ITUPublications

International Telecommunication Union
Radiocommunication Sector

Report ITU-R BT.2506-1

(03/2024)

BT Series: Broadcasting service (television)

Requirements for spatial characteristics of an ideal head-mounted display for immersive video

Foreword

The role of the Radiocommunication Sector is to ensure the rational, equitable, efficient and economical use of the radio-frequency spectrum by all radiocommunication services, including satellite services, and carry out studies without limit of frequency range on the basis of which Recommendations are adopted.

The regulatory and policy functions of the Radiocommunication Sector are performed by World and Regional Radiocommunication Conferences and Radiocommunication Assemblies supported by Study Groups.

Policy on Intellectual Property Right (IPR)

ITU-R policy on IPR is described in the Common Patent Policy for ITU-T/ITU-R/ISO/IEC referenced in Resolution ITU-R 1. Forms to be used for the submission of patent statements and licensing declarations by patent holders are available from http://www.itu.int/ITU-R/go/patents/en where the Guidelines for Implementation of the Common Patent Policy for ITU-T/ITU-R/ISO/IEC and the ITU-R patent information database can also be found.

Series of ITU-R Reports

(Also available online at https://www.itu.int/publ/R-REP/en)

Series	Title
BO	Satellite delivery
BR	Recording for production, archival and play-out; film for television
BS	Broadcasting service (sound)
BT	**Broadcasting service (television)**
F	Fixed service
M	Mobile, radiodetermination, amateur and related satellite services
P	Radiowave propagation
RA	Radio astronomy
RS	Remote sensing systems
S	Fixed-satellite service
SA	Space applications and meteorology
SF	Frequency sharing and coordination between fixed-satellite and fixed service systems
SM	Spectrum management
TF	Time signals and frequency standards emissions

***Note**: This ITU-R Report was approved in English by the Study Group under the procedure detailed in Resolution ITU-R 1.*

Electronic Publication
Geneva, 2024

REPORT ITU-R BT.2506-1

Requirements for spatial characteristics of an ideal head-mounted display for immersive video

(2022-2024)

1 Introduction

Advanced immersive sensory media (AISM) systems enable users to enjoy immersive experiences with an unprecedented degree of presence. An ideal immersive video system should provide the user with an experience similar to that induced by the observation of natural scenes in the real world. One of the most promising approaches for displaying immersive video is the use of a head-mounted display (HMD). In this context, the requirements for an ideal HMD, including spatial and temporal characteristics, range of luminance, colour gamut, tonal variation, physical weight and comfort of wear, need to be clarified. As a first step, this Report presents ergonomic findings related to the spatial characteristics required for an ideal HMD.

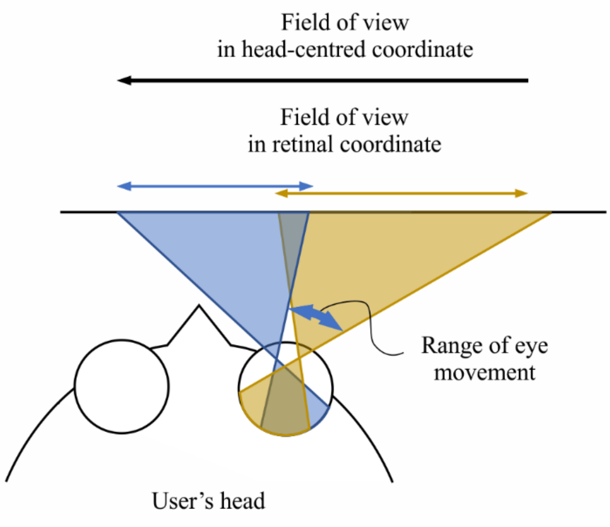
2 Spatial characteristics of an ideal HMD

2.1 Field of view

The field of view in the head-centred coordinate system needs to be assessed by psychophysical experiments to determine if the HMD can present optical stimuli to the entire field of view of the human visual system. Many prior studies on the human visual field [1]-[3] have performed measurements on the retinal coordinate system. However, since users move their eyes freely when wearing an HMD fixed to the head, they see visual objects in the retinal coordinates of any eye position over a wide range of eye movements (Fig. 1). In order to ascertain the required field of view of an ideal HMD, it is necessary to measure the field of view in the head-centred coordinate system. Research on the range of eye movement [4] has suggested that, by linearly integrating the range of eye movement, it may be possible to derive the field of view in the head-centred coordinate system. The result will be adjusted to take account of facial structures such as the nose, forehead and shape of the eye fissure.

FIGURE 1

Relationship between fields of view in head-centred and retinal coordinates



2.1.1 An experiment to assess field of view in head-centred coordinate system

The field of view in the head-centred coordinate system, which contributes to the requirements of an ideal HMD, was measured in a scenario where the head is fixed and the eyes can move freely [5].

