Objective and Subjective Quality Assessments of 3D Videos

A thesis submitted in fulfilment of the requirements for the degree of Doctor of Philosophy

Siu Ming Choy

BEng (ElectEng) – The Hong Kong University of Science and Technology

MSc (ElectEng) – The Hong Kong University of Science and Technology

PGDE – The Chinese University of Hong Kong

School of Engineering

College of Science, Technology, Engineering and Maths

RMIT University

June 2021

Declaration

I certify that except where due acknowledgement has been made, the work is that of the author alone; the work has not been submitted previously, in whole or in part, to qualify for any other academic award; the content of the thesis is the result of work which has been carried out since the official commencement date of the approved research program; any editorial work, paid or unpaid, carried out by a third party is acknowledged; and, ethics procedures and guidelines have been followed.

Siu Ming Choy

School of Engineering, RMIT University

21 June 2021

Acknowledgements

I would like to offer my enduring gratitude to my supervisors, Dr Eva Cheng and Professor Ian Burnett, for their professional guidance and constant support throughout my PhD study. I would especially like to thank Dr Eva Cheng for her motivation and encouragement, valuable comments and suggestions, and novel perspective on my research.

I offer my sincere thanks to Dr Richardt Wilkinson and Professor Michael Austin for their continuous guidance and supervision during my research. I am deeply thankful for their support after my original supervisors left the University.

I am very grateful for having met Dr Heiko Rudolph in Hong Kong and having known him for six years. He conducted a series of multidisciplinary discussions with me that truly opened my mind and changed my perspective, especially at the beginning of my study. I also express my appreciation to Dr Stuart Perry and Mr Ben Simons from the University of Technology Sydney, who conducted several insightful discussions with me and provided me with equipment in the Perceptual Imaging Laboratory and Data Arena during my research in Sydney.

I would like to thank Vocational Training Council in Hong Kong for giving me this opportunity for the application of this part-time mode PhD study. I would also extend my sincere gratitude to my supervisors and colleagues in Vocational Training Council, including Ir Jimmy Tang, Dr Simon Shum, Dr Tsz-Shan Lam, Dr Yu-Hing Shum, Mr Kwok-Leung Chu, Ms Nicole Lung, Ir Chi-Yung Shiu and Ir Chi-Keung Yim, for their encouragement and valuable help with the complicated administrative arrangements for my applications for research study in Australia.

I would also like to thank my schoolmates, including Louie Chan, May Tam, Catter To, Murthy Lai, Billy Chow, Anthony Lai, and Kenneth Pak, with whom they have shared research ideas and pursued studies.

My heartfelt thanks go to my dear family. They have deep knowledge of my happiness and my struggles. My father always encourages me to pursue my studies. My brother always gives me support, advice and his unique perspective on journal publication and thesis writing. I owe them my deepest gratitude.

Finally, I give my deepest gratitude to my wife, Joyce, for her constant support and inestimable love. She has been my companion and anchor and has constantly encouraged me to pursue my studies. She is always my best listener and biggest supporter. I also give my thanks to my daughter, Muk-Yan. Her smile never fails to inspire me to work hard.

Table of Contents

List of Figures

List of Tables

List of Abbreviations

Abbreviation	Meaning
3DTV	3D Television
ACR	Absolute Category Rating
AOI	Area Of Interest
CI	Confidence Interval
DCR	Degradation Category Rating
DERS	Depth Estimation Reference Software
DSCQE	Double Stimulus Continuous Quality Evaluation
DSIS	Double Stimulus Impairment Scale
EEG	ElectroEncephaloGraphy
FDM	Fixation Density Map
fMRI	functional Magnetic Resonance Imaging
FOV	Field Of View
fps	frames per second
Full-HD	Full-High Definition
HEVC	High Efficiency Video Coding
HMD	Head Mounted Display
HVS	Human Visual System
IQA	Image Quality Assessment
ITU	International Telecommunication Union
JND	Just Noticeable Difference
MOS	Mean Opinion Score

Abbreviation	Meaning
MPEG	Moving Picture Experts Group
MR	Mixed Resolution
MVC	Multi-view Video Coding
PC	Pair Comparison
PSNR	Peak Signal-to-Noise Ratio
QoE	Quality of Experience
QoS	Quality of Service
QP	Quantization Parameter
RMSE	Root Mean Square Error
S3D	Stereoscopic Three-Dimensional
SI	Spatial Perceptual Information
SSCQE	Single Stimulus Continuous Quality Evaluation
SSIM	Structural Similarity
SSIS	Single Stimulus Impairment Scale
SSQ	Simulator Sickness Questionnaire
TI	Temporal Perceptual Information
UHD	Ultra-High Definition
VAM	Visual Attention Model
VIMS	Visually Induced Motion Sickness
VQA	Video Quality Assessment
VQEG	Video Quality Experts Group
VSRS	View Synthesis Reference Software

Abstract

Three-dimensional (3D) image viewing is rapidly increasing the popularity of applications such as 3D videos, movies, computer games, and Virtual Reality (VR) environments. Several techniques related to 3D video have been developed, including stereoscopic video, multi-view video, autostereoscopic video and holography. This study focuses on stereoscopic 3D (S3D) video, which is used most widely in the movie industry and in 3D television (3DTV) broadcasting. This broadcasting services have been introduced in several countries in recent years, and video service providers have also offered S3D video services over the Internet. To cope with the market's increasing need for high-quality 3D images, quality assessment of S3D video has been identified as an important area for research and development.

This research proposes novel methods for both objective and subjective quality assessments of S3D video, with a view to extending these methods to the development of S3D quality metrics and multi-view 3D video technologies. The novel methods were verified with software simulations in objective quality assessments and experimental test validations in subjective quality assessments. Firstly, this study conducted objective and subjective assessments to investigate differences in the viewers' Quality of Experience (QoE) between two-dimensional (2D) and S3D videos. Visually Induced Motion Sickness (VIMS) and visual attention were evaluated and verified against previous findings. Secondly, the study was extended to the assessment of the quality of experience of viewers who watched S3D videos in three viewing environments: a flat 3D screen, a panoramic screen and a VR headset. This experiment was performed to identify the effects of VIMS and to analyse 3D fatigue in the three viewing environments. Thirdly, eye-tracking experiments were conducted to obtain eye-gaze data and to develop human saliency maps for the investigation of human visual attention. Existing

human saliency maps and saliency prediction models are compared for benchmarking and validation. These results were evaluated to determine whether visual attention obtained from eye-tracking methods contribute to QoE assessments of S3D videos. Finally, the application of the objective quality metrics and subjective evaluation approaches is proposed as a practical use-case in video coding evaluation.

Comparing the objective and subjective quality assessments between 2D and S3D videos, the findings show that S3D videos create more fatigue for the participants than the 2D video sequences. Also, there is more eye blinking movement when viewing 2D videos compared to S3D videos. When viewers watch S3D videos in three viewing environments, the findings show that the viewers who used a VR device to view the stereoscopic video sequences resulted in higher Simulator Sickness Questionnaire (SSQ) scores. For the panoramic screen, the participants reported the lowest SSQ scores and the highest enjoyment ratings. The difference in results experienced when viewing the same S3D video sequence on different screens reveal that the projection screen is an essential factor that influences the level of visual fatigue, QoE and VIMS. The content of the video sequence and the projection screen used are also key factors that affect the enjoyment rating of the S3D videos. In addition, it is found that eye-tracking is a key method to obtain the visual attention of viewers. Existing saliency prediction models are not accurate enough to predict an actual visual attention model, especially for S3D video sequences. The application of the objective quality metrics and the subjective evaluation in a proposed video coding use-case of S3D video sequences show that a perceptual quality similar to that of uncompressed videos can be provided when the 3D effects are moderate and scenes are stable.

The key contribution of this thesis is in providing more reliable and accurate objective and subjective quality assessments for S3D video, including the QoE, visual fatigue and visual attention from the S3D video in different viewing environments. Finally, this thesis is concluded discussing the recommended future research work in extending into multi-view S3D video technology.

Chapter 1 Introduction

The popularity of three-dimensional (3D) image viewing is rapidly increasing, and it has been extensively used in the production of 3D videos, movies, computer games and immersive multimedia experiences such as Virtual Reality (VR). Several techniques related to 3D video have been invented, including stereoscopic video, multi-view video, autostereoscopic video and holography. These 3D techniques involve the implementation of stereo vision techniques, in which two cameras are used to capture images. Stereoscopic video technology requires greater accuracy and the capacity to deal with the required (real-time) data rate [1]. Moreover, the high data rate can be used to present high resolution images such as medical and precision images [2].

This study focuses on stereoscopic 3D (S3D) videos, which is used most widely in the movie industry and in 3D television (3DTV) broadcasting. 3DTV broadcasting services have been introduced in several countries in recent years, and video service providers have offered S3D video services over the Internet. However, S3D videos contain a large amount of data, which require sufficient storage capacity and bandwidth for transmission. For instance, the data transmission rate of an uncompressed S3D video may be several Gb/s [3]. To cope with the market's increasing need for high quality 3D images in practical applications, quality assessment of S3D has thus been identified as a key area for research and development.

1.1 Motivation

In recent years, 3D video technologies have become popular in commercial markets for both consumer and industry applications. As the technologies in this field have matured, it has been extended to industries such as education, training, entertainment and medical imaging. Due to the breadth of potential applications, there is thus a need to understand and measure viewers' Quality of Experience (QoE) with regard to 3D content.

The most accurate method to assess video quality is a subjective assessment by viewers. However, it is not easily possible to conduct extensive subjective tests for each user due to time, cost, and the breadth of content and applications to be considered. An alternative approach is to develop a human visual system model with objective quality metrics and conduct a series of QoE experiments to predict a model to assess visual quality. Following this approach, many quality assessment models and techniques of two-dimensional (2D) images and videos have been proposed in recent years [4-8]. Objective and subjective quality assessments of 3D video are ongoing areas of research, including the launch of the 3DTV working group for the standardisation of 3D video format and coding and multi-view 3D video coding [9-14], and providing the guidelines of subjective methods for the assessment of 3D videos [15-18] as part of standardisation activities by two international organisations: Moving Picture Experts Group (MPEG) and International Telecommunication Union (ITU). Large-scale perceptual experiments have been conducted according to the standardisation activities to achieve better viewers' experience when watching 3D video.

These challenges have thus motivated the investigation in this thesis of novel objective and subjective assessments of 3D video to allow users to experience improved QoE with 3D videos.

1.2 Aims and Objectives

The aim of this thesis is to investigate and propose novel techniques to enhance current objective and subjective quality evaluations of S3D video. The proposed approaches were evaluated by designing metrics, using software simulations for objective assessments, and implementing experimental validations for subjective assessments.

1.2.1 Research questions

The thesis investigates the following three research questions:

Research Question 1: How to assess the visual fatigue of S3D video in both objective and subjective quality assessments?

- **Research Question 1(a):** How to assess the viewers' perceptual experience and visual attention when watching S3D video?
- **Research Question 1(b):** How to evaluate the perceptual experience and visual attention of viewers whilst watching S3D video in different viewing environments?

Research Question 2: How to assess the visual attention of viewers watching S3D video in relation to the quality of experience?

Research Question 3: How are objective metrics and the quality of experience of viewers affected when implementing a video sequencing model for video coding as a practical application?

The following research study areas and methodology are used to address the research questions and to achieve the objectives of this thesis.

There are three main areas of study in this thesis work:

- Investigate Visually Induced Motion Sickness (VIMS) with Electroencephalography (EEG) signals to correlate 3D fatigue levels with VIMS in two different situations: 1) Comparing 2D and S3D videos using a flat 3D screen; 2) Comparing S3D videos in three different viewing environments.
- Investigate viewers' visual attention using eye-tracking to predict S3D visual attention and develop a human saliency map model.
- Apply the objective quality metrics and subjective evaluation approaches proposed in this thesis work to video coding applications as a practical use-case.

As high-quality S3D video content has been well developed in recent years, the use of objective and subjective assessments was proposed to assess Visually Induced Motion Sickness (VIMS) [19-22] in participants who view S3D videos [23, 24]. Researchers have compared various subjective assessment methods suggested by ITU-R BT2021.1 standard [25] for S3D Video Quality Assessment (VQA) [26-28]. Chen et al. [29] verified five subjective assessment methods and proposed that the Degradation Category Rating (DCR) method was the most stable method for 3D fatigue assessment; however, the assessment of visual fatigue has not yet been fully investigated. In addition, few studies have investigated VIMS from Electroencephalography (EEG) signals using a high-quality video stimulus.

The work in this thesis investigates potential causes of 3D fatigue from VIMS and extends this to QoE assessments of viewers who watched S3D videos in three different viewing environments: a flat 3D projection screen, a panoramic screen and a virtual reality (VR) headset. Most previous studies focused on the evaluation of the quality of the multimedia experience in a viewing environment with a single projection device; few studies have sought to evaluate the QoE in various environments with more than one projection device. Therefore, this study was expanded to identify the effects of VIMS and to analyse 3D fatigue in the three different viewing environments.

Further, researchers are currently focusing on multi-view 3D video technology [30] and the depth map rendering of two views [31-34] to be implemented for the evaluation of 3D multi-view video [35]. For subjective assessments, few models yet exist to assess visual attention in relation to S3D video. Therefore, this thesis investigates viewers' visual attention and improves prediction of S3D visual attention through eye-tracking experiments. The eye-gaze data was investigated to develop a human saliency map model. This model was compared with existing human saliency maps and commonly used saliency prediction models for analysis to evaluate how visual attention from eye-tracking can contribute to the QoE assessment of S3D video.

1.3 Thesis Contributions

The main contributions of this thesis and associated publications are listed as follows:

- Propose objective and subjective quality evaluations to investigate differences in the viewers' QoE and correlate 3D fatigue levels in two different approaches: 1) Comparing 2D and S3D videos viewed with a flat 3D screen; 2) Comparing the viewing of S3D videos in different viewing environments using EEG signals and associated algorithms, enjoyment levels and the use of Simulator Sickness Questionnaire (SSQ) to evaluate VIMS and visual fatigue (presented in Chapter 3 and Chapter 4 with publications [36, 37]).
- Apply objective quality metrics and subjective evaluations to measure the difference in the quality of two stereoscopic views in a proposed video coding approach as a

practical use-case for S3D video evaluation (presented in Chapter 5 with publication [38]).

- Develop a human saliency map model for the human visual attention of S3D video to correlate with eye-gaze data by using eye-tracking equipment. This model is compared with existing human saliency maps and saliency prediction models from other researchers for benchmarking and validating these results. Results were evaluated to determine whether visual attention from the eye-tracking method contributes to the QoE assessment of S3D videos (presented in Chapter 6).

1.3.1 List of publications

- S.-M. Choy, E. Cheng, R. H. Wilkinson, I. Burnett and M. W. Austin, “Quality of experience comparison of stereoscopic 3D videos in different projection devices: Flat screen, panoramic screen and virtual reality headset,” *IEEE Access*, vol. 9, pp. 9584-9594, 2021.
- S.-M. Choy, E. Cheng and I. Burnett, “Hybrid sequencing of uncompressed and compressed 3D stereoscopic video: A preliminary quality evaluation,” In *The International Conference on Electrical Engineering (ICEE)*, Okinawa, Japan, 2016, pp. 1-5.
- S.-M. Choy, K.-H. Chiu, E. Cheng and I. Burnett, “3D fatigue from stereoscopic 3D video displays: Comparing objective and subjective tests using electroencephalography,” In *TENCON IEEE Region 10 Conference*, 2015, pp. 1-4.

1.4 Thesis Structure

Chapter 2 reviews the literature relating to objective and subjective quality assessments in 3D video technologies. First, the background of human S3D perception is presented, then existing approaches to objective and subjective quality assessments are described. Finally, the evaluation of QoE for both objective metrics and subjective evaluation of S3D video are reviewed.

Chapter 3 investigates 3D visual fatigue to identify the effects of VIMS, using both objective and subjective methods to compare the QoE of participants who watch both 2D and 3D videos of the same stimulus content. EEG measurements and a survey analysis using the Simulator Sickness Questionnaire (SSQ) of Kennedy et al. [39] were adopted to consider visual attention. This approach involves the measurement of EEG signals using the Absolute Category Rating (ACR) method to assess 3D fatigue.

Chapter 4 explores the effect of different S3D viewing environments. The QoE is examined when viewing S3D videos comparing three viewing environments: a flat screen, a panoramic screen and a virtual reality headset. EEG detection, SSQ survey analysis and the participants' enjoyment ratings of viewing 3D videos in the various environments were used to investigate the relationship between 3D visual fatigue and the effects of VIMS with the respective devices.

Chapter 5 applies the objective metrics and subjective evaluation approaches to a video coding application as a practical use-case. Hybrid sequencing of uncompressed and compressed content within a single S3D video is proposed as a coding approach; however, such sequencing may affect the correlation between the left and right views required for depth perception in the stereoscopic videos, which may reduce the viewers' QoE. Thus, the hybrid sequencing of stereoscopic video sequences was investigated with both objective and subjective quality evaluations.

Chapter 6 investigates how the human visual attention of S3D videos correlate to the QoE assessments. Existing saliency prediction models are studied, which include factors such as the visual discomfort level, the motion acceleration and the sparsity of the salient regions to ensure a high level of correlation with the eye-gaze data. The model is evaluated with eye-tracking experiments in which the participants viewed a series of S3D video sequences to obtain eye-gaze data and develop human saliency maps for comparison and analysis.

The thesis is concluded in Chapter 7 , summarising findings and contributions as well as discussing future work. In future, the approaches in this thesis may extend to the development of objective metrics and subjective evaluations of multi-view 3D video technologies.

Chapter 2 Literature Review

This chapter reviews the literature on existing objective and subjective evaluations of Stereoscopic 3D Video (S3D), quality of experience of S3D, and eye-tracking techniques and analysis to develop human saliency maps and estimate saliency maps as applied in this research.

2.1 Background

2.1.1 Human stereoscopic vision

Stereopsis is the formal term for depth perception; that is, the human visual perception of objects at various distances along one line of sight as the brain registers the 3D shapes and forms the visual representations [40]. The visual mechanism for depth sensing relies on the input from both eyes.

The binocular geometry of stereopsis is shown in Figure 2.1 [41]. The human eyes create two different views due to the difference in their positions, where information from both eyes is important in the creation of stereoscopic vision. Physiologically, the light received by both eyes falls on the fovea – the back part of the eye with the highest acuity – when the eyes are fixed on the binocular point P. The point Q casts the image away in α degrees in one eye's fovea and β degrees from the other eye's fovea, and the binocular disparity is $(\beta-\alpha)$ degrees. When looking at a new object at a different distance, the point of fixation is altered. The ciliary muscles of both eyes move at the same time, either inward or outward, so that the image of the new object can be centred in the fovea. If the new object is closer, the eyes move inward in a process called convergence, whereas if the new object is further from the eyes, the eyes move outward in a process called divergence.

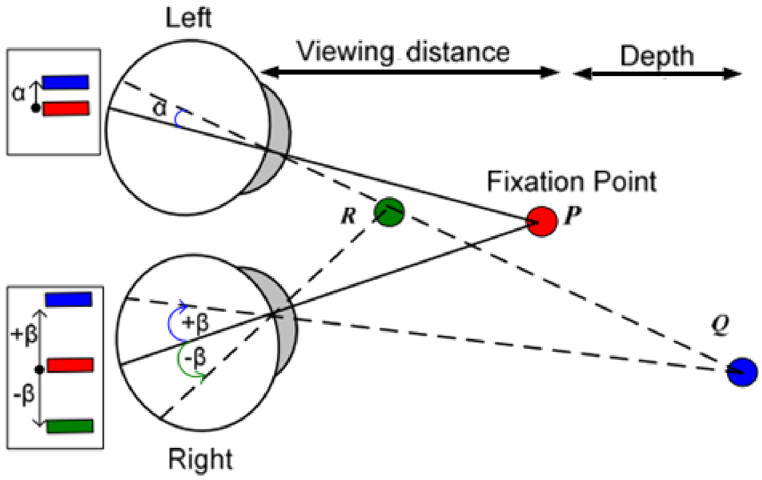


Figure 2.1: Binocular geometry of stereopsis [41].

Other than the binocular geometry of stereopsis, each eye has a single-lens optical system to focus on an object and form an image on the retina. Accommodation is defined as the process where the eye changes its optical power by the adjustment of the eye lens deformed by the ciliary muscles to maintain a clear image [42, 43]. Figure 2.2 (a)-(b) show the accommodation cues when the eye lens adjusts the focal length to focus on objects at different distances. Accommodation is irrelevant to stereoscopic vision, where each single eye is required to accommodate when viewing an object on an S3D screen. Each single eye has a depth of focus so it does not depend on stereoscopic vision. Vergence is a binocular cue to rotate both eyes at the same time in opposite directions around their vertical axis to fixate on the same point of an object [43], as shown in Figure 2.2 (c)-(d) at various distances. Figure 2.3 shows the accommodation and vergence in an S3D screen. Figure 2.3 (a) shows the differences of accommodation and vergence cues [42] i.e., how the eyes accommodate the S3D screen but the eyes also rotate to fix the apparent image at a different position. In Figure 2.3 (b), both eyes of the viewer converge at a distance away from the screen at the depth of an actual 3D point. Also, the S3D screen provides two different light paths perceived in each eye filtered by S3D glasses. Therefore, each eye can observe a different 2D screen point [43]. The accommodation cues match the 2D screen point while the vergence cues approximately match the actual 3D point [44, 45], and the two depth details are contradictory. Therefore, there is a conflict when combining accommodation and vergence cues [44]. Wilkins [46] and Hoffman et al. [47] revealed that one of the possible causes for eye fatigue is this mismatch of accommodation and vergence.

