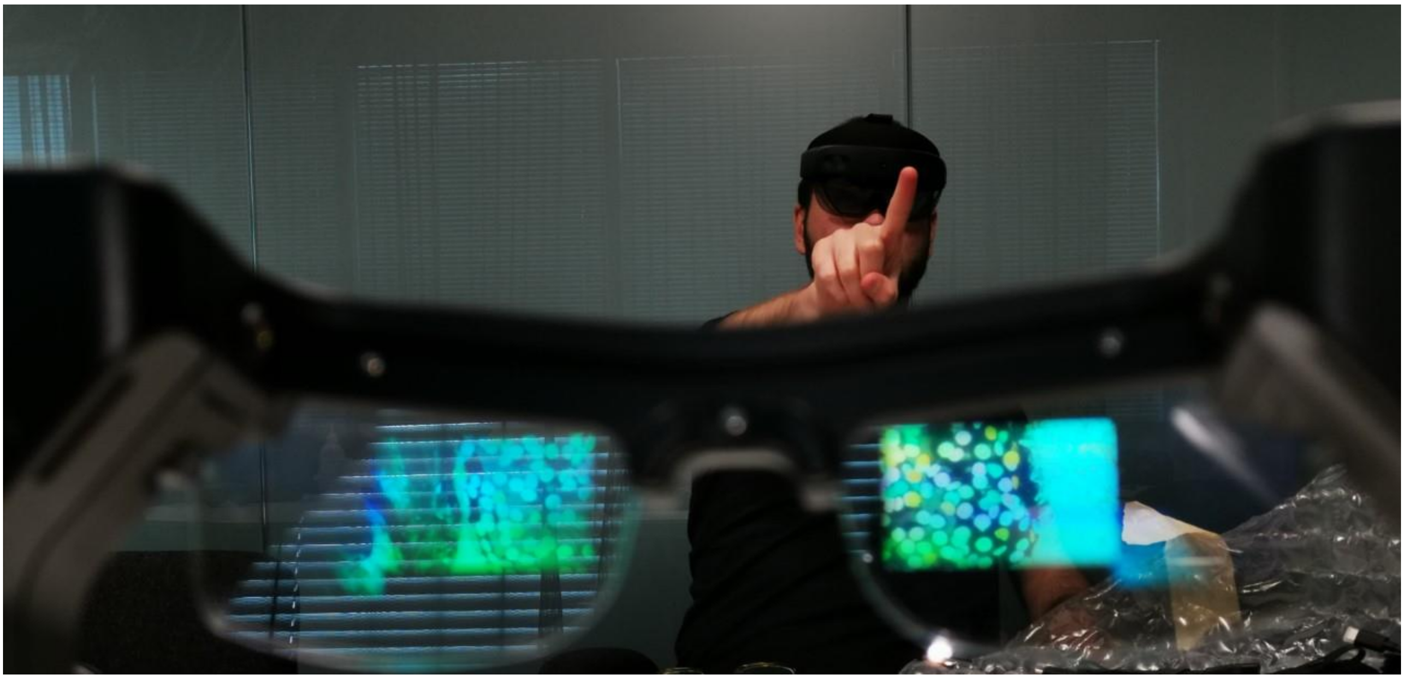
NextGen XR Requirements State-of-the-Art

HumOR (Human Optimized XR) Project



DIPANJAN DAS
Early Stage Researcher
University of Eastern Finland

THOMAS MCKENZIE
Postdoctoral Researcher
Aalto University

UNIVERSITY OF EASTERN FINLAND

Aalto University

PREIN

PREIN FLAGSHIP PROJECT

Author: Dipanjan Das and Thomas McKenzie

Document Change History

Version	Date	Authors	Description of change
01	**1-09-2020**	**Dipanjan Das**	**Formatting done**
02	**7-09-2020**	**Dipanjan Das**	**Formatting done**
03	**15-09-2020**	**Dipanjan Das**	**Revised edition**
04	**6-11-2020**	**Thomas McKenzie**	**Revised edition**

Acronyms used in this Report

AI	Artificial Intelligence
AMOLED	Active Matrix Organic Light Emitting Diode
ANC	Active Noise Cancellation
AR	Augmented Reality
DoF	Degree of Freedom
FOV	Field of View
HUD	Heads-up Display
HumOR	Human Optimized XR
HRTF	Head Related Transfer Function
LCD	Liquid Crystal Display
MR	Mixed Reality
MTF	Modulation Transfer Function
NA	Not Available
OEM	Optical Equipment Manufacturer
OLED	Organic Light Emitting Diode
OS	Operating System
OST	Optical See-through
QoE	Quality of Experience
RAM	Random Access Memory
ROM	Read Only Memory
SBP	Space Bandwidth Product
SLM	Spatial Light Modulator
SMEs	Small and Medium-sized Enterprises
SoA	State-of-the-Art
VAC	Vergence and Accommodation Conflict
VR	Virtual Reality
VST	Visual See-through
XR	Extended Reality

Contents

Forewords

Thanks to **Varjo Oy** and **Dispelix Oy** who are the part of the **HumOR consortium** for allowing us to **interact-one-to-one** with each other for the completion and preparation of this state-of-the-art project report which elaborates the benchmarking specifications of the current **AR/VR** and **whole range of XR products** in the market and how these specifications are implemented or used for the users. This report also describes different device design specifications that serve as the core of the device and help us to better understand performance against **human perception** and **Quality of Experience (QoE)**. This report opens the existing AR / VR / MR terminologies and next generation requirements which will help to better understand the devices. This report is under the work description of the work package 3 (WP3) of the HumOR project and follows up with the first work deliverable (state-of-the-art) in the timeline **M1-M4** and coincides with the timeline of the second work deliverable in the timeline **M9-M22 (**Display Technology Approach**) which is currently ongoing**.

Background

The AR/VR and XR products development process involves design specifications and requirements that are not always satisfactory for the users and introduce some contradictions in terms of huge computing needs, power, wearability, latency, etc. Therefore, to find out the answers to the contradictory questions, which arises to the companies', state-of-the-art analysis is quite important and was studied by undergoing demonstrations and prototypes available with them. Thus, this NextGen XR device requirements state-of-the-art (SoA) report opens to set the performance targets to the industry and helps them to find out the improvement areas and benchmark the specifications.

Methodology

- Methodologies involved in the preparation of this report consists of the company's own official websites, blog posts, journals, books, data sheets, pictures taken during the demonstration (non-confidential and with approval) and visits to the company sites.
- **Varjo Oy** and **Dispelix Oy** are the two companies that are the part of the Humor Consortium were visited for the demonstrations of their own devices and to study the **Quality of Experience (QoE).** This state-of-the-art analysis report preparation also involves **VR lab, Joensuu** to undergo other set of AR/VR and XR products and what are the limitations in using those devices.

1.0 Introduction

This NextGen XR Device requirements State-of-the-art (SoA) elaborately discuss the theoretical foundation for the NextGen XR devices starting from overview. The SoA also illustrates the human factors that are important and how they affect the design of the current XR hardware. The SoA explains in general the specifications of the AR/VR/MR current and evolving devices in the market and explains the terminologies which are used as the design specifications. It describes about how this specification can be benchmarked using the current set of XR products in the market and which areas need such requirements. This SoA also gives a clear answer that helps us to identify how these specifications will be used to study the next generation XR device requirements. In addition to the above, this report studies the specifications available from the companies and what are the challenges that still needs to be implemented to meet the NextGen XR requirements for the industry.

1.1 Reality Overview

The reality overview map can be well understood and illustrated with the help of Milgram and Kishino's Reality-Virtuality Continuum (illustrated in Fig. 1) as explained in [1, 2, 3] and gives us a relation between the real environment (i.e. the world we live in) [2], Augmented Reality (AR), Augmented Virtuality (AV) i.e. bringing real world information into the virtual world and the complete virtual environment or Virtual Reality (VR) world [2, 4]. This whole range of the continuum of reality-virtuality is represented as the Mixed-Reality (MR) spectrum (illustrated in Fig. 2) explained in [3].

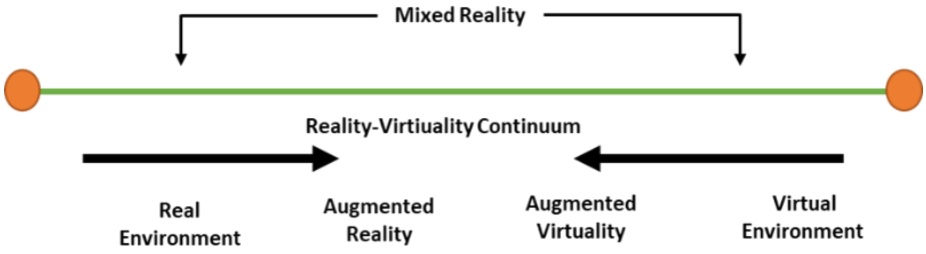


Fig. 1: Reality Continuum 1 [adapted from 1].

The transition from real world to the virtual world in MR continuum zones are subdivided into so called 'Diminished Reality' (which occurs due to inclusion of occlusion which blocks the real-world information partially as explained in [3]), Augmented Reality (Optical-see through MR) and Merged Reality (Video-see through MR) [3].

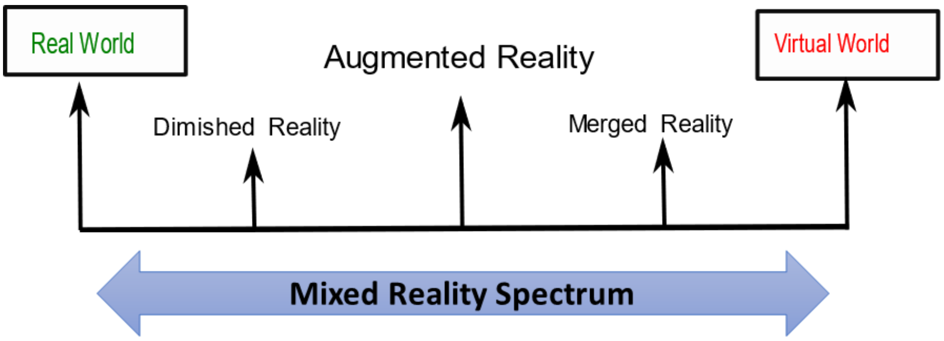


Fig. 2: Reality Continuum 1 [adapted from 1].

Another definition of the Mixed-Reality continuum (see Fig. 3) was further extended by Steve Mann (1962) by introducing a concept of mediated reality and is because one's perception of reality changes or gets modified by the apparatus being worn [1].

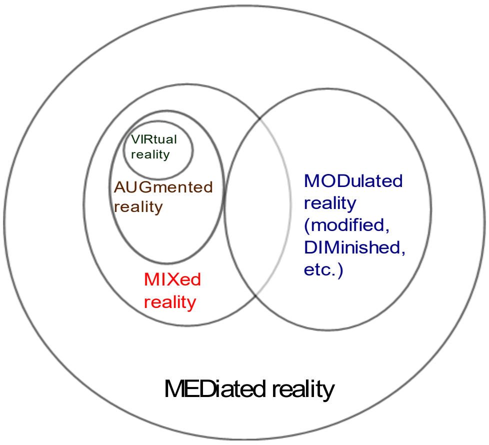


Fig. 3: Mediated Reality [adapted from 1].

The modification can further be interpreted by a taxonomy of reality by Mann (see Fig. 4) where the origin R is the unmodified reality [1], x and y axis denotes the Virtuality axis and mediated reality axis which also explains the transition to produce the combined effect of the augmented and virtual realities.

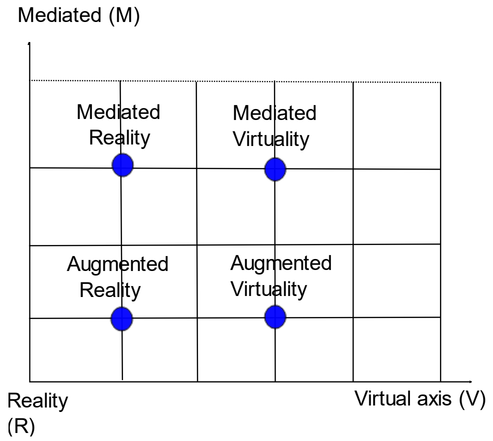


Fig. 4: Taxonomy of Reality from Mann [adapted from 1].

In the next sections, each of the MR continuum sections are explained in detail.

1.2 What is Augmented Reality?

The term Augmented Reality (AR) in general is the overlaying of the virtual information on the real-world scene information also explained in [1, 2]. In other words, we can state the definition of the AR as the superposition or alteration of the digital information or content projected into the user's real-world view.

1.2.1 AR systems

In AR systems (glasses, helmets, HUDS, earphones, etc.) [1] digital information is created with the help of a processing unit and data source or sometimes with the help of a data base as discussed in [1] and is displayed using an AR display which serves as the one of the most important building block of the AR system. The optical building blocks are still needed for AR System to work and they comprise of the optical systems (display engine, combiner engine, sensors, visors, Rx as illustrated in [3]) and Fig. 5 illustrates it.

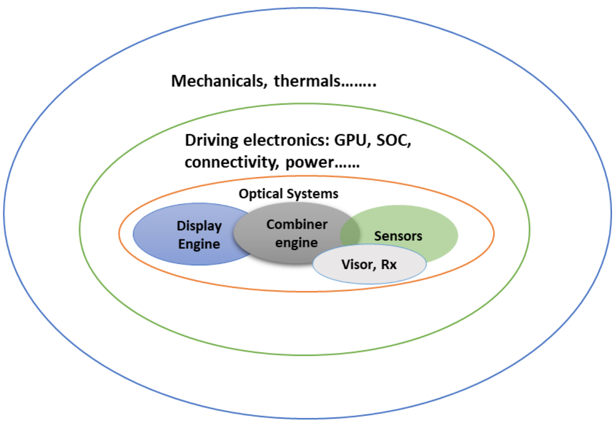


Fig. 5: Optical building blocks of an AR system [1,3].

From Fig. 5, it can be seen that for a visual AR system, the optical system is composed of display engine, combiner engine and the related sensors. The GPU drives the display used in the AR system. The display/projection engine routes or creates the images for user's Field of View (FOV). Related sensors of the AR system refer to the camera, real world position to map the scene in 3D, motion, elevation, eye tracking. Visual AR systems also support audio (microphones, speakers) for communication purposes and the current AR systems also employ spatial sound. In addition to the above, AR systems can help in object identification and categorize where the system is looking into to position the virtual content according to user's selection. Some of the AR systems also use markers for object recognition and some examples are explained in.

Auditory AR devices typically utilize arrays of microphones both inside and outside the earphones. So called 'hear-through' devices place virtual sounds on top of the real world, such that real-world sounds are unaffected. Alternatively, with active noise cancellation (ANC) technology, real-world sounds can be suppressed or altered while desired sounds are amplified, such as in speech enhancement applications. Auditory AR can provide contextual information about the user's surroundings, for context and notifications either in conjunction with a visual AR display or as standalone devices, which avoids disturbing the user's visual environment. Applications include hands-free and vision-free GPS navigation or traffic updates, speech enhancement and real-time language translation.

The applications of AR systems for consumer, commercial and industrial / scientific purposes can be mapped using a simple illustration in Fig. 6. The activity can be categorized as both passive (for general purpose devices like smartphones, tablets, PCs etc.) and interactive (integrated, dedicated) depending on the purpose it serves.

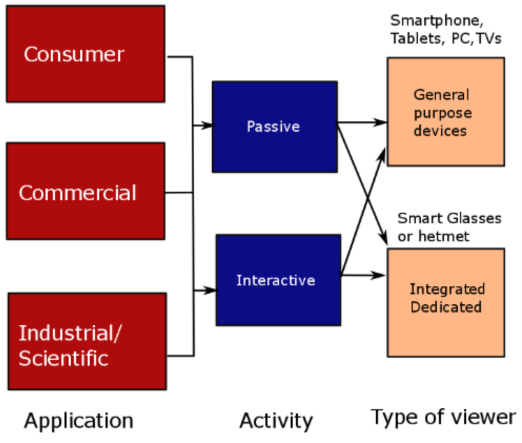


Fig. 6: Augmented Reality applications map [1].

The next section describes about the AR taxonomy and the NextGen AR displays which are currently present or will evolve in the market.

1.2.2 Types of AR systems

The AR systems involves AR displays which can be subdivided into two types: Visible-see through (VST) and optical-see through (OST) display according to [3, 4]. Visual see through AR systems have seven classes and is also explained in [1]. The seven classes are: contact lens, helmet, head-up display (HUD), headset (smart glasses) integrated and add on display, projectors, specialized for other purposes. AR systems can be wearable or non-wearable depending upon the purpose of application also illustrated by the AR taxonomy (see Fig. 7).

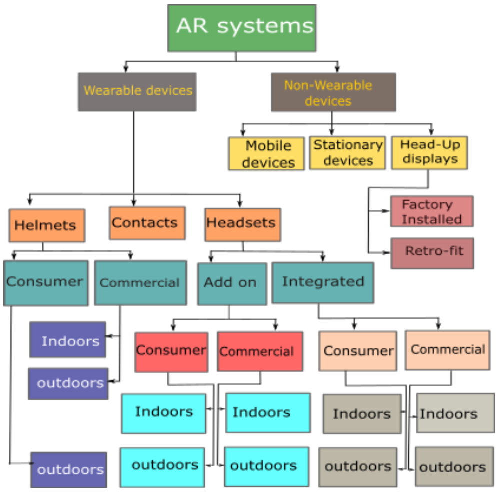


Fig. 7: AR taxonomy [adapted from 1].

1.2.3 Wearable AR Displays

The term 'Wearable AR displays' belong to two classes. The first is 'smart glasses', which are capable of providing a discrete display for user experience with can show only graphical information, mostly pictograms and basic text and is not showing any stereoscopic 3D scene or image. The other class actually represents 'stereoscopic display systems' which can represent 3D stereoscopic content and ensure the spatial tracking which makes things complex and system bulkier as explained in [5]. But these two categories suffer from several challenges relating to how the graphical information (virtual image) are displayed to the user. These challenges might be related to display the information or the content at short distances in front of the user Field of View (FOV) (both horizontal and vertical) or it might be due to the brightness levels of looking into the content in different lighting conditions as explained in [5]. Mostly, this system is monocular (for one eye) and 3D scene cannot be presented in that case. To represent the 3D scene, binocular vision is quite important and is possible with the class of the 'Stereoscopic head mounted AR displays' which is described in the next section.