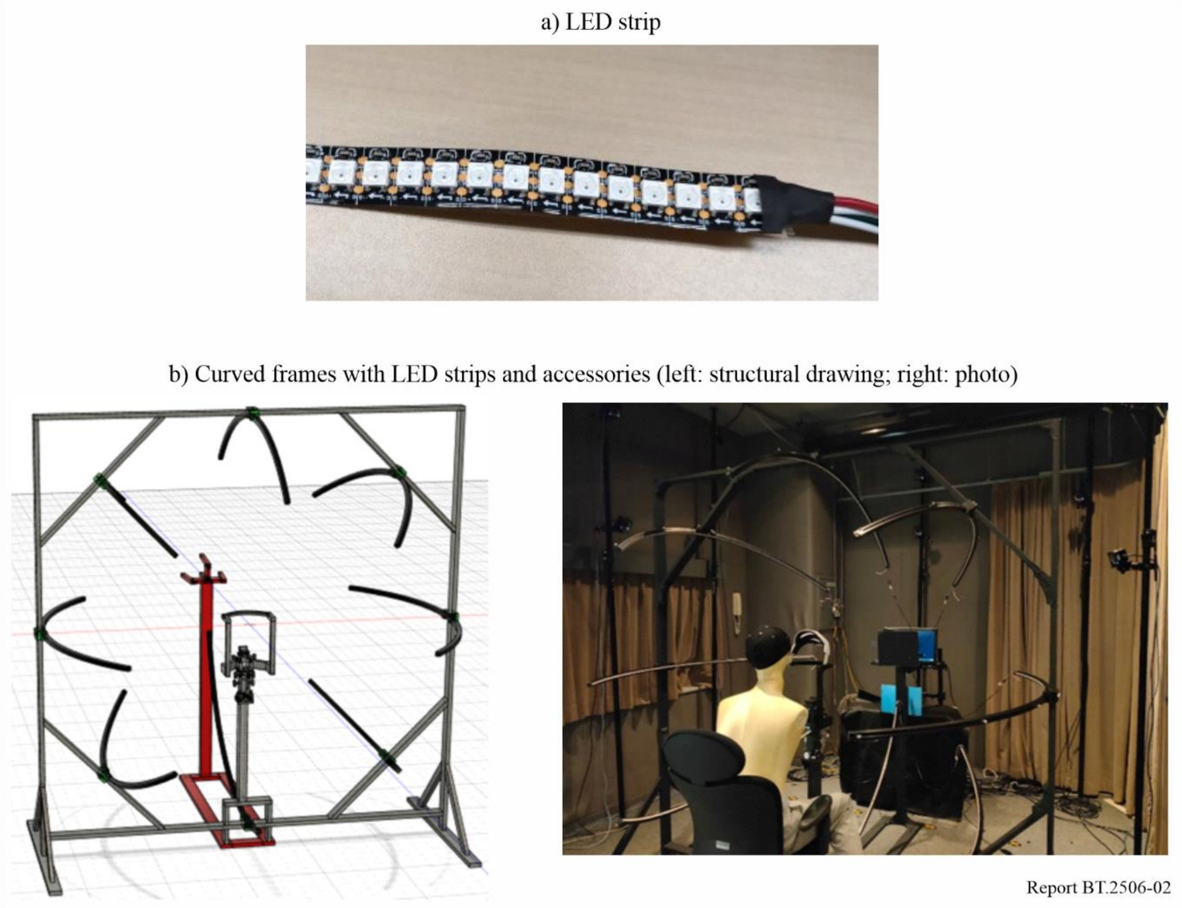
2.1.1.1 Method

Figure 2 shows the apparatus used in the experiment. The stimuli were blinking LED lights lined up in strips (Fig. 2(a)) containing 144 individual RGB LEDs (5×5 mm) per metre. Eight strips of 1.57 m in length were used.

The LED strips were attached to the inner surface of black curved aluminium frames with a radius of 1 000 mm (Fig 2(b)). The curved frames were installed to the inside of a square black aluminium frame measuring 2 080 mm per side. Each participant's head was located at the centre of the square by placing the chin on a chinrest equipped with an xyz stage.

FIGURE 2

Apparatus for an experiment assessing field of view



A structure consisting of four adjustment tubes directed toward the centre of the square was installed in front of the participant's head. Participants were asked to position their right eye at the centre. When the eye was correctly centred, the participant would see the same luminance of the adjustment lights set behind the structure through the tubes.

The positions of the participant's head, the adjustment tubes and the eight curved frames were measured using a three-dimensional position measurement system in order to calculate the angular positions of the individual LEDs and to monitor the movement of participant's head.

The experiment was performed in an almost completely dark room where participants could identify the position and shapes of the curved frames and other structures of the apparatus but not the position of individual LEDs.

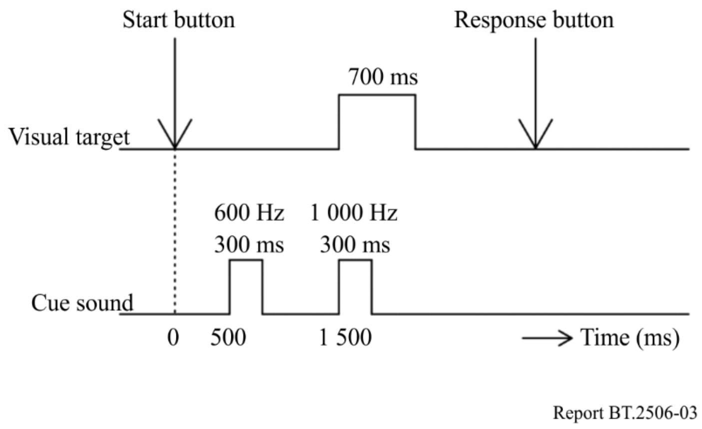
Nine adults, four men and five women with ages ranging from 20 to 33 (average 24.9 years) with normal uncorrected vision, participated in the experiment. Their left eye was shielded by eye patches and the experiment was performed on their right eye.