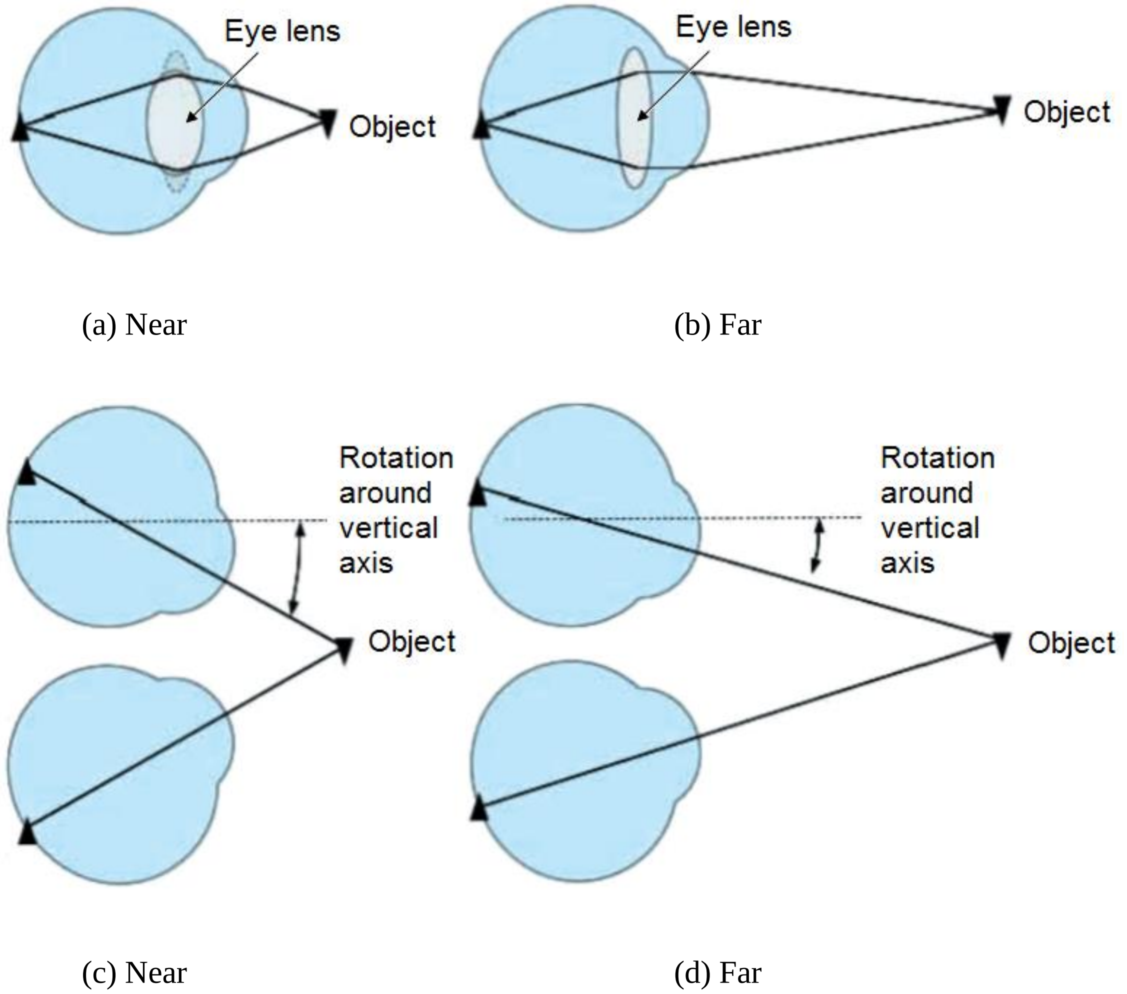
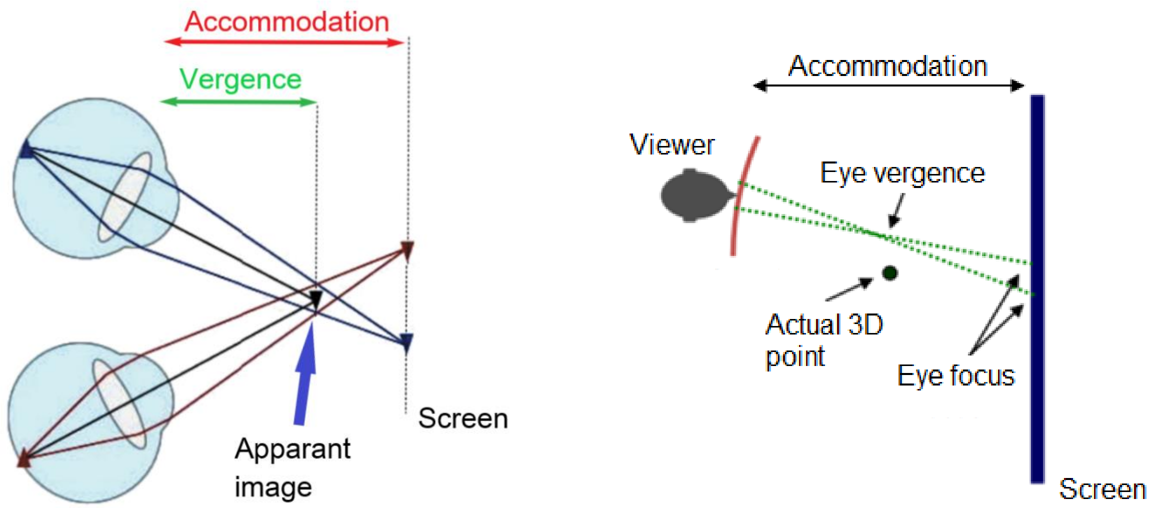


Figure 2.2: Accommodation (a)-(b) and vergence (c)-(d) cues when viewing an object at various distances [42].



(a) (b)

Figure 2.3: Accommodation and vergence in an S3D screen (a) Differences between the two cues [42], (b) Vergence cues approximately match the actual 3D point [43].

Taking advantage of human binocular vision, S3D create the illusion of depth in moving images by displaying different images for each eye. S3D simulates generalised stereo vision, where the perception can vary between different people. Thus, there is a need to understand S3D QoE.

2.1.2 Stereoscopic video

Stereoscopic video is used in various media such as movies, 3DTV and games. Three-dimensional (3D) gaming is increasing in popularity as several major companies, such as Cubix, Whimsy Games and Argentics, offer 3D gaming and video watching experiences [48]. In 2010, S3D was among the best-selling categories of mass consumer products. Other application fields in 3D technologies include 3D cinema, entertainment in theme parks, cultural heritage and medical surgery. Recently, the development of VR headsets such as Facebook Oculus Rift [49], HTC VIVE [50] and Sony PlayStation VR [51] include stereoscopic support as one of the features of their products. Multi-view video is also an extension of S3D video for future development [52].

Companies who manufacture 3D consumer products and content producers are challenged to create a natural experience along with eye comfort for the audience [53]. Several tools and techniques are used for watching stereoscopic videos, and audiences in a cinema or a game room cannot perceive the 3D effect unless they wear the proper equipment for watching 3D. Special glasses are used to view S3D videos, including active lenses, passive polarised lenses and red-blue anaglyph 3D glasses, as shown in Figure 2.4 [54]. The display technology depends on the optical filter which filters the correct image for the left and right eyes and directs the light to each eye. In particular, passive glasses do not require electronics and batteries. These optical filters sort the correct image for each eye based on the polarisation of the light. In contrast, active shutter technology does require powered glasses and is generally used for LED and plasma televisions [55]. However, autostereoscopic display technology does not require the use of glasses to achieve the 3D effect. This technology likewise depends on an optical filter that divides the incoming light and directs the correct part of the light to each eye. These 3D images then arise from the incoherence of the inputs to the left and right sides. However, regardless of the stereoscopic rendering technique, human visual perception can conflict with the limitations of display devices or content generation that can potentially cause viewer discomfort.

Figure 2.4: Glasses to view S3D videos (left to right): red-blue anaglyph 3D glasses, passive polarised lenses, active lenses [54].

Compared with 2DTV, S3D can cause more stress to the eyes and can be uncomfortable for prolonged watching for some viewers. However, the main benefit of 3DTV is greater depth perception, which leads to a better perception of sharper images and a sense of presence and naturalness. Surveys have shown that people are more likely to watch S3D images than 2D images [56]. In gaming, stereoscopy increases immersion and spatial presence. There are various advantages of stereoscopic videos and image systems. For example, in interactive environments, various actions are aided by stereo vision, such as throwing, catching or hitting a ball; driving or parking a car, and performing medical surgery [57] to improve the accuracy of depth judgment.

2.1.3 Visual fatigue and visual discomfort in stereoscopic video

Visual fatigue and visual discomfort are two different terms used in both psychological and subjective perception when viewers watch S3D videos.

Visual fatigue is caused by the accumulation of excessive perceived visual effects, and generally disappears after taking a rest for a certain period of time. The severity of the visual fatigue depends on the intensity and temporal properties such as duration and appearance time [58]. It is a common phenomenon in adults and it might have a correlation with age and the decline of mobility, and increase in likelihood of falls. The fatigue can impact the peripheral and central parts of the brain. Objective fatigue is referred to as the decline of operational performance and attention of the viewer [59]. Subjective fatigue depends on the age, sex, mental state and health condition of the viewer [60]. Both the subjective and objective fatigue of the viewer can be measured with EEG as it is non-invasive to the participants [61]: the amplitude and wavelength of the EEG signals denotes the participants' level of brain fatigue. Detailed EEG literature will be discussed further in Section 2.4.4.

Visual discomfort is a self-assessed physical and/or psychological state signalling a degree of annoyance when the viewer takes part in a visual task, or perceives negative sensation related to the task [58]. Generally, visual discomfort has a shorter rise and fall time than visual fatigue. Visual discomfort disappears when the visual task is either interrupted by other tasks or the task is completed. The main causes of visual discomfort are the Vergence-Accommodation conflict, disparity distribution, binocular distortion and the motion of the S3D video [58]. Visual discomfort can be measured by the visual evaluation reported by participants [62].

Generally, both visual fatigue and visual comfort can be assessed via questionnaires to collect the symptoms that have occurred [62].

2.1.4 Visual attention in relation to quality of experience (QoE)

Visual attention involves perception mechanisms in the selection of the perceived stimuli in the human visual system when the participant views 3D content. For instance, the object of interest perceived from the 3D content is related to the depth-of-focus; the duration and the content to be perceived from the depth of 3D objects are relevant to the vergence load; and, salient objects are intended to be more easily perceived in the central vision area. Nevertheless, salient objects perceived when participants view an S3D video may be related to visual fatigue and visual discomfort [62]. Lambooij et al. [63] identified that visual discomfort is the counterpart of visual fatigue from subjective evaluations, where visual discomfort can reflect some functions of QoE. Sohn et al. [64] and Lee et al. [65] proposed models to investigate the correlation between 3D visual attention and visual discomfort. Each model contains a database providing 3D content with different colours, motion and depth for analysis. Also, numerous researchers have adopted eye-tracking techniques and evaluation indicators of saliency map models to assess the visual attention of participants, which will be further discussed in Section 2.5.

2.2 Overview of Objective and Subjective Evaluations of 3D Video

Image quality assessment is an approach to measure the quality of an image by combining the effects of image distortion, quantification and accumulation into a single score, and it can be either subjective or objective. Figure 2.5 shows various methods for the assessment of image quality. Objective quality assessment includes three categories: Full-reference [66] such as Peak Signal-to-Noise Ratio (PSNR) [67, 68] and Structural Similarity Index Measure (SSIM) [69], Reduced-reference [66], and No-reference [66] quality assessment methods. Subjective quality assessment includes two categories. The first one is called Single Stimulus method such as Absolute Category Rating (ACR) [70], SSCQE (Single Stimulus Continuous Quality Evaluation) [70] and SSIS (Single Stimulus Impairment Scale) [70] methods, whilst the second one is called Double Stimulus method including Degradation Category Rating (DCR) [7], DSCQE (Double Stimulus Continuous Quality Evaluation) [7] and DSIS (Double Stimulus Impairment Scale) [7] methods. Some methods will be discussed further in Sections 2.3 and 2.4, respectively. Figure 2.5 highlights three objective and subjective evaluations adopted in the thesis, which are further discussed in Chapters 3-5.

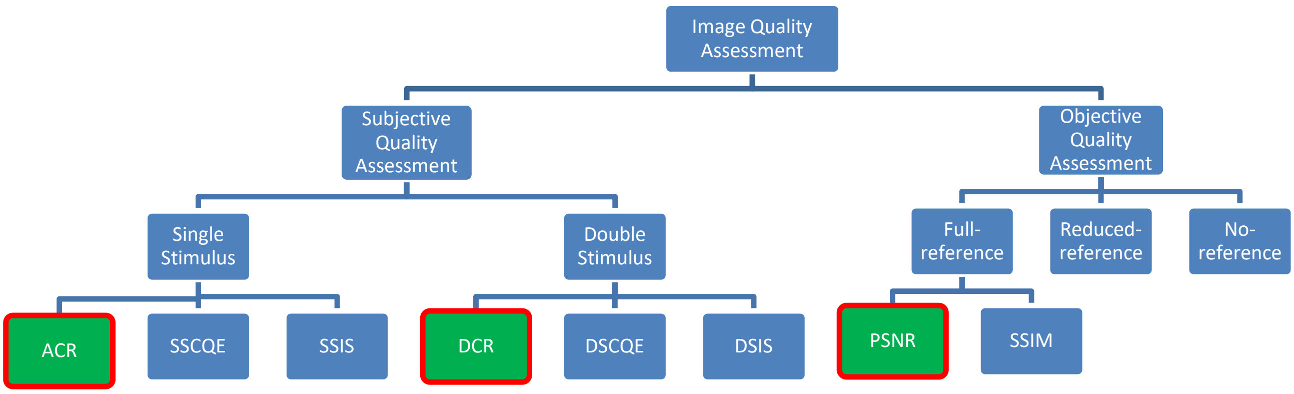


Figure 2.5: Image quality assessment methods.

Extending on image quality assessment methods, current objective and subjective techniques address various shortcomings in Video Quality Assessment (VQA) methods. Researchers have investigated various methods for subjective VQA [26, 27, 71]. Chen et al. [29] verified five subjective assessment methods and proposed that the Degradation Category Rating (DCR) method was the most stable for 3D fatigue assessment, but it has not yet been fully investigated.

Several objective assessment techniques, such as detection of blocking artefacts [23], Just Noticeable Difference (JND) [72], and No reference Stereoscopic Parallax based Distortion Metric (NOSPDM) image quality metrics [73] have been applied to compare the video quality; however, they do not include the subjective assessment of 3D fatigue. Some researchers have investigated 3D fatigue by video frequency [74] and VIMS in 2D modelling [71], which is most closely related to the research presented in this thesis.

2.3 Objective Evaluation of 3D Video

Existing approaches for the objective evaluation of 3D videos have been identified. According to the literature, there are four major factors [33, 34, 70, 75] that affect the threshold of the stereo effect and determine the visual quality seen by viewers. These four factors are as follows:

1. Temporal Masking: One of the characteristics of the Human Vision System (HVS) that is specific to watching videos. This effect causes the temporal distortion of visibility [33, 70].
2. Spatial Masking: An increase in the spatial non-uniformity of the background luminance causing a reduction in the visibility of the stimuli [34, 70].
3. Binocular Masking: One stimulus exerts an influence on the other stimulus when two monocular stimuli are viewed with the corresponding retinal locations of the two eyes [70].

4. Luminance Contrast: Human visual perception is sensitive to luminance contrast rather than an absolute luminance value [70, 75].

2.3.1 Human stereo perception evaluation

From the viewpoint of human stereo perception, the quality of stereoscopic images is affected by the degree to which the two images are distorted, and also by the experience of binocular perception. Most methods have focused on the correlation of stereoscopic image distortion and binocular perception, such as asymmetric assessment [23] and Just Noticeable Difference (JND) threshold measurement [72]. However, these assessment methods are mainly extensions from 2D quality assessment methods. Quan et al. [74] showed that the size of the disparity between the left and right views of a stereoscopic image is more important in the determination of visual comfort. Although some researchers have studied stereoscopic display with respect to visual fatigue, research into 3D QoE is still ongoing. In particular, few of the methods suggested by Quan et al. [74] account for the disparity between the two views of a stereoscopic image pair. The properties of a stereoscopic image that are essential for QoE include statistical characteristics, such as global similarity and local discrepancy [76].

2.3.2 Objective metrics

There are existing approaches for objective metrics including the detection of the difference of the video quality between the two stereoscopic views, detection of the blocking artefact and detection of blurring in the edge region [70, 77, 78].

The Peak Signal-to-Noise Ratio (PSNR) is the most widely used objective metric due to its low complexity and clear physical meaning [67, 68]. It quantifies the image (or video frame) quality by measuring the difference in the intensity between two images.

The PSNR is calculated as in Equations (2.1) and (2.2):

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left|S(i,j) - C(i,j)\right|^2 \tag{2.1}$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{2.2}$$

where the Mean Squared Error (MSE) measures the quality, m and n represents the image pixel, $S(i, j)$ is the reference image, $C(i, j)$ is the degraded image and MAX_I is the maximum value of a pixel. If the pixels are represented at 8 bits per pixel, the MAX_I value is 255.

Another commonly used objective metric, the Structural Similarity Index Measure (SSIM) [69], assumes that the HVS is highly adapted to extract structural information from the field of view. However, the ability of the HVS to adapt to S3D imagery is still under study. Wang et al. [79] proposed another SSIM measure for depth map variation using Equation (2.3); where $\overline{D}$ and $\overline{D}'$ are the estimated and actual depth maps respectively, σ_D and $\sigma_{D'}$ are the variance of the estimated and actual depth maps respectively and c_1, c_2 are constants.

$$\mathrm{SSIM} = \frac{(2\overline{D}\overline{D}'+c_1)(2\sigma_D\sigma_{D'}+c_2)}{((\overline{D})^2+(\overline{D}')^2+c_1)(\sigma_D^2+\sigma_{D'}^2+c_2)} \tag{2.3}$$

Other than the objective metrics mentioned in this chapter, researchers have thus begun to focus on functional changes in the human body that may occur when viewing 3D stereoscopic images or videos [74, 80], such as subjective quality assessments and human perceptual terminologies.

2.4 Subjective Quality Assessments of 3D Video

Several subjective quality assessment methods of 3D video have been developed. The International Telecommunication Union (ITU) has standardised some subjective methods for VQA. ITU-T Recommendation P.910 [81] standardises the subjective VQA methods for multimedia applications in evaluation procedures, experimental design, test methods and statistical analysis, such as the Double Stimulus Continuous Quality Scale (DSCQS), the Pair Comparison (PC) method, the Degradation Category Rating (DCR) method, the Absolute Category Rating (ACR) method and the ACR method with Hidden Reference (ACRHR), as previously shown in Figure 2.5 in Section 2.2. Various subjective methods have also been standardised in different voting sequences and testing requirements [27, 28]. Tominaga et al. [71] compared various subjective methods, such as the DSCQS, the ACR and the DCR. Researchers have compared various subjective assessment methods for 2D and 3D video quality [27-29] and have suggested that the ACR method is the most suitable for assessment stability and participants' assessment time of 3D videos [28]. In the ACR method shown in Figure 2.6, a video is played once for 10 s, and a voter is required to assign a score from a discrete scale of five grades (excellent = 5, good = 4, fair = 3, poor = 2 and bad = 1). After watching the 10 s target video, the voter must award a grade within 5 s. Further studies in this thesis using the ACR method have identified the characteristics of 3D videos with variations in video quality [82].

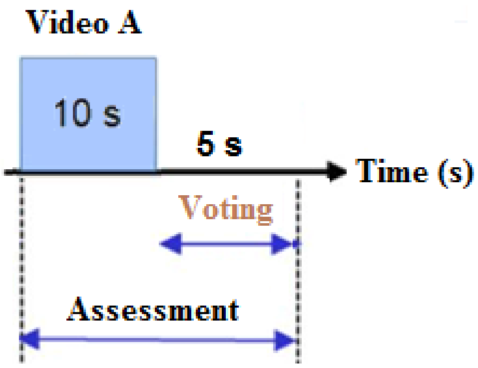


Figure 2.6: ACR method for QoE assessment.

2.4.1 Specification of viewing angle and distance

To provide a suitable viewing location for viewers, key parameters such as the viewing distance, screen width and viewing angle are considered. For research evaluations, MPEG and ITU international organisations provide viewing conditions for S3D video [16, 25]. Also, two main standards are used in the commercial environment for cinemas, the Society of Motion Picture and Television Engineers (SMPTE) [83] and the Tomlinson Holman Experiment (THX) [84] specifications. While most researchers are focused on the viewing conditions of S3D video defined in MPEG and ITU standards, few studies involve the investigation of the viewing conditions in the commercial environment [85], where watching 3D videos in cinema (or home theatre) environments is a common viewing environment for users.

Figure 2.7 shows the viewing locations recommended by SMPTE and THX standards [86], where W and H are the width and height of the screen respectively, and the viewing angle is measured by the viewer who is located at the central position to watch the screen. THX specifications have been adopted in experiments to provide a better environment for visual fatigue tests [87]. In the THX specifications, to view an acceptable video quality, the maximum viewing angle is 36° and the maximum viewing distance is 1.54 times the screen width.