1.2.4 Stereoscopic head-mounted AR displays

The 'stereoscopic display systems' as discussed above for displaying 3D scene includes display systems which are called 'stereoscopic head-mounted AR displays' which gives 3D spatial information or awareness to the user for understanding the professional tasks that might be overlaid with the real world [5]. The term so called 'stereoscopic imaging' which uses human binocular vision (using two eyes at the same time to present an image) is also not free from binocular disparities which gives the 3D depth information. The binocular disparities give rise to 'stereopsis': depth perception produced by the brain through comparison of the visual stimuli from both eyes (explained more in [6]). The image formation in stereoscopic imaging is done in such a manner that the image is formed at the fixed focal distance and this image formation also needs fixed accommodation. But, the vergence of the human vision system keeps on varying and has the freedom. This causes the conflict between the naturally occurring accommodation and vergence results in vergence and accommodation conflict also called as 'VAC' (see Fig. 8) and is compensated using different kind of displays described as the 'next generation 3D displays' in the next section.

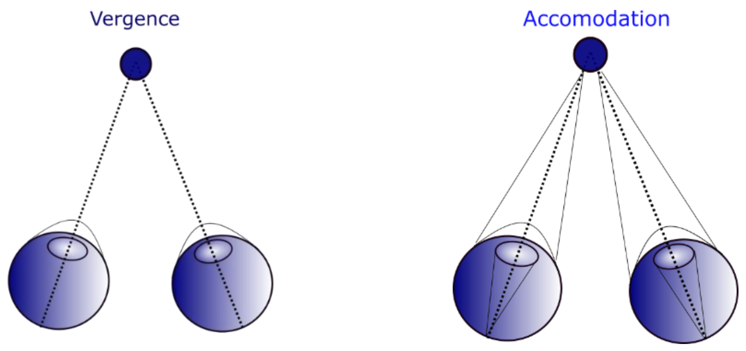


Fig. 8: Mitigation of VAC and FOV [adapted from 6].

1.2.5 Next generation 3D displays

The problem with the AR displays discussed in the previous section to represent a 3D scene is a big challenge and 3D displays are required to eradicate this problem and the overview is discussed in [5].Vergence-Accommodation conflict (VAC) is the root cause of the conventional stereoscopy and can be eradicated with the help of the next-generation technologies discussed in [5]. 3D scene reconstruction requires to recreate a true wavefront of the light that emanates from the scene and this can be only achieved with the help of holographic displays and this technology is used by the companies such as 'VividQ' and they use the technology in their AR system. The reason is that true wavefront carries a large amount of information and can in future help for real-time data rendering, transfer of the data and for other purposes as well [5]. Thus, the next section describes how we can compensate the problem with the wearable and stereoscopic 3D AR displays that will solve the limitation of VAC problem, but also additional problems related to the next generation of AR devices [5].

1.2.6 Varifocal displays

Varifocal displays are the special kind of AR displays that introduce eye tracking in order to determine the gaze direction of the viewer. It determines the angle of vergence of both the eyes to change the focal distance of the image plane to take into consideration that vergence always matches with the accommodation as discussed in [1, 6] in order to represent the correct monocular cues. The main reason for using varifocal displays is continuously actuating the focal distance and avoiding the problem of the VAC. The varifocal designs are quite common in the AR display world and the companies like Magic Leap have implemented varifocal designs in the AR headsets and the specifications have been listed in the collection of specifications of the AR/VR devices section discussed in this State-of-the-art report. Thus, the problem of VAC is mitigated from their headsets and set to use two discrete focal planes and the switching between the focal planes depends on the user's eye tracking. The implementation of the second focal plane improves the quality of the image and make it quite close to give a sharp 3D image. Varifocal design is implemented, and prototypes are presented and discussed in [6] to explain its effect for near-eye see-through AR systems. In the next section, the multi-focal displays are explained which gives another improvement in the field of the next generation improvement in the AR display technology in contrast to the varifocal displays.

1.2.7 Multifocal displays

The Multifocal displays are kinds of AR displays which does not need eye-tracking to provide true monocular cues to the user. This can be implemented by making of human perception to focus on the multiple focal planes by the using the method of the time-multiplexing discussed in [5, 6, 7] (multifocal displays) or by using dense collection of focal planes as described in [7]. But the main issue in focusing to the multiple stacks of focal planes is the reduction of the maximum image brightness as explained in [5, 7]. If the total number of focal planes are 's' then a single focal plane is shown for only '(1/s) fraction of time and this hampers the brightness and full color capability issues. To implement such multi-focal approach in the AR products in the market has been challenging. The companies like 'Avegant' has already implemented such designs for the next generation displays that can be used and are also listed in the collection of specifications of the AR products in this report. The same multi-focal approach was also explained by another company called 'LightSpace Technology' which utilizes Volumetric technology approach. In the next few sections, the improvement in image brightness, stereoscopic imaging is discussed with the help of new class of displays which will bring more life to the previous versions and eradicate the limitations [5].

1.2.8 Light field displays

The Light-field displays are kind of next generation displays that employs both the feature of adding monocular cues and angle-dependent lighting effects such as glare as explained in [5]. Angle-depended properties can be simulated in this type of the display would be give a high advantage as the position of the human eye is fixed and typically changes with movement of the head and the body. The rendering pipelines with the help of positional and tracking systems gives the track of head and eye movements and can interpret the rest process related to analyze the angle-dependent properties. But still this light-field displays has number of computational challenges explained in [5] and [10]. The number of the discretized images or views needs to be very high in order to compensate the challenge and this might also lead to the interference with the real-time capability. The next challenge is the high-bandwidth for light-field data transfer which in this era can be possible as data compressibility is possible for full set of the light field information. But the physical interpretation of the light fields is still an issue that is existing. Companies like Nvidia have already presented such physical interpretation by utilization of the micro-display with the overlaid lenslet array and is explained in [5] and usually left with issue of low image resolution (due to pixel sharing into sub-portions). To compensate the problem of the low resolution, startup companies like 'Creal3D' has also presented their time-multiplexed multiple pin-light sources for illumination of a reflective Spatial Light Modulator (SLM) technology to get the full-scale resolution. Thus, this concept serves as the emerging concepts to curb the image resolution issues and can serve in the list of the next generation requirements for the AR display technology. But this advanced feature still not give the information about the true wavefront and can only be described with the help of the next section which focuses on the holographic displays.

1.2.9 Holographic displays

The holographic displays will be the actual aim of the next generation requirements for the wearable technologies to come and for future. But this involves lots of challenges to employ this technology as it has an image source problem according to [1]. Some challenges have also been addressed in [12, 13] and one of them is the limited space bandwidth product (SBP) and how we can handle the challenges is the key to employ this technology for the future AR products in the market [5]. The next phase of the challenge introduced is the challenge of wavefront reconstruction where phase information is recorded, and this now requires completely coherent light and is also discussed in [14, 15]. This challenge was partly approached with the help of the integral 2D imaging as explained in [1]. For recent years, the creation and calculation required for the computer-generated 3D content was not a big problem and possible holograms can be generated though it turns out to be resource intensive step. Graphical data processing has also improved in recent times and companies like Microsoft has demonstrated real-time holographic image rendering pipeline as explained in [5]. But problems still exist with those systems also these systems use the laser system which are not speckle free and create image artifacts. In addition to this, another area where improvements to be done is the need of an ultra-high resolution Spatial Light Modulator (SLM) which limits the viewing angles and the eye box for the next generation wearable displays. If the above challenges can be implemented then holographic systems are great for the emerging future of the wearable display as they can give information about depth cues, monocular focus cues, computation nature of these displays make them free from optical aberration correction.

1.2.10 Comparison of next generation 3D displays for AR systems

This section compares the whole set of the 3D displays for implementation in the next generation to understand the mitigation of VAC, monocular focusing Eye tracking, computational requirements and reuse of the existing stereoscopic information. This comparison (Fig. 9) enables us to take a deep dive into the next generation requirements and is adapted from [5].

	Conventional Stereoscopic	Varifocal	Multifocal	Light field	Holographic
Vergence-accommodation Conflict	Yes	No	No	No	No
Monocular focusing cues	No	Computed	Yes	Yes	Yes
Eye-tracking	No	Yes	No	No	No
Image resolution	Best Good	Best	Good	Good-fair	Fair
Computational Requirements	Minimal	Moderate-large	Minimal	Large	Large
Reuse of existing stereoscopic content	Yes	Partial	Yes	No	No

Fig. 9: NextGen 3D displays for AR systems [modified from: 5].

The previous sections in this state-of-the-art explained the basic and general overview of the term called Augmented Reality (AR), AR systems and what are the different AR displays. The previous sections mentioned in this report also explained about how introduction of the next generation 3D display will play an important role in transforming the AR vision for the future and what are the requirements that needs to be present for great user experience for using AR systems. The next topic in this report relates to the term called Virtual Reality (VR), types of the VR devices, human factors affecting VR and its applications.

1.2.11 Auditory AR technology

Arrays of multiple microphones process the audio from outside, allowing for varying degrees of noise suppression. In ANC technology, artificial intelligence (AI) algorithms are often employed in the sound classification process and speech enhancement applications, such as in Sony headphones [118]. Hear-through technology attempts to compensate for the acoustic effects of wearing headphones such that the auditory experience nears that of when not wearing any headphone at all. This paves the way for playback of additional audio, which can be overlayed on the real-world audio. By using microphone arrays with several directional microphones, beamforming technology can be utilized to analyse the direction of incoming sounds and therefore isolate sound from the desired direction while reducing the volume of other directions. Real-time language translation devices use speech recognition algorithms, language translation, and text-to-speech technology, in conjunction with ANC, to deliver near real-time language translation.

Auditory AR devices can also utilize binaural audio to place virtual sounds in real-world scenarios and environment (see Section 1.3.8).

1.3 What is Virtual Reality (VR)?

The term 'Virtual Reality' (VR) is defined with different definitions from the literature. The 'VR' defined by [16] states that, VR is defined as inducing target behavior in an organism by using artificial sensory stimulation, while the organism has little or no awareness of the interference. This further breaks the definition of the VR into four subsections in the definition itself and these are the targeted behavior, organism, artificial sensory stimulation and awareness. Then, another definition from [17] breaks VR in two terms: 'virtual', which means near, and 'reality', which means what we can express as human beings. If we take together from the second definition it combines to become ‘near-reality’. This suggests that it can mean anything and can make something called reality emulation according to [18]. In technical terms, we can also say that VR is a term which is used to describe ‘a 3D computer generated environment. This virtual environment can be explored and interacted by a person who uses this virtual environment.

VR environment creates a feeling of immersion and is generated by the combination of headsets, omni-directional treadmills, loudspeakers or headphones and special kind of gloves and this combination can stimulate our senses. After the definition of the ‘Virtual Reality’ was clear, we can elaborate on the types of the VR systems. So, in the next section we will discuss about types of the VR system that can be used to describe the next generation device requirements.

The VR system can be explained with the help of a simple flow diagram (Fig. 10), adapted from [2].

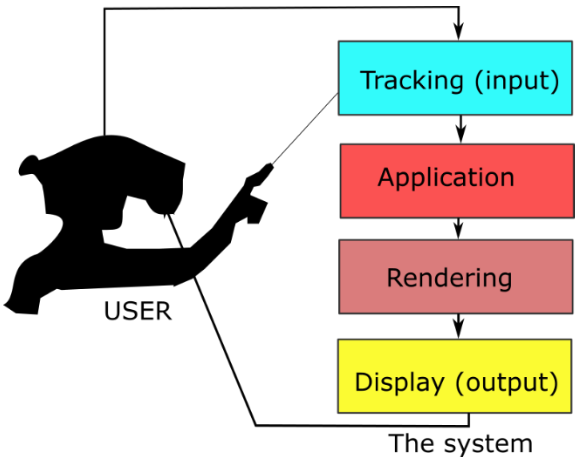


Fig. 10: Schematics of a simple VR system [adapted from 2].

The VR systems according to [2] can be used for direct and indirect communication. The direct communication can be inserted by the VR developers as an intermediary environment that can directly interact and is subdivided into structural and visceral communication. On the other hand, for indirect communication VR environment creators can make indirect stories that indirectly affects the user experience.

1.3.1 Types of VR systems

The VR platform can be integrated to either the mobile platform like explained in [19] but can also be tethered with the PC as explained in [20]. There are pros and cons of both the systems and can be expressed in the next section where it is discussed in detail.

1.3.11 Mobile VR

The mobile VR Platform already described by the name itself uses mobile phone or handset for immersive experience in front of the user’s eyes and the companies have used this platform to provide immersive experience to the user in terms of different applications like gaming, etc. The pros and cons of the mobile VR platform are discussed here and taken from [18].

Pros of Mobile VR:

1. Reliable, flexible to use anytime.
2. Ease of use with no additional setup required.
3. Battery powered sleek.
4. Ultra-light design.

Cons of Mobile VR:

1. Limited by power and thermal requirements.

1.3.12 Tethered VR

The tethered VR Platform works with what is known as the PC-based interfaces and uses strong computing requirements for the operations and have the following pros and cons for its operation. Tethered VR operations can provide virtual trainings for education, flight simulator for training of the pilots and other applications discussed in.

Pros of tethered VR:

1. Not Limited by power and thermal requirements.

Cons of tethered VR:

1. Expensive for high-end experiences.
2. Limited by use of wires and block intuitive actions and immersion.
3. Usage limited to a fixed location.

After covering the type of the VR platform, now it is important to go through VR device applications and address the issues related to the VR devices in the next two sections.

1.3.2 VR device applications

There are usually wide variety of applications for VR as explained in and the list includes not only architectures, sports, medicine, arts, entertainment but also flight simulator experience, trainings, entertainment etc. There are other applications, but this is not described in detail here. The focus is to cover the next generation requirements and specifications of the VR displays and the problems that can be mitigated so that the current devices meet the VR requirements. But to see the details into the problem, the next section which contains information about the issues that are related to the VR systems.

1.3.3 Issues in VR devices

There are several issues that have been accounted as described in [2] which states the current problems and, in this section, we focus on the common issues seen in a very general way. The issues stated with the VR here are the current challenges and can be met with the improvements in the next generation device specifications to curb them.

1.3.4 VR sickness - human factor for VR

The term ‘VR sickness’ is usually an issue which resembles as a human factor for the VR users as explained in [2]. According to [21], the VR sickness describes basically the ‘motion sickness’ which occurs due to the poor ergonomics of the device itself. This affects the user of the VR device in the different ways and is described in [2]. The root cause of the motion sickness comes into action when the user’s head remains in a static position with respect to the moving VR world. But many of the current VR devices also have tracking systems that can automatically adjust this shift in the motion and adjust the image accordingly. Thus, there misorientation or the mismatch creates the feeling of nausea in the user and this phenomenon is called ‘Cybersickness’. There has been extensive research work related to the cybersickness reduction in the next generation devices and the methods of the cybersickness reduction is not discussed in detail for the state-of-the-art-report. Cybersickness is also created by a time lag also known as ‘latency’ and will be discussed in detail in the next section.

1.3.5 Eye strain, seizures, aftereffects

Eye strain, seizures and aftereffects can also be caused by the non-moving stimuli as explained in [2]. This can even cause serious level of discomfort and might lead to the adverse heath deterioration.

Some of the problems includes VAC, binocular-occlusion conflict and flicker as explained. The aftereffect is the phenomenon which occurs after spending considerable time with the VR devices. It has been found that the 10% of the those using simulator experience suffers from the aftereffects. There is a certain span of time after which the aftereffect can be readapted, and the user comes back to his normal world [2].

1.3.6 Hardware challenges for VR

The hardware challenges of using VR headsets still persists due to the fact that the device itself can cause physical fatigue, headset might not be a good fit for the user, it might also cause injuries while wearing it and can affect the human hygiene. Thus, a lot of improvements have also been made by the companies who are making the VR devices and a lot has been done to improve the ergonomic design so that the physical weight of the display is reduced and make the user feel comfortable wearing the device. The weight balance of the device is quite important in such a case and the weight of the collected device specification is one of the specifications that is important to note for this state-of-the-art report. Due to the phenomenon of cybersickness which results as a result of the immersive VR experience, this can result in injuries related to human vision and is termed as 'physical trauma' by [2]. In addition to the physical trauma the immerse VR experience can also result in Repetitive strain injuries for the repetitive use of carpal and metacarpals movements as explained in [2]. Human health and hygiene are also affected by different user of the same VR equipment as the phase masks used for the devices are more prone to bacteria, germs and microbes. Thus, there is a quite a lot of challenges which needs attention from the VR community and the people who are working with the VR devices in the future.

1.3.7 Latency

The term 'latency' related to can defined as the 'time' that a system takes to respond to a user's action and is expressed as the true time from the start of the motion to the time a pixel resulting from the motion responds. The type of the display used in the VR device decides the latency of the device. This latency can be reduced using prediction warping according to [2] and the aim will be to stabilize the scene in the real world, keeping in the actual latency is always greater than zero. The latency also causes motion sickness or cybersickness which are already explained in the previous sections. There are several negative effects of latency as explained in [2]. The negative effects include degraded visual acuity which results in the motion blur and can be hard for the user distinguish between the moving objects. Thus, this becomes one of the challenging areas where different companies who have their VR headsets is trying to achieve a very low latency to produce a high-quality VR experience [2]. Audio latency is less sensitive than visual, as around 20 ms is considered the just-noticeable-difference.

1.3.8 Auditory VR technology

Auditory VR devices utilize binaural audio through head-related transfer function (HRTF) convolution, which imparts the directional cues of the human auditory system to the sounds of the virtual world played back through headphones, such that virtual sounds originate from outside of the head in a specified position in space. Changes to the user's position in space and head orientation, from a position tracking system, are compensated for in the audio engine. Binaural audio can utilize artificial reverberation to make sounds appear to originate in the same acoustic space as the real-world sounds, which improves immersion [119].

1.4 What is Mixed Reality (MR)?

Mixed Reality (MR) according to the continuum explained in the previous section 'overview of the reality' is defined by the full spectrum of the devices that includes the transition to real world on one side of the continuum spectrum and completely virtual world on the other side of the spectrum and is explained also in and also is a key requirement for the future of the next generation devices. In general, the MR block diagram is explained by Fig. 11, which illustrates the schematics of the MR device that involves the terminologies like head tracking, gesture sensing and depth mapping (for tracking of the human experience and user interaction with the MR world). The block diagram (see Fig. 11) also explains in general the display light engine which serves as the core heart of the MR device and is also explained in [3] greater details, imaging optics, exit pupil expansion which can be achieved with the help of the combiner optics.