The task was to detect a blinking light spot with eye movement only, not head movement and report it by pressing a button.

At the beginning of the experiment, participants were asked to place their heads facing forward on the chinrest. The position of the chinrest was adjusted by using adjustment tubes to manipulate the xyz stage on which the chinrest was mounted (Fig. 2(b)). The experiment consisted of eight blocks, each of which was used to estimate the threshold light position along a single curved frame. The order of the frames was randomized. At the beginning of each block, the curved frame containing the target point light was indicated by a PC monitor placed behind the adjustment tubes. The trial started by pressing a button.

FIGURE 3

Time course of a trial



NOTE – Participants moved their gaze toward the curved frame containing the target point light during the first cue sound. The target point light and the second cue sound appeared at the same time.

Each block consisted of a series of trials. At the beginning of each trial, participants adjusted the position and direction of their heads and pressed the start button (Fig. 3). Following the button press, the first cue sound (600-Hz tone) was beeped. The participants moved their eyes toward the curved frame containing the target point light during the sound. The second cue sound (1 000-Hz tone) beeped at the same time as the onset of the target point light. The target point light was a green light ($x = 0.212$, $y = 0.755$) flickering at 10 Hz for 700 ms. Its maximum luminance was 128 cd/m^2 and its area was 0.08 square degrees. The position of the target light was controlled by using an adaptive psychophysical method (Best-Pest rule [6]). After the target point light disappeared, the participants reported whether they could detect the target by pressing one of two buttons, yes or no. If the number of reversals of direction of the target position displacement reached 14, the block was terminated. Each block consisted of 24.1 trials on average.

The experiment was repeated two or three times for each participant.

2.1.1.2 Results

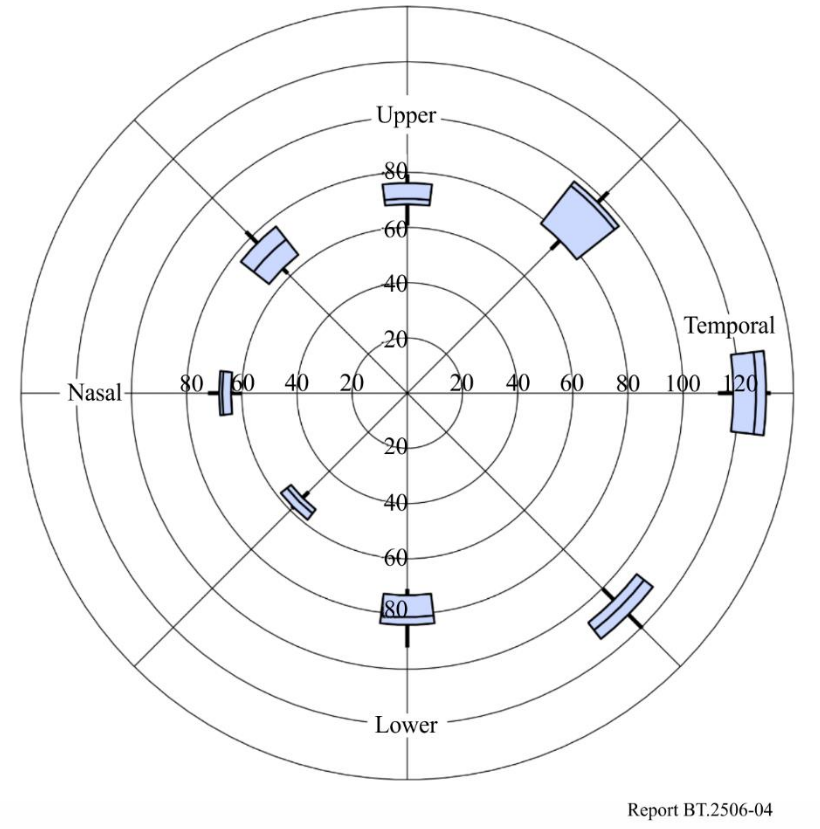
The average maximum head displacement for each participant was 6.5 mm in distance, 0.93 degrees in pitch, 1.32 degrees in yaw, and 1.40 degrees in roll. No recalibration of LED light positions was performed.

Figure 4 shows the thresholds of the angular position of the detected target stimuli, which were calculated as follows. First, three-dimensional stimulus positions were converted into angular positions. Second, the data for all repetitions of the experiments were pooled separately for each participant and curved frame, and a psychometric function was estimated from the pooled data by maximum likelihood [6], with a threshold defined as the position corresponding to a 50% response rate. This was repeated for each curved frame and each participant.

The medians of thresholds were 67.1 degrees on the nasal side, 126.4 degrees on the temporal side, 70.4 degrees on the upper side, and 81.1 degrees on the lower side. Previous studies [2,3] have reported that the normal visual field extends approximately 60, 100, 60 and 75 degrees or 60, 90, 60 and 70 degrees on the nasal, temporal, upper, and lower sides, respectively. The current findings differ in that eye movement extended the visual field significantly on the temporal side, about 25 to 35 degrees, but not on the nasal, upper, or lower sides.