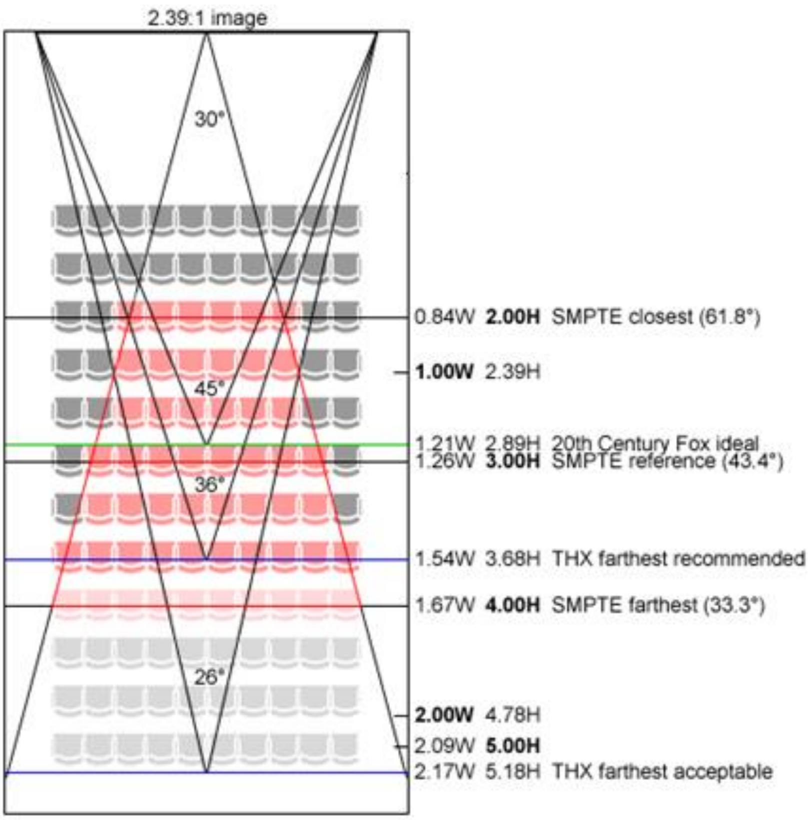


Figure 2.7: Requirements of viewing locations for two different cinema specifications: SMPTE and THX [86].

2.4.2 Visually Induced Motion Sickness (VIMS) evaluation

Visually Induced Motion Sickness (VIMS) is a category of motion sickness. It is related to nausea, strain, disorientation, general discomfort, vertigo and more. It happens when a viewer gets a perception of being in motion whilst being in a stationary state. This condition stands as an obstacle in front of any gamer involved in a 3D and virtual reality gaming experience. VIMS can affect the participants and is uncomfortable for many users. VIMS can also be caused by watching many 3D movies. Viewers might have an increased risk of health problems like photosensitive epilepsy in the long run after excessive viewing of 3D movies [88]. The recovery of VIMS might vary from meditation to simple exercises. One of the techniques is called natural delay in reducing VIMS. However, it can take up to a single day to fully eliminate symptoms. The alternative technique is a hand-eye coordination task. This technique is developed in an attempt to reduce the 3D illusion effect rapidly [89]. It involves a hand-eye coordination task which includes ocular focus and perception of body movement. The above two methods allow the user to reduce the VIMS effect.

VIMS is a subjective evaluation that involves the widely used Simulator Sickness Questionnaire (SSQ) to assess motion sickness [72], which has 16 parameters for the assessment of 3D visual fatigue. Zou et al. [90] performed VIMS for viewers watching 3D movies, and the results indicated that viewers might have an increased risk of potential health problems, such as photosensitive epilepsy when watching an increasing number of 3D movies.

2.4.3 Simulator Sickness Questionnaire (SSQ)

The most common subjective evaluation of VIMS is the Simulator Sickness Questionnaire (SSQ) developed by Kennedy et al. [24, 39]. The SSQ was developed to assess motion sickness, with different ratings for a selection of 16 major symptoms as shown in Table 2.1. Each symptom item is rated on four levels: none (0), slight (1), moderate (2) and severe (3). Each sub-scale of a symptom item have a different rating; the rating scheme of the sub-scale weighting for the items was based on an analysis of simulator sickness experienced by American pilots using 10 different flight simulators [91]. The three sub-scales, Nausea (N), Oculomotor (O) and Disorientation (D), are based on the greatest varimax-rotated loading structure which the varimax factor used to indicate the presence of a general factor for identification [39]. The symptom items of each sub-scale have varimax loading factors of at least 0.3. Some items include more than one sub-scale factor, such as difficulty focusing, nausea, difficulty concentrating and blurred vision. The Total Severity (TS) score is then computed as the weighted sum of the three sub-scales based on the types of symptoms and the scoring level, with a multiplication factor.

Table 2.2 shows the potential score ranges of SSQ scores of each symptom level at three sub-scales. For example, if all participants have “moderate” Oculomotor (O) symptoms, the resulting O score is 106.1. Another example is that if all participants have ‘‘slight’’ symptoms related to Disorientation (D), the resulting D score is 97.4. The total SSQ score, referred to as the TS score, can range from 0 to 235.6. Experiments conducted by Solimini et al. [92] examined VIMS in viewers watching 3D videos, and the results indicated that viewers might have an increased risk of potential health effects, such as photosensitive epilepsy when watching an increasing number of 3D videos.

Table 2.1: Sixteen symptom items with ratings and three sub-scales in SSQ [39].

SSQ Symptoms	**Nausea (N)**	**Oculomotor (O)**	**Disorientation (D)**
General discomfort	1	1	
Fatigue		1	
Headache		1	
Eye strain		1	
Difficulty focusing		1	1
Increased salivation	1		
Sweating	1		
Nausea	1		1
Difficulty concentrating	1	1	
Fullness of head			1
Blurred vision		1	1
Dizzy (eyes open)			1
Dizzy (eyes closed)			1
Vertigo			1
Stomach awareness	1		
Burping	1		
Total	**[1]**	**[2]**	**[3]**

$$\begin{aligned} \text{Nausea} &= [1] \times 9.54 \\ \text{Oculomotor} &= [2] \times 7.58 \\ \text{Disorientation} &= [3] \times 13.92 \end{aligned} \tag{2.4}$$

Table 2.2: Potential score ranges of SSQ scores.

Symptom	**Nausea**	**Oculomotor**	**Disorientation**	**SSQ Score**
None	0	0	0	**0**
Slight	66.8	53.1	97.4	**78.5**
Moderate	133.6	106.1	194.9	**157.1**
Severe	200.3	159.2	292.3	**235.6**

2.4.4 Biosignal-based evaluation

To quantify the self-reported subjective fatigue and emotional engagement of viewers when watching 3D videos, Electroencephalography (EEG) can be used. The Quality of Multimedia Experience (QoMEX) community has previously used EEG for analysis of multimedia experience [93-95]. The QoMEX community is leading experts from industrial partners and academic professionals to discuss the current and future research on multimedia quality, quality of experience (QoE), quality of service (QoS) and user experience [96]. EEG records the electrical activity generated by the brain by using electrodes placed on the scalp [97]. The position of electrodes in EEG equipment is shown in Figure 2.8 [98], where one electrode is typically used as the reference position at the earlobe or mastoid location such that the measured potential difference of the electrical signal is just equal to the voltage drop from the measured electrode to the reference electrode. The EEG signals are classified into delta, theta, alpha, beta and gamma waves according to the following frequency bands [72]:

- Delta waves (1-3 Hz): relate to deep sleep
- Theta waves (4-8 Hz): correlate with emotional stress for adults
- Alpha waves (9-14 Hz): reflect physical relaxation, meditation and creative visualisation
- Beta waves (15-30 Hz): reflect emotional state and focus level
- Gamma waves (above 30 Hz): can be used for the diagnosis of some brain illnesses

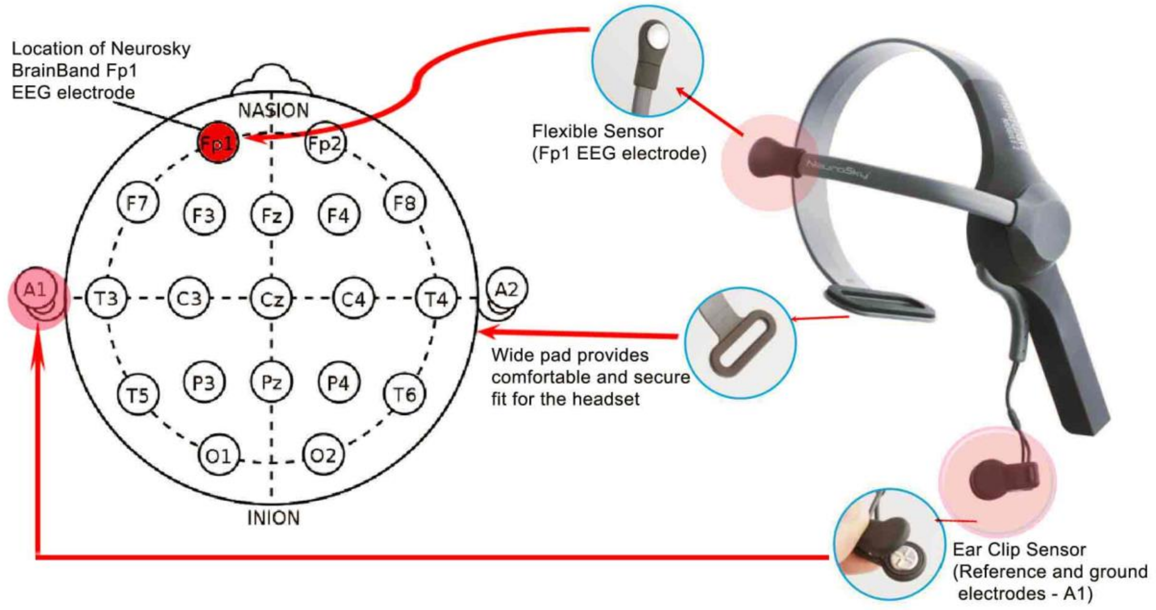


Figure 2.8: An example of consumer EEG headset equipment to connect the two electrodes in the respective locations [98].

A typical medical-grade EEG apparatus can collect brain waves from around 16 to 20 contact points simultaneously [99], as shown in Figure 2.9 [61]; however, the experimental setup is non-trivial. Therefore, some consumer-grade EEG apparatus are available in the commercial market because of the higher mobility of the device, participants, and lower price point than the typical medical-grade equipment. Various commercial EEG equipment such as NeuroSky [100], Muse [101], EMOTIV [102] and OpenBCI [103] vary from 1-8 electrode channels. Bateson et al. [104] compared 30 mobile EEG headsets adopted by previous researchers for the categorisation of the device, participant mobility, the system specification, and the number of channels. Such categorisation of the mobile EEG equipment can thus help to quantify the EEG equipment in a standardised way for research equipment selection.

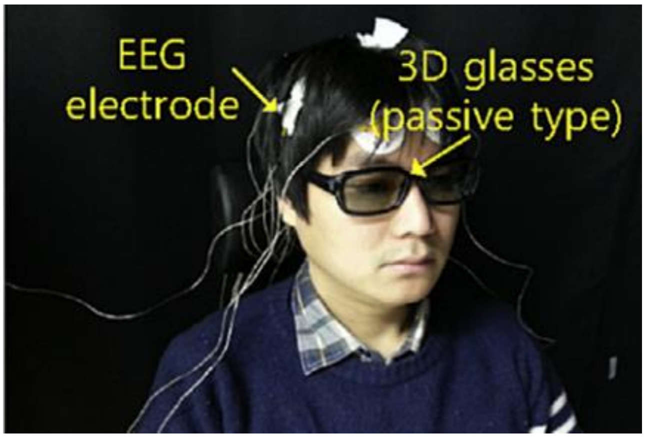


Figure 2.9: A typical medical-grade EEG apparatus for measurement [61].

Zou et al. [90] proposed nine EEG indices to assess stereoscopic visual fatigue, including three basic indices and six ratio indices. Six indices (α, β, α/β, α/θ, θ/(α+β) and (θ+α)/β) were found to be significantly different before and after the viewing period. Three of these indices (α, α/β, α/θ) were confirmed to show temporal variation after verification by Grey Relational Analysis (GRA). GRA is an alternative indicator for stereoscopic visual fatigue [105], and its grade can be calculated with Equation (2.5):

$$\begin{aligned} X_R(k) &= \{x_R(s) \mid s = 1, 2, \ldots, n\} \\ X_A(k) &= \{x_A(s) \mid s = 1, 2, \ldots, n\} \end{aligned} \tag{2.5}$$

where $X_R(k)$ is the reference sequence, $X_A(k)$ is the alternative sequence, and s is the number of samples of the nine EEG indices. The Relational Coefficient (RC) from the reference sequence to each alternative sequence was calculated with Equation (2.6) [90]:

$$\mathrm{RC}_{R,A}(k) = \frac{\Delta_{\min} + \tau\Delta_{\max}}{\Delta_i(k) + \tau\Delta_{\max}} \tag{2.6}$$

where:

$$\Delta_i(k) = |\Delta_R(k) - \Delta_A(k)|$$
$$\Delta_{\max} = \max_i \max_k \Delta_j(k)$$
$$\Delta_{\min} = \min_i \min_k \Delta_j(k)$$

where $\Delta_i(k)$ is the difference between two sequences for comparison, τ is a coefficient between 0 and 1, and Δ_{max} and Δ_{min} are the maximum and minimum values of the differences between the alternative sequences, respectively. $RC_{R,A}(k)$ indicates the closeness of two sequences for comparison. Comparisons show that an alpha wave is a potential optimal visual fatigue index of S3D video due to its effective identification of the fatigue level and sufficient sensitivity for an EEG index for visual fatigue detection.

The International Federation of Clinical Neurophysiology (IFCN) provides a list of brain wave energy testing methods for the measurement of 3D fatigue and has proposed several formulae for calculating brain wave energy from various measured parameters [99, 106]. The results show that a brain wave power ratio of more than 0.05 reveals an increase in the fatigue level.

2.5 Eye-tracking Techniques in Human Vision System Analysis

Eye-tracking systems are used for following eye positions of the observer in real time. They can measure the gaze point or the motion of the eye [107]. Eye-tracking is a widely used technique in the research of the vision system and Human-Computer Interaction (HCI).

Typically, the raw eye-tracking data is recorded using specialized eye-tracking equipment by applying an algorithm, or series of algorithms to the raw data. Then, fixations and saccades are extracted. Both fixations and saccades can be described in the spatial and the temporal domains; the variability of where exactly the viewer focused on (fixation), can be used for assessing the consistency of eye movements (saccades). The eye positions are sampled uniformly in time (typically at 60 Hz). This sampling method does not require pre-processing of eye-tracking data into discrete fixation-saccade sequences, as is usually adopted in eye movement analysis [108].

Gaze point evaluation is one of the methods for eye-tracking analysis as it can convey the fast-changing focus of the visual interest from viewers [109]. A major challenge for gaze input is the limited accuracy and precision, and the noise in measuring the gaze point. Accuracy means how the measured gaze point matches the real gaze position, whilst precision is the amount of variation of samples succeeded within a fixation. However, there is always some noise in measuring our gaze, even during a fixation [110]

The distribution of fixations across an image can be represented by presenting a heat map: a spatial density plot showing how frequently the viewer focused on the scene. The discrete fixations are first transformed into a continuous distribution, convolving a binary map with a Gaussian function. Figure 2.10 shows an image of a scene with fixations marked as red dots and a heat map where the peaks are shown in colour, which is proportional to the height of the heat map.

Figure 2.10: An example to present fixations on the image scene (a) marked in red dots, (b) presented in heat map [108].

2.5.1 Calibration

The calibration procedure is a critical factor in spatial accuracy, and can provide a mapping function that allows raw eye-tracking data to be converted to the coordinates of the viewing screen. Calibration usually involves presenting targets at a sequence of known locations. A simple approach would be to use a linear regression; that is to use a simple straight line equation ($y = mx + c$) describing the relationship [108, 111].

Gegenfurtner et al. [112] defined the comparison of perceptual performance and eye movement in the context of speed discrimination. In their study, the definition of perception was based on human interaction with the environment, further defining the perception of velocity as based on the error signal and a reference signal. These signals are obtained from the signals generated by the oculomotor plant, and it was shown that the relationship between the noise sources can be analysed. Turano et al. [113] also showed that constant displacements are associated with eye movements that change the velocity of the retinal image. Furthermore, this study also showed that eye movements must be compensated. The distal-motion model was analysed along with the retinal-motion model, in which the speed was analysed to identify the perceived speed of the distal stimulus changes, which are used to evaluate the direction and speed of eye movements. The study showed that the distal motion can be applied to examine fully compensated eye movements. Taking a different approach, O'Connor et al. [114] and Bennett et al. [115] studied the influence of ageing on eye direction discrimination and speed discrimination. O'Connor [114] demonstrated that both eye direction discrimination and speed discrimination have an important influence, whilst Bennett [115] showed that age did not have a significant effect.

2.5.2 Eye-tracking methods for gaze estimation

Ferhat and Vilariño [116] studied visible light gaze tracking in a 2D setting to analyse techniques such as gaze estimation, calibration methods and the variation of head pose. The study was based on an experiment in which a single camera was set up to perform remote gaze tracking. They also conducted some calibration experiments under a controlled environment to optimise methods to perform eye-tracking in a 2D setting. The findings revealed that high-performance techniques are suitable when eye-tracking can be performed in the desired manner. Furthermore, the study found that convolutional neural networks can be used to obtain accurate results for eye-tracking.

Hanhart et al. [117] proposed a gaze estimation method based on a 3D model, which is calculated based on the fitting of an eye model and the contour of the iris extracted from an image of the eye. However, it remains unclear whether the refraction phenomenon will affect their basic premise, namely, that the iris contour can be modelled as a simple perspective projection of a circular 3D iris. This seems to be a reasonable assumption for a near-eye camera installed almost directly in front of the eyes; however, the geometry of a headset with head-mounted eye-tracking usually requires a more inclined camera angle, and this assumption is not satisfied in such a case. Diekes et al. [118] proposed an approach to explain the effects while ignoring the corneal refraction can result in angular errors of several degrees. Such an approach used eye images from a single camera for refraction modelling for gaze estimation.

Narcizo et al. [119] reviewed current remote eye-tracking systems for research purposes, such as Human-Computer Interaction (HCI), data visualisation and human behaviour based on ocular activities. The system suggested by Hennessey et al. [120] comprised a single camera and a single LED to determine the centre of the cornea. They assumed refraction in the cornea and estimated the centre of the 3D pupil via back-projection points from the 2D pupil contour. As the radius of the back-projection pupil is estimated from the measured value of the pupil in the 2D image, their method is an approximate solution. Other researchers have proposed similar approximate calculation methods but used an eye-tracker with a single camera and multiple blinking points [121]. Cognolato et al. [122] further reviewed the current eye-trackers with gaze estimation systems. The results showed that each of the commercially available eye-trackers allowed for unobtrusive capture of real-time visual information.

2.5.3 Gaze point accuracy and area of interest determination

The fixed gaze point is a key indicator to obtain the human visual attention in visual science research, and is more specific than gaze estimation. Ciancio et al. [123] used eye-tracking equipment to measure the gaze point of fixation as the output measure of interest. The gaze point reveals the target to which the eye is looking at. An eye-tracker collecting data at a sampling rate of 1000 Hz will collect 1000 individual sampling points per second, known as gaze points. If a series of sampling points are very close in time and space, they form a gaze cluster constituting a fixed gaze point, which indicates that the eye is locked on the target. Expressions for accuracy estimate calculations were summarised by Kar et al. [124], as shown in Equations (2.7) – (2.11).

Gaze point coordinates:

$$\begin{aligned} POG\_X &= mean\left(\frac{POG\_X_{left} + POG\_X_{right}}{2}\right) \\ POG\_Y &= mean\left(\frac{POG\_Y_{left} + POG\_Y_{right}}{2}\right) \end{aligned} \tag{2.7}$$

Gaze displacement (*POG_d*):

$$POG\_d = \sqrt{(POG\_X)^2 + (POG\_Y)^2} \tag{2.8}$$

Pixel accuracy (*Pixel_acc*):

$$Pixel\_acc = \sqrt{(\text{target} \cdot X - POG\_X)^2 + (\text{target} \cdot Y - POG\_Y)^2} \tag{2.9}$$

On-Screen Distance (*OSD*):

$$OSD = \text{pixelsize} \times \sqrt{\left(POG\_X - \frac{x_{pixel}}{2}\right)^2 + \left(y_{pixel} - POG\_Y + \frac{\text{offset}}{\text{pixelsize}}\right)^2} \tag{2.10}$$

Angular accuracy (*Angular_acc*):

$$Angular\_acc = \frac{\text{pixelsize} \times Pixel\_acc \times \cos\left[mean\left(\tan^{-1}\frac{OSD}{dist}\right)\right]}{mean\_dist} \tag{2.11}$$

The mean gaze coordinates of the eyes' point of gaze (POG) are represented as *POG_X* and *POG_Y*, whilst *dist* and *mean_dist* are the distance between the eye and projection screen, and the mean distance between the eye and eye-tracker respectively. The pixel shifts of the x and y pixels move along the x and y directions, and the offset is the distance of the eye-tracker sensor from the lower edge of the projection screen. Further details of the calculation formulae are explored in the literature [125].

The eye movement between two fixations are called saccades. It has been found that the eyes become locked because the central "visual range" of the human eye is limited. The "visual range" refers to the number of words visually read before and after the current fixed words. The physiological explanation deals with the fovea, parafovea and the nerves around the eyeball [126]. The interpretation of saccades reflects how the trajectory of the eye movement changes and whether the position of the gaze changes. This is a spatial indicator of eye-tracking [126].

Determination of the Area of Interest (AOI) is done by selecting a displayed stimulation target area and extracting eye movement indicators in this area for statistical analysis. It defines the area from which the eye movement indicator metric is calculated. The display indicators for each area, such as the time from the beginning of the stimulus to the time the participant views the area, the amount of time the participant spends in the area and the number of people who move their eyes away and then back, are useful when evaluating the performance of two or more areas in the same video, picture, website or program interface [127]. The first fixation time is the duration of the first fixation point that falls on the AOI and serves as an important reference index of time in the eye movement index, especially in reading research. It reflects the early characteristics of vocabulary access when the eyeball is watching and reading [128].