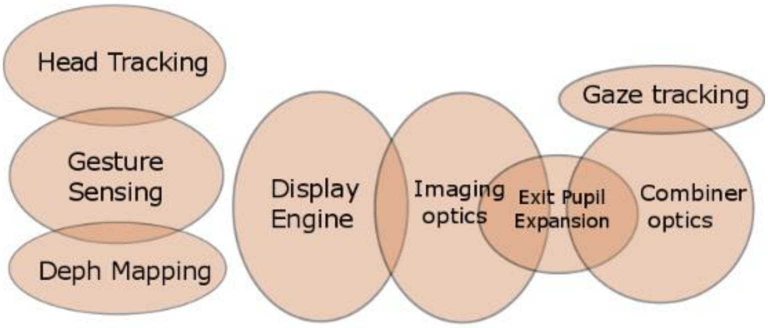


Fig. 11: MR block diagram [adapted from 3].

One of the most important blocks of the above MR block diagram is about 'gaze tracking' which is also one of the key areas where several works have been presented. To have the ultimate MR experience there should be combination of two important aspects one is the 'comfort' and the other 'immersion' and the specifications to achieve the next generation MR experience are presented in the Fig. 12**.** The 'comfort' category is grouped into angular resolution and reducing screen door effect, size/weight and controlling the center of gravity or the weight balance, VAC mitigation, brightness and contrast, ghosting, pupil swim wide IPD coverage and the prescription correction, untethered, Motion to photon latency, fast eye tracking, 3D displays.

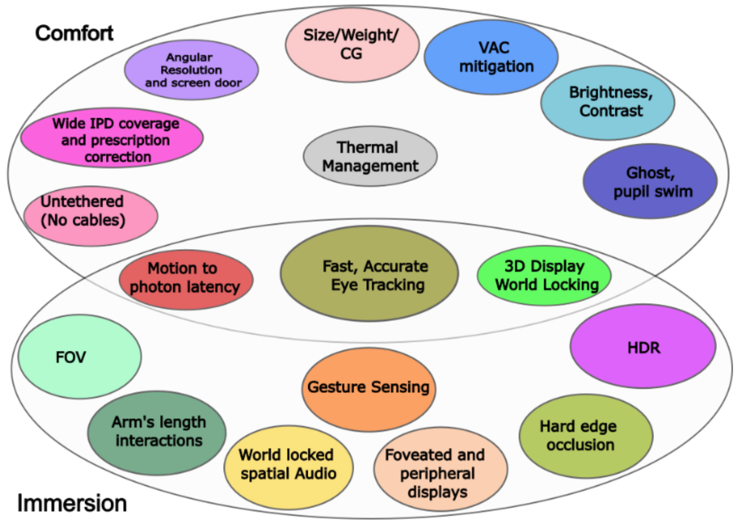


Fig. 12: Comfort and immersion requirements for Ultimate MR experience [adapted from 3].

In the similar manner, the other category 'immersion' is grouped into FOV, gesture sensing, HDR, arm's length interactions, world locked spatial audio, foveated and peripheral displays and hard occlusion. The specifications tabulated in the next sections of the report contains information about the next generation devices and the devices which are in the development phase and trying to implement the above. But there is also a requirement to understand the human factors (like the human visual system and human vision FOV) for the next generation MR requirements.

1.4.1 Human factors for NextGen MR requirements

The human factors are mainly responsible for looking into the viewing performance of the system and there is a need to understand the basics of the human visual system which consists of the line of sight and the optical axis, lateral and longitudinal chromatic aberrations correction by the human eye, visual acuity (the ability of the human eye to resolve small features), stereo acuity and stereo disparity and specifics of the human vision system, the eye model and the specifics of the human vision which is quite important to study one of the main targets of the next generation XR devices. Human-vision FOV is explain in the next section to better understand it.

1.4.2 Human vision FOV

Human-vision FOV can be divided into two: the 'horizontal FOV' which gives the horizontal angular extend of the binocular human vision (shown in Fig. 13a) and the vertical FOV (see Fig. 13b). The span of the H-FOV is 220° but the binocular vision is only 120°. The stereopsis is limited to 40° in relation to the fixation as explained from [3]. On the other hand, the vertical FOV is 15° off below the line of sight (see Fig. 13b) and further relaxation of the head gaze lowers the line of sight by 10°. Human-vision FOV can be defined as a dynamic concept because of the unconstrained and constrained movement of the eye. The mechanical motion of the eye can be with ±40° H but the unconstrained movement of the eye with reflex is only limited to ±20° H. Thus, this gives a total static foveated region ranging from 40-45° H. The Fig. 14 shows the overlap of the right field and the left fields and the human binocular FOV and is horizontally and vertically asymmetric. The small black circle in the Fig also gives the foveated region in which the sustained gaze angle is possible and matches with the current FOV of the AR/MR devices. It has been found that for a given gaze angle, the color recognition occurs at ±60° FOV, shape and text-recognition occur at ±40° FOV and ±30 ° FOV respectively.

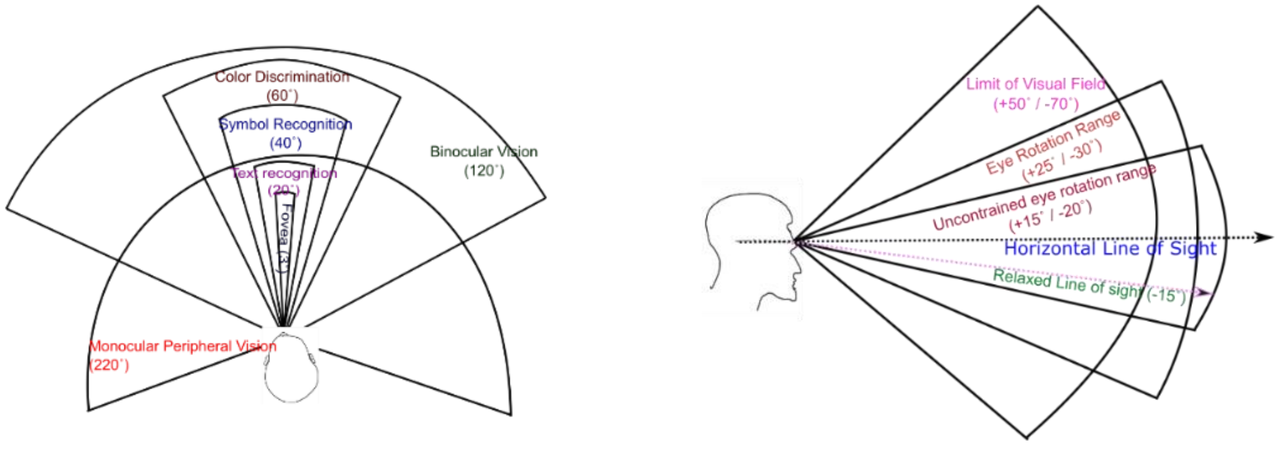


Fig. 13: (a) Horizontal and (b) Vertical FOV of human vision [adapted from 3].

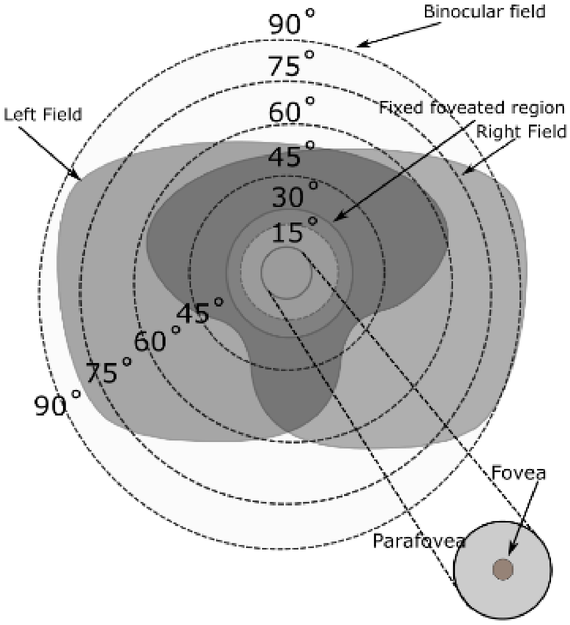


Fig. 14: Human binocular FOV - overlap of the left and the right fields [adapted from 3].

AR/VR headsets and mobile phones have wide variation of the of FOV and can be understood with the help of Fig. 15 where the display diagonal FOVs are mapped around fixed foveated display region which generally gives a high resolution, color depth and also allows sustained gaze as well as accommodation.

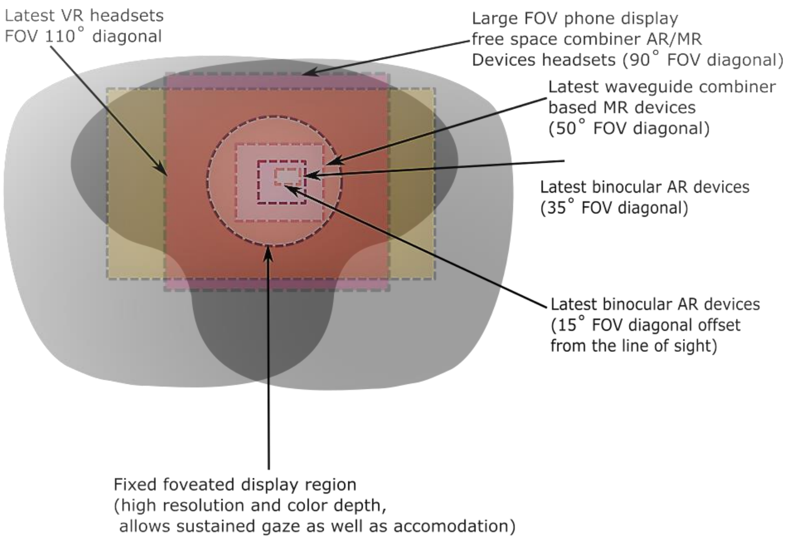


Fig. 15: Typical FOV of the current SoA smart glasses overlapped with the binocular Human vision [adapted from 3].

1.4.3 Display FOV and see-through FOV for various AR/VR smart glasses

For the current AR/VR devices the display hardware adaptation to human binocular vision is quite important and for this one needs to understand two terminologies for the AR/VR devices. The two terminologies are the Display FOV and the other is the see-through FOV. In case of VR devices, see-through is not important and display diagonal FOVs (D-FOVs) are important. Fig. 16 illustrates the overlap of the human binocular vision showing the asymmetric cross-section (left and the right fields) together with the see-through and display FOVs.

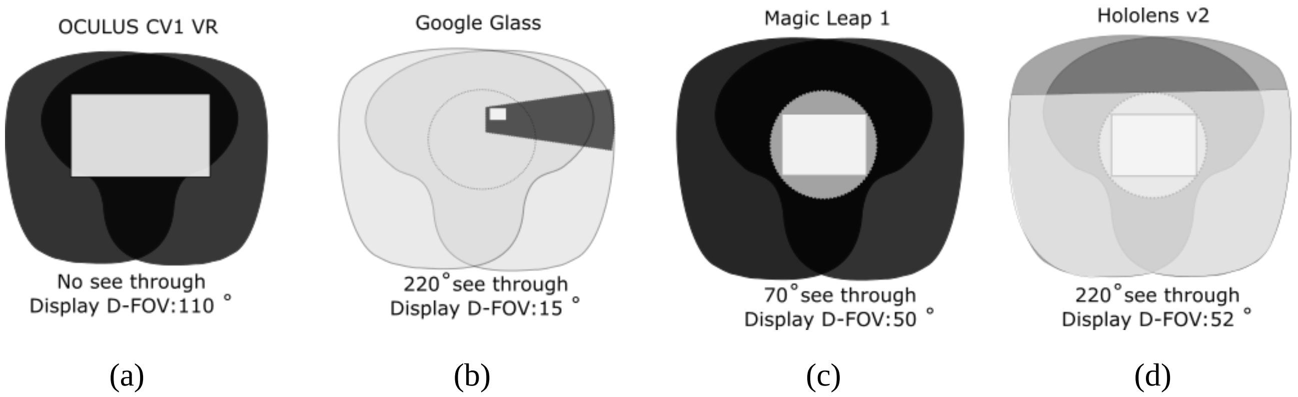


Fig. 16: Display FOV and see through FOV for various AR/VR smart glasses [adapted from 3].

After coming across AR, VR and the MR devices it is important to discuss the next section which represents the whole ecosystem of the next generation of the AR, VR and MR experience and is called the Extended Reality (XR).

1.5 What is Extended Reality (XR)?

Extended Reality (XR) is defined as the whole sum of the AR, VR and MR device properties into the next generation stand-alone devices that serve as the ultimate user experience. This is also present in the XR platform idea from Qualcomm [22], which bunches up the entire concept and explains the current issues with the XR world and what we need to do in order for the device to serve as a XR device which will serve as the only one platform to interact and change the whole digital word and transform it with XR.

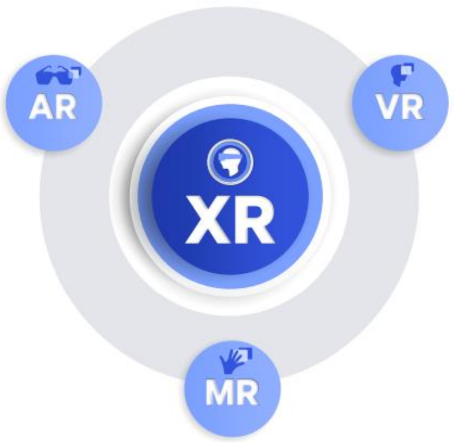


Fig. 17: XR Platform [adapted from 22, Image Courtesy: Qualcomm].

According to the current analysis presented, it has been found that XR community will be huge and it will be the one of the platforms connected to the whole ecosystem which can be called as an XR ecosystem. In the next section the XR ecosystem and its wide scale of applications are discussed in the next section.

1.5.1 XR ecosystem and the next generation XR device requirements

The XR ecosystem as described in the above section can be pictorially represented in Fig. 18 which covers the whole range of areas where the next generation XR devices can be implemented. The application areas (see Fig. 18) are industrial and manufacturing hubs, to the field of engineering, retail, marketing and advertising, entertainment world, in case of emergency response, in military and defense applications, training and for medical field also.

In the regard of the XR ecosystem, it will grow even bigger in the coming future and the chain can go even longer.

Fig. 18: XR ecosystem [Image Courtesy: 22]

But we still need one of the future XR glasses that will become a reality. One similar model is presented in this section (see Fig. 19) where it is illustrated that a next generation XR glass consists of the several features such as new optics and projection with semi-transparent displays, eye tracking cameras, Ambient light sensors, passive and active cameras with the fish eye and telephoto lens, optoelectronic night vision and thermal imaging sensors, bone conduction transducers, rear camera, directional speakers, tracking and recording cameras, inertial, haptic, environmental and health sensors, multiple sensitivity audio microphones. But to include these features into one will be the challenge and in the next section the challenges in the XR technology is described.

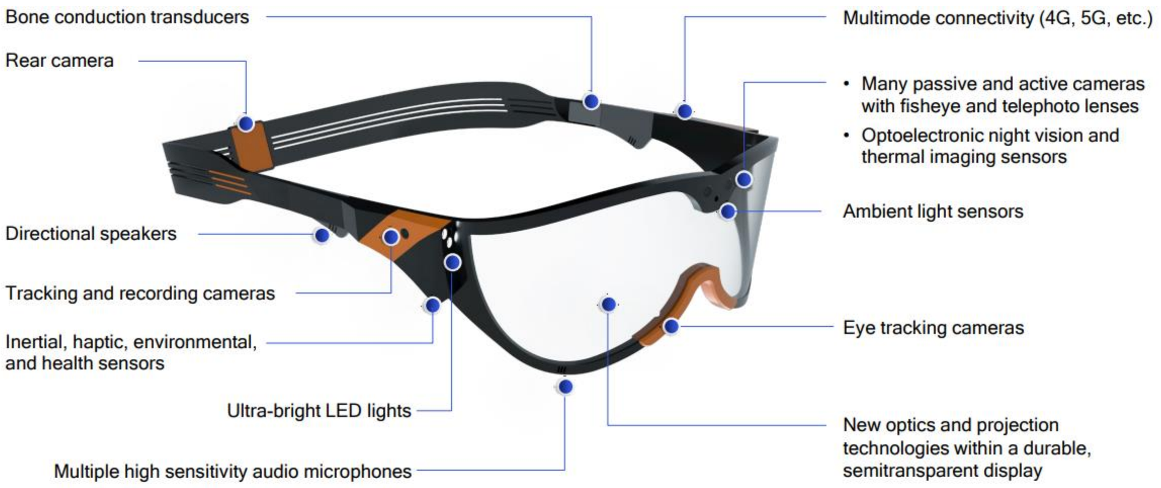


Fig. 19: NextGen XR Glasses [adapted from 22].

1.5.2 Display challenges

The display challenges here mean to convey richer visual content and switching between fully and partial virtual world [22].

1.5.3 Common illumination

The concept of the common illumination means making of the virtual objects in the augmented world indistinguishable from the real objects with the same view [22].

1.5.4 Motion tracking

The term motion tracking here means complete intelligent device embedded tracking for intuitive head, hands and eye interaction [22].

1.5.5 Power and thermal efficiency

This is also crucial to use the device for full day so that the operation continues for a longer period, keeping in mind sleek, light versions with no external cooling devices like fans [22].

1.5.6 Connectivity

In order to render the information faster and to avoid the latency issues faster networks such as 5G can be implemented for cloud computing (see Fig. 20 and Fig. 21) [22].

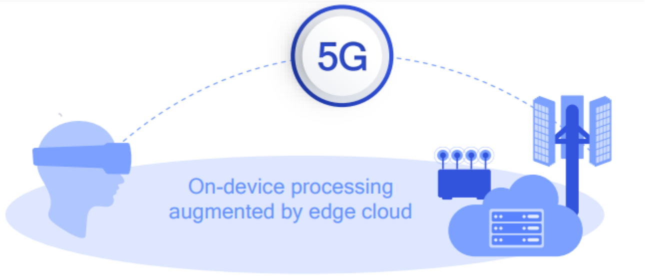


Fig. 20: Role of 5G with XR on device processing [adapted from 22].

Fig. 21: NextGen XR device [adapted from22].

1.5.7 Current immersive Virtual Reality extreme requirements and challenges

Immersive Virtual Reality (VR) extreme requirements require to focus on three pillars as stated in the [22]. These three pillars are visual quality, sound quality and intuitive interactions (see Fig. 22).