FIGURE 4

Distribution of angular thresholds of detected target



NOTE – Numbers indicate eccentricity in units of degrees in head-centred coordinate system.

Widths of field of view, which are the sums of the angular thresholds of the two sides across the midpoint, were 189.4 degrees horizontally and 153.3 degrees vertically. This suggests that an HMD with a horizontal field of view of 189.4 degrees can cover the entire monocular field of view. In terms of the binocular field of view, an ideal horizontal field of view would be 252.8 degrees, which can be obtained by doubling the data of the temporal side (126.4 degrees).

These values are based on the medians of data. To cover the field of view of 75% of the participants, monocular and binocular horizontal fields of view would increase up to 197.1 degrees and 260.8 degrees, respectively, and vertical field of view would increase up to 157.9 degrees.

2.1.1.3 Requirements of field of view for an ideal HMD

The results suggest that the field of view of about 200 degrees horizontally and 160 degrees vertically would be sufficient to provide users of HMD with an entire monocular field of view. In terms of binocular viewing, the horizontal field would increase up to about 260 degrees, which is much wider than the widest horizontal field of view (210 degrees) of the currently available consumer product [7].

2.1.2 Assessment in a more natural situation

The assessment described in § 2.1.1 was performed in unnatural situations where the participant's head was fixed, and the participants were asked to move their eyes to the peripheral side to the maximum. In daily life, the gaze is shifted toward peripheral visual targets. When this shift is large, the gaze is shifted by a combination of eye and head movements.[8] This combination is called eye-head coordination. Eye movement occurs during gaze shifting toward the far peripheral target in eye-head coordination, which is one of the largest eye movements in daily life and is smaller than the maximum eye movement. [9] Therefore, if eye movements that occur in daily life, rather than the maximum range of eye movement, are considered in the design of HMDs, it would be sufficient to measure the field of view required during eye-head coordination movements in the head-centred coordinate system. To address this issue, a sufficient field of view was measured during the gaze shift by eye-head coordination. [10]

2.1.2.1 Method

Figure 5 shows the apparatus used in the experiment, consisting of twenty 4 K panels (LG, 65EV5E) arranged cylindrically (Fig. 5A). The radius of the cylinder was 2.6 m and the participants sat on a chair positioned at its centre. Sixteen adults (eight men and eight women) with normal uncorrected vision participated in the experiment. Their mean age was 24.4, with a standard deviation (SD) of 4.1. Motion trackers (OptiTrack, Prime X13) were used (Fig. 5B) to obtain the position and direction of the participants' heads. The participants wore a bicycle helmet with retroreflective markers attached.

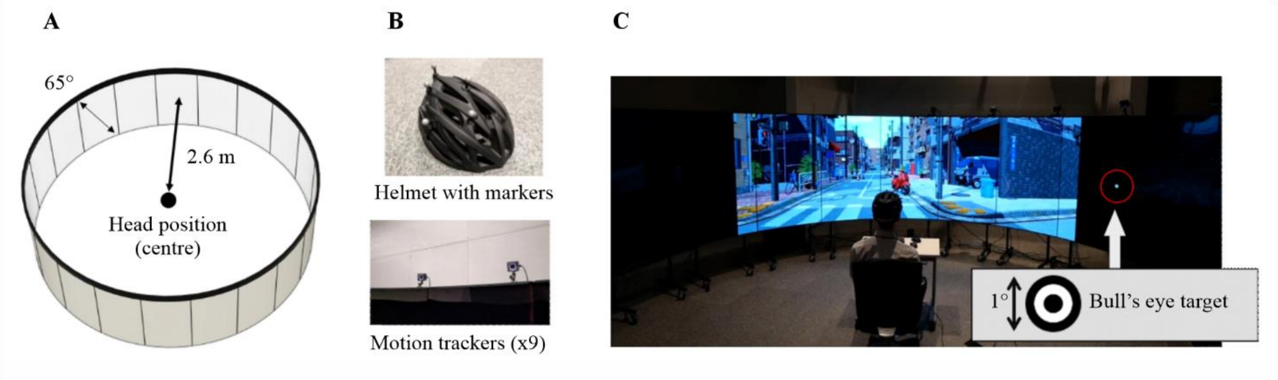
Computer graphics images were generated as visual stimuli by rendering three three-dimensional scenes – street, forest, and classroom – into 43 200 × 3 840 pixels images using Unity (Fig. 6). Unlimited (Fig. 7, left) and limited (Fig. 7, right) images were used in the experiment. An unlimited image was displayed on all 20 panels. A limited image was displayed in only a limited region. The horizontal size of the displayed region of the limited image was changed from trial to trial and ranged from 225 to 275 degrees in 11 steps. The displayed region moved horizontally along the direction of the participant's head, and the centre of the displayed region was the same as that of the head.

In the experimental trial, unlimited and limited versions of the images were presented sequentially and the order of presentation was randomised for each trial. The participants were asked to move their gaze toward a peripheral target and report which image was the limited version.

Before the presentation of the stimulus image, participants were asked to orient their heads in the initial direction. After participants oriented their heads, they initiated the trial by pressing a button. Five hundred milliseconds after the onset of the image, the bull's eye target was presented, accompanied by a beep sound signalling the onset of the target, and the position of the target was 60 degrees rightward from the initial direction. Participants were asked to shift their gaze toward the target by moving their eyes and head, and not moving their torso. The duration of the presentation of the visual stimulus image and target was 2 000 ms. This process was repeated twice for the limited and the unlimited versions of the images. After the presentation of the two images, participants reported their answers by pressing a button. Each participant performed 330 trials consisting of three scenes, eleven steps, and ten repetitions.