Ahn et al. [129] studied visual scene stimulation and found that participants will scan by saccades due to the difficulty of visual processing. The ratio in the visual scene refers to the proportion of the gaze information from the participants' gaze in the target AOI to the rest of the non-target area. This approach is often used in advertising and Web design to quantify viewers' focus, identifying which parts of a visual scene are more attractive to different participants. During reading, the saccade rate can be referred to as the skipping rate because it refers to the probability that the target AOI will be skipped. Studies have found that the length and height of a word in the sentence predictability have an important influence on the saccade rate [130]. In reading research, the saccade rate can reveal the degree of familiarity of the words [131]. The number of "look-backs" reflects the participant's reprocessing of previous information and provides information about the number of times the participant's gaze returns to a specific target defined by the AOI. This information allows the researcher to check which areas repeatedly attracted participants (good or bad) and which areas are seen and then moved from. Although eye tracking cannot tell researchers how the participants feel when they look at something, it can provide gaze and visual attention focus [132].

2.5.4 Current saliency prediction models

Compared with a long history of the gaze point detection method, only a short history in salient object detection tasks has developed as it involves several computer vision tasks, such as visual tracking, image captioning and the segmentation of images. Itti et al.[133] and the Graph-Based Visual Saliency (GBVS) model by Havel et al. [134] are the earliest works in this field. These saliency object detection models, referred to the saliency prediction models, are mainly based on a bottom-up method, using different low-level visual features, such as colour and edges, due to the significant object detection. Therefore, the salient object detection model also referenced basic theories of the human visual attention mechanism, including the contrast hypothesis and centre-periphery hypothesis. Jiang et al. [135] then proposed a salient object detection task, which can be regarded as an extension of the visual attention mechanism of the object segmentation task. The proposed model includes a prediction function of the visual comfort to differentiate an S3D image as a low- or high- level comfortable stereo viewing and derive the saliency maps for different visual comfort, this saliency map can provide better visual comfort assessment. Liu et al. [136] used a multi-scale contrast, centre-surround histogram, and centre-surround histogram at different scales. After the three saliency measures of colour spatial distribution, conditional random fields were used to integrate these saliency features, and the first significant object detection data set is also proposed [135, 137, 138]. Three important evaluation indices are proposed: precision, recall and *F*-value (*F*-measure), where two variables, *F*-defined value and evaluation index, have become the most commonly used evaluation indices in the field of significant object detection in vision science research. He et al. [139] proposed a salient object detection method by obtaining the centre coordinates of the salient object using several detection algorithms, and the proposed model has better detection effects and higher detection rates. The research of Iatsun et al. [140] conducted follow-up work in the direction of salient object detection, which proposed a model to consider video characteristics, such as disparity range, motion activity, and the previous visual fatigue state.

2.5.5 Eye-tracking for S3D video quality of multimedia experience

Researchers have already adopted eye-tracking to understand how the human visual system responds to visual attention when watching 3D video content [141]. Hanhart et al. [141] and Fang et al. [142] created an eye-tracking dataset providing the eye-tracking information, such as fixation points and fixation density maps and developed a visual attention model to understand the visual mechanism which can enhance the QoMEX of S3D video. Then, Banitalebi-Dehkordi et al. [143] evaluated existing saliency detection methods and proposed a new 3D visual attention model to validate and benchmark these methods.

2.6 Current Developments in Multi-view 3D Video

Multi-view video includes multiple video sequences captured by several cameras at the same time, but in different locations [144]. The view direction and the viewpoint can be changed within the range captured by cameras. Various applications of multi-view video coding include free-viewpoint TV and 3D video applications for home entertainment and surveillance [9].

Multi-view 3D video increases the transmission bandwidth because it typically provides 16 cameras with different viewing angles [35]. Therefore, these video sequences contain a large amount of data, but there are limitations on data distribution applications [144]. However, it is critical to limit the bandwidth while minimising the total distortion between two different views. Several schemes for bit allocation between the texture and depth map for a multi-view 3D video have been proposed to ensure that the synthesised view has nearly the same quality as the original view. Depth Image-Based Rendering (DIBR) approaches eliminate the requirement to deliver each view of a multi-view video if the left and right views of the desired view have been transmitted to other viewers [145-147]. Moreover, the number of left and right views in a multi-view video must be constrained to ensure the quality of the synthesised view [148], such that the desired view of each user can be synthesised with good quality.

With the development of advanced video quality, significant changes have occurred in frame rate, dynamic bit depth and colourimetry [30]. In the current development of 4K resolutions (also called UHD-1), the video should include ITU BT.2020 colourimetry, frame rates up to 120 fps and bit depth up to 12 bits for the high dynamic range. Figure 2.11 shows how the viewing angle affects the advanced video quality, where increasing the viewing angle provides higher-resolution images.

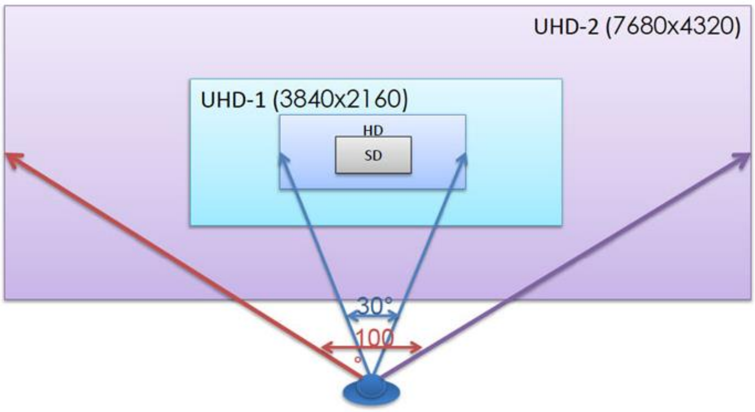


Figure 2.11: Viewing angle with the advanced video quality [30].

Researchers have proposed methods for multi-camera processing arrangements and evaluations, and the software framework of free viewpoint TV is currently in MPEG standardisation [10, 17]. However, few methods of multi-view video coding have been proposed for S3D [149].

2.7 Summary

This chapter has reviewed key literature on objective and subjective quality assessments of S3D video. Objective evaluation is a long-standing approach to image and video assessment, whereby the key approaches are detection of blocking artefacts, Just-Noticeable Difference (JND), and NOSPDM image quality metrics. However, objective metrics are mainly an extension from 2D quality assessment methods, and are generally used without considering the subjective assessments of 3D fatigue for S3D.

Various subjective quality assessment methods of 3D video are available and related to ITU and MPEG standardisations. Most commonly, the SSQ is used to assess VIMS and brain fatigue. Recent research has adopted biosignal-based evaluation for continuous-time feedback, including EEG to assess neurological activities and eye-tracking to assess the human visual attention in 2D video assessment. Using eye-tracking, human saliency maps and saliency prediction models can thus be developed for 2D video. However, there is currently limited investigation on the subjective evaluations of commercial environments by other standards, such as SMPTE and THX, and comparison of subjective evaluations in different viewing environments. Further, few studies involves both surveys (e.g., SSQ) and biosignals (e.g., EEG) to assess VIMS, or complement with eye-tracking to assess the human visual attention of S3D video and correlate human visual attention with QoE assessments for S3D video. Working with these key approaches to address research gaps from previous studies, this thesis investigates VIMS with EEG signals to correlate 3D fatigue level with VIMS by comparing 2D and S3D videos in a flat 3D screen; and compares S3D videos in three different viewing environments, with the use of commercial standard: THX. Also, the thesis investigates visual attention of viewers using eye-tracking to predict S3D visual attention and develop a human saliency map model.

The presentation and evaluation of the QoE of a flat 3D screen and different viewing environments are discussed in Chapter 3 and Chapter 4 respectively, to address the challenges outlined in Sections 2.2–2.4. The application of the quality metrics and subjective evaluations as a use-case in S3D is then evaluated in Chapter 5 to reflect the methods raised in Sections 2.3 and 2.4. The discussion and the current QoE assessments of S3D video using eye-tracking technology from Section 2.5 are then presented in Chapter 6 .

Chapter 3 Comparison of Objective and Subjective S3D Evaluation using EEG Biosignals

3.1 Introduction

In S3D image and video, each eye is presented with a different image, and the combination of these two images in the brain applies depth to generate a 3D image. However, it has been reported that long 3D movies or visual effects may cause health effects in the audience, such as eyestrain, nausea, headaches or visual fatigue [22].

It is important to understand the QoE between 2D and 3D image and video content from the perspective of end-users [150]. Using objective approaches, one method to differentiate 2D and 3D video is based on the stereoscopic concept. Subjective evaluations generally focused on Visually Induced Motion Sickness (VIMS) [151]. All existing studies confirm that S3D videos produce greater symptoms of VIMS to participants [23], compared to 2D image and video content.

To augment users' self-reported VIMS experience, an alternative and complementary method is to use continuous-time biosignals. Few researchers have investigated VIMS augmented with biosignals such as EEG signals. Therefore, this chapter focuses on the comparison of QoE measuring self-reported VIMS and recording EEG biosignals when participants view both 2D and S3D videos of the same, non-stimulating consumer content. The experimental approach had two main aims:

- To investigate 3D visual fatigue to identify the effects of VIMS, using EEG measurement and the Simulator Sickness Questionnaire (SSQ), to consider visual attention and assess 3D visual fatigue.
- To understand whether 2D or 3D video content caused more visual fatigue and to what extent [53].

3.1.1 Electroencephalography (EEG)

Electroencephalography (EEG) is a method of recording the bioelectrical activities in the brain. It is classified into five wave types. As described in Section 2.4.2, several researchers have used and recommended the use of EEG signals to evaluate VIMS. According to Choi et al. [152], there are five waves associated with EEG signals and these include alpha (α), beta (β), delta (δ), theta (θ) and gamma (γ) waves. The five types of waves correspond to particular frequency bands which correlate with various emotional states of the viewers. Hence, it is possible to understand the emotions of the viewer using the EEG signals by correlating the emotions corresponding to the particular frequency ranges of the detected waves.

Table 3.1 below lists the various EEG components and the frequency ranges used in the cognitive neuroscience system. The delta wave is a temporal component that occurs in the frequency band of 1-3 Hz which reflects deep sleep, and the state of unconsciousness of the mind. The second temporal component is the theta wave that occurs in a frequency band between 4 and 8 Hz and reflects light sleep and emotional stress. The third component, the alpha wave, is an oscillatory component that occurs between 9 and 14 Hz and reflects physical relaxation, creative and meditation visualisation [153, 154]. The beta wave oscillatory component occurs in the 15 to 30 Hz frequency band and reflects the emotional state. Lastly, the gamma wave, an oscillatory component, occurs above 30 Hz and may reflect a diagnosis of some brain illnesses.

According to Young and Cashwell [155], the brain wave power can be measured in varying frequency bands consistent with the alpha, beta, gamma and theta waves. These waves can be used to identify fatigue by comparing the ratio of fast waves (α and β) to slow waves (δ and θ) as a function of time. Cognitive neuroscience shows that the comparisons of the behaviours of temporal components on two occasions namely before and after the event can be used to measure mental fatigue. In addition, the frequency bands show different levels of alertness. On the one hand, when the alertness increases, there is also an increase in the proportion of low-frequency bands such as the α, θ and δ waves [156]. In contrast, when the alertness level decreases, there is a decrease in the proportion of high-frequency bands namely the γ and β waves [155]. Human emotion recognition and bioelectrical activities can thus be measured using the EEG signals as proposed by Petrantonakis et al. [106] and Liu et al. [72].

Table 3.1: Characteristics of major EEG components.

Component	**Name**	**Frequency Band (Hz)**	**Reflection**
Temporal	Delta (δ)	1 – 3	Deep sleep, unconscious of the mind
Temporal	Theta (θ)	4 – 8	Light sleep, emotional stress
Oscillatory	Alpha (α)	9 – 14	Physical relaxation, meditation
Oscillatory	Beta (β)	15 – 30	Emotional state
Oscillatory	Gamma (γ)	>30	Diagnose brain illness

Table 3.1 also shows a reflection for each component. The reflection gives an indication of the emotional status of the display user [152]. Oscillatory components show the neural states of participants [61]. Recent research indicates that a brain wave is the response of the emotional

pattern associated with the alpha wave properties [90], and this type of association can be observed by multi-channel EEG recording.

The EEG signal of a viewer changes depending on emotions arising from the viewer's viewing experience. From the EEG, one can classify the viewing emotions into six categories; sadness, happiness, disgust, anger, surprise and fear [157]. Hence, in real-time, the monitoring of the user's viewing experience of the S3D video can be monitored by recording the EEG signal throughout the video. Such approaches have been used by researchers such as Liu et al. [72] to determine 3D visual fatigue. During the experiment, viewers in a binocular parallax condition watched both the 2D and 3D videos. The study revealed that the EEG signals are significantly correlated with the subjective measurement of 3D visual fatigue, and can be used to determine the dominance level of emotional interaction in the human brain [72]. Although the results were positive, Liu et al. [72] used a method that relied on auditory stimuli only without considering the visual stimuli. The researchers compared both 2D and S3D videos while choosing an EEG method to record the signals. Zou et al. [90] further investigated whether the nine types of EEG indices are effective in assessing the visual fatigue of the viewer. In their study, alpha waves produce better results in detecting stereoscopic visual fatigue. Another study conducted by Hu et al. [158] relied on functional Magnetic Resonance Imaging (fMRI) to bridge between low-level and high-level semantics for video classification. They have also studied EEG signals to establish their success rate in determining user emotions. In their study, Vinhas et al. [159] extended this work to consider additional biosignals, e.g., respiration rate and volume, galvanic skin response, heart rate and skin temperature, in regard to the multimedia content and delivery based on the user emotions.

3.2 Experimental Methodology

3.2.1 EEG for visual fatigue assessment

The viewing of S3D video and the fatigue due to excessive viewing can impact the bioelectrical signals in the brain. EEG can be used to measure changes in the brain during specific activities and for that reason, EEG can be used to measure fatigue whilst watching videos. It has been found in research that an individual cannot keep their attention for a long time without causing fatigue. The amount of fatigue experienced whilst watching 2D and S3D videos is not well understood, and therefore this research has been conducted to measure the impact and amount of fatigue during the watching of 2D and S3D videos of the same content [160]. In the case of physical fatigue, it is measured or analysed by exhaustion; in contrast, mental fatigue is generally measured through reduced mental activity. The proposed approach of using EEG to measure visual fatigue whilst watching 2D and S3D videos is investigated in this chapter, as it provides a more accurate human activity assessment than the VIMS approach alone [26, 73].

3.2.2 EEG for eyeblink measurement

Wang et al. [161] advocated the use of eye fatigue assessment models to evaluate eye fatigue in regard to the eye blink data and eye movement data. Such assessment can be achieved with the aid of an eye-tracker. However, an eye-tracker is not the only device or method that can be used to assess eye fatigue. Previous studies have showed that health monitoring and eye blinking movement are correlated, where eye movement, eye-gaze and pupil contraction are related to fatigue [162, 163]. It can indicate physical condition, mental workload, stress, sleep disturbances, depression, Parkinson’s disease, neurodegenerative diseases and different social activities [70, 108]. In particular, eye blinking can be observed when an individual is fatigued.

Different studies have shown that increased mental fatigue results in a change in blink velocity and blink interval [163]. Thus, in this work, in addition to measuring visual attention, the eye blinking is also measured using EEG data capture.

3.2.3 EEG equipment used for data acquisition

A consumer-grade EEG apparatus called the NeuroSky MindWave [97] was used for the current experiments. The NeuroSky system [100] was chosen due to its benchmark against the Biopac system, a widely used medical- and research-grade EEG system [164]. The benchmarking compared the EEG signals measured by the two types of apparatus; electrodes for the two systems were placed as close as possible to the same location to avoid interference.

Figure 3.1 and Table 3.2 show the raw EEG signals and the correlation coefficient of the two EEG systems. Both systems show a similar pattern of waveform and the eye blink sensitivity. The average Discrete Fourier Transform (DFT) power spectrum results showed that the EEG signals of the NeuroSky system are comparable to those of the Biopac system [164]. Therefore, the NeuroSky system was used in the 3D fatigue experiments of this thesis.

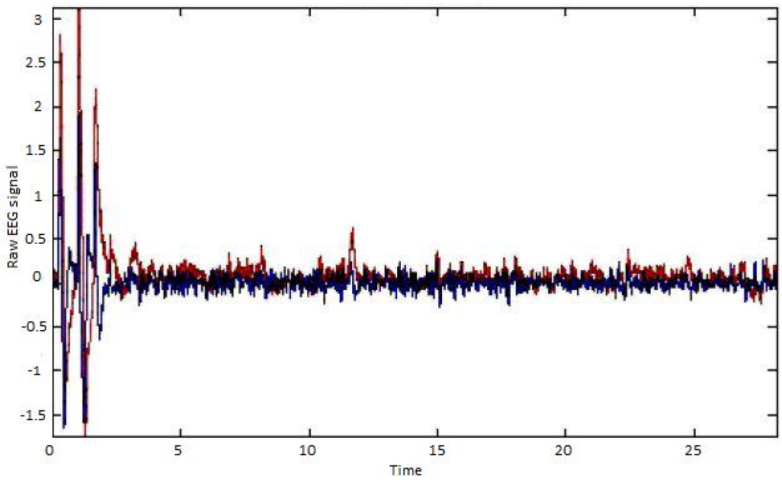
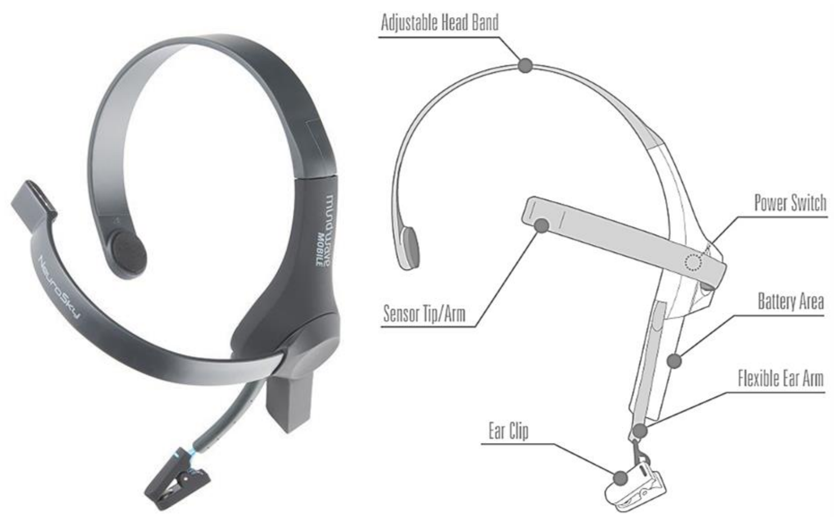


Figure 3.1: Raw EEG signals of NeuroSky and Biopac systems over time (blue line: NeuroSky, red line: Biopac) [164].

Table 3.2: Correlation coefficient values between NeuroSky and Biopac systems [164].

Time (seconds)	**Correlation Coefficient**	**Meditation Rating**	**Attention Rating**
15 – 16	0.771	73	61
17 – 18	0.712	78	56
18 – 19	0.858	80	45
19 – 20	0.567	88	41
20 – 21	0.564	93	43
21 – 22	0.581	86	38
22 – 23	0.321	95	41
23 – 24	0.685	88	43
24 – 25	0.751	85	45
25 – 26	0.842	76	56
Average power spectrum	**0.715**	**-**	**-**

Figure 3.2 shows a diagram and design of the NeuroSky MindWave EEG headset used in the experiment for data acquisition of brainwaves. The headset consists of different parts, such as sensor tip and arm, ear clip, flexible ear arm, adjustable head band, power switch, battery area and ThinkGear ASIC chipset. For the principle of the data acquisition of brainwaves, two sensors are used to detect and filter the EEG signals. The sensor tip detects bioelectrical signals from the forehead and an ear clip sensor acts as a reference ground for the ThinkGear chipset to filter the electrical noise (ambient noise generated by human, computers and other electrical appliances and devices) [97].



(a) (b)

Figure 3.2: NeuroSky MindWave EEG headset (a) Diagram, (b) Design [97].

This headset can measure raw EEG signals such as alpha, beta, delta, theta and gamma waves at a sample rate of 512 Hz. The headset can also detect attention level, meditation level and eye blink rate, using proprietary algorithms supplied by the manufacturer [97, 165]. The manufacturer's specifications of the NeuroSky MindWave EEG equipment state that the meditation (similar to relaxation) and the attention (similar to concentration) level of a NeuroSky MindWave wearer are reported by an "eSense" meter to characterise the wearer's mental states. The eSense meter has a scale that ranges from 1 to 100. Any value lying at the middle of the scale from 40 to 60, is said to be a neutral condition [97]. Hence, by using the eSense meter, the attention and meditation levels were measured and found to have an average level of 50 with all the devices. As concluded, a reading of between 40 and 60 indicate a neutral condition; hence, participants experienced a neutral condition. These findings correspond to the results of the studies conducted by [166]. A study conducted by Andreu-Sánchez et al. [167] found that in the presence of chaotic and fast audiovisual material, conscious processing decreases while the attention scope increases.

3.2.4 Experimental procedures and conditions

The experiment has been approved by the Universities' Human Research Ethics Committee, based on the guidelines of "National Statement on Ethical Conduct in Human Research" developed jointly by the National Health and Medical Research Council, the Australian Research Council and Universities Australia, as shown in Appendix 4. Further, all the EEG experiments in this thesis were conducted according to the International Federation of Clinical Neurophysiology standards [168].