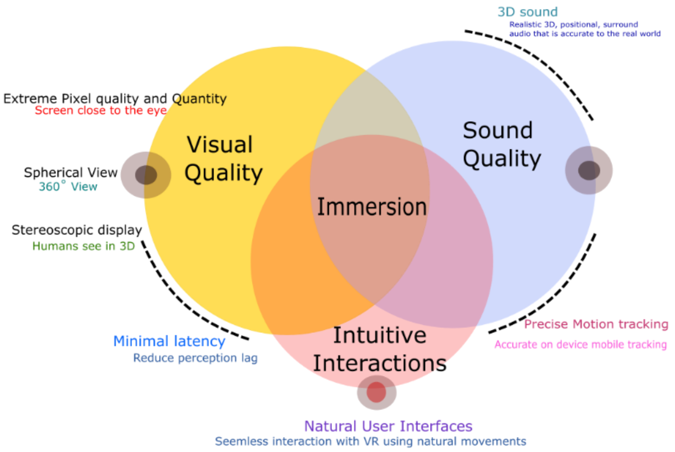


Fig 22: Extreme requirements for immersive Virtual Reality [adapted from 22].

1.5.71 Visual quality

To have the ultimate visual quality, there should be extreme pixel quality and quantity as the screen is very close to the eye. There must be spherical view which will enable the users to have full 360° coverage. For immersive VR, our entire FOV needs the virtual world. There should be utilization of 145° horizontal FOV. The fovea of the human eye can see 60° PPD but comprises less than 1% of the retinal size [22]. As the device comes in proximity of the eye, there is introduction of the screen-door effect and the screen takes most of the FOV, there is a need for increasing the pixel density. The next thing is the term called – 'foveated rendering' which needs to be implemented in the next generation XR devices as human eye can see high resolution only where the fovea is focused. Thus, there is no need to render the entire image with the high resolution but rendering in the region where the eye is fixated. This selection in the image quality is usually done with the help of a GPU that renders a small rectangle with high resolution and the region around it with a lower resolution. Lens correction in terms of fixing distortion (introducing Barrel warp to compensate pincushion distortion) and chromatic aberration (by image processing or using Pancharatnam Berry phase lenses as stated in [23 -34]). VR headsets must support multiple 360° video formats (cube map format) for encoding and decoding video to determine pose and show appropriate view of 360°. The next generation immersive

VR experience must also involve stereoscopic display for representing stereoscopic image (Fig. 23) as the binocular vision helps the human brain to determine the depth.

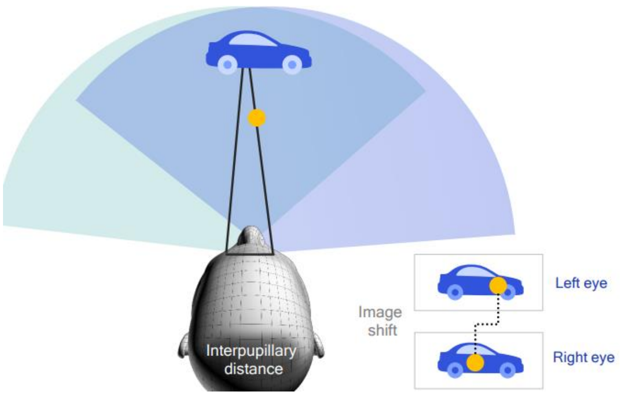
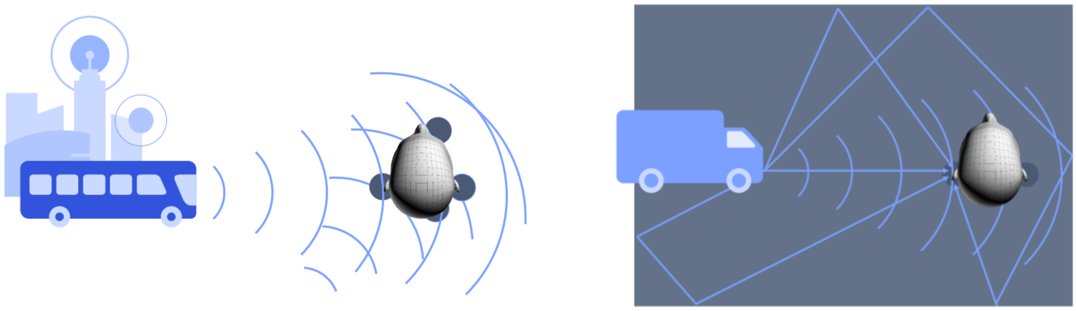


Fig 23: Stereoscopic 3D imaging [Image Courtesy: Qualcomm, adapted from 22].

1.5.72 Sound quality

Another important requirement for the fully immersive VR experience is the sound or audio quality which needs to meet the expectation of binaural human hearing capabilities (see Fig. 24a). For audio quality to be sufficient in XR applications, it has to be 'plausible' (realistic when compared to the user's expectations), and not necessarily 'authentic' (realistic when compared to an explicit real-world reference) [120]. A head-related transfer function (HRTF) contains the directional cues for a sound coming from a specific point in space to the point of the eardrums. Convolution of a sound with the HRTF and reproducing the result over headphones can give the impression of the sound originating outside the head and from the specified point in space, a process known as binaural synthesis. Reproducing sounds in more than one direction or distance requires multiple measurements of HRTFs. In VR applications, the user should be able to rotate their head freely and the virtual audio sources remain anchored in the real world position, which requires dynamic binaural synthesis. This involves fast switching of the HRTFs in convolution. A dense distribution is required for a perceptually seamless transition between measurements, with a necessary resolution as fine as 2° in the azimuth and 2° in the elevation plane reported [121], though this resolution is reported to be as fine as 1° in some cases [122]. Obtaining HRTFs between measurement positions requires interpolation. A common approach used in VR audio is to use Ambisonic technology [123], which involves spatial sampling and reconstruction of a soundfield using spherical harmonics. Ambisonic format audio can be rotated to compensate for head rotations.

As every individual's body is a different shape and size, individualised HRTFs produce the most natural and believable binaural experience. However, the measurement of high resolution HRTF datasets is laborious, expensive and time consuming. Generic HRTF sets, measured using dummy heads, are often used instead, though these can produce issues such as front / back and up / down confusions.



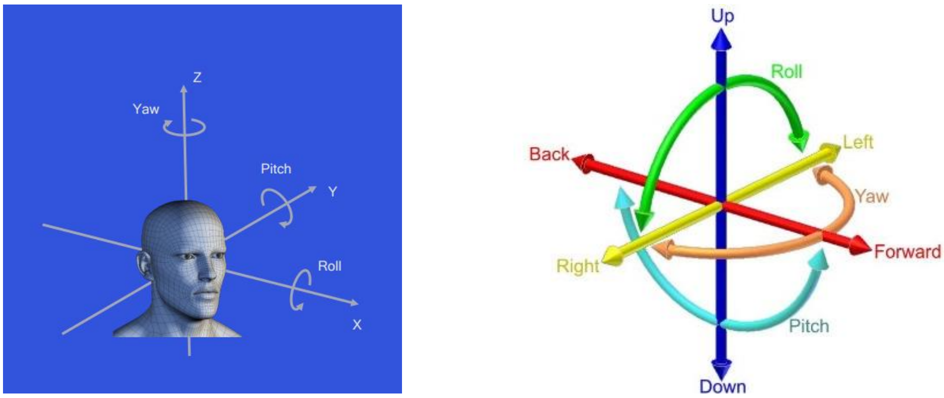
(a) (b)

Fig 24: 3D audio (a) 3D positional audio (b) Sound reverberation [Image Courtesy: Qualcomm, adapted from 22].

The above context explains the position of the sound, but the experience becomes more immersive when there is reverberation of sound (see Fig. 24b). Reverberation affects distance perception, externalization of sound sources, and geometry of the space. If accurate reverberation can be achieved, then the immersion quality will increase [119].

1.5.73 Intuitive interactions

The final block of the extreme immersive requirements involves so called 'Intuitive Interactions'. Intuitive interactions involve precise motion tracking of the head movements, natural user interfaces to seamlessly interact with VR using natural movements that are wireless and minimum latency to avoid perceptible lag. For precise motion tracking, there should be a control of movement with 3 degrees of freedom (3-DOF) and 6 degrees of Freedom (6-DOF) and this is illustrated by Fig. The 3 DOF tracking illustrates the directional view of the user from a fixed point and participate in rotational movements (yaw, pitch and roll) as shown in Fig. 25a and 6-DOF tracking, on the other hand illustrates both the rotational and the translation movements (see Fig. 25b). The 6-DOF intuitive motion tracking should be present in the next generation XR devices and has a clear advantage over the 3-DOF intuitive motion tracking in determining the accurate position, orientation and free translational movements [22].



(a) (b)

Fig. 25: (a) 3DOF tracking [Image Courtesy: Qualcomm, adapted from 22] and (b) 6DOF tracking [adapted from 22]

2.0 Reasonable Specifications

The reasonable specifications for this state-of the-art report includes AR/VR/MR device and design specifications and gather the benchmarking specifications for the next generation XR ecosystem. This starts with the AR/VR/MR optical device specifications used for the design of the AR/VR architecture and focusses on the optics, display, FOV, resolution, power, tracking, computing, ergonomics, environmental understanding and cost of the devices. This also includes the specification comparison of the current AR/VR/MR products based on the benchmarking specifications. An audio specification is also included, before information of the AR/VR/MR product offerings to the current markets and which applications they serve is presented.

2.1 AR/VR/MR Optical Device Specifications

The optical specifications for the AR/VR/MR devices include important parameters such as field of view (FOV), aspect ratio, eye-relief, eye-box, exit pupil, eye-relief and vertex distance, pupil swim, display immersion, stereo overlap, brightness, angular resolution, foveated rendering and optical foveation.

2.1.1 Field of view

Field of view (FOV) is defined as the angular range over which an image can be projected in the near or the far field. It is measured as an angular parameter and its unit is measured in degrees and resolution over FOV in pixels per degree (PPD) and is also discussed in [3]. In case of the next generation devices FOV is mentioned as the diagonal measure of the rectangular aspect ratio also mentioned in the next section. In the above, sections information about the horizontal and vertical FOV was mentioned and how for different AR/VR products the display diagonal FOV and see-through FOV varies and overlaps with the human vision. FOV can vary with the size of the eye-box and for achieving the larger FOV, the aspect ratio can change its shape from rectangular to square or may be elliptical.

2.1.2 Aspect ratio

Aspect ratio is defined ratio which defines the display shape and determines the FOV of the next generation AR/VR/MR device. The most common aspect ratio used is 16:9 for rectangular shape, though there might be variation in the aspect ratio.

2.1.3 Eye-box

Eye-box is defined as the volume that starts at the AR/VR/MR exit pupil and extends back towards the eye. If the position of the eye is placed around the eye-box then the viewer sees the entire FOV [35]. Eye-box can also be defined as a spatial range where the eye can be located or positioned which sees the entire FOV [36, 38]. Eye-box as defined by [3] is the 3D region between the combiner and the human eye pupil over which entire FOV fills the human pupil. The size of the eye box varies. The eye box can be measured with the help of the vignetting criteria (see Fig. 26).

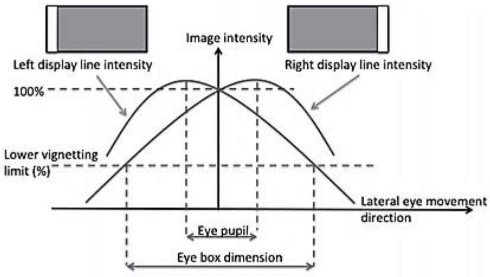


Fig. 26: Definition of eye box through vignetting criteria [adapted from3].

The inter-pupillary distance (IPD) also varies with the size of the eye box and is different for male and female. Therefore, it is also required that the next generation AR/VR/MR devices has sensors so that there is automatic IPD adjust and the combiner is fixed according to the eye box size of the user. This is also one of key things that needs to be implemented in the next generation devices.

2.1.4 Exit pupil, eye relief and vertex distance

Exit pupil is defined as the area of the region of the image or beam of light formed by the AR/VR/MR device optics[3, 36]. The typical range of the exit pupil is 15 to 20 mm wide eye relief is defined as the distance from the exit pupil to the nearest surface of the AR/VR/MR display or device. The eye relief is designed to be larger than 20 mm to allow the users who wear spectacles. The vertex distance is defined as the distance between the base surface of the lens at its verted to the tip of the cornea. Typically, vertex distance varies from 12 mm to 17 mm. 3D region can be filled in different manner for different AR/VR/MR configurations. This method is quite important to achieve the goal of the maximum perceived resolution by filling up the entire FOV [3].

2.1.5 Stereo overlap

It is a phenomenon which occurs due to vergence reflex of the human eye by presenting two different images to the same eye. Stereopsis can fuse this image and sometimes this leads to the VAC issue explained in the previous section of this report. VAC mitigation is also one of requirements to form a high-quality stereoscopic image.

2.1.6 Brightness

Brightness can be defined as the luminance and illuminance level for the AR/VR/MR display. The brightness comes from the efficiency of the light engine used which is the output nits/total input watt. A thorough study was made in [3] to calculate the efficiency of the light engine.

2.1.7 Angular resolution

Angular resolution of the AR/VR/MR is defined as the number of pixels within a unit angle. The maximum angular resolution of human eye is about 60 PPD for central vision and 30 PPD for peripheral vision. The current AR/MR/MR are trying to get a PPD close to 60 PPD for high angular resolution and maximum FOV but are not able to build those due to limitations in the optical system design as explained in [3, 36]. MTF measures the angular resolution of the display. Typically, 30% MTF is used to measure the angular resolution.

2.1.8 Foveated rendering and optical foveation

Foveated rendering phenomenon was explained in the previous sections in order to minimize power. But to obtain this foveation various techniques can be followed to provide a high-resolution experience and these techniques are static digital foveation without gaze tracking display (range is 40 deg over FOV cone), gaze-contingent dynamic digital foveation (dynamic range is 15 to 20 deg).

2.1.9 Pupil swim

The optical distortion variation which occurs due to change in distortion as a function of lateral position of the human eye pupil in the eye box. The pupil swim can also occur in prescription glasses. This is a major problem if we go towards higher FOV [3]. This complete the set of reasonable optical specifications and in the next section we will go through in general what are the benchmarking specifications of the AR/VR/MR devices.

2.2 AR/VR/MR Auditory Device Specifications

The audio specifications for future AR/VR/MR devices include hardware choices for headphone design and requirements for digital signal processing systems.

2.2.1 Hardware

Device transducers should aim for a playback frequency range that equates roughly to the range of human hearing: from 20Hz to 20kHz. For ANC applications, headphones should feature microphones on both the outside and inside of the ear canal. Inside microphones are used to equalize pressure and tweak the frequency response of the transducers. Outside microphones are used to measure external sound pressures, as well as for auditory AR features as described in section 1.2.11 such as language translation and selective noise suppression. Multiple directional microphones can use beamforming to spatially isolate sound from different directions. Latency should be kept below 20 ms.

2.2.2 Software and compression

Rendering choices are open, but Ambisonics is currently the audio format of choice. First-order Ambisonics, which uses 4 channels of audio, is accurate to around 700Hz. It is currently implemented in Google Resonance (Youtube) and Unity. Facebook and wWise use higher orders of Ambisonics (second- and third-order, respectively), which offer higher fidelity sound, more accurate timbre and more precise localization, but come at the cost of higher channel count (9 and 16, respectively) and thus more storage and bandwidth requirements. MPEG-H and OPUS codecs offer spatial audio compression options.

3.0 Benchmarking Specifications of the AR/VR/MR and Next Generation Devices used in this State-of-the-art Report

The benchmarking specifications or parameters of the AR/VR/MR devices can be well understood by going through specifications of display, light engine, mechanical, environmental maximum ratings or safety, tracking, computing and processing unit, audio and speech, connectivity, ergonomics. Thus, in the next sections we will go through the qualitative description of each of the specifications and give a general overview of them in relation to the specifications compared in this state-of-the-art report.

3.1 Display Specification

The display specifications for the AR devices and the displays includes important parameters such as display type, efficiency, MTF, transmission and contrast.

3.1.1 Display type

Display type for the AR/VR/MR includes a variety of display type from LCD, OLED, AMOLED in the list of the benchmarking specifications used for the comparison. For light engine specifications the microdisplays used are Liquid Crystal on Silicon (LCoS) or Digital Light Processing (DLP) type depending upon the system design of the projector or the light engine. The LCoS displays are used with the light engine having polarization conversion explained in [High efficiency light engine for AR projectors] and source as red-green-blue (RGB) LEDs. On the other hand, DLPs are used for the projector design with RGB lasers. Display type used in the AR/VR/MR devices is also explained well in [36, 37].

3.1.2 Efficiency

The efficiency of the display comes in the nits/Lum and is not clearly mentioned as some of the AR/VR/MR devices have brightness specification. The efficiency was clearly mentioned for the OEM AR manufacturing companies mentioned in this report.

3.1.3 Resolution

Resolution qualitatively gives a measure of the fine details of the image in most of the AR/VR/MR devices are specified in pixels in [horizontal (H) x vertical (V)] format and the benchmarking specifications comparison states the resolution in most of the AR/VR/MR devices in the pixels format. The greater the number of pixels, the more is the pixel density, and this is required in order to obtain a high-resolution image. Most of the benchmarking specification while comparing gave this figure in pixels format. Sometimes, it also exists that the total number of pixels are enumerated.

3.1.4 Modulation Transfer Function

Modulation Transfer Function (MTF) is a direct measure of the image quality and is a very important factor to decide the quality of the display. It also gives integrated polychromatic performance of the system according to [39]. It is very crucial to determine the MTF in order to understand the performance of the system and can be measured in a variety of ways according to [40, 41]. It also describes how the image contrast of the system varies with the spatial frequency. The MTF data is not clearly mentioned during the comparison of the benchmarking specifications except in the cases where AR OEM manufacturers mentioned them in their product data sheet.

3.1.5 Transmission (For AR/MR experience)

This gives the measure of the see-through specially for AR devices because here transmission and transparency is very important. To overlay the virtual information on the real-world object principle of occlusion is used. Thus, transparency is required to achieve hard edge occlusion. For VR experience transparency is not a big issue as the see-through feature is not readily required.

3.1.6 Contrast

In most of the current AR/VR/MR devices mentioned in the report uses ANSI contrast and the checker box contrast method to determine the contrast. This is also which is used to measure the MTF of the system by using edge identification method explained in.

3.1.7 Light engine specifications

The specifications for the AR/VR/MR devices and the displays includes important parameters such as projector type, FOV, image orientation, display resolution, projection type, flux at the exit pupil, dimensions of the light engine, weight of the light engine.

3.1.8 Mechanical specifications

The mechanical specifications for the AR/VR/MR devices include important parameters such as dimensions, thickness, weight and the shape outline of the device. This decides the ruggedness of the device and strength against environmental factors.