FIGURE 5

Apparatus and stimulus of an experiment assessing required field of view in semi-natural situation



Report BT.2506-05

FIGURE 6

Images used as visual stimuli

Report BT.2506-06

FIGURE 7

Unlimited and limited visual stimuli

Report BT.2506-07

2.1.2.2 Results

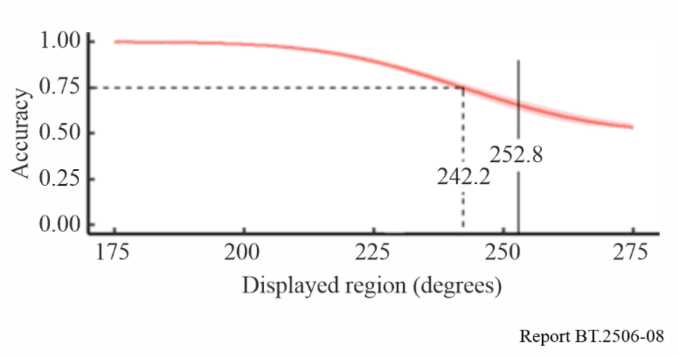
To estimate the threshold that presence and absence of the limitation could be distinguished, a model consisting of a cumulative distribution function (CDF) for normal distributions was assumed: $p_i(Correct|x_i, \mu, \sigma) = 1 - NormalCDF(x_i, \mu, \sigma)/2$, where the correct responses were obtained with probabilities p_i when given the size of the displayed region x_i. The mean μ, and SD σ, of the normal distribution were estimated by Bayesian estimation, using Markov chain Monte Carlo method.

Figure 8 shows the results of the experiment regarding the accuracy of the task against the size of the displayed region for a limited version of the stimulus. According to the estimated function, 242.2 degrees was estimated as the threshold with a 75% correct response ratio, which means that a displayed area larger than 242.2 degrees could not be distinguished from 360 degrees.

This number is smaller than the result of 252.8 degrees derived from the experiment performed with a fixed head and maximum eye movement, as mentioned in § 2.1.1.2. This difference seems reasonable because the eye movements that occur during eye-head coordination are expected to be smaller than the maximum eye movement range.

FIGURE 8

Results of an experiment assessing the required field of view in the semi-natural situation



2.1.2.3 Requirements of field of view for an ideal HMD for practical use

The results of the experiment suggest that if the HMD is designed based on the size of the practical eye movement that occurs in daily life, approximately 240 degrees of the horizontal field of view would be sufficient.

2.2 Spatial resolution

It is not necessary to fill the fields of view mentioned in § 2.1.1.3 or § 2.1.2.3 with pixels sharing the same uniform density as the photoreceptors in the human fovea because the spatial acuity of the human visual system is not homogenous. Since peripheral and central areas have qualitatively and quantitatively different spatiotemporal characteristics, and the peripheral area basically has a lower visual acuity than the central area [11], [12], an HMD should include pixels with the maximum density only for the area the fovea can scan, which is consistent with the range of eye movement. Outside of this area, a smaller number of pixels would be sufficient. Therefore, reducing pixel density with eccentricity without being noticeable to users is a challenge. A similar concept, known as “foveated rendering” [13], [14] has been well established, focusing on saving rendering precision for the peripheral areas within the field of view. Assessing the extent to which hardware and computational resources can be conserved without compromising the users' experience would prove beneficial in saving the resources for image rendering.

2.2.1 Experiment to assess spatial resolution in peripheral visual field in head-centred coordinate system

The characteristics of the spatial resolution in the peripheral visual field in the head-centred coordinate system, which contribute to the pixel density requirements of an ideal HMD, were measured in a scenario where the head was fixed and the eyes could move freely [15].

2.2.1.1 Method

Figure 9 shows the apparatus used in the experiment, which consisted of four curved displays. Each display had a curvature of R = 1 000 mm and the displays were arranged around a virtual sphere with a radius of 1 000 mm in four directions: upward, downward, nasal, and temporal. The displays were