For the experiment, the participants were required to complete an SSQ after watching each 2D and 3D video sequence. The captured biosignal data and SSQ scores were then analysed to study the fatigue level, total SSQ scores, and total eye blinking frequency (measured by the EEG headset) across all the video sequences.

3.2.4.1 Experimental procedures

Figure 3.3 illustrates the experimental methodology used to measure visual fatigue whilst watching 2D and S3D videos. The study was conducted on 12 males and 3 females aged between 18 and 38 years old. All participants are seated for the 3D experiment, which is based on the ITU-R BT. 2021 standard [25].

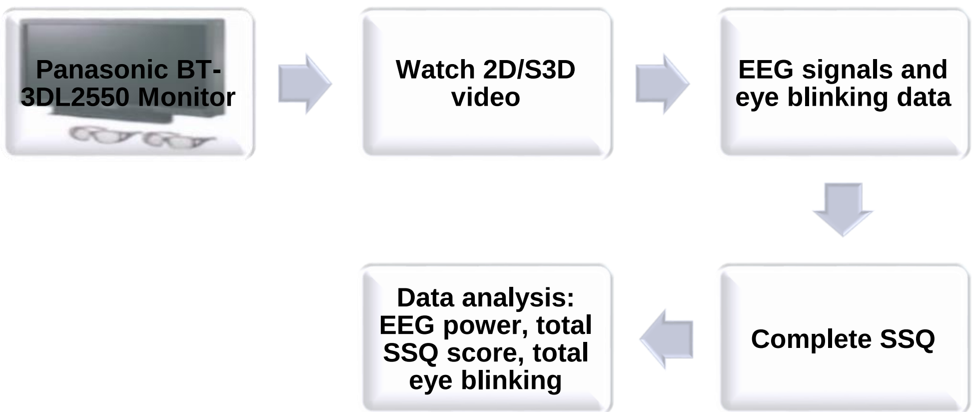
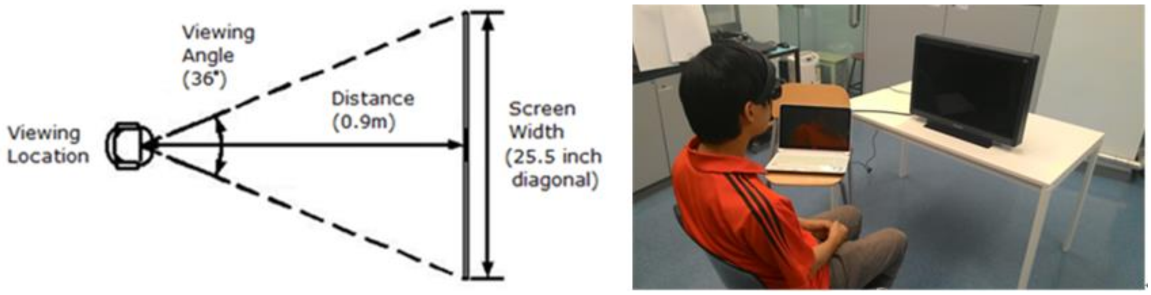


Figure 3.3: Experimental methodology of QoE test in S3D flat screen.

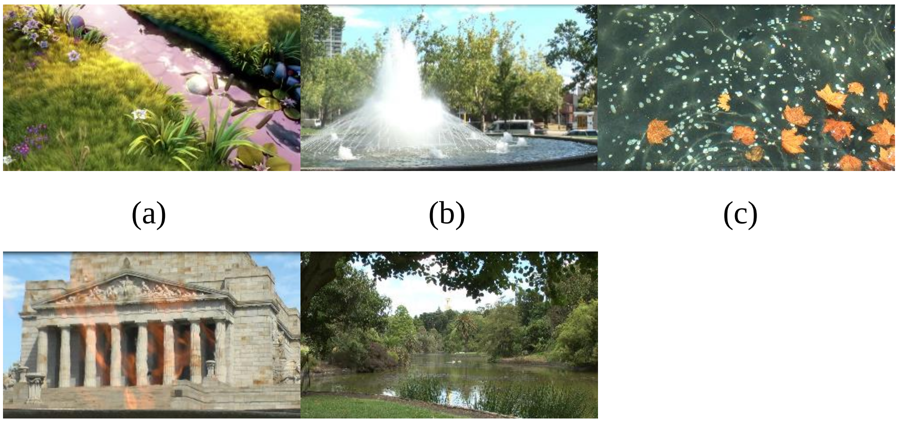
The research was conducted in controlled laboratory conditions at 23℃±1℃ and 50%±8% relative humidity. The S3D video sequence was presented on a 25.5’’ Panasonic BT-3DL2550 LCD S3D screen and all 2D and S3D video sequences were displayed in 1920×1080 HD resolution [169]. Figure 3.4 below illustrates the equipment setup with the viewing dimensions of the experiment. All participants are seated for the 3D experiment, which is based on the ITU-R BT. 2021 standard [25]. The participants were seated in front of a 3D monitor at a 0.9 m viewing distance with a 36° viewing angle to watch 2D movies, as specified by the THX Cinema Certification specification [87]. The controlled environment with the THX specification settings is used to simulate a real-life 3D environment to test whether VIMS exists when participants view both 2D and S3D videos.



(a) (b)

Figure 3.4: Experimental setup of viewing location from THX farthest recommendation (a) Viewing dimensions (b) Actual setup.

All the video sequences were from the RMIT3DV [75] and Big Buck Bunny [170] databases, as shown in Figure 3.5. All video sequences were natively recorded in S3D, with the left view selected as the 2D video shown to participants for each sequence. Sequence BBB is from the Big Buck Bunny [170] database, while sequences water fountain, wishing well, flame and garden is from the RMIT3DV [75] database.



(a) (b) (c)

(d) (e)

Figure 3.5: Video sequences used (a) BBB [170], (b) Water fountain [75], (c) Wishing well [75], (d) Flame [75], (e) Garden [75].

Table 3.3 below summarises the characteristics of each video sequence. Two factors were used as the selection criteria for the five video sequences. One of the selection criteria used was the variety of 3D effects, and the second criteria was the comparison of 3D video experiences between animation and outdoor scenes. The first criterion was used to test whether the variation of 3D effects affected VIMS, while the second criterion was used to test if there were any significant differences in VIMS between animation and outdoor scenes.

Table 3.3: Characteristics of the S3D video sequences.

Name of Movie	**Type**	**Description**	**3D effect**
(a) BBB	Animation	Cartoon characters with the garden background	Weak
(b) Water fountain	Outdoor scene	Water, trees	Moderate, water fountain
(c) Wishing well	Outdoor scene	Warped water moment	Strong warped water movement
(d) Flame	Outdoor scene	Chaotic flame	Moderate at the chaotic flame
(e) Garden	Outdoor scene	Lake, Birds, trees	Weak

All video sequences were presented for 5 minutes; if the sequence was shorter than the required duration, the sequence was repeated to fulfil the time requirement. The reason is that starting a vision-related activity for 5-10 minutes, participants may experience visual fatigue [171]; therefore, 5-minute period is the minimum time to experience visual fatigue. Firstly, the 2D video was presented and the participant's eye blinking and EEG data were measured and monitored by the NeuroSky software. After 5 minutes of rest, the participant watched the same video rendered in S3D for another 5 minutes with their eye movement and EEG data recorded. Then, each participant completed an SSQ after each 2D/S3D video sequence pair [172]. The viewing of each video pair was conducted in a random order to reduce the sequence bias, and the average duration of each viewing experiment was 1.5 hours. The reason for a long duration is that some participants indicated tiredness and additional rest pauses were included.

3.3 Results and Discussion

3.3.1 SSQ results

The results from the SSQ are shown in Figure 3.6. The SSQ scores of the 3D videos are higher than any of the scores of 2D videos. In particular, the “Wishing well” video sequence causes the greatest increase by 50% change in VIMS between the 2D and 3D sequence, potentially due to the chaotic and frequent movement of the water throughout the sequence. Similarly, the “Water fountain” sequence contains regular water movement, and the “Flame” contains chaotic flame movement; both sequences also result in increased VIMS between the 2D and 3D sequences. In contrast, the “Garden” sequence features the least movement being a near-still sequence of a floral scene, resulting in the least SSQ variation (20% change) between the 2D and 3D sequence. Overall, for the video sequences shown, the VIMS is comparatively lower for 2D conditions than for the 3D conditions. Sequences containing more scene movement causing greater increases in VIMS, likely due to more engaging and higher cognitive scenes.

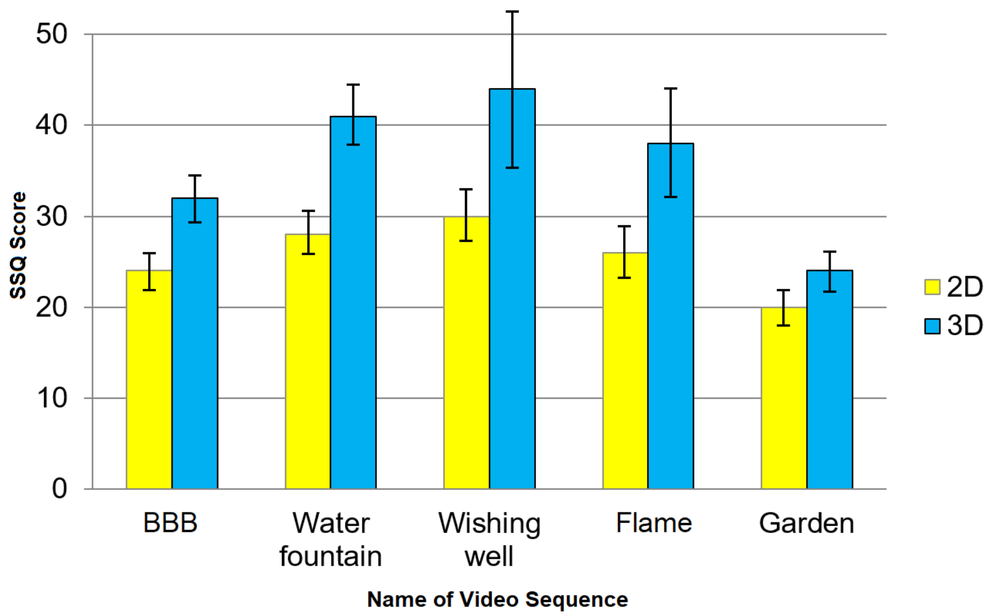


Figure 3.6: Average SSQ score for 2D and 3D sequence of videos (95% confidence interval).

3.3.2 Eye blinking movement

Table 3.4 shows the measured eye blinking movement rates of the participants. It can be seen from Table 3.4 that amongst the five videos, the "Wishing well" video caused the least eye blinking which suggests that it is the most fatiguing video. Congruent with the SSQ result in Figure 3.6, the video is very engaging as it includes rapid water movement, unstable flashes of light, a shadow from the water, and different interactions underwater [173]. A similar effect due to movement in the video can be seen for the "Water fountain" video. These results indicate that movement videos are likely very engaging; however, movement causes the viewer to be attentive, thus reducing eye blinking [174], and causing the most fatigue due to higher cognitive activity. In contrast, as in the findings shown in Figure 3.6, the "Garden" video is the least fatiguing video due to the stable scene.

Table 3.4: Total eye blinking for 2D and 3D video sequences.

Name of Videos	**Count of eye blinking in 5 minutes**	
	2D	***3D***
BBB	56.4	54.3
Water fountain	53.7	49.2
Wishing well	50.4	45.3
Flame	57.8	55.2
Garden	63.4	60.1
Average	**56.3**	**52.8**

The experiment also showed that eye blinking whilst watching a 2D video is overall higher than the same sequence in S3D, as shown in Table 3.4. The result of the average eye blinking for 2D videos is 56.3 whereas the average for S3D videos is 52.8. These average eye blink results indicate that the visual fatigue is likely to be higher for 3D video sequences compared to 2D video sequences.

3.3.3 EEG brain activity measurement

Hirvonen et al. [175] proposed test methods which use brain wave power ratios of EEG signals for the measurement of 3D visual fatigue according to the International Federation of Clinical Neurophysiology (IFCN). Equations (3.1) and (3.2) show two brain wave power ratios using different frequency bands of EEG signals. The threshold power ratio level between low and high fatigue is 0.05, as shown in Equation (3.3).

$$Power\ Ratio_1 = \frac{\theta + \alpha}{\beta} \tag{3.1}$$

$$Power\ Ratio_2 = \frac{\theta + \alpha}{\alpha + \beta} \tag{3.2}$$

where:

$$Power\ Ratio_{1,2} \begin{cases} > 0.05 \text{ (High fatigue level)} \\ < 0.05 \text{ (Low fatigue level)} \end{cases} \tag{3.3}$$

The measured brain wave power ratio for participants watching the five video test sequences in 2D and S3D is represented in Figure 3.7, where P1 and P2 indicate the power ratios given by Equations (3.1) and (3.2) respectively. The brain wave power ratios are the averages across all of the 15 participants, where it can be seen that the average brain wave power ratios for the 3D videos are consistently higher than the power ratios for the 2D videos. In Figure 3.7, the axis of the brain wave power ratio is centred at 0.05 which indicates the threshold level for high fatigue [168]. Figure 3.7 also indicates that 3D videos can be above the high fatigue threshold level, depending on the content in the video. The "Wishing well" video, consistent with the results shown in Figure 3.6 and Table 3.4 causes the highest level of fatigue. Further analysis of the brain wave power ratios for each participant across all 3D videos found that over 90% of participants were above the high fatigue threshold level when watching 3D videos (this analysis is not shown in Figure 3.8).

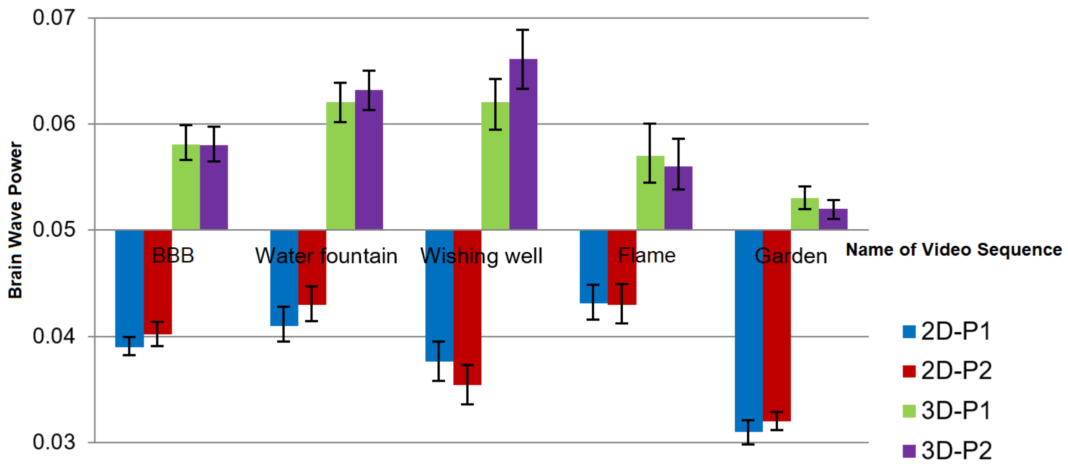


Figure 3.7: Brain wave power ratio distribution.

Figure 3.8 shows the brain wave power ratios of each participant viewing the "Wishing well" video sequence. This video was chosen for further analysis due to the image distortion through water movement in the content, depth perception change, time-varying and rapid higher frequency motion. Therefore, it leads to more fatigue for the viewers. The higher brain wave power ratio for the 3D video can be clearly seen from Figure 3.8. Further, the brain wave power ratios for 2D and S3D videos significantly differ, which suggest different emotion recognition, as discussed in [173]. Congruently, the highest level of nausea and VIMS were recorded for the "Wishing well" video sequence in the SSQ (Section 3.3.1). The SSQ score for each participant viewing the "Wishing well" video sequence is shown in Figure 3.9. The result shows that the SSQ score for the S3D video is generally higher than that for 2D video. Also, more than half of SSQ scores for 2D and S3D videos differ significantly. This result aligns with the brain wave power ratios analysis shown in Figure 3.8.

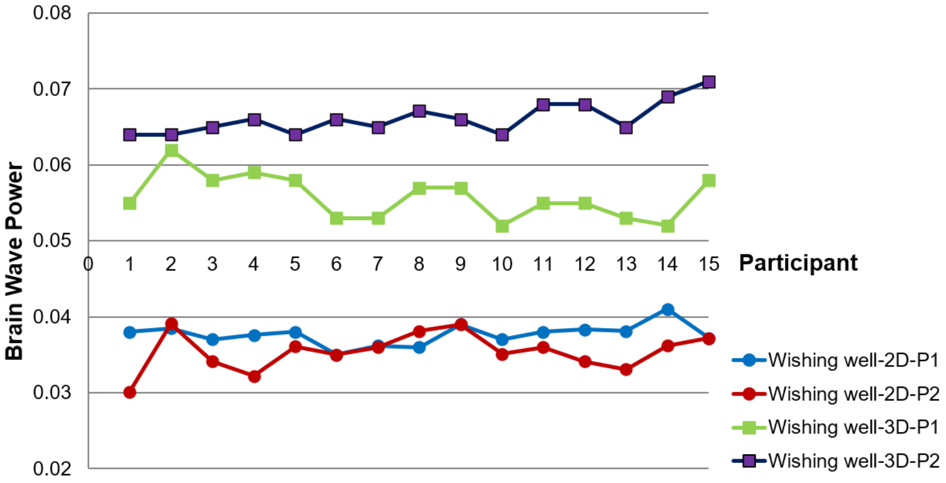


Figure 3.8: Brain wave power ratio distribution for the “Wishing well” video sequence.

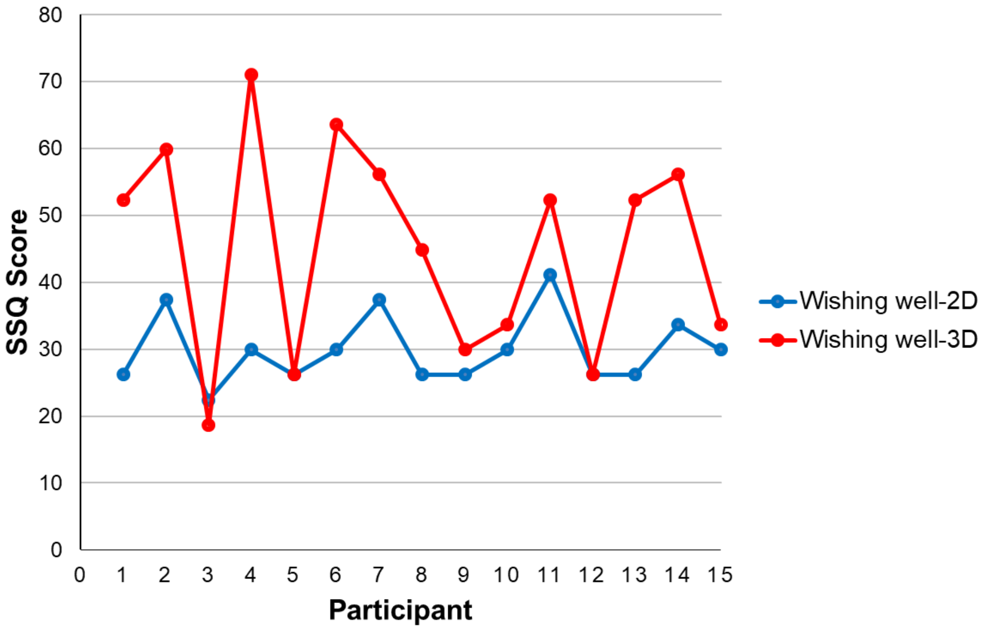


Figure 3.9: SSQ score distribution for the “Wishing well” video sequence.

3.3.4 Discussion

Previous studies have proposed a series of methods to assess eye fatigue whilst watching videos [137, 161, 176]. However, eye fatigue has been less investigated for S3D videos. Previous studies have focused on how an S3D projection screen causes VIMS to viewers. In this thesis, SSQ has been used to assess VIMS symptoms through an experimental evaluation. Results presented in Section 3.3.1 show that none of the participants felt any VIMS after the experiment because all the SSQ scores are less than 50, indicating none of the symptoms according to Table 2.2 in Section 2.4.3. However, the majority of the participants exhibited eye fatigue when participants watched video sequences for a prolonged period of time. Thus, in this work EEG signals are used to quantitatively measure eye fatigue by assessing visual information transmission [177].

3.4 Conclusion

This chapter proposed and investigated the use of EEG biosignals to measure visual and eye fatigue to augment VIMS measurement through participants' self-reporting using the SSQ. Experiments were conducted to compare fatigue caused by 2D and S3D video sequences of the same video content. Experimental results indicate that the 3D video sequences caused more fatigue for the participants than the 2D video sequences. Further, participants exhibited more eye blinking movement for 2D videos, indicating less eye and visual fatigue. Congruently, the brain wave power ratio results from the EEG biosignals showed that 3D videos caused higher power ratio values than 2D videos, and larger than the fatigue threshold value of 0.05.

Previous researchers have shown that 3D videos cause the most VIMS for viewers, and the results in this chapter confirm these same outcomes. Most VIMS and fatigue were seen whilst watching the "Wishing well" video sequence.