3.1.9 Computing and processing unit specifications

The computing and processing units for the AR/VR/MR devices includes important parameters such as processor used, random access memory (RAM), read only memory (ROM) and external memory, weight of the computing unit. In order to meet the current requirements of the AR/VR/MR devices for running training session like flight simulator, interacting with the immersive world and many other tasks. It also consists of the operating system (OS) and have a variety of them listed in the benchmarking specification comparison chart in this state-of-the-art report.

3.1.10 Processor

This part is crucial for any AR/VR/MR device as this serves as the brain for processing the computational load and works in support of the memory and other units related to the processing.

3.1.11 RAM and ROM

RAM and the ROM acts are the memory elements in order to support the processing of the data or the rendering of the image or video while the system is operating.

3.1.12 Weight of the computing unit

This is also taken as quite important if you consider that the AR/VR/MR devices needs the processor and memory element and is also a good fit for the ergonomics of the full package itself. The weight of the computing unit is measured in grams and they take most of the device weight and in future might need the requirement of integration to the next generation AR/VR/MR device itself.

3.1.13 Tracking specifications

Tracking basically means motion tracking and used for sensing the environment and is used mostly with the AR/VR/MR devices and they are basically grouped into two types, one is the optical method of tracking and other is the non-optical method of tracking. The tracking specifications are mentioned in [42].

3.1.131 Optical methods

The optical method of tracking involves using imaging sensors [42].

3.1.132 Non-optical methods

The second method is the no optical methods that do not involve the use of optical imaging sensors requirements for AR/VR/MR devices involves the MEMS accelerometers, gyroscopes and magnetometers installed in them according to [42]. The basic sensor requirements are met with these and they help these devices to detect the position, orientation, velocity, rotation of the device.

3.1.14 Audio and speech specifications

Most of the devices mentioned in the benchmarking specification comparison of the specifications of AR/VR/MR devices have audio integrated or external microphone to process the audio information. Both audio input and out can be integrated with the next generation devices. But in some of the devices spatial audio is integrated for 3D sound effects. Standalone audio AR devices can also be of great benefit, while integrating into people’s existing habits of widespread headphone usage.

3.1.15 Power and battery back-up specifications

Power requirements are the huge concern for the current AR/VR/MR devices as there is a loss of optical power through displays in showing a high-resolution image without foveated rendering and battery backups are also an essential part to determine the up time for the device. The power consumption can be reduced by incorporating foveated rendering to ensure that the image is highly focused and fixated around the fovea and occupies the full FOV of the human vision.

3.1.16 Connectivity specifications

The connectivity specifications for the devices in the benchmarking specification comparison chart of the current AR/VR/MR devices include USB Type, WiFi, tethering with the smartphone, antenna specific information. Though the future connectivity requirements should have access to 5G as mentioned in the extreme requirements as the future requirements want the current devices to work without the help of cables. This will help in the processing of the data very fast.

3.1.17 Latency specifications

Latency also explained in the previous sections of this report and accepted as key parameter to reduce the sicknesses which occurs due to interaction with the virtual or simulated world is also one of the specifications which is mentioned in the comparison for AR/VR/MR devices. The current devices have shown latency less than 20 ms. For audio rendering, latency below 20 ms is vital for both sound rendering and scene rotation.

3.1.18 Ergonomic specifications

The ergonomics specifications of the current AR/VR/MR devices include the weight of the device, the balance of the device (centre of gravity), how comfortable the device is to wear and whether the device has any thermal issues related to it. In the comparison chart of the AR/VR/MR devices the device ergonomic criteria are based on these factors.

3.1.19 Environmental understanding specifications

Environmental understanding of the device is very important criteria to understand (specially for the AR devices) and is used to get the physical understanding of the space and can be also used to place feature and interact with the environment. In the comparison, simultaneous localization and mapping is the most popular technique which was found in the current AR/VR/MR devices.

In the next section the specification comparison is presented for the AR devices.

4.0 Benchmarking Specifications and Comparison of Specification of Current and Developer Editions of AR Devices or Dev Kits

(*Note*: Reference to this section are shared in the* References)

Device Name	Cost	Display Type and optics	Resolution	FOV	Audio and Speech	Power	Connectivity	Computing and processor	Tracking	sensors	Ergonomics	Environmental understanding/Rugged characteristics
Future Pacific	**$1099**	optics- Coaxial surfaces, Display not mentioned, Brightness info- 3000lm, camera- Dual global shutter fisheye cameras +RGB Camera	1920 x 1080	52°	Not clearly specified, capability for voice recognition	NA	GPS, WIFI, USB 3.1, Bluetooth	OS- Android 7.1, Qualcomm snapdragon 835, Qualcomm kyro 280 CPU with 8 cores	AMreal Light Field tracking, Ai3D Human posture tracking	Gyroscope, gravity sensor and magnetometer	95/88g, available in 3 different colors, adjustable	Object recognition, Gesture recognition, Voice recognition, Image recognition

kura Gallium
$1199
8k75, 4K144, Adjustable resolution per eye with VESA display stream compression, HDR, True Black, Novel structure waveguide geometric eyepiece, customized microLED projection, Display brightness 4000 nits, prescription compatible Transparency: 95%
8K per eye, unlimited depth of FOV (10cm to infinity)
150°
NA
6 hours
Display port2.0 interface (backward compatible, USB 3.1 gen2
Qualcomm snapdragon 855/8gb
6-DOF head tracking
2QVGA eye tracking cameras
weight - 80g
Not Clearly mentioned

Device Name	Zebra	Retissa-Visirium
Cost	NA	$ 931
Display Type and optics	Display- Full color OLED	contrast- 100000:1, Camera- Photo: 8 MP, Video: 1080P FHD, privacy: LED indicator, Auto focus, 35° angle of depression
Resolution	640 X 400	640 x 400
FOV	20.3°	20° (diagonal)
Audio and Speech	Integrated Microphone	Digital microphone with noise cancellation, output: Mono speaker
Power	NA	NA
Connectivity	USB 2.0, Host device connectivity Android 5.0+ and windows 10	Plug and play, USB type-C 2.0
Computing and processor	NA	Intel® Movidius™ Myriad 2
Tracking	head tracker	Custom button, Gesture control, Voice control
sensors	9-axis integrated sensor	6 axes
Ergonomics	Weight <30 g (without cables)	GLXSS weights 33.4 g, comfortable wearing
Environmental understanding/ Rugged characteristics	Sealing IP67, Drop specification- 5Ft/1.5m, -4 to 122 °F, -20 TO 50 °C	NA

Device Name	Iristick	Third Eye X2
Cost	**$2499**	**$1950**
Display Type and optics	Display- WQVGA, Imager- 5MP 4:3 aspect ratio, video- FULL HD 30 FPS, Camera- zoom 5X optical zoom with liquid lens and autofocus	Thin waveguide display, wide peripheral vision, Smallest form factor, Camera-13 MP Light Transmission- -90% of outside light
Resolution	428 x 240	1280 x 720
FOV	13°	42°
Audio and Speech	Integrated speaker near right ear, microphone- Quad microphone array for improved voice commands	Audio is optional
Power	12WH, changing time -2 hours	1750mAh
Connectivity	NA	Integrated CPU/GPU, 4GB RAM 64 GB storage, additional memory option via USB-C, integrated WIFI Android version- 8.1 and Bluetooth
Computing and processor	NA	
Tracking	NA	NA
sensors	9 axis sensors (accelerometer, gyroscope, compass)	
Ergonomics	Three versions 71/71/62g, height- 4.5 cm, width- 17cm/15.2 cm,	170 g
Environmental understanding/Rugged characteristics	length – 20cm, Safety standards- USA (ANSI-Z87) EU (EN-166S	Vision Eye SLAM SDK integration, Audio and Gaze UI,

Device Name	Glass up F4	NEDplusAR
Cost	**$2367**	**NA**
Display Type and optics	LCD display driving method- Color filter Active matrix LCD, LCD display size- 0,3, aspect ratio- 4:3, -full color, IPD-55mm – 71mm range, lenses- light brown/Gray/ Transparent, Video Camera- Full color, 5 MP, 15 FPS	Virtual screen size- 165 de@3m, exit pupil diameter- 8mm, Exit pupil distance- 64mm, Transmittance- 50%-70% Thickness- 2mm, Material- optical polymer 1Display panel size- 5.5-inch, lens-optical polymer
Resolution	640 X 480 (VGA)	1080p
FOV	22°	70°
Audio and Speech	NA	NA
Power	Battery- 4000mah, 6 h battery backup	NA
Connectivity	WIFI, Bluetooth	NA
Computing and processor	OS on board- Linux based, SDK- C#, python, Java, (Android and Desktop), C++, JavaScript, objective, Perl, PHP, QT5, Ruby, Swift	NA
Tracking	NA	NA
sensors	9 axis accelerometers gyro, compass	NA
Ergonomics	NA	weight- 22g
Environmental understanding/Rugged characteristics	IP 31, Operating temperature 5-35 degree, OSD- English, Apps- Default application language, Manual- Italian, English, Spanish, Chinese.	NA

WESTUNITIS-Infolinker	**Device Name**
NA	Cost
Display- WQVGA Aspect Ratio- 16:9	Display Type and optics
428 X 240	Resolution
14°	FOV
NA	Audio and Speech
NA	Power
NA	Connectivity
OS- Android 4.2.2, CPU- Ti OMAP 4470 dual-core MAX1.5GHz	Computing and processor
NA	Tracking
Touch sensor- 1, 3-axis accelerometer 3 axis gyro, 3 axis magnetic sensor	sensors
weight- 56g	Ergonomics
NA	Environmental understanding/Rugged characteristics

WESTUNITIS-Picolinker
NA
Display- WQVGA Aspect ratio- 16:9
640 x 360
16.7°
NA
NA
NA
NA
NA
NA
30g
HDMI input (1280 x 720p at 60 fps, HDCP not supported)/HDMI type A male connector, Waterproof performance- IPX4

Device Name	LYNX (MR)	HMDmd
Cost	**$1500**	**NA**
Display Type and optics	Display- Immersion and comfort Optics- 4-fold catadioptric freeform prism, Dual LCD, Refresh rate- 90 Hz	Display- Realtime 2D/3D video, Data overlay- Overlay of case-specific data, Image Source- OLED, Optics- Aspheric, injection molded, Brightness- 35-foot lamberts, Update Rate- 30 H
Resolution	1600 X 1600	video resolution- 2K HD (1920 X 1080)
FOV	FOV/ PPD- 90 DEG/18 PPD	FOV- 35 deg x 20 deg
Audio and Speech	Audio- Communication and immersion speakers (stereo speakers), microphone arrays (2 channels)	NA
Power	Power- Low power consumption for long battery life	NA
Connectivity	Fluid and precise tracking with 6 or USB-C. The USB-C port can also be used to add a 5G modem or an additional sensor.	Interface to video Source- Wireless, Voice Command system- Kopin Wisper Chip, Support Electronics- Waist -Mounted
Computing and processor	Processing system- Powerful and Energy efficient (Based on the Qualcomm Snapdragon XR2 platform, the lynx is at least twice as powerful as the previous generation headsets), SOC-Qualcomm Snapdragon XR2, Memory 6GB LPDDR5,	NA
Tracking	6 DoF Positional tracking- B and W cameras, hand tracking (two-handed, gesture recognition), eye tracking- low latency tracking. , Eye tracking- IR cameras	NA
sensors	NA	NA
Ergonomics	comfortable	NA
Environmental understanding/Rugged characteristics	SLAM with word anchors	Weight- 150 g

Device Name	TRIVISIO- ARS.31	Rhino X-Ximmerse
Cost	**NA**	**$199**
Display Type and optics	Micro displays- 2 x SXGA LCD , color- 24-bit RGB color input, Aspect ratio- 5:4(12mm x 9.6mm active area display), Luminance (RGBW)- 180 cd/m2,Frame rate- 60 Hz, pixel response time- <8 ms, see-through optics- 70% image reflection, 30% image transmittance, Focus plane- 1m (can be adapted on order request, eg.0.5m, 10m, 150m), Eye relief- 30mm (Regular glasses can be worn), Eye Motion box – 8mm (h) x 6 mm(v), Camera(optional)- 5MP module with autofocus. Operating temperature- 0 to 60 deg (Operating temp display) Weight- TBD, Dimensions- 115mm x 90mm x 56mm (without headband and hinges)	CPU- snapdragon 835. Display- 5.5, LCD VFOV/HFOV- 57/50 controller
Resolution	1280 X 1024	1440 x 2560
FOV	30° (diagonal)	NA
Audio and Speech	Audio(optional)- Built-in microphone, mono speaker and headset	NA
Power	NA	Power- 368mAh
Connectivity	NA	NA
Computing and processor	NA	9.OS- Android 7.1 +
Tracking	NA	Tracking - 6DOF headset and
sensors	NA	NA
Ergonomics	NA	NA
Environmental understanding/Rugged characteristics	NA	Weight- 590 g Memory- 6GB/64GB

	Vuzix	Micosoft Hololens 2
Cost	M400 - $1799 Blade - $799	$3500 -H2
Display Type and optics	WVGA color display, see through wave guide optics, display type is DLP, Aspect Ratio: 16:9, Brightness:1000 nits, 24-bit color	Display Optics See-through holographic lenses (waveguides)3:2 light engines, Holographic density >2.5k radiant (light points per radian) Eye-based rendering, Display optimization for 3D eye position
Resolution	854 x 480	2k 3:2 light engines
FOV	28°	horizontal FoV of 43° and a vertical of 29°
Audio and Speech	Integrated speaker, Voice control- customizable and supports customizable languages, triple noise cancelling microphones BT Audio: HSP/A2DP	Audio and speech Microphone array, 5 channels Speakers, Built-in spatial sound
Power	135mAh internal battery, 1000 mAh head worn USB-C external battery with 3 level indicators	NA
Connectivity	USB 3.1 Gen 2 on USB type C Wifi 2.4/5GHz 802.11a/b/g/n/ac Bluetooth 5.0 BR/EDR/LE	Wifi-5(802.11ac 2x2), Bluetooth- 5, USB-C
Computing and processor	8 Core 2.52 Ghz Qualcomm XR1, Gps Glonas 6GB LPDDR4 RAM 64GB internal memory	Qaulcomm snapdragon 850 compute platform second-generation custom-built holographic processing unit
Tracking	3 DOF tracking	Head tracking ,4 visible light cameras, Eye tracking, 2 IR cameras
sensors	3 axis gyro, 3 axis accelerometer, 3 axis magnetometers	1-MP time-of-flight (ToF) depth sensor IMU Accelerometer, gyroscope, magnetometer Camera
Ergonomics	3.3 oz in weight, eye glass frames, safety glasses, headband mount	Single size fit over glasses Weight- 566 g
Environmental understanding/Rugged characteristics	Certifications- IP67 Drop safe to 2 meters	Human Understanding, environment understanding

Device Name	**Ride On**
Cost	**NA**
Display Type and optics	Display- viewable area, brightness- 3000 nits, Lens- Dual anti-fog, anti-scratch UV400, CE EN174 certification, Camera- HD video recording, 8MP with video and aerodynamic shell, CC certified
Resolution	1080p
FOV	24°
Audio and Speech	NA
Power	2500mAh 7 usage hours, 24 standby hours
Connectivity	microUSB for charging and data transfer.
Computing and processor	NA
Tracking	NA
sensors	NA
Ergonomics	Ergonomics and fit- 3 layers of foam for an ergonomic fit, High density face foam 70 cm length adjustable strap. Weight- 240 g, light weight
Environmental understanding/Rugged characteristics	safety- CE and F

Device Name	Epson Moverio BT-300	Recon
Cost	**$699**	**NA**
Display Type and optics	Si-OLED (Silicone - organic LED) Driving method Mono Crystalline Silicon Active Matrix Display size- 0.43-inch-wide panel (16: 9), Pixel Number-921, 600 pixels RGB Screen Size (Projected Distance) 80 inches at 5 m - 320 inches at 20 m Color Reproduction-24-bit color (16.77 million colors), Update frequency 30 Hz	Display: Color, Recon Jet Pro - Smart glasses - 8 GB - Wi-Fi, Bluetooth, ANT+/ANT - 3 oz - black
Resolution	1280 X 780	NA
FOV	23 °	NA
Audio and Speech	NA	NA
Power	NA	NA
Connectivity	Wireless Network IEEE 802.11a / b / g / n / ac with WiFi Direct and WiFi Miracast (Source / Sink) Bluetooth Bluetooth Smart-ready USB 2.0	IEEE 802.11b/g/n, Bluetooth 4.0, ANT+/ANT Input Device-
Computing and processor	Processor Intel® Atom ™ x5, 1.44GHz Quad Core RAM memory	NA
Tracking		NA
sensors	Camera 5 million pixels GPS Yes, in the controller (A-GPS) Compass	NA
Ergonomics	Comfortable	NA
Environmental understanding/Rugged characteristics	Operating Temperature 5 ° C - 35 ° C, 41 ° F - 95 ° F, 20% - 80% Humidity Battery Life 6 hrs (Video mode with Android at 25 °) Dimensions: Headset	Storage: 8 GB

Garmin
NA
Dimensions: 2.4" x 0.77 x 0.78" (60.0 x 19.7 x 19.8 mm)
428 x 240
NA
NA
8 hours
ANT+® connectivity, Touch Panel
NA
Ambient light and accelerometer
1.1 oz (29.7 g)
Water rating: IPX7 Mounting: Ambidextrous

Device Name	Nreal	Magic Leap 1
Cost	**$586 +$295** (tethered with smartphone)	**$2295**
Display Type and optics	Combined lightguide	Display Refresh Rate-120 Hz
Resolution	1080p resolution per eye	1.3 M pixels per eye
FOV	52°	50°
Audio and Speech	NA	spatial Audio
Power	About 3 hours with computing unit	3.5 hours battery backup
Connectivity	USB-C	NA
Computing and processor	CPU- Qualcomm Snapdragon 845 Operating System Android OS	LUMIN OS 6 core CPU design, NVDIA 256 cuda graphics, 64-bit architecture, 1.7GHz processing speed
Tracking	Headset: 6DoFController: 3DoF	6-DOF tracking with controllers, 6 hours of battery life, Head and eye tracking
sensors	NA	
Ergonomics	Weight Headset: 0.2 pounds (88g) Computing unit: 0.4 pounds (170g) Detachable controller: 0.1 pounds / 23g	316g, prescription inserts can be applied
Environmental understanding/Rugged characteristics	NA	Simultaneous Localization and Mapping

