
198. Abiri, R.; Borhani, S.; Sellers, E.W.; Jiang, Y.; Zhao, X. A comprehensive review of EEG-based brain–computer interface paradigms. *J. Neural Eng.* **2019**, *16*, 011001. [[CrossRef](#)]
199. Bernal, G.; Yang, T.; Jain, A.; Maes, P. PhysioHMD. In Proceedings of the 2018 Association for Computing Machinery International Symposium on Wearable Computers, Singapore, 8–12 October 2018; Association for Computing Machinery: New York, NY, USA, 2018; pp. 160–167.
200. Vourvopoulos, A.; Niforatos, E.; Giannakos, M. EEGlass: An EEG-eyeware prototype for ubiquitous brain-computer interaction. In Proceedings of the 2019 Association for Computing Machinery International Joint Conference on Pervasive and Ubiquitous Computing, London, UK, 11–13 September 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 647–652.
201. Luong, T.; Martin, N.; Raison, A.; Argelaguet, F.; Diverrez, J.-M.; Lecuyer, A. Towards Real-Time Recognition of Users Mental Workload Using Integrated Physiological Sensors Into a VR HMD. In Proceedings of the 2020 Institute of Electrical and Electronics Engineers International Symposium on Mixed and Augmented Reality (ISMAR), Online, 9–13 November 2020; pp. 425–437.
202. Barde, A.; Gumilar, I.; Hayati, A.F.; Dey, A.; Lee, G.; Billingham, M. A Review of Hyperscanning and Its Use in Virtual Environments. *Informatics* **2020**, *7*, 55. [[CrossRef](#)]
203. Losey, D.M.; Stocco, A.; Abernethy, J.A.; Rao, R.P.N. Navigating a 2D virtual world using direct brain stimulation. *Front. Robot. AI* **2016**, *3*, 72. [[CrossRef](#)]
204. Lee, W.; Kim, H.C.; Jung, Y.; Chung, Y.A.; Song, I.U.; Lee, J.H.; Yoo, S.S. Transcranial focused ultrasound stimulation of human primary visual cortex. *Sci. Rep.* **2016**, *6*, 34026. [[CrossRef](#)]
205. Farooq, U.; Grudin, J. Human-computer integration. *Interactions* **2016**, *23*, 26–32. [[CrossRef](#)]
206. Sra, M.; Jain, A.; Maes, P. Adding Proprioceptive Feedback to Virtual Reality Experiences Using Galvanic Vestibular Stimulation. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems, Glasgow, UK, 4–9 May 2019; Association for Computing Machinery: New York, NY, USA, 2019; pp. 1–14.
207. Spicer, R.P.; Russell, S.M.; Rosenberg, E.S. The mixed reality of things: Emerging challenges for human-information interaction. In *Next-Generation Analyst V*; International Society for Optics and Photonics: Bellingham, WA, USA, 2017; Volume 10207, p. 102070A; ISBN 9781510609150.
208. Mueller, F.F.; Lopes, P.; Strohmeier, P.; Ju, W.; Seim, C.; Weigel, M.; Nanayakkara, S.; Obrist, M.; Li, Z.; Delfa, J.; et al. Next Steps for Human-Computer Integration. In Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, 25–30 April 2020; Association for Computing Machinery: New York, NY, USA, 2020; pp. 1–15.
209. Hainich, R.R. *The End of Hardware: Augmented Reality and Beyond*; BookSurge: Charleston, SC, USA, 2009.
210. Bariya, M.; Li, L.; Ghattamaneni, R.; Ahn, C.H.; Nyein, H.Y.Y.; Tai, L.C.; Javey, A. Glove-based sensors for multimodal monitoring of natural sweat. *Sci. Adv.* **2020**, *6*, 8308. [[CrossRef](#)]
211. Lawrence, J. Review of Communication in the Age of Virtual Reality. *Contemp. Psychol. A J. Rev.* **1997**, *42*, 170. [[CrossRef](#)]
212. Hendaoui, A.; Limayem, M.; Thompson, C.W. 3D social virtual worlds: Research issues and challenges. *Inst. Electr. Electron. Eng. Internet Comput.* **2008**, *12*, 88–92. [[CrossRef](#)]
213. Wann, J.P.; Rushton, S.; Mon-Williams, M. Natural problems for stereoscopic depth perception in virtual environments. *Vis. Res.* **1995**, *35*, 2731–2736. [[CrossRef](#)]
214. Ahmed, S.; Irshad, L.; Demirel, H.O.; Tumer, I.Y. A Comparison Between Virtual Reality and Digital Human Modeling for Proactive Ergonomic Design. In Proceedings of the International Conference on Human-Computer Interaction, Orlando, FL, USA, 26–31 July 2019; Springer: Cham, Switzerland, 2019; pp. 3–21.