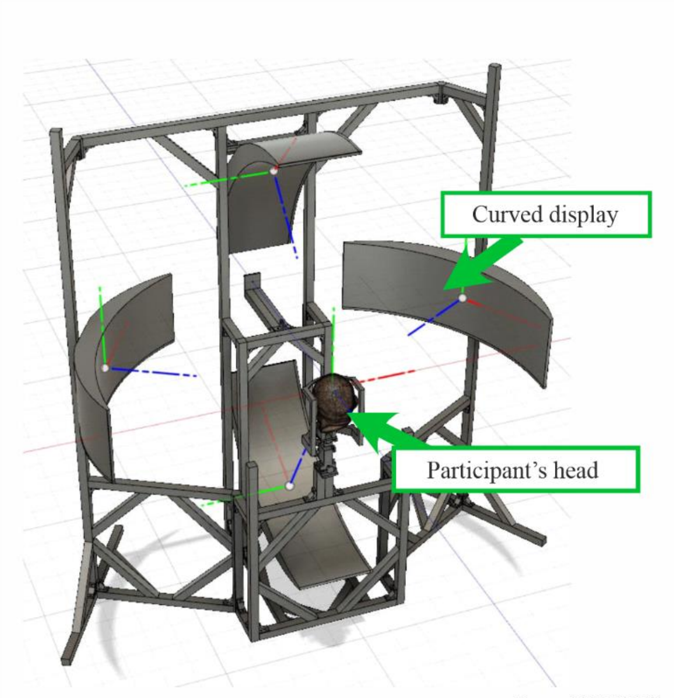
Odyssey Neo G9 49", Samsung, with 5 120 × 1 440 pixels. the screen width was 119.8 cm, which is equivalent to a visual angle of 68.8 degrees. The displayed area ranged from 10.6 to 79.4 degrees of eccentricity from the origin, which is in the front direction of the participants' head, hereafter referred to as FoV eccentricities, in contrast to retinal eccentricities, in the upward direction, and 25.6 to 94.4 degrees in the other three directions. Eight adults, five men and three women with ages ranging from 21 to 33 years (average 24.5) with normal uncorrected vision participated in the experiment. The left eye was shielded using eye patches, and the experiment was performed on the right eye. As in the experiment mentioned in § 2.1.1, the participants' heads were fixed at the centre of the device, and their eyes were allowed to move freely.

Two patterns of visual stimuli were used in the experiment, intact two-dimensional pink noise patterns and their modified versions. The modified version consisted of two regions: the intact pink noise region and the low-pass-filtered pink nose region, whose cut-off spatial frequencies were 1 and 8 cycles per degree (cpd) (Fig. 10). In the modified pattern, the filtered region was arranged on the peripheral side. The participants discriminated between the intact and modified patterns.

Each trial began with the participants pressing a button. At the beginning of the trial, a black solid line against a grey background was presented to indicate the position of the edge between the intact and filtered regions. The participants shifted their gaze toward the indicated position. Subsequently, the visual stimuli, the intact and modified patterns, were sequentially presented for 500 ms each with an inter-stimulus interval of 500 ms, and the order was randomised in each trial (Fig. 11). The participants were asked to report the stimulus that included the filtered region. Following an adaptive psychophysical method (QUEST rule [13]), the edge between the intact and filtered regions was moved from trial to trial. The trials were repeated 40 times in each session. The entire experiment was performed in separate sessions in four directions and two cut-off spatial frequencies. Using these procedures, the threshold eccentricities of the edges at which the patterns could be discriminated at 82% of the correct response ratio, farthest from the front, were determined.

FIGURE 9

Apparatus for an experiment assessing spatial resolution of peripheral visual field



Report BT.2506-09

FIGURE 10

Visual stimuli of the experiment assessing spatial resolution of peripheral visual field

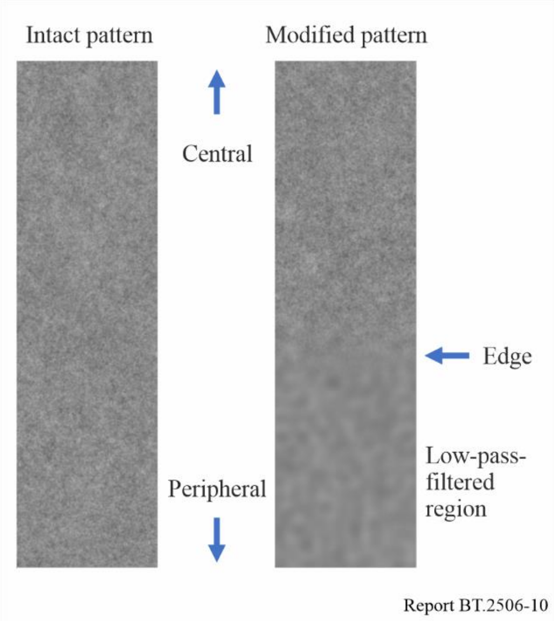
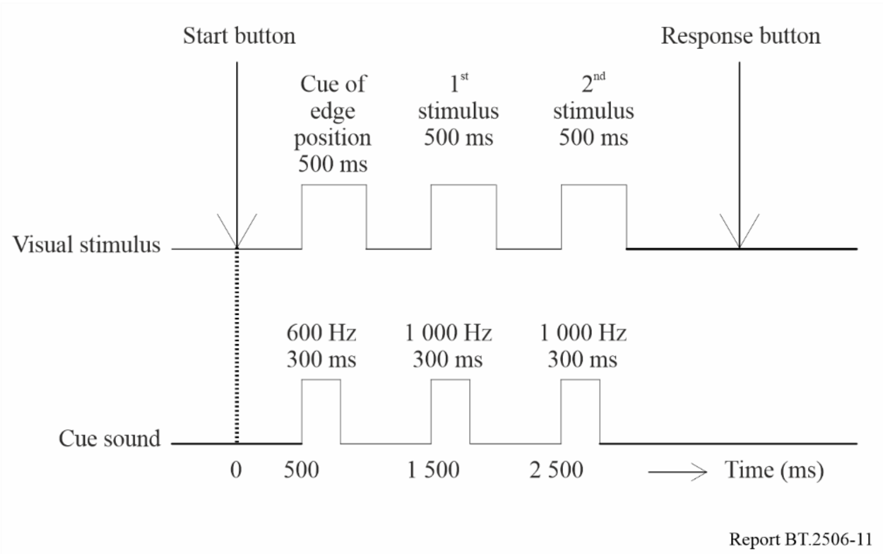