In future work, an extensive SSQ should be conducted to evaluate and analyse the fatigue level with the help of biosignals [178]. In addition, a medical-grade EEG apparatus is capable to operate at least 16 contact points for simultaneous brain wave collection. As a future extension to this work, this may give more accurate results for the EEG recordings and fatigue analysis. Also, future studies may consider other factors on stereoscopic effect in human health, such as Zernike polynomials score, Modulation Transfer Function Index (MTFI) analysis and higher-order root mean square [179-181]. The sequence of videos watched can also impact the fatigue level of a viewer. Therefore, more investigations can be conducted to see if a broader range of content and shorter video sequences has an impact on the visual fatigue levels. Furthermore, further extended studies may include the SMPTE and THX specifications for the 3D fatigue assessment of UHD-1 video quality and corresponding parameters. Regarding the impact of the viewing device, the study in this chapter used a passive flat screen device. The study did not include any projection or VR devices, and this is explored in the work of the next chapter.

Chapter 4 Comparison of S3D Video QoE in Different Viewing Environments

4.1 Introduction

Engagement in 3D video content has increased since the use of virtual reality (VR) devices such as Head Mounted Display (HMD) headsets, especially among university and college students [182]. VR refers to the environment created by a computer simulation in which a person can interact with the environment through electronic devices such as goggles enabled with a screen. In such immersive multimedia environments, stereoscopic imagery further enhances the sense of immersion and reality, compared to using only 2D imagery alone.

Despite the wide use of projectors in immersive multimedia environments, research has shown that the majority of viewers experience some degree of visual fatigue [80]. This problem has also been associated with exposure to stereoscopic images and videos. Viewers have reported having experienced headaches, eye strain, dizziness and nausea, which are the symptoms of visual fatigue [183-185]. Prolonged use of a screen is known to induce ocular discomfort [137, 186], and excessive utilisation of VR with an HMD may result in a major clinical disorder [187]. In this regard, various QoE assessments towards 3D content have been developed, hence motivating further research.

A study conducted by Quan et al. [74] evaluated the viewers' experience and perception of 3D content. The researchers hypothesised that 2D content can be used to differentiate the perception of 3D content. Studies have also suggested the use of a panoramic screen to obtain better visual attention for 3D content perception. Lee and Kim [186] found that compared to curved screens, flat screens cause greater visual fatigue effects. Hence, projecting 3D videos on a curved screen can reduce adverse effects on the eyes. Wegner et al. [18] suggested the projection of 3D videos on a curved circular panoramic screen as the best design specification, as shown in Figure 4.1. The challenges associated with such a design is the requirement of subjective quality experiments that must be performed before it is more widely implemented.

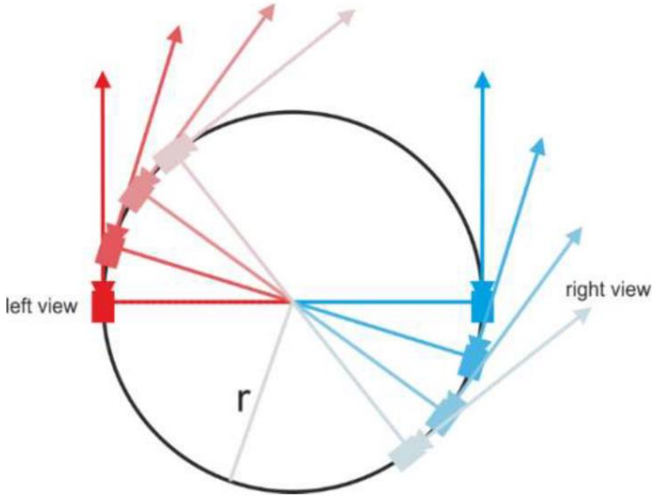


Figure 4.1: Circular projection by the rotation of cameras.

For more personalised use, HMDs have become popular due to their ability to display both 360° and 3D video content. Examples of HMDs include Facebook's Oculus Rift [49] and the HTC VIVE [50], both of which can project omnidirectional video contents. Due to the increased popularity of these HMDs, several researchers have conducted experiments to determine the QoE associated with viewing S3D videos on these platforms [188-190]. These previous studies used Simulator Sickness Questionnaires (SSQ) to evaluate the Visually Induced Motion Sickness (VIMS) by viewing S3D videos. Although some studies on QoE have been conducted, there are no studies to date that evaluate the QoE of S3D videos on two or more different viewing environments such as a panoramic screen [191], flat screen or VR headset.

To investigate viewers' QoE in different viewing environments, this chapter investigates and evaluates the effects of viewing S3D videos projected on different screens. The research sought to assess the associated 3D fatigue with the use of more than one projection screen ranging from a research-grade flat 3D screen, a 3D panoramic screen [191] and a consumer-grade VR headset for S3D videos. The effects of VIMS was measured through a series of subjective and objective evaluations. For the subjective evaluation, participants were recruited to view a set of S3D videos. As the participants viewed the videos, both their eye blinking signals and EEG signals were recorded, following the methodology outlined in Chapter 3 . The participants were then asked to complete an SSQ in which they rated the enjoyment level and the quality of the S3D videos that they watched. This QoE data analysis and evaluation is thus discussed in-depth in this chapter.

4.2 Background

4.2.1 Existing QoE experimental methodologies

A number of researchers [192-195] have conducted experiments to understand the QoE whilst viewing 3D videos on different types of screens and viewing environments. Such experiments can be categorised into three groups including adding stimulus for comparison, subjective evaluations among participants and image quality assessment models. Upenik et al. [192] used subjective evaluations to understand the quality of images projected on HMD displays, recommending the use of a testbed to perform the subjective tests in an omnidirectional environment on distinct projection screens. In another study, Sun et al. [193] suggested the use of a novel Image Quality Assessment (IQA) model from 360° image databases. The viewpoint images were used as the input and the output was omnidirectional content. This content was composed of six viewpoint images [193]. The evaluations identified that this model produces the best performance in the assessment of image quality. Narciso et al. [194] used an HMD to study how the video and sound format affects the user's cybersickness and sense of presence, where cybersickness is the feeling of dizziness or nausea while presence indicates the "sense of being there" in a VR environment [196]. However, the statistical results did not reveal any significant difference in the cybersickness and sense of presence between the sound and video variables. To further investigate the phenomenon, Narciso et al. [195] recommended adding smell as an additional sensory stimulus to measure how the various VR environments affect fatigue, stress, sense of presence, transfer of knowledge, and cybersickness of the viewer. However, the results found that smell and the measured variables were not correlated. Proposing the use of biosignal-based evaluations, the study by Duan et al. [197] was motivated by the knowledge that omnidirectional videos and images offer an immersive experience in a VR environment. Hence, the researchers sought to track the movement of the eye and the head during a subjective quality rating experiment. Duan et al. [197] suggested the establishment of an omnidirectional IQA database comprised of 320 distorted images and 16 source images. A subjective quality evaluation was then conducted on the database in a VR environment. A comparison was then made between the existing IQA databases and the created database. The results showed that high frequency content and image content in VR environments were more preferable to viewers, and the loss of the image detail affected the VR visual experience.

Singla et al. [190] conducted a study to compare the QoE of omnidirectional content displayed and viewed on various HMDs. The researchers identified three classifications of omnidirectional content namely: low, medium and high degrees of content and camera motion. The authors further evaluated the three content types and established that they contribute significantly to the quality rating of the video content and resolution. To further understand cybersickness, feeling of presence, and perceptual quality of omnidirectional content in VR environments, Anwar et al. [196, 198] conducted and compared the subjective evaluation of two high priority QoE factors: the interest and the familiarity of a user in a VR environment. The two factors were evaluated using the Absolute Category Rating (ACR) method. The impact of these factors was evaluated using the QoE prediction methods from Artificial Neural Network (ANN) under different situations on the cybersickness level of users. The results found that the prediction accuracy rate of the proposed method is 90% compared to other existing prediction models. Higher scores on QoE levels are rarely achieved since cybersickness is a common barrier that creates discomfort [199, 200]. The cybersickness and strain arise as a result of the user being fully immersed in the content.

In regard to the quality assessment of S3D videos, Zhang et al. [201] performed two assessments on panoramic videos. The assessments were subjective and objective quality assessments of videos encoded on distinct bitrates and noise at different resolutions. Data on the subjective quality assessments showed a variance of subjective perceptions between the panoramic videos and normal videos when the bitrate was varied. In their study, Appina et al. [202] suggested the use of 288 test videos obtained from 12 “pristine” S3D videos; the term “pristine” referred to a video that was uncompressed. The researchers used novel subjective quality and objective quality prediction tests, where pristine videos were chosen for the subjective quality tests with reference to spatial information, motion information and disparity. A 6-point scale rating was used in the experiment; however, the QoE assessments were not further elaborated.

4.2.1.1 VIMS measurement

As explored in Chapter 3 , exposure to S3D videos causes VIMS and visual fatigue. Owing to this observation, Naqvi et al. [23] proposed methods to assess VIMS caused by S3D videos and made a comparison of the ratio of high- and low-frequency components in the video content. For the comparison, the high-frequency component was classified as a frequency that exceeded 0.15 Hz while the low-frequency component was classified as a frequency below 0.15 Hz. The high- and low-frequency components defined above are corresponding to parasympathetic activities and sympathetic modulations respectively. However, the proposed method included the S3D flat screen only.

Sickness arising from exposure to screens such as VR HMDs has been evaluated by various researchers using SSQ [203]. In this regard, VIMS can be assessed using SSQ and the results correlated to visual fatigue. Tychsen and Foeller [203] found that the impact of being exposed to VR displays can be studied more effectively using the SSQ method. The results also found that young children may view immersive 3D video game content using current HMDs without deleterious effects on visuomotor property. In contrast, Duan et al. [19] conducted an experiment to evaluate the VIMS of immersive videos of actual scenes by controlling visual oscillations. Experimental results showed that the increase of the frequency of visual oscillations can increase the level of VIMS. Furthermore, Wang et al. [176] investigated eye fatigue compared to the VIMS caused by prolonged exposure to an HMD. The researchers used eye-tracking methods to assess eye fatigue and SSQ scores to assess VIMS. After the analysis of eye fatigue was completed, Wang et al. [176] also developed an assessment model to assess eye fatigue using HMDs.

4.2.1.2 Statistical analysis approaches

In previous studies [160, 168, 204], ANalysis Of VAriance (ANOVA) has been suggested for the statistical analysis of EEG signals. Trejo et al. [156] compared and analysed EEG signals using an ANOVA test to evaluate the cognitive fatigue during sustained mental work in a controlled environment. To study the features that characterise EEG signals, Salami et al. [205]

used a Common Spatial Pattern (CSP) algorithm through a Brain-Computer Interface (BCI). They utilised the Unbalanced Factorial ANOVA to understand the feature vectors of EEG signals, and the ANOVA table was used to obtain the F-distribution parameter through linear regression.

Mouzé-Amady and Horwat [168] identified three Hjorth parameters namely complexity, mobility and activity. Mehmood et al. [160] processed the above three parameters by performing emotion recognition experiments that involved recording the EEG signals. The signals were then analysed using a one-way ANOVA and the peak EEG features were selected for different EEG frequency ranges. Measurement of the EEG signals allowed the researchers to obtain information on eye blinks, attention, meditation and brain wave powers. Zou et al. [90] then proposed brain wave power equations that were used to calculate the visual performance level. It was established that EEG signals yield more accurate results than typical subjective rating in the assessment of visual fatigue.

4.2.2 Different S3D viewing environments

4.2.2.1 Flat screen 3D monitor

As discussed in Chapter 3 , 3D flat screens may cause visual fatigue to the viewer, especially those that are affected by epileptic response and the screen can cause eye fatigue and headaches due to the viewing requiring glasses [20, 206]. Additionally, most images displayed on 3D screens required a higher resolution which means that they need more space for data storage and also they take a longer period of time to process this data for display compared to 2D videos [207]. Even though there are drawbacks to viewing S3D videos on flat screens, according to Meitzler et al. [208], when it came to rating the quality of experience when watching videos from 3D screens, the viewers had an increased sense of enjoyment as compared to other viewing devices. It was noted that the enjoyment resulted from longer viewing distances, larger viewing screens and a large space in the viewing area.

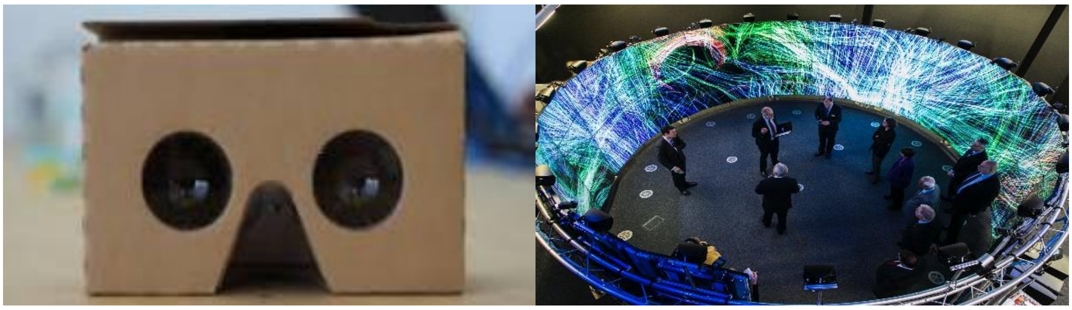
4.2.2.2 Curved and panoramic 3D screens

Urakami et al. [209] posited that curved screens create a better feeling of immersion than flat screens. However, since no studies had yet been conducted to evaluate the effects of curved screens on immersiveness, Urakami et al. [209] sought to compare the immersion experience of curved and flat screens. It was found that curved screens provide a better speed and accuracy than flat screen displays in a visual search task. Further, Kyung and Park [210] evaluated the effects of screen curvature radius and the screen size with the visual search accuracy, visual fatigue and visual search speed. They compared the experience using a curved screen and a flat screen, showing that when the screen size is increased, the readability of the texts and the degree of visual fatigue on a flat screen was affected. However, it was not affected in a curved screen. In this regard, they proposed that large curved screens are better and more appropriate for use in multi-purpose applications. In future, more visual devices such as split screens and the layered effects and transitions will emerge as hypothesised by Bizzocchi [211].

Panoramic screens differ from traditional flat or curved screens because they are able to present panoramic views and have a wide screen. Zhang et al. [212] described panoramic display screens as the most significant media content in VR. The screens represent a video that displays continuous content in every direction. Images and videos of a panoramic nature are taken simultaneously using more than two cameras [213]. In that regard, each camera captures a scene in a certain direction separately with specific geometric constraints. All the images are then stitched together to create a panoramic scenario. A large panoramic image with a wide field of view is thus seamless and comprises many overlapping narrow Field Of View (FOV) images [214].

To display a 360° FOV, the imagery is then rendered assuming that the human head is at the central point. Thus, using a panoramic display screen, the viewer can see the image in several directions, thus enabling a 3D-like experience. In that regard, panoramic screens are part of multi-projector display systems that generate high-resolution displays for wide field of view images.

Currently available panoramic screens include consumer headsets comprised of a regular screen and a plugin to allow navigation with a mouse. The panoramic screen can thus be an HMD that allows the viewer to be immersed in a virtual world. Other devices to display panoramic video include the low-end single-viewer Google Cardboard [213], to high-end multi-viewer Cave Automatic Virtual Environment (CAVE) [215, 216] shown in Figure 4.2. However, the majority of panoramic videos and content are currently viewed on HMDs [215].



(a) (b)

Figure 4.2: Devices to display panoramic video (a) Google Cardboard, (b) CAVE-like environment [216].

Panoramic screens present several benefits to viewers. Firstly, they cover a wide area of view, even from far distances. Folen et al. [217] hypothesised that a panoramic immersive video provides the viewer with an efficient yet cost-effective environment. Although the study found a difference in outdoor and indoor experiences, the use of Oculus Rift VR equipment showed no difference [218]. In another study conducted by Sabarudin and Tiau [219], digital panoramic systems produce better and improved images. Folen et al. [217] examined the psychophysiological effects from a panoramic screen compared to a flat screen, where participants wore an HMD with a panoramic screen display. The results showed that the omnidirectionality of panoramic videos allow viewers to view the video by focusing on what interests them the most. However, in such videos, it is not possible to view the video entirely due to the wide FOV.

The limitation of panoramic videos is the limited transmission bandwidth because of the large amount of data. Due to this limitation, at present the videos must be compressed before being sent to viewers. The application of existing coding approaches then becomes a challenge when mapping the videos. One has to map the panoramic video onto a plane in regard to the geometric transformation rules. After mapping, the panoramic video is then rendered into a sphere to create a display suitable to the viewer. On the other hand, panoramic screens have to display panoramic video in its entirety regardless of the varying resolutions, which is different from the traditional flat screen. In addition, panoramic display screens give the viewer an immersive experience, whilst an HMD maintains a fixed and consistent position with respect to the human eye. If the fixed and consistent position is missing, the viewer's experience becomes affected, and VIMS may develop [212].

4.2.2.3 VR HMDs

Desai et al. [220] defined VR as a computer-simulated environment that allows the user to experience that environment. Juanes et al. [221] described VR as a computer system that creates an artificial world where the user possesses the sense of being in that world and the ability to move and manipulate present objects. VR headsets are often made of a stereoscopic HMD, head motion tracking sensors and stereo sound. The tracking sensors consist of accelerometers, gaming controllers, structured light systems, magnetometers, eye-tracking sensors, and gyroscopes. The virtual reality of the VR headset is facilitated and enhanced by interactive devices such as helmets and gloves. VR devices have gained wide application in video games, and simulators.

In recent research, Urakami et al. [209] indicated that curved displays can create a feeling of immersiveness that is similar to a VR environment. For VR related devices, their suitability is based on the capacity to take the user into the world immersively. Immersiveness is an illusion that gives the user the impression that they are drawn into the image or that they are present in a scene that is far away from them.

However, there are some challenges with VR headsets. When VR content is shown in HMDs where the user can notice a latency in the VR system, this can cause either a headache or nausea [222, 223]. Further, the weight of an HMD may be an issue for some viewers [223]. As stated by Craig et al. [223], other disadvantages of HMDs are that their FOV is narrow and the user is isolated from the people surrounding them, which makes the present viewing experience challenging.

The choice of VR equipment used when watching S3D videos is thus critical. Various commercial VR hardware platforms such as the Facebook Oculus Rift [49], HTC VIVE [50], FOVE [224] and Google Cardboard [225] provide varying levels of VR experience. Hence, different VR hardware results in a different user experience. Table 4.1 shows the specifications of the resolution per eye, refresh rate and FOV from the different commercial VR devices mentioned above [226]. In Table 4.1, Google Cardboard was not considered for use due to the resolution constraint of the required smartphone, relatively low refresh rate and FOV compared to the other VR devices. The HTC VIVE and the Oculus Rift are the two leading PC VR headsets on the commercial market [227]. Singla et al. [190] also compared the QoE and SSQ of these two VR headsets. Further, the FOVE VR headset has an embedded eye-tracking function that can automatically record eye-gaze data to be used for QoE in a VR environment; however, the headset is now rare in the commercial market.

Table 4.1: Specifications of different commercial VR equipment.

Brand of VR headset	**Resolution per eye**	**Refresh rate (Hz)**	**Field Of View (FOV) (degrees)**
HTC VIVE	1080×1200	90	110
Oculus Rift	1080×1200	90	110
FOVE	1280×1440	70	100
Google Cardboard	Depends on smartphone	60	80

4.2.2.4 Oculus Rift and user experience

Parisi [228] described the Oculus Rift headset as a spectroscopic display that has a built-in head motion tracking sensor. Figure 4.3 shows the Oculus Rift equipment which consists of a headset, an infrared sensor and a joystick controller. It can create stereoscopic vision that possesses excellent depth, parallax and scale. Juanes et al. [221] created a virtual reality experience on the platform with artificial environment capabilities in a medical operating room. The aim was to expose the user to the perception of virtual situations with the aid of Oculus Rift to depict its immersive potential. The VR device possesses stereoscopic vision, which allows for the training of students to tend towards reality. The virtual space allows for the creation, modification, and arrangement of various 3D objects. The software development tools of the Oculus Rift can develop virtual environment platforms for both research and educational purposes.

Figure 4.3: Oculus Rift headset with infrared sensor and controller [49].

Lee et al. [229] considered Oculus Rift to be HMD technology suitable for VR games. According to Bailenson [230], the VR technology of Oculus Rift has led to an improvement of the video quality and the creation of an immersive medium in which the user experiences a sensation of being fully transported into a virtual 3D world. Compared to screen-based media, Oculus Rift gives the user a more instinctive experience. Tan et al. [231] highlighted that Oculus Rift gives users an improved VR experience compared to previous HMDs; the device has the potential to immerse the user into emotional levels and a feeling of sense of reality.

Tan et al. [231] also found that users can naturally look around without using controls. However, after exploring the experience of ten participants, it was found that the users experienced cybersickness despite heightened experiences. Also, the user experienced a deeper immersion and a higher degree of “flow”. Users also experienced a large range of head motion to navigate through the game environment; however, the gaming experience was negatively affected by such excessively demanding movements [230].

4.2.2.5 HTC VIVE and user experience

The HTC VIVE equipment has a headset, two infrared laser emitter units (Lighthouses) and two controllers, as shown in Figure 4.4. The laser emitters are the tracking technology that the device relies on. They alternately send both vertical and horizontal beams that span 120° in each direction [227]. The device can control the motion using motion controllers and external sensors for the entire VR room. With the aid of built-in motion sensors and external sensors, the headset accurately tracks the head motion [227]. HTC VIVE also provides the user with realistic and precise graphics with 360° headset tracking capabilities. The device also offers HD video with directional audio for interaction in the virtual environment [232]. Niehorster et al. [233] described HTC VIVE as an appropriate device that gives the user a large FOV and high-resolution content. Also, the device has a room-scale tracking system, which is sufficient in both precision and accuracy [233].

Figure 4.4: HTC VIVE headset with infrared sensor and controller [50].