215. Bonner, E.; Reinders, H. Augmented and Virtual Reality in the Language Classroom: Practical Ideas. *Teach. Engl. Technol.* **2018**, *18*, 33–53.
216. Royackers, L.; Timmer, J.; Kool, L.; van Est, R. Societal and ethical issues of digitization. *Ethics Inf. Technol.* **2018**, *20*, 127–142. [[CrossRef](#)]
217. Welch, G.; Bruder, G.; Squire, P.; Schubert, R. *Anticipating Widespread Augmented Reality*; University of Central Florida: Orlando, FL, USA, 2018.
218. Smits, M.; Bart Staal, J.; Van Goor, H. Could Virtual Reality play a role in the rehabilitation after COVID-19 infection? *BMJ Open Sport Exerc. Med.* **2020**, *6*, 943. [[CrossRef](#)] [[PubMed](#)]
219. Huang, H.M.; Rauch, U.; Liaw, S.S. Investigating learners' attitudes toward virtual reality learning environments: Based on a constructivist approach. *Comput. Educ.* **2010**, *55*, 1171–1182. [[CrossRef](#)]
220. Siricharoen, W. V The Effect of Virtual Reality as a form of Escapism. In Proceedings of the International Conference on Information Resources Management, Auckland, New Zealand, 27–29 May 2019; p. 36.

221. Pesce, M. AR's Prying Eyes. *Inst. Electr. Electron. Eng. Spectr.* **2021**, *19*.
222. DeCarlo, C.C. *Toward the Year 2018*; Foreign Policy Association, Ed.; Cowles Educational Corp.: New York, NY, USA, 1968.
223. Aati, K.; Chang, D.; Edara, P.; Sun, C. Immersive Work Zone Inspection Training using Virtual Reality. *Transp. Res. Rec. J. Transp. Res. Board* **2020**, *2674*, 224–232. [[CrossRef](#)]
224. Sowndararajan, A.; Wang, R.; Bowman, D.A. Quantifying the benefits of immersion for procedural training. In Proceedings of the IPT/EDT 2008—Immersive Projection Technologies/Emerging Display Technologies Workshop, Los Angeles, CA, USA, 9–10 August 2008; Volume 2, pp. 1–4.
225. Nigay, L.; Coutaz, J. A design space for multimodal systems. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems—CHI'93, Amsterdam, The Netherlands, 24–29 April 1993; Association for Computing Machinery Press: New York, NY, USA, 1993; pp. 172–178.
226. Covarrubias, M.; Bordegoni, M.; Rosini, M.; Guanziroli, E.; Cugini, U.; Molteni, F. VR system for rehabilitation based on hand gestural and olfactory interaction. In Proceedings of the 21st Association for Computing Machinery Symposium on Virtual Reality Software and Technology, Beijing, China, 13–15 November 2015; Association for Computing Machinery: New York, NY, USA, 2015; Volume 13, pp. 117–120.
227. Yeh, S.C.; Lee, S.H.; Chan, R.C.; Wu, Y.; Zheng, L.R.; Flynn, S. The Efficacy of a Haptic-Enhanced Virtual Reality System for Precision Grasp Acquisition in Stroke Rehabilitation. *J. Healthc. Eng.* **2017**, *2017*, 9840273. [[CrossRef](#)]
228. Manuel, D.; Moore, D.; Charissis, V. An investigation into immersion in games through motion control and stereo audio reproduction. In Proceedings of the 7th Audio Mostly Conference on A Conference on Interaction with Sound—AM'12, Corfu, Greece, 26–28 September 2012; Association for Computing Machinery Press: New York, NY, USA, 2012; pp. 124–129.
229. Shaw, L.A.; Wuensche, B.C.; Lutteroth, C.; Buckley, J.; Corballis, P. Evaluating sensory feedback for immersion in exergames. In Proceedings of the Australasian Computer Science Week Multiconference, Geelong, Australia, 30 January–3 February 2017; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1–6.
230. Triantafyllidis, E.; McCreavy, C.; Gu, J.; Li, Z. Study of multimodal interfaces and the improvements on teleoperation. *Inst. Electr. Electron. Eng. Access* **2020**, *8*, 78213–78227. [[CrossRef](#)]