FIGURE 11

Time course of a trial



2.2.1.2 Results

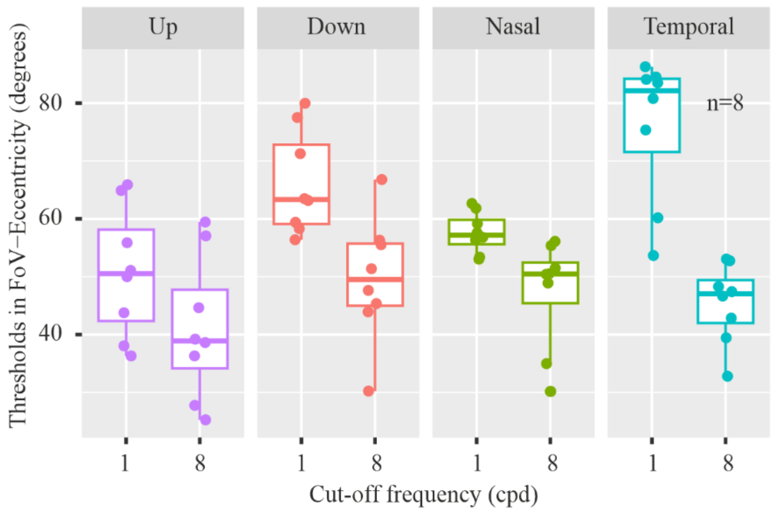
Thresholds inducing an 82% correct response ratio in FoV eccentricities are presented in Fig. 12. The medians of the thresholds for cut-off frequencies of 1 and 8 cpd were 50.5 and 38.9 degrees in the upward direction, 63.3 and 49.5 degrees in the downward direction, 57.2 and 50.5 degrees in the nasal direction, and 82.2 and 47.0 degrees in the temporal direction, respectively. This indicates that the visual area, for example, outside 50.5 degrees in the upward direction, was insensitive to spatial frequencies higher than 1 cpd. Therefore, it is suggested that in this area, the absence of spatial frequency components higher than 1 cpd would not deteriorate the visual experience and that pixel densities higher than 2 pixels per degree (ppd) would not contribute to inducing a sensation of high definition. Consequently, it is possible to save hardware and software resources without deteriorating the visual experience of users in areas beyond the eccentricities shown in Fig. 12.

The effect of the cut-off frequency was relatively more salient in the temporal and downward directions than in the upward and nasal directions. This contrast seems to be related to the visual vignette which may be caused by the anatomical structure of the face around an eyehole, such as the nose. Therefore, a design considering the results of this experiment would bring more benefits to saving resources on the temporal side than on the nasal side.

The area sensitive to higher spatial frequency components was narrower than the linear sum of the characteristics revealed in previous studies performed using the retinal coordinate system [11], [12] and the range of eye movement [16]. This implies that an assessment performed in the head-centred coordinate system is required for the HMD design.

FIGURE 12

Detection thresholds of low-pass filtered region in FoV eccentricity



Report BT.2506-12

2.2.2 Requirements of spatial resolution for an ideal HMD

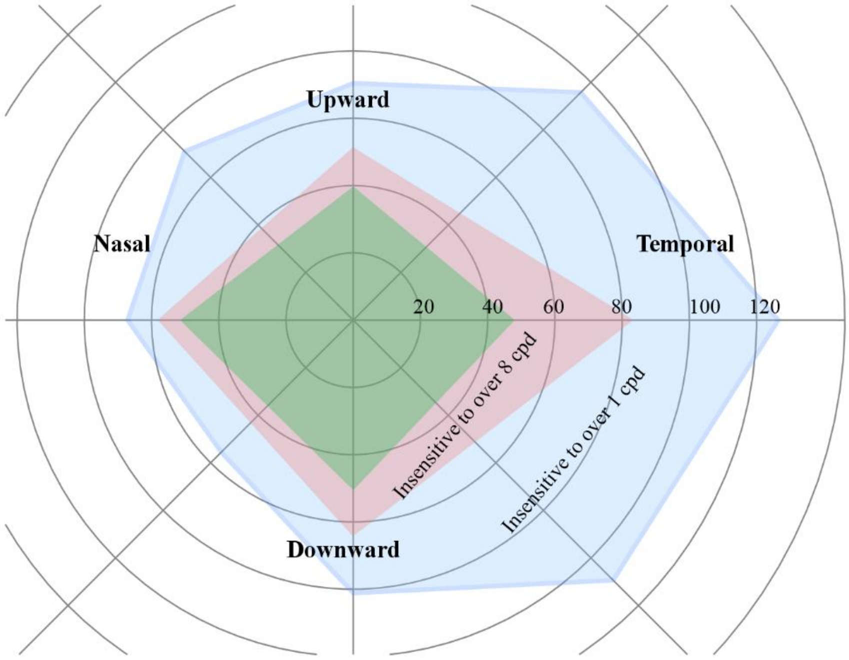
Figure 13 depicts the comparison between the field of view mentioned in § 2.1.1 and the spatial resolution data mentioned in § 2.2.1. Within an entire area with sensitivity to the point light, only a limited part of the area requires a high density of pixels. Since most of the green area, with a horizontal width of about 100 degrees and a vertical width of about 90 degrees, can be scanned by the fovea, it is reasonable to assume that this area should be filled with pixels of sufficiently high density, namely, 60 ppd. Because the red area, with a horizontal width of about 140 degrees and a vertical width of about 110 degrees, beyond the green area is not sensitive to spatial frequencies higher than 8 cpd, a pixel density of over 16 ppd is not required. Similarly, the blue area beyond the red area does not require more than 2 ppd.