4.2.2.6 Selection of VR device: HTC VIVE vs Oculus Rift

Both HTC VIVE and Oculus Rift are the biggest competitors in the market and both devices are tethered VR headsets. Borrego et al. [234] conducted a study to compare the HTC VIVE and Oculus Rift. The researchers found that although both VR devices had an excellent and comparable performance at sitting height (1.3 m), HTC VIVE had a working area two times larger than that of the Oculus Rift. The VIVE device had a larger working range of 7 m [234]. HTC VIVE has also been found to provide an experience with a higher degree of immersion [235]. In their study, De Paolis and De Luca [236] also investigated the QoE of both Myo (a gesture control armband controller for the Oculus Rift, which is shown in Figure 4.5) and the VIVE controller. The findings revealed that the VIVE controller has better usability results compared to the Myo. Since these studies, newer technological developments such as the VIVE Pro may offer viewing experiences surpassing Oculus Rift in future, already shown to have the highest comfort and clarity [227].

Figure 4.5: Myo – Gesture control armband controller for the Oculus Rift [237].

4.3 Experimental Methodology

This section is grouped into three sub-sections which relate to three different viewing environments that are commonly used as the viewing environments of S3D videos: a VR headset device, a panoramic screen and a flat screen. Thus, in this thesis, two types of panoramic screens were used in the experimental work: a composition of several curved screens to form a panoramic screen, with multiple projectors to project S3D videos (described in Section 4.3.2); and, a VR HMD (described in Section 4.3.3).

Figure 4.6 below illustrates the methodology used to conduct the experiment. The participants were given five sets of S3D video sequences, watching each sequence for a duration of one minute each. Sequence bias was minimised by playing the S3D videos in a random order. The random order of the S3D videos was generated using the Stat Trek [238] Random Number Generator.

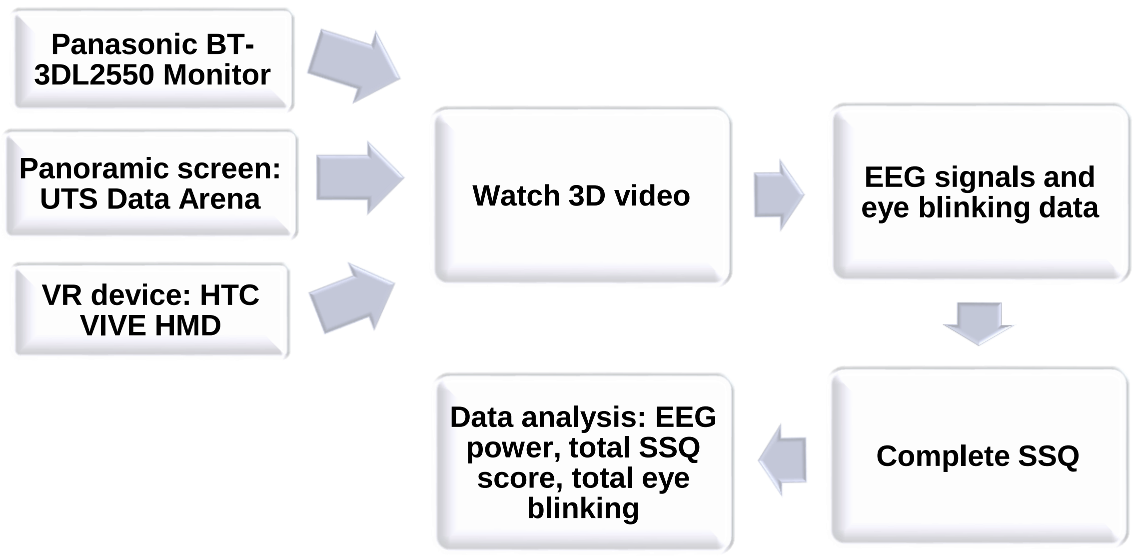


Figure 4.6: Block diagram of the experimental methodology.

The previous work in Chapter 3 has identified that watching S3D videos can cause motion sickness and fatigue; thus, participation time was divided into three sessions, each lasting approximately 30 minutes with a 10-minute break between each session. The average duration of the whole experiment for each participant was 1.5 hours, inclusive of the rest breaks. Such an average duration was essential in minimising both fatigue and VIMS. For ethical experimental conduct, participants who suffered from dizziness, physical or motion sickness were given the freedom to discontinue the study.

4.3.1 Device 1: Flat 3D screen

The flat 3D screen used in the study was a 25.5'' Panasonic BT-3D L2550 Full HD (1920×1080) LCD 3D screen, the same used as in Chapter 3 . The five S3D video sequences were played on the flat screen and the participants' responses obtained. According to the THX Cinema Certification [87], the participants should sit 0.9 m in front of the S3D screen and a viewing angle of 36° when viewing 3D videos, as shown in Figure 4.7. The equipment and the view of S3D video settings are the same for the flat 3D screen experiments conducted in Chapter 3 in order to maintain a controlled experimental environment.

Figure 4.7: A participant wearing the NeuroSky MindWave EEG headset to watch 3D movies on a flat 3D screen.

4.3.2 Device 2: 3D video projection on panoramic screen

In this experiment, the panoramic screen used was the Data Arena at the University of Technology Sydney (UTS) [191]. The room is hemispherical in shape with a diameter of 10 m and consists of a large panoramic screen and six 3D stereo panoramic projectors arranged in a circular position as shown in Figure 4.8. The S3D videos were projected to six 3D stereo panoramic projections from one large panoramic screen. Table 4.2 below shows the specifications of the Data Arena used in this study. Each of the participants wore active shutter 3D glasses and stood in the middle of the panoramic screen to view the stereoscopic videos, as shown in Figure 4.9.

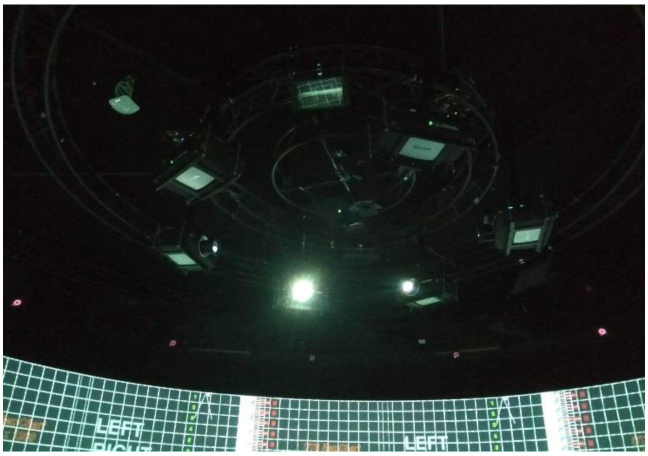
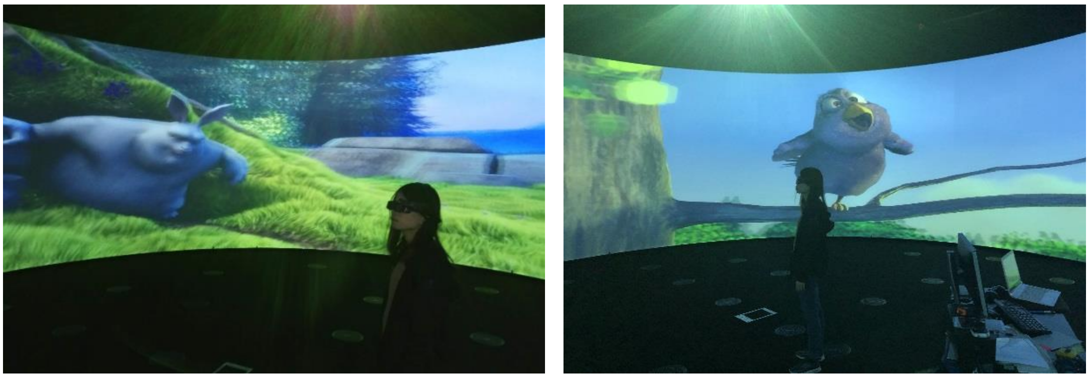


Figure 4.8: Six 3D stereo panoramic projectors arranged in a circular position.

Table 4.2: Specifications of the UTS Data Arena [191].

Dimension	Hemisphere
	Diameter: 10 m
	Height: 4 m from the floor to projectors
Screen Dimension	Height: 3.5 m, panoramic
Resolution of video projector	1920×1200



(a) (b)

Figure 4.9: A participant wearing the NeuroSky MindWave EEG headset and active shutter glasses to watch 3D movies in the Data Arena (a) Moving, (b) Standing at the centre.

The video walls and the six S3D video projectors were driven by a high-performance computer graphics system. The rendering process, graphic computation and reconstruction of the 3D images were achieved through the use of software tools such as Houdini [239], Equalizer [240] and FFmpeg [241]. For this experiment, the 360° surround panoramic screen used six projectors that displayed three identical, edge-blended 1920×1080 S3D video sequences. A work station, shown in Figure 4.10, was the main control for the experiment within the arena space.

Figure 4.10: A work station to control the panoramic screen in the Data Arena.

4.3.3 Device 3: 3D video on 360° in a VR headset

Taking into consideration the findings from previous researchers' studies comparing the HTC VIVE and Oculus Rift as discussed in Section 4.2.2.6, the HTC VIVE was adopted for the experimental work in this thesis. The HTC VIVE headset has a FOV of 110°, a refresh rate of 90 Hz and a total resolution of 2160×1200 pixels. The manufacturer suggested setup of the HTC VIVE is shown in Figure 4.11 for a standing-only setting. Therefore, body movements and the synchronicity of the stimuli were restricted to ensure that the participants sat down when watching the S3D video sequences. However, the participants were allowed to turn their heads to achieve varying viewing positions. Hence, while the limited body movement can affect the user experience, the movement of the head does not interrupt but enhances the viewing experience. The experimental setup is shown in Figure 4.12.

Figure 4.11: The experimental setup of HTC VIVE using a standing-only setting [50].

Figure 4.12: A participant wearing the HTC VIVE HMD with NeuroSky MindWave EEG headset to watch 3D movies using a VIVE controller.

4.3.4 EEG devices

The experiment required that participants use different viewing environments and rate the experience in each environment. In addition, participants' EEG biosignals were recorded for continuous-time QoE measurement. A NeuroSky MindWave brainwave headset was used, which was used in Chapter 3 and described in Section 3.2.2, and both the eye blink rate and the EEG signals were recorded using the NeuroSky software provided.

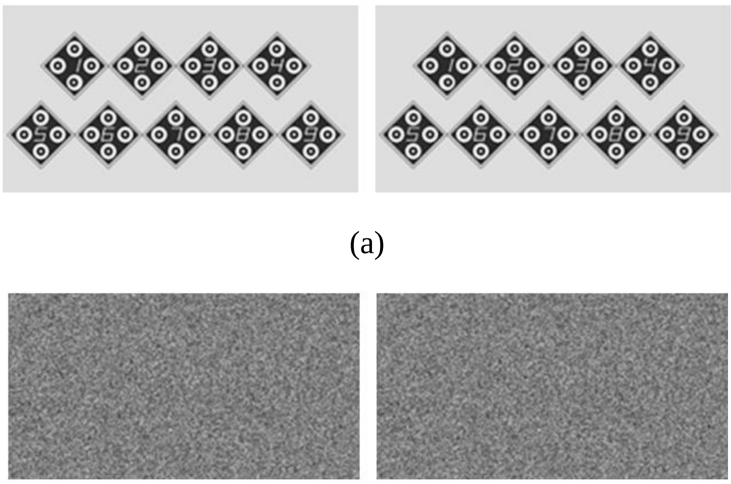
4.3.5 Video stimulus and metrics

All S3D video sequences chosen for this experiment are the same as used in the flat 3D screen experiments conducted in Chapter 3 for consistency in the experiential video sequences tested.

To measure the VIMS, each of the participants was required to complete the SSQ after viewing each of the five S3D video sequences, as per the methodology in Chapter 3 . In addition to the SSQ and EEG biosignal measurement as described in Chapter 3 , the Absolute Category Rating (ACR) method was added to perform the subjective assessment of the S3D video quality in accordance with ITU-T P.910 [81] recommendations. A scale of five levels (1: not enjoyable at all, and 5: very enjoyable) was used to rate the enjoyment of the video watched. Based on this, participants were asked to rate the perceived enjoyment quality of the S3D video on the scale provided.

4.4 Results and Discussion

A total of fifteen people participated in the viewing experiment, consisting of 4 females and 11 males. The participants' ages were between 18 and 46 years and the mean and standard deviation of the years of the participants' ages were 29.1 years and 8.67 years respectively. Before the start of the experiment, each participant was asked about their fatigue level to ensure that everyone recruited was not fatigued. Participants were then required to complete a short 3D vision test to ensure that their stereo vision was appropriate for the study. The vision test for screening included fine stereopsis and dynamic stereopsis tests of 3D vision, as shown in Figure 4.13. The stereoscopic test materials for the test are available in the ITU-R BT2021.1 standard [25]. The questionnaire of this experiment can be referenced to Appendix 2.



(a)

(b)

Figure 4.13: 3D vision tests (a) Fine stereopsis, (b) Dynamic stereopsis, showing the left and right images.

4.4.1 SSQ results

The results of the SSQ scores from each viewing environment are shown in Figure 4.14 and Figure 4.15 (95% confidence intervals shown). Figure 4.14 shows the average total SSQ score for each video sequence and for each viewing environment. Figure 4.15 shows the SSQ scores with three sub-scales for each video sequence for the three different viewing environments. The SSQ sub-scales are for the disorientation, oculomotor, and nausea produced by each viewing device when used to watch each of the S3D videos provided. The results from Figure 4.14 and Figure 4.15 also show that both SSQ scores and their associated sub-scales for the "Water fountain" and "Garden" video sequences viewed in the Data Arena are zero.

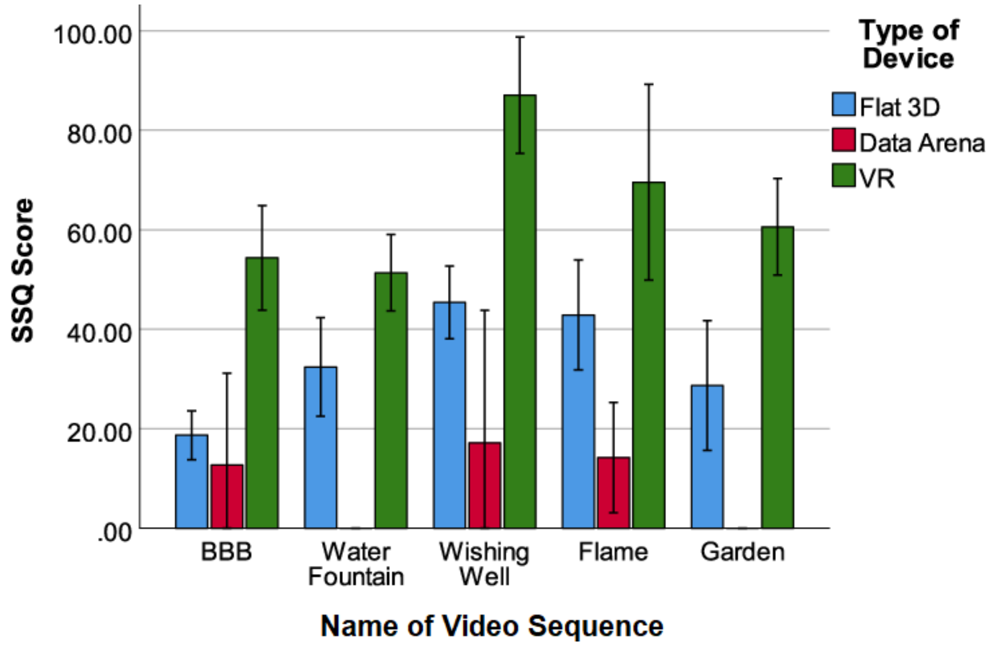


Figure 4.14: Total SSQ scores.

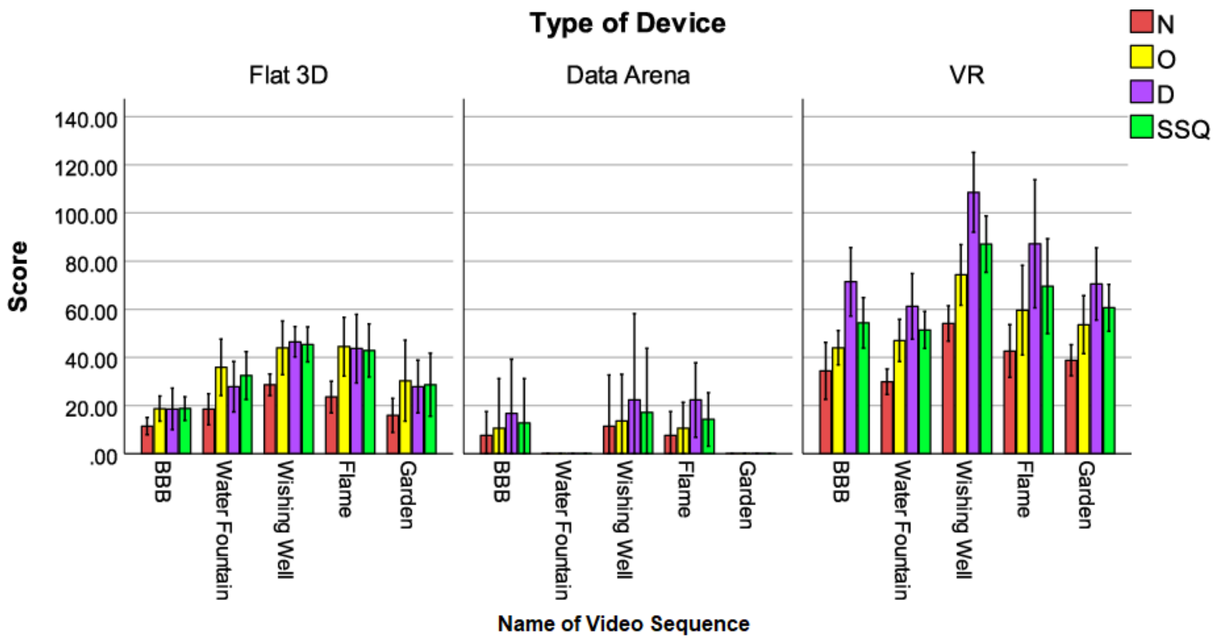


Figure 4.15: Total SSQ scores with 3 sub-scales: Nausea (N), Oculomotor (O) and Disorientation (D).

From Figure 4.14, it can be noted that the VR device had the highest total SSQ scores followed by the flat 3D screen and lastly the panoramic screen. For all video sequences the total SSQ score for the VR device was more than 50, whereas the scores for the flat 3D screen were between 20 and 50 and the panoramic screen had scores lower than 20. Since SSQ scores depict the likelihood of VIMS, it implies that the VR device has a higher possibility of causing VIMS than the panoramic or flat 3D screens. Panoramic screens have the least capacity to cause VIMS with the "Water fountain" and "Garden" video sequences having the lowest scores at zero.

For the "Flame" and "Wishing well" video sequences from all the viewing environments, the SSQ scores were higher compared to the "Garden" and "Water fountain" video sequences. These results suggest that chaotic movement of the "Flame" and "Wishing well" video sequences may correlate with higher SSQ scores. After a single-factor ANOVA was conducted for the SSQ scores, it was revealed that the factor was F = 2.78 and $p < 0.01$. Therefore, the factor was significant. However, none of the participants identified as experiencing motion sickness symptoms during the study for any of the viewing environments.

4.4.2 EEG brain activity

Similar to the work in Chapter 3 , the brain wave power ratios for the five video sequences and the three different projection devices were recorded and analysed. The results of the analysis are shown in Figure 4.16 below. The graph shows P1 and P2 which represent the calculation results using Equations (3.1) and (3.2) from Section 3.2.2 respectively.

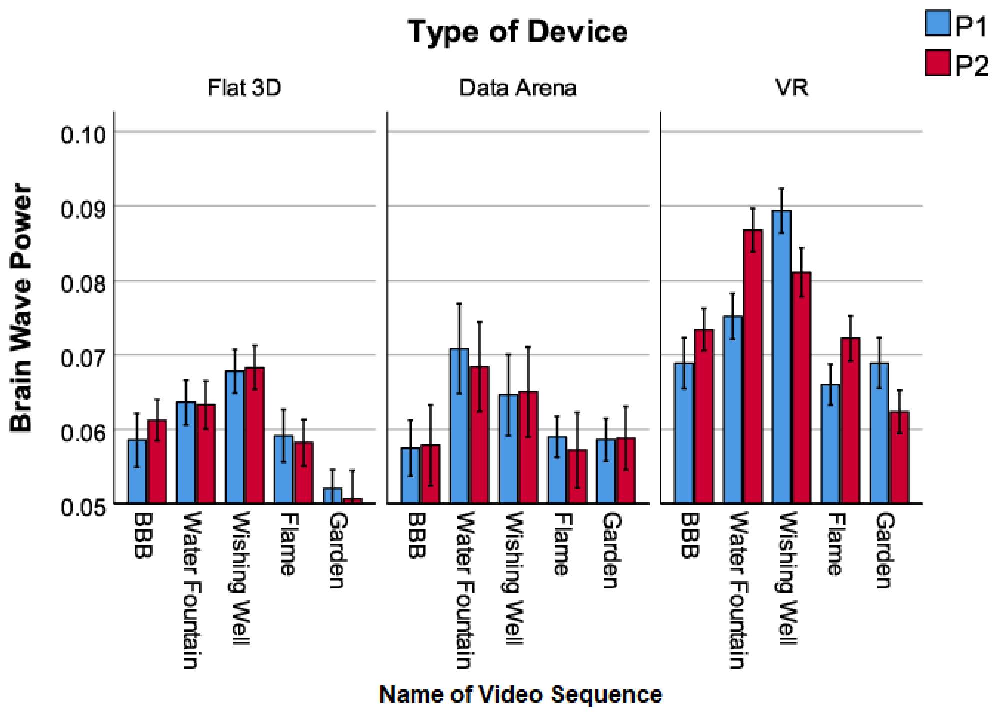


Figure 4.16: Measured brain wave power ratios for various viewing environments and S3D videos.