Device Name	Toshiba
Cost	$1900
Display Type and optics	OPTICS- 0.26" Diagonal Pixel Pure™ Display, CAMERA-5 Megapixel, F2.8, Focus Range: 10cm to infinity, Auto Focus, Video Capture: -Up to 1080p @ 30fps, Flash LED
Resolution	Display Resolution: 640 x 360, Content Resolution: 1280 x 720
FOV	NA
Audio and Speech	Dual Microphones, AUDIO, Built-in Speaker
Power	
Connectivity	CONNECTORS, USB Type-C™
Computing and processor	Navigation Controls 3 Control Buttons, Touchpad
Tracking	NA
sensors	NA
Ergonomics	Weight- 310g / .68lbsVersatile Mounting options available, AR100 Lens-less Frame, AR100 Safety Frame, AR100 Safety Helmet Mount, AR100 Headband Mount
Environmental understanding/Rugged characteristics	Security- Trusted Platform module-(dTPM) v2.0 Fingerprint Reader (Optional) Slot for Security Lock

Olympus-EyeTrek INSIGHT EI-10	Every sight Raptor	Device Name
$1500	$649	Cost
OLED- nHD+ , Camera Resolution- 1992 × 1216 effective pixels	BEAMTM Display Projector, WVGA+ OLED Display, Ultra High Brightness and environments, Display equivalent to a 65, screen size (virtual viewing distance 12 feet), Adjustable brightness level, digitally adjustable screen positioning (using Every sight' s Smartphone App) Camera- 13.2	Display Type and optics
640 X 400	Video Modes: 1080p 30Hz / 720p 60Hz / 480p 30Hz, Still Image Resolution: 13.2 MP / 3.2 MP Quick Shutter	Resolution
Diagonal 13° (approx)	**75°**	FOV
Microphone, 16bit monaural	Internal Speaker, Low noise microphones for video recording and voice control	Audio and Speech
Li-ion polymer, Battery Life:30-60 min (approx., depending on the applications used), Battery Voltage / Power-DC 3.7V / 1.1Wh, Battery Capacity- 300mAh	Up to 8-hour battery life, Rechargeable via USB Micro-B Low power voice DSP, Low power sensors fusion processor running proprietary line-of-sight algorithms	Power
Voice Input / Output,4-pole mini jack,Bluetooth 4.1 (A2DP / AVRCP / SPP / HID) Supported frequency band: 2402MHz-	Ant+™ Bluetooth 2.0, BLE 4.1 (Bluetooth Smart) Wi-Fi - 802.11b/g/n 2.4GHz Micro USB 2.0 for battery charging and data transfer, Controls Multi-Touch Pad for use with or without gloves Voice control, Bluetooth remote control, Power/ Display Sleep button Privacy LED to indicate that camera is recording	Connectivity
CPU-TI OMAP 4470, OS-Android 4.2.2 RAM (main memory)-1GB, ROM (user memory)-8GB	Qualcomm® Snapdragon™ 410E, 1.2GHz Quad-Core ARM Cortex-A53, Qualcomm® Adreno™ 306 GPU 2GB SDRAM, 16GB or 32GB Internal Storage, Storage specifications refer to capacity before	Computing and processor
NA	NA	Tracking
3-axis accelerometer, 3-axis gyro, 3-axis magnetic field	NA	sensors
Weight- 66 g (approx. including battery)	Frame & Visor, Grilamid® TR-90 frame for a durable yet lightweight experience, Impact-resistant interchangeable visor, Adjustable nose piece for better fit. Weight- approx. 3.45oz / 98g	Ergonomics
Operation environment (incl. accessories) Temperature: 5°C (41°F) - 35°C (95°F), Humidity: 30 - 90% (no dew condensation) Storage environment(incl. accessories)	Environmental, Dust & Water Resistance, Operating ambient temperature: 32° to 104° F (0° to 40° C), Nonoperating temperature: 14° to 140° 10° to 60° C), Relative humidity: 5% to 95% noncondensing, Light Tint visor for low light conditions, Everysight Smartphone Companion App o Compatible with iOS version 9 and above or Android version 4.4 and above	Environmental understanding/Rugged characteristics

Device Name	Konica - Minolta AIRe lens	HiScene	North (Acquired by Google)
Cost	NA	NA	Without Prescription: $599/ $799 CAD With Prescription: $799 / $999 CAD
Display Type and optics	NA	2*LCOS, Aspect ratio- 16:9, Backlight brightness: greater than 10000cd/m^2 camera 2 million pixels	Full Color Display
Resolution	NA	1280 x 720	110 X 110
FOV	NA	72°	14° FOV diagonal
Audio and Speech	NA	Audio Double microphone noise reduction, loudspeaker	Integrated microphone, speaker
Power	NA	NA	Battery- Focals: 700 mAh Loop: 0.8mAh 18 hours backup
Connectivity	NA	Mobile communication connection Support 4G full Netcom Data communication connection Wi-Fi 802.11ac, BT4.1 Positioning Beidou, GPS, SIM card Single card (Nano) T-Flash card Maximum support 128G expansion, I/O interface Type C full interface, support USB3.0 OTG,	Focals smart glasses: Bluetooth 4.2Loop ring: Bluetooth LE Android 5.0 and newer iPhone 5S and newer with iOS 11+ and newer
Computing and processor	NA	NA	Qualcomm APQ8009w with Arm Cortex A7 (32-bit) at 1.09GHz and Qualcomm Adreno 304 GPU
Tracking	NA	NA	9-axis IMU
sensors	NA	NA	ambient light sensor, proximity sensor, Indicator LED , Ambient light sensor
Ergonomics	NA	NA	Weight 65- 80 g
Environmental understanding/Rugged characteristics	NA	NA	Ambient Temperature range -10ºC (14ºF) to 40ºC (104ºF), Optimal working- 5ºC (41ºF) and 40ºC (104ºF) , Display less clear and hue may appear -10ºC (50ºF) and above 40ºC (104ºF)

Device Name	Kopin	Innovega	Guangli	Dream world	Vivo AR glass
Cost	**$899**	NA	NA	NA	NA
Display Type and optics	LCD	NA	NA	NA	NA
Resolution	400 x 240	NA	NA	NA	NA
FOV	10°	NA	NA	NA	NA
Audio and Speech	NA	NA	NA	NA	NA
Power	NA	NA	NA	NA	NA
Connectivity	NA	NA	NA	NA	NA
Computing and processor	NA	NA	NA	NA	NA
Tracking	NA	NA	NA	NA	NA
sensors	NA	NA	NA	NA	NA
Ergonomics	Weight (g) 65 g, Dimensions SoC Qualcomm Quad Core,	NA	NA	NA	NA
Environmental understanding/Rugged characteristics	NA	NA	NA	NA	NA

Polaris	**Augmedics**	**Lemnis**	**Form**	**Tesseract**	**Avegant**	**OPPO**	**Pareal**	**Device Name**
NA	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	**NA**	Cost
NA	NA	NA	NA	NA	NA	NA	NA	Display Type and optics
NA	NA	NA	NA	NA	NA	NA	NA	Resolution
NA	NA	NA	NA	NA	NA	NA	NA	FOV
NA	NA	NA	NA	NA	NA	NA	NA	Audio and Speech
NA	NA	NA	NA	NA	NA	NA	NA	Power
NA	NA	NA	NA	NA	NA	NA	NA	Connectivity
NA	NA	NA	NA	NA	NA	NA	NA	Computing and processor
NA	NA	NA	NA	NA	NA	NA	NA	Tracking
NA	NA	NA	NA	NA	NA	NA	NA	sensors
NA	NA	NA	NA	NA	NA	NA	NA	Ergonomics
NA	NA	NA	NA	NA	NA	NA	NA	Environmental understanding/Rugged characteristics

Device Name	Combine Reality	Canon	All-sight	LLvision	Nvidia	Ajnalens	LetinAR	Sony	Solos
Cost	NA	NA	NA	NA	NA	NA	NA	NA	$499
Display Type and optics	NA	NA	NA	NA	NA	NA	NA	NA	NA
Resolution	NA	NA	NA	NA	NA	NA	NA	NA	NA
FOV	NA	NA	NA	NA	NA	NA	NA	NA	NA
Audio and Speech	NA	NA	NA	NA	NA	NA	NA	NA	NA
Power	NA	NA	NA	NA	NA	NA	NA	NA	NA
Connectivity	NA	NA	NA	NA	NA	NA	NA	NA	NA
Computing and processor	NA	NA	NA	NA	NA	NA	NA	NA	NA
Tracking	NA	NA	NA	NA	NA	NA	NA	NA	NA
sensors	NA	NA	NA	NA	NA	NA	NA	NA	NA
Ergonomics	NA	NA	NA	NA	NA	NA	NA	NA	NA
Environmental understanding/Rugged characteristics	NA	NA	NA	NA	NA	NA	NA	NA	NA

Device Name	Snap Spectacle 3	Vue Glasses	Level	Echo Frames (Audio)
Cost	$380	$299	$328	$180
Display Type and optics	NA	NA	NA	NA
Resolution	NA	NA	NA	NA
FOV	NA	NA	NA	NA
Audio and Speech	NA	NA	NA	NA
Power	NA	NA	NA	NA
Connectivity	NA	NA	NA	NA
Computing and processor	NA	NA	NA	NA
Tracking	NA	NA	NA	NA
sensors	NA	NA	NA	
Ergonomics	NA	NA	NA	NA
Environmental understanding/Rugged characteristics	NA	NA	NA	NA

Google glass Enterprise Edition2
NA
Camera 8Mp, Optical Display Module
640x360
80°
Audio out Mono Speaker, USB audio, BT audio Microphones 3 beam-forming microphones
Battery 820mAh with fast charge
Wi-Fi 802.11ac, dual-band, single antenna Bluetooth, Bluetooth 5.x AoA,
SoC Qualcomm Quad Core, 1.7GHz, 10nm OS Android Oreo, Memory & Storage ,3GB LPDDR4 / 32GB eMMC Flash
IMU Single 6-axis Accel/Gyro, single 3-axis Mag, Power-saving features On head detection sensor, and Eye-on screen sensor
Touch Multi-touch gesture touchpad, Charging & Data, USB Type-C, USB 2.0 480Mbps, LED, Battery 820mAh with fast charge Privacy (camera), power (rear)
Weight~46g (pod)
Ruggedization Water and dust resistant

Device Name	ODG	Varjo XR-1 (Developer Edition)
Cost	**$1000/$1800**	**$11911 (€9995)**
Display Type and optics	Dual 16:9 stereoscopic see-thru displays at up to 80fps	**Camera & Optics** - 2 x 12 megapixel at 90 Hz. Fixed focus, 1/3" sensor size with 1.55µ pixel size. 94° circular viewing angle. **Imaging pipeline** -4 streams (2 per eye): 1008 x 1008 downscaled from the full 12 Mpx + foveated 834 x 520 full resolution crop. **VR display system** -Bionic Display™ with resolution at over 20/20 vision (over 60 PPD / 3,000 PPI), flicker free screen refresh at 60/90 Hz. Infinite contrast ratio (over 10,000:1)
Resolution	720p	1920 x 1080 low persistence micro-OLEDs and two 1440 x 1600 low persistence AMOLEDs.
FOV	NA	87°
Audio and Speech	NA	NA
Power	NA	NA
Connectivity	Bluetooth 4.1, 802.11ac, GNSS (GPS/GLONASS)	10 m active optical Thunderbolt cable, Video output- 2 x DisplayPort 1.2 or 2 x Mini DisplayPort, USB port - 1 x USB-A 3.0 port or newer, not using any Hub, Thunderbolt port- 1 x Thunderbolt 3 port
Computing and processor	Qualcomm Snapdragon 805 2.7GHz quad-core Processor 3GB Pop LP-DDR3 RAM 64 GB Storage	Processor-Intel Core i7-7820X, GPU - NVIDIA GeForce® RTX 2080 Ti NVIDIA Quadro® RTX 6000 Memory-32 GBStorage space-2 GB
Tracking	NA	Industrial-grade, sub-degree accuracy integrated 100 Hz eye tracking delivering unmatched accuracyandprecision across real and virtual environments, even when wearing glasses or contact lenses. Compatible with Steam VR™ and ART™
sensors	NA	An active IR sensor system consisting of 2 wide-angle cameras and LEDs. The depth system is synchronized with video pass-through to provide low latency and robust depth map.
Ergonomics	NA	XR-1 can also be worn by people who wear glasses.It features an active airflow system, adjustable and detachable headband and head strap, comfortable usage with two different sized, easily replaceable face cushions and automatic interpupillary distance (IPD) adjustment for easy and accurate
Environmental understanding/Rugged characteristics	NA	NA

5.0 Specifications from Optical Equipment Manufacturer (OEM) AR Companies

(*Note*: Reference to this section are shared in the* References)

OEM Manufacturers	Design Characteristics	Light Engine specifications	Mechanical Specifications	Environmental Resistance
Dispelix Oy	Efficiency (nits/Lum) – 200 MTF (lp/degree)- 15 (>45 % contrast) Transmission (At normal incidence) - 82 % Contrast- 60:1	Projector Type - DLP FOV (Degrees) - 30 (Currently 50) Image orientation- 16:9 Display Resolution - 854 x 480 Projection Type – color sequential Flux at the exit pupil- (lm/watt)- Dimensions (mm)- available Weight (g)- 9	Dimensions – NA Thickness of the waveguide -1.3mm Weight of the waveguide - 10.5 g Shape outline - Customizable (No specific shape documented)	No data available

Waveoptics	Efficiency (nits/Lumen) For three types of design - 400, 175, 160 MTF (cycles /degree) - 18 (>30%contrast) Transmission (At normal incidence) – For two designs - 80 % and 85% Contrast- 40:1	Projector Type - LCOS FOV (Degrees)- Three versions – 28, 40 55 Image orientation- 1:1, 16:9 (both available) Display Resolution (pixels)- Projection Type – Color sequential Flux at the exit pupil - (lm/watt)- Dimensions (mm)- Not available Weight (g)- Not Available	Dimensions- Available Thickness of the waveguide – (Two versions available) - 3.1 mm, 1.15mm Weight of the waveguide (three types available) - 14.1 g, 12 g, 7g Shape outline - T01, D08 Outline	Thermal recycling- Low temperature storage-

In the next section the specification comparison is done for the VR devices.

6.0 Benchmarking and Comparison of Specifications of Different VR Products: Current and Developer Editions

(*Note*: Reference to this section are shared in the* References)

Name	Cost	Display Type	Resolution	FOV	Audio	Refresh Rate	Latency	Optics	Inputs	type
Oculus Go	**$199**	Fast switch LCD	1280 x 1440 x 2	110°	Internal speaker of own Headphones	TBD	TBD	TBD	TBD	Standalone Mobile VR
Powis popup	**$500**	Smartphone supported	Smart phone supported	Smart phone supported	Smart Phone supported	Smartphone supported	Smart phone supported	Not adjustable lenses	Smart phone supported	Smart phone HMD

HP Windows MR headset Developer	Lenovo Explorer Windows MR Headset	Name	Samsung Oddysey	Vuzix iwear	HTC Vive Pro
$450	**$400**	Cost	**$500**	**$230**	**$1100**
LCD	LCD	Display Type	AMOLED	LCD	OLED
1440 x 1440 x 2	1440 x 1440 x 2	Resolution	1440x1600x2	1280 X720 X2	1440 X 1600
105°	105°	FOV	110°	55°	110°
Built in Audio-Out	Built in Audio-Out	Audio	Integrated AKG Headphones	Noise Isolating headphones	Integrated Headphones
90Hz	90Hz	Refresh Rate	90Hz	Unknown	90 Hz
Not Published	Not Published	Latency	Unknown	Unknown	unknown
Adjustable IPD via Software	Adjustable IPD via Software	Optics	IPD Adjustable	Floating Display Design	IPD adjustable
Wireless Motion Controllers	Wireless Motion Controllers	Inputs	USB, HDMI	HDMI, USB	Wireless Motion controllers
Tethered HMD	Tethered HMD	type	Tethered HM	Video HMD	Wireless HMD

Powis VR viewer 3	Oculus Rift	Deepoon E2	Name	Sony Play station VR	Acer Windows MR	Dell Visor Windows Mixed Reality Headset
$30	$599	$220.00	Cost	$350	$400	$450
Smart Phone supported	OLED	AMOLED	Display Type	OLED	LCD	LCD
Smart Phone supported	1080x1200 x2	1080p	Resolution	960 x1 080 x 2	1440 x 1440 x 2	1440 x 1440 x 2
Smart Phone supported	110°	120°	FOV	110°	105°	105°
Smart Phone supported	Built in Audio	None	Audio	360 Audio via included headphones	Built in Audio-Out	Built in 3.5mm jack
Smart Phone supported	90Hz	75Hz	Refresh Rate	120Hz or 90Hz	90Hz	90Hz
Smart Phone supported	<20 ms	<1ms (Claimed)	Latency	<18 ms (Estimated)	Not Published	Not Published
IPD and Focal Adjustment	IPD Adjustment	Japanese Non-Spherical lenses	Optics	Fixed Lenses, Compatible with Glasses Users	Adjustable IPD via Software	IPD Adjustable via Software
Smart Phone supported	HDMI 1.3	HDMI, USB 2.0	Inputs	PS4 Controller, PlayStation Move Wands	Motion Controllers	Wireless Motion Controllers
Smartphone HMD	Tethered HMD	Tethered HMD	type	Tethered HMD	Tethered HMD	Tethered HMD

Ritech Riem 3	Figment VR	Samsung Gear	Name	Homido Grab	Merge VR googles	Royole Moon
$19.99	$39.99	$129.00	Cost	$40	$60	$850
Smart Phone supported	Smart Phone supported	Super AMOLED	Display Type	Smart Phone supported	Smart Phone supported	2x AMOLED
Smart Phone supported	Smart Phone supported	2960 x 1440	Resolution	Smart Phone supported	Smart Phone supported	1920x1080x2
NA	NA	101°	FOV	100°	96°	Unknown
Smart Phone supported	Smart Phone supported	Smart Phone supported	Audio	Smart Phone supported	Smart Phone supported	Professional-grade Noise Cancelling Headphones
Smart Phone supported	Smart Phone supported	60 Hz	Refresh Rate	Smart Phone supported	Smart Phone supported	60Hz
Smart Phone supported	Smart Phone supported	NA	Latency	Smart Phone supported	Smart Phone supported	0.01ms
No Focus Adjustment required, IPS Adjustment available	Non-Adjustable	Focus, IPD not needed	Optics	37mm Biconvex lenses	IPD Adjustable 42mm Lenses	Adjustable IPD and Focus
Touch Button	Smart Phone supported	Touchpad, Wireless Controller	Inputs	Phone Dependent, has external button	Left and Right Buttons	HDMI, USB, WiFi
Smartphone HMD	Smartphone HMD	Smartphone HMD	type	Smartphone HMD	Smartphone HMD	3D Video HMD