In the experiment mentioned in § 2.2.1, two different spatial frequencies, 1 and 8 cpd, were used as the experimental parameters. These values were selected arbitrarily, and assessments with other values such as 2 and 4 cpd are also expected to be useful.

As of 2024, HMDs that combine two display panels with different spatial resolutions are available in the market [17]. If the number of steps of spatial resolution is increased from two to more than two, it is possible to save more resources, and the spatial efficiency will be improved. However, the optical design would inevitably become more complicated, and the vulnerability to noise in assembling parts and misalignment in wearing would also be unacceptable.

FIGURE 13

Comparison of field of view and spatial resolution



Report BT.2506-13

2.3 Spatial aspects of requirements for an ideal HMD

Since most prior research has only focused on the characteristics of the human visual system without focusing on the requirements for image-displaying systems, they have performed assessments mainly on a retinal coordinate system, and the required specifications for an ideal HMD are not necessarily fully understood. Therefore, further findings based on a head-centred coordinate system can help to address this issue.

References

[1] H. M. Traquair, "The Normal Field of Vision," in *An Introduction to Clinical Perimetry*, 3rd ed., London: Henry Kimpton, 1938, pp. 4-5.

[2] R. H. Spector, "Visual Fields," in *Clinical Methods: The History, Physical, and Laboratory Examinations*, 3rd ed., H. K. Walker, W. D. Hall, and J. W. Hurst, Eds. Boston: Butterworths, 1990, ch. 116, pp. 565-572.

[3] L. Racette, M. Fischer, H. Bebie, G. Holló, C. A. Johnson, and C. Matsumoto, "What is Perimetry?" in *Visual Field Digest: A Guide to Perimetry and the Octopus Perimeter*, 7th ed. Könitz, Switzerland: Haag-Streit AG, 2018, ch. 2, pp. 7-23.

[4] S. Hanif, F. J. Rowe, and A. R. O'Connor, "A Comparative Review of Methods to Record Ocular Rotations," Br. Ir. Orthopt. J., vol. 6, p. 47, Aug. 2009, doi: 10.22599/bioj.8.

[5] M. Harasawa, Y. Miyashita, and K. Komine, "Perimetry on Head-centered Coordinate System for Requirements of Head-mounted Display," 28th International Display Workshops, VHF5-4L, 2021.

[6] A. Pentland, "Maximum Likelihood Estimation: The best PEST," *Percept. Psychophys.*, vol. 28, no. 4, pp. 377–379, Jul. 1980, doi: 10.3758/BF03204398.

[7] https://www.starvr.com/product/

[8] D. Guitton, and M. Volle, “Gaze control in humans: eye-head coordination during orienting movements to targets within and beyond the oculomotor range,” Journal of Neurophysiology, vol. 48, no. 3, pp. 427-459, 1987.

[9] S. Hanif, F. J. Rowe, and A. R. O’connor, “A comparative review of methods to record ocular rotations,” British and Irish Orthoptic Journal, vol. 6, no. 47, 2009.

[10] Y. Miyashita, M. Harasawa, K. Hara, Y. Sawahata, and K. Komine, “Estimation of Required Horizontal FoV for Ideal HMD Utilizing Vignetting under Practical Range of Eye Displacement,” IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Shanghai, China, pp. 691-692, 2023.

[11] J. Rovamo, V. Virsu, and R. Näsänen, “Cortical Magnification Factor Predicts the Photopic Contrast Sensitivity of Peripheral Vision,” Nature, vol. 271, 5640, pp. 54-56, 1978.

[12] D. Regan, and K. I. Beverley, “Visual fields described by contrast sensitivity, by acuity, and by relative sensitivity to different orientations,” Investigative Ophthalmology & Visual Science, vol. 24, pp. 754–759, 1983.

[13] B. Guenter, M. Finch, S. Drucker, D. Tan, and J. Snyder, “Foveated 3D graphics.” ACM Transactions on Graphics, vol. 31, issue 6, Article No. 164, pp. 1-10, 2012.

[14] G. Tan, Y.-H. Lee, T. Zhan, J. Yang, S. Liu, D. Zhao, et al., “Foveated imaging for near-eye displays,” Opt Express, vol. 26, issue 19, pp. 25076-25085, 2018.

[15] M. Harasawa, Y. Miyashita, and K. Komine, “Required Specifications on Spatial Aspects for Ideal HMD Estimated on Head-Centered Coordinate System,” 30th International Display Workshops, VHF4-1L, 2023.

[16] A. B. Watson, and D. G. Pelli, “QUEST: A Bayesian adaptive psychometric method,” Perception & Psychophysics, vol. 33, pp. 113–120, 1983.

[17] https://varjo.com/products/xr-4/