VRIZZMO Revolt	VROID T1	Virtoba X5	LG 360	Utopia 360	Sybol Colorcross
$60	$69	$49.99	$199.00	$39.99	$6.99
Smart Phone supported	Smart Phone supported	Smart Phone supported	IPS LCD	Smart Phone supported	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported	920 x 720	Smart Phone supported	Smart Phone supported
110°	NA	120°	80	NA	Smart Phone supported
Smart Phone supported	Smart Phone	Smart Phone supported	Via Headphone Jack	Smart Phone	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported	NA	Smart Phone supported	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported	NA	Smart Phone supported	Smart Phone supported
Double Lens System	Individual IPD Adjustment Per Eye	Diopter and IPD	Adjustable Focus Only	IPD and Focal Length Adjustable	No Adjustment
Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
Smartphone HMD	Smartphone HMD	Smartphone HMD	Smartphone HMD	Smartphone HMD	Smartphone HMD

VRIIZZMO	Name	VR cat	zeiss VR one Plus
$60	Cost	$12.99	$129
Smart Phone supported	Display Type	Smart Phone supported	Smart Phone supported
Smart Phone supported	Resolution	Smart Phone supported	Smart Phone supported
120°	FOV	Smart Phone supported	100°
Smart Phone supported	Audio	Smart Phone supported	Smart Phone supported
Smart Phone supported	Refresh Rate	Smart Phone supported	Smart Phone supported
Smart Phone supported	Latency	Smart Phone supported	Smart Phone supported
Adjustability Unknown	Optics	33.5 mm	Adjustment Free Premium Zeiss Custom Lenses
Smart Phone supported	Inputs	Smart Phone supported	Smart Phone supported
Smartphone HMD	type	Smartphone HMD	Smartphone HMD

Name	Homido Mini	Leji VR mini	Freefly VR	Homido v2
Cost	$15	$13	$50	$75
Display Type	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
Resolution	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
FOV	100°	96°	120°	100°
Audio	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
Refresh Rate	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
Latency	Smart Phone supported	Smart Phone supported	Smart Phone supported	Smart Phone supported
Optics	Not Adjustable	Focus and IPD Adjustable	IPD Adjustable, no Diopter Adjustment	Adjustable IPD. Near and Farsightedness Support
Inputs	Smart Phone supported	Smart Phone supported	Capacitive Touch Triggers	Smart Phone supported
type	Smartphone HMD	Smartphone HMD	Smartphone HMD	Smartphone HMD

Stimuli 2VR	Dodocase smart VR	DreamZ VR
$30.00	**$40**	**$59.00**
Smart Phone supported	Smart Phone supported	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported
NA	NA	120°
Smart Phone supported	Smart Phone supported	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported
Smart Phone supported	Smart Phone supported	Smart Phone supported
34mm Lenses, Not Adjustable	Not Adjustable	IPD Adjustable
Smart Phone supported	Smart Phone supported	Smart Phone supported
Smartphone HMD	Smartphone HMD	Smartphone HMD

Name	Cost	Display Type	Resolution	FOV	Audio	Refresh Rate	Latency	Optics	Inputs	type
I am CARDBOARD XGVR	**$69.00**	Phone supported	Smart Phone supported	100°	Smart Phone supported	Smart Phone supported	Smart Phone supported	Adjustability Unknown	Bluetooth Button	Smartphone HMD
Beenoloculus	**$50**	Smart Phone supported	Smart Phone supported	N/A	Smart Phone supported	Smart Phone supported	Smart Phone supported	NA	Smart Phone supported	Smartphone HMD
Fibrum Pro	**$79**	Smart Phone supported	Smart Phone supported	110°	Smart Phone supported	Smart Phone supported	Smart Phone supported	No Adjustment, 30 mm Lenses	Smart Phone supported	Smartphone HMD
Diodlo Glass H1	**$79**	Smart Phone supported	Smart Phone supported	100-120°	Smart Phone supported	Smart Phone supported	Smart Phone supported	IPD and Focal Length	Touchpad, Volume Buttons	Smartphone HMD
Durovis Dive 7	**$50.00**	Tablet Dependent	Tablet Dependent	110°	Tablet Dependent	Tablet Dependent	Tablet Dependent	Focal Length and IPD Adjustable	Tablet Dependent	Tablet HMD

Name	Cost	Display Type	Resolution	FOV	Audio	Refresh Rate	Latency	Optics	Inputs	Type
Google daydream View	**$79.00**	Smart Phone supported	Smart Phone supported	90°	Smart Phone supported	Smart Phone supported	Smart Phone supported	No Adjustment	Smart Phone supported	Smartphone HMD
BoboVR Z3	**$22**	Smart Phone supported	Smart Phone supported	80-95°	Smart Phone supported	Smart Phone supported	Smart Phone supported	IPD and Focal Length	Smart Phone supported	Smartphone HMD
Auravisor	**$200**	LCD	1920x1080	100°	Yes	60Hz	NA	Adjustable, no glasses required	Bluetooth, WiFi, HDMI	All-In-One HMD
Viewmaster VR	**$39.99**	Smart Phone supported	Smart Phone supported	NA	Smart Phone supported	Smart Phone supported	Smart Phone supported	Limited Adjustment	Smart Phone supported	Smartphone HMD

Name	Varjo VR2/VR2-PRO	Oculus Rift	Xiaomi Mi VR play	Durovis Dive 5	Speck Pocket VR
Cost	VR2- $ 5919 (€4995) VR2PRO- $7106 (€5995)	$599	$14.99	$30	$69.95
Display Type	Next generation Bionic Display, low persistence micro-OLEDS and AMOLEDs	OLED	Smart Phone supported	Smart Phone supported	Smart Phone supported
Resolution	1920 x 1080 for OLEDS 1440 x 1600 for AMOLEDS	1080 x1200 x2	Smart Phone supported	Smart Phone supported	Smart Phone supported
FOV	87°	110°	N/A	~90°	N/A
Audio	Audio jack available	Built in	Smart Phone supported	Smart Phone supported	Smart Phone supported
Refresh Rate	60/90 Hz	90Hz	Smart Phone supported	Smart Phone supported	Smart Phone supported
Latency	<20ms	<20 ms	Smart Phone supported	Smart Phone supported	Smart Phone supported
Optics		IPD Adjustment	N/A	Focal Length and IPD	Not Adjustable
Inputs	Light optical fiver cables with USB-C, link box easy connection with PC	HDMI 1.3	Smart Phone supported	Smart Phone supported	Smart Phone supported
Type	Tethered HMD	Tethered	Smartphone Case	Smartphone HMD	Smartphone HMD

Name	Cost	Display Type	Resolution	FOV	Audio	Refresh Rate	Latency	Optics	Inputs	Type
Panasonic VR Glasses	**NA**	NA	NA	NA	NA	NA	NA	NA	NA	NA

VR Developer Editions

(*Note*: Reference to this section are shared in the* References)

Name	Cost	Display Type	Resolution	FOV	Audio	Refresh Rate	Latency	Optics	Inputs	Type
OSVR HDK 2	**$399.99**	OLED	1080x1200x2	Not Published	Additional Headphones Supported	90Hz	Unknown	Per-eye Diopter and IPD	USB	Tethered HMD
Meta 2 (Developer kit)	**$1495**	AR Projection	2432x1366x2	90°	Four Surround Speakers	60Hz	Unknown	AR Projection Visor	Video and Data Cable	Tethered AR/MR HMD
Sulon Q	**NA**	OLED	2560x1440	110°	External Headphones	TBA	TBA	TBA	Spatial Sensors, Cameras	Wireless Mixed Reality HMD

Pinch VR	Altergaze VR	Atheer air smart glasses	Project Alloy	Name	StarVR
NA	**$60.00**	**$3950**	**NA**	Cost	**NA**
Smart Phone supported	Smart Phone supported	Projection	TBA	Display Type	TBA
Smart Phone supported	Smart Phone supported	1280x720	TBA	Resolution	2560x1440 Per Eye
TBA	110°/90°	50°	TBA	FOV	210° Horizontal, 130° Vertical
Yes - Phone Dependent	Smart Phone supported	Yes	TBA	Audio	via Headphones
Smart Phone supported	Smart Phone supported	60Hz	TBA	Refresh Rate	TBA
Smart Phone supported	Smart Phone supported	NA	TBA	Latency	TBA
TBA	50 mm Custom Lens System	See-through Projection Display	TBA	Optics	Duel Fresnal Lenses, Adjustable
TBA		USB-C, Wi-Fi, Bluetooth, 3D Gesture	TBA	Inputs	HDMI
Smartphone HMD	Smartphone HMD	All-In-One HMD	Wireless Merged Reality HMD	type	Tethered HMD

Name	Pico G2 4KS	TRIVISIO-SXGA.62 (VR)	AirVR	Blue Sky Pro	Neurogoggles
Cost	NA	NA	NA	NA	NA
Display Type	NA	SXGA LCD	Designed for iPad Mini with Retina Display	TBA	TBA
Resolution	4K	1280 X1024	Varies - Retina PPI	1080x1920 x2	TBA
FOV	NA	45° per eye	100°	123°	TBA
Audio	NA	Built-in Microphone mono speaker and headset	From iPad	Yes - Integrated	NA
Refresh Rate	75Hz	60Hz	60Hz	TBA	NA
Latency	NA	<8ms	Unknown	30ms	NA
Optics	NA	No lens info shared Backlight luminance 180 cd/m2 Focalplane accommodation-2130 mm Stereo camera for video see-	NA	Adjustable - No Glasses Needed	NA
Inputs	NA	NA	None	HDMI	NA
type	Tethered HMD	Tethered HMD	Tablet HMD	Tethered HMD	Tethered HMD

Primax (8K X)	XTAL
$1399	NA
Customized low persistence liquid (CLPL),	Two Quad HD high-density OLED displays
3840 x 2160 x 2	5120 x 1440 (2560 x 1440 per eye)
200°	180°
	Spatial 3D sound from built-in Sound card
Microphone	Built in Microphone and voice recognition brings voice
75(native)/110 Hz(upscale)	
MTP latency-15ms	Extremely low latency: no numbers mentioned
lens- fresnel x2	Custom built optics with exceptionally clear patented Non-Fresnel VR lenses
IPD-55mm- 75mm	
Headphone, USB X 2,, DPX1, Audio jack, type cx1	NA
Tethered HMD	Tethered HMD

(*Note*: Reference to this section are shared in the* References)

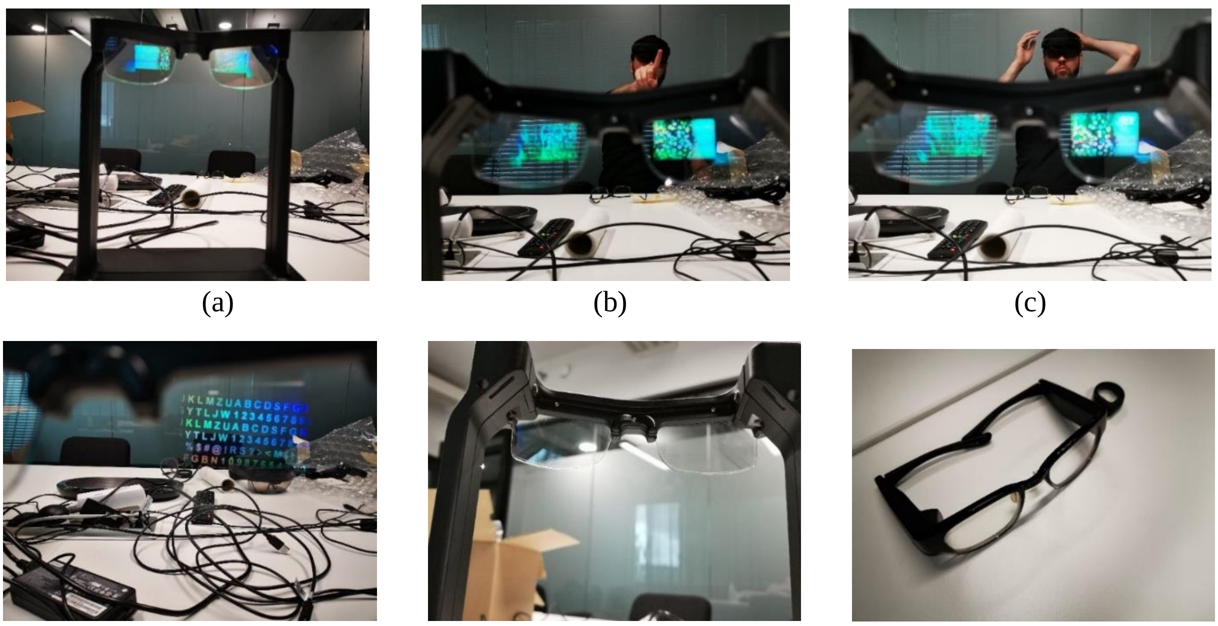
7.0 Demonstration of XR Products from HumOR Consortium Partners and Facilities

7.1 Dispelix Oy (SME, AR OEM)

- Demonstration of the commercially available AR devices in the market such as Magic Leap 1, Microsoft HoloLens 2, North Glasses (not working), Rucon, Vuzix (monocular version not available).
- Study about hologram interaction, haptics, sound, Voice guidance, gestures, eye tracking calibration, ergonomics (comfort, weight balance), device calibration through the above set of commercially available devices.

7.1.1 Dispelix glasses

Demonstration of in-house made **Dispelix** own glasses, **image** and **text readability (**see Fig. 28) of **moving pictures** taking into consideration latency issues, color, problems with the customized versions and room for improvement etc.



(a) (b) (c)

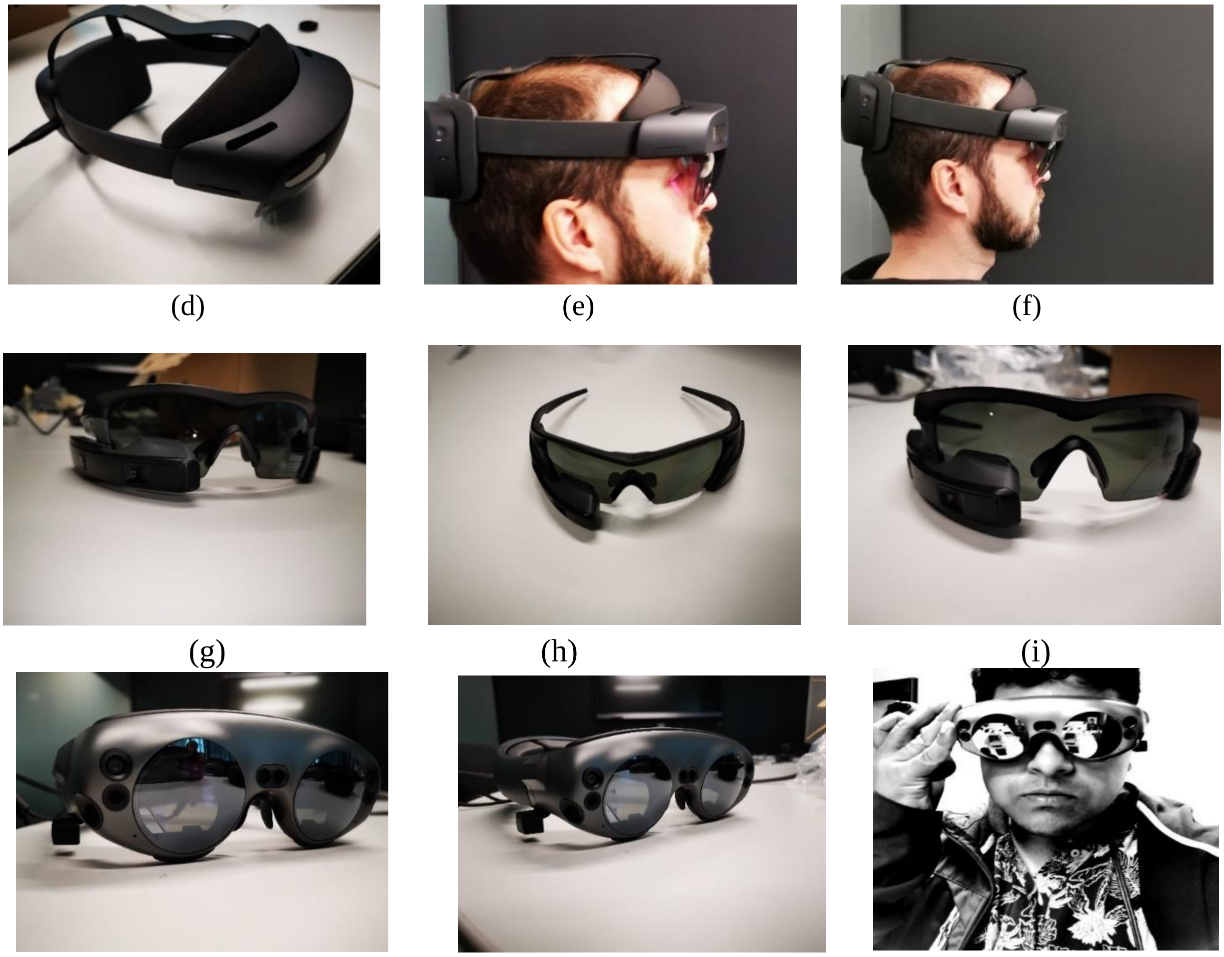
(d) (e) (f)

Fig 28: Demonstration of the text readability through Dispelix glasses**.**

7.1.2 Other AR devices ergonomics and comfortability check

Several other devices were also demonstrated for ergonomics, comfortability check during the visit to company Dispelix Oy and are illustrated in the Fig. 29.

(a) (b) (c)



(d) (e) (f)

(g) (h) (i)

(j) (k) (l)

Fig. 29: Available AR devices comfortability and ergonomics check

7.2 Varjo Oy

- Demonstration of the commercially available Varjo VR products such as VR-2, VR-2 PRO using DCS flight simulator, leap motion used cases, to understand hand tracking and virtual object interaction.
- Demonstration of developer edition of XR-1 and Mixed Reality experience for medical applications such as surgery, cockpit view of flight, architecture examples such as customized rooms and user interaction with the real environment while being in the virtual world.
- Thorough discussion about the specification and internal optics with the device schematics, eye tracking, latency of the devices, computing, weight balance, haptics, video see-through optics, display.
- Discussion about the in-house facility, capabilities and specifications sharing (public information)

7.2.1 Pictures taken during the demonstration

(a)

(b)

(c)

(d)

(e)

(f)

(g)

Fig 30: Pictures taken during the demonstrations: Disassembly of XR-1 Developer Edition.

9.0 Conclusions

The state-of-the-art report describes the next gen XR requirements and benchmark specifications that can be used in future next generation devices in the XR ecosystem. The report clearly explains the basic terminologies about the diverse topics used in the current or existing AR/VR/MR and defines how we merge everything to XR.

During the preparation of the report, challenges arose in finding the specifications from company's webpages and blog posts and other open sources. In some of the cases, like HMDmd, separate requests for specifications were made as the sources were not just enough to state any information.

The report discusses the broad challenges in the display technology and how 3D next generation display will come soon to compensate the human vision related limitations to make. The state-of the-art into future XR devices and immersive requirements that seems to be challenging. Additionally, what needs to be considered for future XR auditory technology has been presented.

This SoA also opened two ideas - the first one is the next generation display technology approach that will follow up in the future to help companies understand the optical quality and integration to the current generation XR devices. The second one is to identify the objective of the cross-modal effects like audio, tracking, haptics into the next generation XR devices.

In addition to the above, this report also helped to identify which specific areas of the current generation devices needs improvement and how to make use of the benchmarking specifications and find answers to the extreme requirements of the XR ecosystem and industry.

10.0 Future Works

In the next phase of the project, optical technology approach analysis of the current AR/VR devices provided by the companies (Dispelix Oy, Huawei technologies Oy (CO.) Ltd. and Varjo Oy) will be presented and the current state-of-the art will be examined. For examining the above, display parameter testing (measuring parameters like display MTF, etc.) will be done and limitations will be sorted out to meet the future XR device requirements.

References

[1] John Peddie, ‘Augmented reality - where we all live’, *Springer*, 2017.

[2] Jason Jerald, ‘The VR book: human-centered design for Virtual Reality’, *Morgan & Claypool*, 2015.

[3] Bernard Kress, ‘Optical architectures for Augmented-, Virtual- and Mixed Reality headsets’, *SPIE*, 2020.

[4] Peter Wozniak, Oliver Vauderwange, Avikarsha Mandal, Nicolas Javahiraly, Dan Curticapean, "Possible applications of the LEAP motion controller for more interactive simulated experiments in augmented or Virtual Reality," *Proc. SPIE 9946, Optics Education and Outreach IV*, 99460P (27 September 2016); https://doi.org/10.1117/12.2237673

[5] Zabels, R.; Osmanis, K.; Narels, M.; Gertners, U.; Ozols, A.; Rūtenbergs, K.; Osmanis, I. “AR Displays: Next-Generation Technologies to Solve the Vergence,” *Accommodation Conflict. Appl. Sci.* 2019, 9, 3147.

[6] Kaan Akşit, Ward Lopes, Jonghyun Kim, Peter Shirley, and David Luebke. 2017. Near-eye varifocal Augmented Reality display using see-through screens. *ACM Trans. Graph.* 36, 6, Article 189 (November 2017), 13 pages. https://doi.org/10.1145/3130800.3130892

[7] Jen-Hao Rick Chang, B. V. K. Vijaya Kumar, and Aswin C. Sankaranarayanan. 2018. Towards multifocal displays with dense focal stacks. *ACM Trans*. Graph. 37, 6, Article 198 (November 2018), 13 pages. https://doi.org/10.1145/3272127.3275015

[8] Zhan, T., Xiong, J., Zou, J. et al. Multifocal displays: review and prospect. *PhotoniX* 1, 10 (2020). https://doi.org/10.1186/s43074-020-00010-0

[9] Jannick P. Rolland, Myron W. Krueger, and Alexei Goon, "Multifocal planes head-mounted displays," *Appl. Opt*. 39, 3209-3215 (2000)

[10] Wu, J.-Y., Chou, P.-Y., Peng, K.-E., Huang, Y.-P., Lo, H.-H., Chang, C.-C., and Chuang, F.-
[11] M. (2018) Resolution enhanced light field near eye display using e-shifting method with birefringent plate. *Jnl Soc Info Display*, 26: 269– 279. doi: 10.1002/jsid.665.

[12] Stahl, R. and Jayapala, M. (2011), Holographic displays and smart lenses. *Optik & Photonik*, 6: 39-42. doi:10.1002/opph.201190328

[13] Wang, D., Liu, C., Shen, C. *et al.* Holographic capture and projection system of real object based on tunable zoom lens. *PhotoniX* **1,** 6 (2020). https://doi.org/10.1186/s43074-020-0004-3

[14] Di Wang, Chao Liu, and Qiong-Hua Wang, "Method of chromatic aberration elimination in holographic display based on zoomable liquid lens," *Opt. Express* 27, 10058-10066 (2019)

[15] Zehao He, Xiaomeng Sui, Guofan Jin, and Liangcai Cao, "Progress in Virtual Reality and Augmented Reality based on holographic display," *Appl. Opt.* 58, A74-A81 (2019)

[16] Steven M. LaValle, ‘Virtual Reality’, *Cambridge University Press*, 2019. http://lavalle.pl/vr/

[17] Virtual Reality Blog website, ‘https://www.vrs.org.uk/virtual-reality/what-is-virtual-reality.html’

[18] Qualcomm boundless XR presentation, ‘https://www.qualcomm.com/media/documents/files/boundless-mobile-xr-over-5g.pdf’

[19] VRS webpage_samsung_oddesy, ‘ https://www.vrs.org.uk/headsets/samsung-odyssey’

[20] VRS webpage_htcvive_pro, ‘ https://www.vrs.org.uk/headsets/htc-vive-pro/’

[21] VRS webpage_main, ‘https://www.vrs.org.uk/

[22] Qualcomm_mobile_future of XR_pdf,
‘https://www.qualcomm.com/media/documents/files/the-mobile-future-of-extended-reality-xr.pdf’

[23] Yaqin Zhou, Yue Yin, Yide Yuan, Tiegang Lin, Huihui Huang, Lishuang Yao, Xiaoqian Wang, Alwin M. W. Tam, Fan Fan & Shuangchun Wen (2019) Liquid crystal Pancharatnam-Berry phase lens with spatially separated focuses, *Liquid Crystals*, 46:7, 995-1000, DOI: 10.1080/02678292.2018.1550820

[24] Tao Zhan, Yun-Han Lee, Guanjun Tan, Jianghao Xiong, Kun Yin, Fangwang Gou, Junyu Zou, Nannan Zhang, Dongfeng Zhao, Jilin Yang, Sheng Liu, and Shin-Tson Wu, "Pancharatnam–Berry optical elements for head-up and near-eye displays [Invited]," *J. Opt. Soc. Am.* B 36, D52-D65 (2019)

[25] Tao Zhan, Jianghao Xiong, Yun-Han Lee, Ran Chen, and Shin-Tson Wu, "Fabrication of Pancharatnam-Berry phase optical elements with highly stable polarization holography," *Opt. Express* 27, 2632-2642 (2019)

[26] Kun Gao, Hsien-Hui Cheng, Achintya K. Bhowmik, and Philip J. Bos, "Thin-film Pancharatnam lens with low f-number and high quality," *Opt. Express* 23, 26086-26094 (2015)

[27] Tao Zhan, Yun-Han Lee, Guanjun Tan, Jianghao Xiong, Kun Yin, Fangwang Gou, Junyu Zou, Nannan Zhang, Dongfeng Zhao, Jilin Yang, Sheng Liu, and Shin-Tson Wu, "Pancharatnam–Berry optical elements for head-up and near-eye displays [Invited]," *J. Opt. Soc. Am.* B 36, D52-D65 (2019)

[28] Tao Zhan, Jianghao Xiong, Yun-Han Lee, Ran Chen, and Shin-Tson Wu, "Fabrication of Pancharatnam-Berry phase optical elements with highly stable polarization holography," *Opt. Express* 27, 2632-2642 (2019)

[29] Kun Gao, Hsien-Hui Cheng, Achintya K. Bhowmik, and Philip J. Bos, "Thin-film Pancharatnam lens with low f-number and high quality," *Opt. Express* 23, 26086-26094 (2015)

[30] Lee, Y., Tan, G., Zhan, T., Weng, Y., Liu, G., Gou, F., Peng, F., Tabiryan, N. V., Gauza, S., & Wu, S. (2017). Recent progress in Pancharatnam–Berry phase optical elements and the applications for virtual/augmented realities, *Optical Data Processing and Storage*, *3*(1), 79-88. doi: https://doi.org/10.1515/odps-2017-0010

[31] Hao Yu, Ziyuan Zhou, Yongle Qi, Xinfang Zhang, and Qi-Huo Wei, "Pancharatnam–Berry optical lenses," *J. Opt. Soc. Am.* B 36, D107-D111 (2019)’

[32] Tao Zhan, Yun-Han Lee, and Shin-Tson Wu, "High-resolution additive light field near-eye display by switchable Pancharatnam–Berry phase lenses," *Opt. Express* 26, 4863-4872 (2018)

[33] Moon, S., Lee, C., Nam, S. *et al.* Augmented Reality near-eye display using Pancharatnam-Berry phase lenses. *Sci Rep* **9,** 6616 (2019). https://doi.org/10.1038/s41598-019-42979-0

[34] Yan Li, Yueda Liu, Sida Li, Pengcheng Zhou, Tao Zhan, Quanming Chen, Yikai Su, and Shin-Tson Wu, "Single-exposure fabrication of tunable Pancharatnam-Berry devices using a dye-doped liquid crystal," *Opt. Express* 27, 9054-9060 (2019)

[35] Gamma scientific webpage, https://www.gamma-sci.com/sites/default/files/white-paper/ned-testing-whitepaper.pdf.

[36] Ho Jin Jang, Jun Yeob Lee, Jaeyun Kim, Jeonghun Kwak & Jae-Hyeung Park (2020) Progress of display performances: AR, VR, QLED, and OLED, *Journal of Information Display*, 21:1, 1-9, https://doi.org/10.1080/15980316.2020.1720835.

[37] Vieri, C., Lee, G., Balram, N., Jung, S. H., Yang, J. Y., Yoon, S. Y., and Kang, I. B. (2018) An 18 megapixel 4.3″ 1443 ppi 120 Hz OLED display for wide field of view high acuity head mounted displays. *Jnl Soc Info Display*, 26: 314– 324. doi: 10.1002/jsid.658

[38] Li, K. and Lake, A. (2019), 31-3: Eyebox Evaluation in AR/VR Near-eye Display Testing. SID Symposium Digest of Technical Papers, 50: 434-437. doi:10.1002/sdtp.12949

[39] Optikos pdf for MTF, https://www.optikos.com/wp-content/uploads/2013/11/How-to-Measure-MTF.pdf.

[40] Optikos webpage, https://www.optikos.com/measure-mtf/.

[41] Brian M. Deegan, Patrick E. Denny, Vladimir Zlokolica, Barry Dever, Laura Russell, "Addressing challenges of modulation transfer function measurement with fisheye lens cameras," Proc. SPIE 9403, *Image Sensors and Imaging Systems* 2015, 94030F (13 March 2015); https://doi.org/10.1117/12.2083347

[42] vrs_web_vr_motiontracking, 'https://www.vrs.org.uk/virtual-reality-gear/motion-tracking/'

[43] Dispelix Oy webpage, DPX 30° waveguide display specs, https://www.dispelix.com/product/.

[44] Waveoptics webpage, 'T01-28°, D08-40°, KATANA-28° Waveguide display specs', https://enhancedworld.com/products/waveguides/.

[45] Varjo webpage, 'Varjo XR-1, VR-2 and VR-2 pro tech specs', https://varjo.com/products/ .

[46] DAQRI webpage, 'DAQRI smart glasses technical specifications', https://daqri.com/products/smart-glasses/

[47] Vuzix Webpage, 'Products m-series', https://www.vuzix.com/products/m-series.

[48] Vuzix Webpage, 'Products blade - smart-glasses', https://www.vuzix.com/products/blade-smart-glasses.

[49] Hololens webpage- https://www.microsoft.com/en-us/hololens/hardware#.

[50] Oculus webpage - https://www.oculus.com/compare/.

[51] Vuzix - https://www.vuzix.com/products/m400-smart-glasses.

[52] Garmin - https://buy.garmin.com/en-US/US/p/530536#specs.

[53] nreal - https://www.nreal.ai/specs.

[54] Magic Leap - https://www.magicleap.com/en-us/magic-leap-1.

[55] Varjo - https://varjo.com/products/vr-2-pro/.

[56] Everysight Raptor - https://everysight.com/about-raptor/.

[57] VividQ - https://vivid-q.com/display-applications/ .

[58] Toshiba - https://us.dynabook.com/smartglasses/products/index.html .

[59] Olympus - https://www.getolympus.com/smartglasses .

[60] Iristik - https://iristick.com/products/iristick-z1-premium .

[61] Konica Minolta - https://vrgal.net/ar-glasses-from-konica-minolta/ .

[62] Hiscene - https://www.hiscene.com/detail-glasses/ .

[63] Zebra - https://www.zebra.com/us/en/products/mobile-computers/wearable-computers/hd4000.html .

[64] Guangli - https://www.prnewswire.com/news-releases/guangli-rolls-out-its-first-light-field-ar-glasses-300870444.html .

[65] North (acquired by Google) - https://www.bynorth.com/ .

[66] Hmdmd - http://hmdmd.com/ .

[67] Ximmerse - https://www.ximmerse.com/en/RhinoX.

[68] RideOn - https://www.rideonvision.com/new/ski-goggle.php#goggle-specs .

[69] Kura - https://www.kura.tech/products .

[70] Innovega - https://www.medgadget.com/2019/06/emacula-augmented-reality-for-vision-impaired-interview-with-steve-willey-ceo-of-innovega.html.

[71] Dispelix - https://www.dispelix.com/ .

[72] Waveoptics - https://enhancedworld.com/ .

[73] Huawei - https://www.gearbest.com/blog/new-gear/huawei-vr-glass-review-using-experience-hands-on-test-9681 .

[74] Dreamworld - https://www.dreamworldvision.com/ .

[75] Kopin - https://www.kopin.com/kopin-microdisplays/ .

[77] Pacific Future - https://www.pacificfuture.co/portal/index/amglass.html .

[78] Vivo - https://www.theverge.com/2019/6/26/18759404/vivo-ar-glass-announced-mwc-shanghai-2019.

[79] Trisvisio - https://www.trivisio.com/hmd-nte .

[80] Pareal - http://www.pareal.com/.

[81] Tilt five- https://tilt-five.backerkit.com .

[82] Avegant - https://www.avegant.com/light-field .

[83] Tesseract - https://tesseract.in/ .

[84] Realmax - https://www.realmax.fi/about-realmax .

[85] Firefly - https://fireflydimension.com/ .

[86] Form - https://www.formswim.com/.

[87] Lemnis - https://www.lemnis.tech/ .

[88] Augmedics - https://augmedics.com/ .

[89] Oppo - https://www.oppo.com/en/newsroom/stories/augmented-reality-glasses-are-about-to-change-the-game-/ .

[90] Westunitis - https://www.westunitis.com/hardware/infolinker2/ .

[91] VRgineers - https://vrgineers.com/xtal/ .

[92] Kc-textil - https://www.kc-textil.com/products-2-2/ .

[93] ThirdEye - https://www.thirdeyegen.com/x2-smart-glasses .

[94] Polaris - https://www.polaris-ar.com/ .

[95] Combine Reality - https://combinereality.com/ .

[96] Epson - https://tech.moverio.epson.com/en/bt-30c/ .

[97] Canon - https://virtualrealitytimes.com/2020/02/05/canon-md-20-canon-reveals-the-ar-successor-to-the-mreal-md-20/ .

[98] All sight - https://wearabletechnologymall.com/best-smart-glasses-2020-allgsight-s1-ar-best-smart-glasses-2020/ .

[99] NEDplusAR - http://www.nedplusar.com/en/index/chanpin/cid/79.html .

[100] Lynx - https://lynx-r.com/specs.html .

[101] Qdlaser - https://en.retissa.biz/ .

[102] LLvision - https://www.llvision.com/en/glxss-me.html .

[103] Nvidia - https://bit-tech.net/news/tech/peripherals/nvidia-shows-off-next-generation-ar-glasses-prototypes/1/.

[104] Samsung - https://www.digitalbodies.net/augmented-reality/samsungs-ar-glasses-revealed-at-ces-2020/ .

[105] Ajnalens - https://www.ajnalens.com/ajnalite/ .

[106] LetinAR - https://letinar.com/.

[107] Sony - https://developer.sony.com/develop/sed-100a-holographic-waveguide-display/specifications .

[108] Glassup - https://www.glassup.com/en/f4/#f4-scheda.

[109] Oculus - https://www.oculus.com/?locale=fi_FI.

[110] HTC Vive - https://www.vive.com/us/product/vive-cosmos/specs/ .

[111] Pico 2 - https://www.pico-interactive.com/eu/en/compare_products.html .

[112] Panasonic - https://news.panasonic.com/global/press/data/2020/01/en200107-5/en200107-5.html .

[113] Halomini - http://www.shadowcreator.com/halomini/halomini.html .

[114] VLA VR - https://www.vlavrone.com/product-vla-vr-one.

[115] North Focal Specs - https://support.bynorth.com/assets/Important_Information_Guide.pdf

[116] North Focal Specs1 - https://www.tomshardware.com/reviews/north-focals-smart-glasses-ar,5968.html.

[117] North Focal Specs2 - https://www.technipages.com/north-focals-reviews-and-specs .

[118] Sony headphones - https://www.sony.com/electronics/headband-headphones/wh-1000xm4.

[119] D. R. Begault, E. M. Wenzel, and M. R. Anderson, “Direct comparison of the impact of head tracking, reverberation, and individualized head-related transfer functions on the spatial perception of a virtual speech source,” *Journal of the Audio Engineering Society*, vol. 49, no. 10, pp. 904–916, 2001.

[120] J. Blauert, “Concepts behind sound quality: some basic considerations,” *Internoise*, vol. 9, pp. 72–79, 2003.

[121] Lindau, Alexander, Hans Joachim Maempel and Stefan Weinzierl (2008). “Minimum BRIR grid resolution for dynamic binaural synthesis”. In: *Acoustics Conference*. Paris, pp. 3851–3856.

[122] Blauert, Jens (1997). Spatial hearing: the psychophysics of human sound localization. *Cambridge, MA: MIT press*.

[123] Gerzon, Michael A. (1973). “Periphony: with-height sound reproduction”. In: Journal of the Audio Engineering Society 21.1, pp. 2–10.