For all the video sequences under test, the average brain wave power ratios ranged from 0.05 to 0.09. These results were greater than 0.05, hence showing that all the participants exhibited a higher level of 3D fatigue [36]. From the results in Figure 4.16, it is evident that the VR device resulted in brain wave power ratios ranging between 0.0623 and 0.0893. These are the highest levels of brain wave power ratios stimulated across all five video sequences. On the other hand, the panoramic screen and flat 3D screen caused similar brain wave power ratios: the brain wave power ratios for the panoramic screen ranged from 0.0507 to 0.0683 while for the flat 3D screen they ranged between 0.0572 and 0.0708. When comparing the video sequences that produced the highest fatigue levels, it can be seen that the "Wishing well" video sequence resulted in a brain wave power ratio of 0.0893 for the VR device and 0.0683 for the flat 3D screen. The results show that the "Wishing well" video sequence caused the highest fatigue level. For the panoramic screen, the "Wishing well" video caused the second highest and highest fatigue levels for the "Water fountain" video sequence with fatigue levels of 0.0650 and 0.0708, respectively. The most significant differences between the brain wave power ratios were recorded when participants were watching "Wishing well" with a VR device. This result was reflective of the various emotional responses exhibited by the participants [175]. The two brain wave power ratios were analysed by performing a single-factor ANOVA. The analysis gave the factors from the two brain wave power ratios as F = 1.85 and $p < 0.01$ and F = 3.23 and $p < 0.01$ respectively, showing that the two brain wave power ratios were found to be significant.

4.4.3 Eye blink frequency, attention and meditation levels

In Figure 4.17 below, the average eye blink frequency, as calculated from the EEG of the participants is shown when watching the different video sequences on each viewing device.

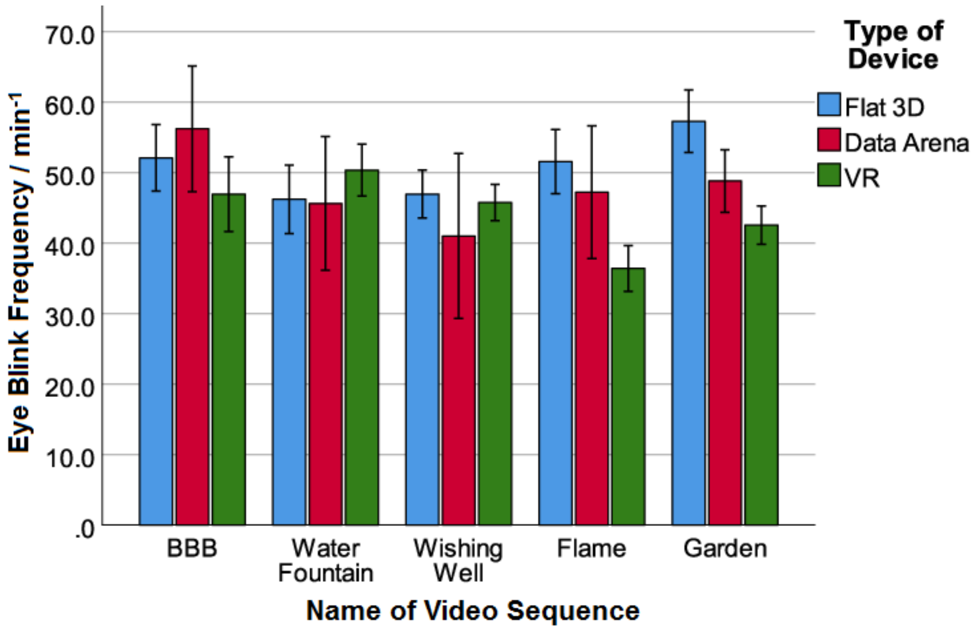


Figure 4.17: Eye blink frequency.

For all the participants, the flat 3D screen produced an eye blink frequency over a one-minute duration ranging from 46.2 to 57.3 times per minute. These values were greater than 36.4 to 50.3 times recorded when the VR device was used to measure the same parameter. The SSQ scores in Figure 4.14 and the brain wave power ratios in Figure 4.16 correlate well with the eye blink results, indicating that the flat 3D screen may cause a lower level of visual fatigue compared to the VR device. Also, the eye blink frequency was greater with the flat 3D screen than the Data Arena except for the "BBB" video sequence. This result may be due to a lower level of visual fatigue for the different scene content, as the "BBB" video sequence is an animation. The results also revealed that the confidence intervals from the Data Arena were larger than for the other devices (at least more than ±5 for each video sequence), which may also affect the accuracy.

The lowest average frequency of eye blink was recorded for the “Wishing well” video sequence, and may be due to the movements of the objects underneath water, the natural lighting conditions and the unstable water movements (consistent with the results presented in Chapter 3). When watched with the VR device, the “Flame” video sequence had the lowest frequency of eye blink at 36.4 times per minute. This may be a result of the rapid movement of the flame which is the main focus point in the scene. Several researchers have conducted studies to investigate if there is any implication of eye movements and eye blinks when viewing chaotic scenes [108, 242].

The attention and meditation levels of the participants whilst they were watching the five video sequences on the three different viewing devices were recorded with the NeuroSky headset and calculated using the NeuroSky software as shown in Figure 4.18 below.

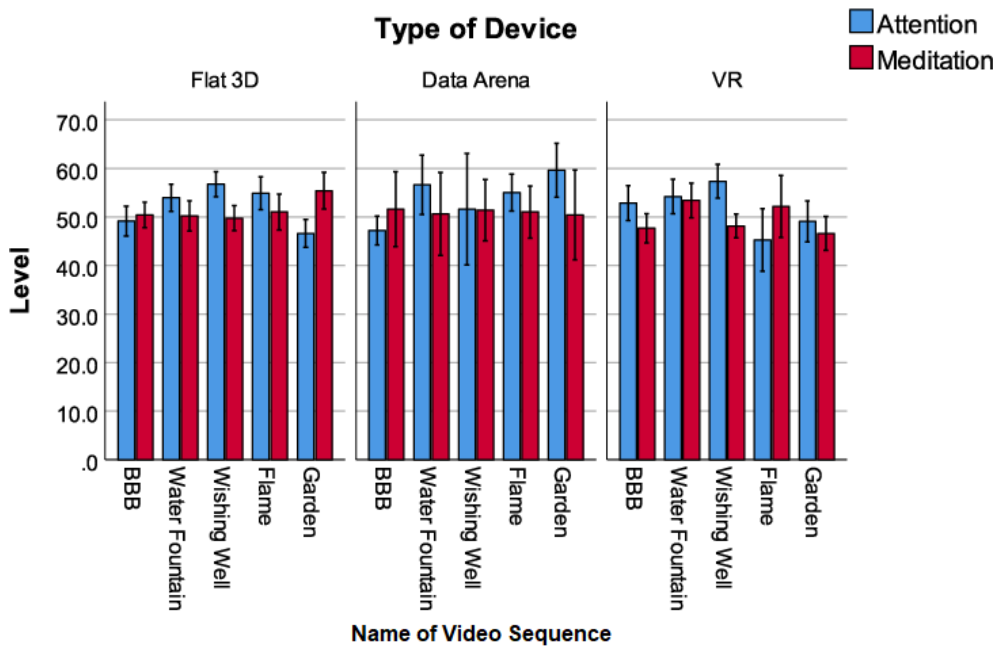


Figure 4.18: Attention and meditation levels.

The manufacturer's specification of the NeuroSky MindWave headset recording the meditation and attention levels was previously described in Section 3.2.3. From the results obtained in the current work, the attention level of the "Wishing well" and "Flame" video sequences had the highest readings of 56.7 and 54.9, respectively, when viewed on the flat 3D screen. When the other two environments, the VR device and the Data Arena, were used, a contrary trend was noticed. These results reveal that the viewing environment affects the correlation between attention and consciousness of viewers. A single-factor ANOVA of the attention and meditation levels and the eye blink frequency was conducted, and it was found that the video sequence and the viewing device are not significant factors affecting attention and meditation levels. This is consistent with Ulker et al. [243], who found that attention and meditation levels from the participants were in acceptable ranges because the viewers can recognise the content.

In addition, it is known that VR technology increases the levels of meditation and attention. Navarro-Haro et al. [244] explored VR technology to capture the attention of participants and to give them the illusion of being there in a 3D computer-generated environment, such as a panoramic screen or VR headset. Studies have found that before participants use VR environments, they should be prepared through mindfulness or meditation. According to [244], VR technology can also improve the state of mindfulness and reduce negative emotional states. Kosunen et al. [245] found that a VR environment induces deeper relaxation, a deeper level of meditation and a feeling of presence.

4.4.4 Enjoyment rating of S3D video

In Figure 4.19 below, the enjoyment ratings obtained using the ACR from all the participants are reported. For all participants, the enjoyment rating of viewing video sequences using a flat S3D screen was found to be between 2.93 and 3.80. This range was higher than as found for the VR device, which was between 2.13 and 3.33. Also, the results show that the panoramic screen resulted in an enjoyment rating from 3.40 to 4.60. These results suggest that compared to the flat 3D screen and the VR device, the panoramic screen offers the highest enjoyment rating. The high enjoyment ratings can be attributed to a longer viewing distance, a larger viewing screen and perhaps also the novelty of the unique immersive space provided by the UTS Data Arena. In contrast, the participants reported the lowest enjoyment ratings for the VR device that can be attributed to a shorter viewing distance, a smaller viewing screen and the poor experience exposed to immersive environments by the VR device. This is because they might be influenced by a lower sense of spatial presence and realism in the virtual world [246]. For the "Flame" and "Wishing well" video sequences viewed with the VR device, the enjoyment ratings were the lowest compared to the "BBB", "Garden" and "Water fountain" video sequences. These results suggest that the chaotic movement of the "Flame" and "Wishing well" video sequences may be correlated with the lower enjoyment rating. From the single-factor ANOVA on enjoyment rating, a factor of F = 12.2, $p < 0.01$ was found to be significant.

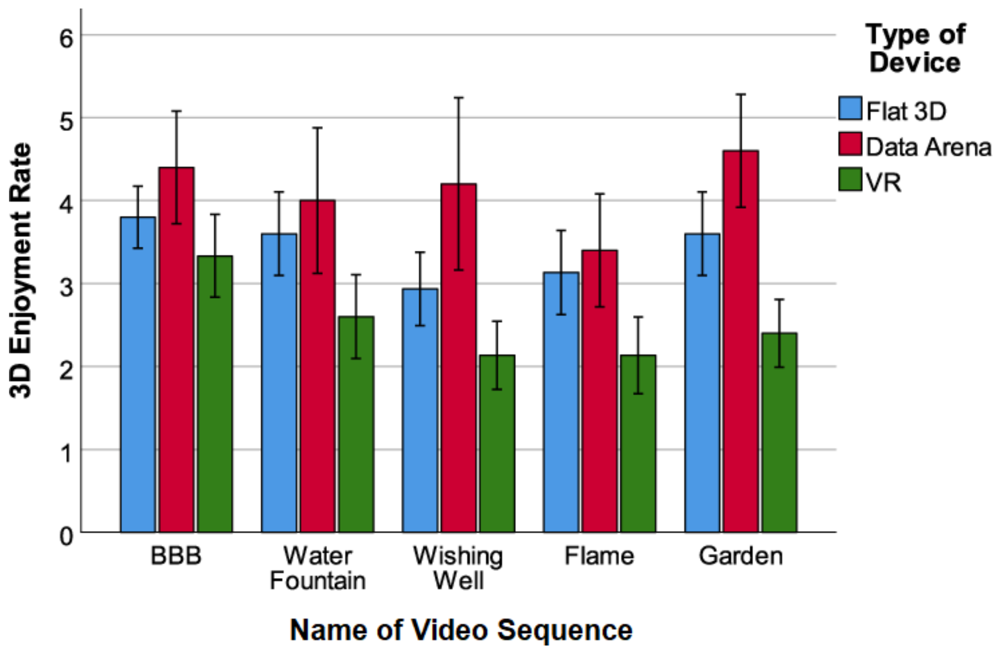


Figure 4.19: Enjoyment rating of participants.

4.4.5 Statistical analysis of QoE experimental factors

The data obtained from the QoE evaluation metrics – SSQ scores, ACR enjoyment rating, and EEG signals (analysed for visual fatigue and total eye blink frequency) – were recorded for each of the five video sequences. The collected data was analysed using Microsoft Excel and Statistical Package for Social Sciences (SPSS). Descriptive statistics is divided into two, namely the measures of variability and measures of central tendency. For this study, the measures of central tendency and variability which were determined included the standard deviation and mean values of each data set. Inferential analysis of data is also important in a study as it aids in understanding and unravelling the correlations between the study variables. As part of the inferential analysis, ANOVA was also analysed to identify the significance of the QoE factors for the three various viewing environments and five sets of video sequences.

Two-factor within-subjects ANOVA analyses were carried out to statistically evaluate the impact of experimental factors and further understand the correlation existing among the eye blink frequency, the brain wave power ratios and the QoE assessment metrics of SSQ score with the three different devices. The study included the analysis of three critical experimental factors namely the ACR enjoyment rating, the video sequence and the viewing device. The three factors analysed are as follows:

1. Enjoyment rating: The participants viewed S3D videos and their enjoyment levels were recorded.
2. Video sequence: The study involved the use of five different S3D video sequences.
3. Viewing environment: The experiment was intended to determine viewers' experience after watching S3D videos. However, the videos were viewed in different environments on different screens. The three screens used were a 3D flat screen, a panoramic screen and a VR headset.

The results for the statistical analysis of the QoE experimental factors are recorded in Table 4.3, Table 4.4 and Table 4.5.

Table 4.3: Results from within-subjects ANOVA on SSQ.

Factor	**DF**	**Mean Square**	**F ratio**	***p*-value**
Device	2	8260	22.7	<0.01
Video sequence	4	1270	3.50	0.01
Enjoyment rating	4	1910	5.26	<0.01

Table 4.4: Results from within-subjects ANOVA on brain wave power ratios (a) Power Ratio$_1$, (b) Power Ratio$_2$.

(a)

Factor	**DF**	**Mean Square**	**F ratio**	***p*-value**
Device	2	0.001	38.6	<0.01
Video sequence	4	0.001	25.1	<0.01
Device × video sequence	8	0	3.48	<0.01

(b)

Factor	**DF**	**Mean Square**	**F ratio**	***p*-value**
Device	2	0.002	64.8	<0.01
Video sequence	4	0.001	28.5	<0.01

Table 4.5: Results from within-subjects ANOVA on eye blink frequency

Factor	**DF**	**Mean Square**	**F ratio**	***p*-value**
Device	2	772	14.5	<0.01
Video sequence	4	247	4.64	<0.01
Device × video sequence	8	317	5.94	<0.01

The results of the within-subjects ANOVA revealed the presence of significant correlations among the eye blink frequency, the brain wave power ratios and the total SSQ scores. The above ANOVA tables only include the *p*-values that are less than 0.05. Based on the *p*-values, the attention and meditation levels had insignificant effects. However, in all the QoE assessments the significant impact was $p \leq 0.01$ for both the video sequence and the device. The eye blink frequency demonstrated a significant cross-influence factor between the video sequence and the device as the *p*-value was less than 0.01. Similarly, the *p*-value for SSQ scores was less than 0.01 indicating that it was significant with the enjoyment rating. The statistical cross-influence analysis conducted during the experiment and the results obtained in Tables 4.3-4.5 revealed the following insights about the QoE.

4.4.5.1 Type of viewing environment

In every QoE assessment, when viewing the S3D videos, the type of device used in the viewing environment is significant for every parameter being investigated: the enjoyment rating (ACR), brain wave power ratios, and SSQ. This study was limited to only three viewing environments, hence future work can enhance the study by using additional screens and testing them with different video content with various 3D effects. Such studies will provide further exploration of the correlation among the QoE parameters.

4.4.5.2 Eye blink frequency, attention and meditation level

The one-factor ANOVA of individual factors – the eye blink frequency, and the attention and meditation levels – did not depict any significant effect in the different types of device or viewing environment used. However, when the two-factor ANOVA of eye blink frequency was conducted, there was a significant difference between the video content and the device used to view the content. Based on these findings, the eye blink frequency has a relationship with the video content and the stereoscopic depth. The results are therefore insufficient to establish the existence of any difference when static or dynamic 3D video content is used. Hence, future work should investigate if there is any significant difference for different video content that is static or dynamic.

4.4.5.3 Enjoyment rating (ACR)

The enjoyment ratings and the SSQ scores were found to be significantly correlated. In addition, the two correlated subjective indicators – enjoyment rating and SSQ scores – were found to correlate with the quality perception and VIMS. The above correlations can be explored further in future work to understand the effects when more video sequences are used in the experiment.

4.5 Conclusion

In this chapter, the aim of the study was to compare the QoE of S3D videos in different viewing environments: namely, a flat screen, panoramic screen, and VR headset. A series of experiments was performed to evaluate the occurrence of 3D fatigue that viewers may experience when in three different environments. The evaluation of the viewer experience whilst watching S3D video sequences used five metrics: ACR enjoyment rating, SSQ (measuring VIMS), and EEG – measuring attention and meditation levels, eye blink detection, and the brain wave power ratios in different brain wave frequency bands. These metrics were evaluated with a variety of content and the results revealed that the viewers who used a VR device to view the stereoscopic video sequences resulted in higher SSQ scores. Higher SSQ scores imply that the visual fatigue was higher for the VR device compared to the other two screen display types. Also, the participants reported the lowest enjoyment ratings for the VR device. This is because they might be influenced by a lower sense of spatial presence and realism in the virtual world [246]. In addition, the experimental results indicated that participants who watched S3D video sequences were more likely to exhibit higher VIMS. For the panoramic screen, the participants reported the lowest SSQ scores and the highest enjoyment ratings when viewing S3D video sequences. The difference in results experienced when viewing the same S3D video sequence on different screens reveal that the screen and viewing environment are an essential factor that influences the level of visual fatigue, QoE and VIMS. The results also revealed that the chaotic movement in some video sequences had an effect on the SSQ, level of visual fatigue and enjoyment ratings. Hence, such findings show that the content of the video sequence and the viewing screen used are key factors that affect the enjoyment rating of the S3D videos.

This study was limited by the sample size of video sequences and the number of participants. Hence, it is recommended for future work to use a larger sample size of video sequences and participants to further assess the correlation of ACR enjoyment rating with both the brain wave power ratios and SSQ scores.

Chapter 5 S3D Objective Quality Metrics and Subjective Evaluation Methods applied to Video Compression Evaluation

5.1 Introduction

In order to transmit an S3D video from one point to another, it needs to be compressed so that it can be transported over a transmission medium. The compression of 3D videos refers to the compression of two stereoscopic views. As stated in Nasralla et al. [247], with the rapid advancements in video stereoscopy, network transmission can be a challenge due to the capacity or bandwidth needed and also the types of errors that can occur during the transmission process. These issues are related to the transmission and storage of S3D videos, and the methodology used to compress the data.

Video evaluation approaches are often applied to test video coding approaches. Thus, this chapter investigates the application of S3D evaluation approaches to testing a novel video compression technique. The compression technique proposed for evaluation uses a hybrid sequencing of uncompressed and compressed stereoscopic 3D (S3D) video, with the evaluation studying how the sequence would impact perceptual videos subjectively judged by viewers and how 3D videos can be scaled to fit available bandwidth. Immediate insight is given into the differences between the compressed and uncompressed video content, and the changes in the perceived quality through the S3D evaluation approaches used.

5.2 Background

5.2.1 Video coding sequences of 3D video

Current approaches for video coding generally use scalable stereo video coding to control the bitrate. Videos with different bitrates and content have also been assessed as an alternative to the computationally complex scalable video coding techniques, and to cater to stereoscopic videos with scenes that may be susceptible to potential artefacts introduced into the left/right views from compression [26]. Zou et al. [90] compared different 3D representation formats and coding architectures to evaluate the performance of various 3D video compression approaches, and suggested that the sequencing of compressed content may affect the 3D video quality. Asan et al. [248] conducted to evaluate the QoE of videos based on the content type and the switch of the resolution with respect to the Mean Opinion Score (MOS) degradation for different video adaptation patterns. The results show that the content type and the change of resolution pattern have a significant impact on the perception of the change in resolution. Mallik et al. [249] proposed a mixed-resolution multi-view video codec in High Efficiency Video Coding (HEVC) for limited transmission bandwidth, resulting in a higher coding performance to traditional HEVC codec at low bitrate by objective and subjective evaluations.

5.3 Experimental Methodology

5.3.1 S3D video compression

There are limitations on the stereoscopy that can effectively mask binocular fusion such that the perceived output is similar to the output with higher quality. These limitations can be experienced in subjective experiments by introducing just one form of 3D video (e.g., left or right view of S3D video). Also, in subjective experiments, the analysis of uncompressed Mixed Resolution (MR) stereoscopic video allows certain limitations in the resolution between two views and removes the challenge of evaluating subjective effects, due to several forms of asymmetry in S3D images [250]. In addition, the aim of MR is that the perception of stereoscopic video is not affected when one view is higher quality and the other view is lower quality, which is widely used in S3D video [251]. This assumption is based on the binocular suppression theory that the HVS combines two images such that the quality of the perceived image is similar to the higher quality view [252]. The novel compression approach proposed in this thesis as a use-case for evaluation investigates how various resolutions of 3D video content are compressed and the types of techniques used together with the type of codec used to utilise the transmission due to limited bandwidth. The research in this chapter focuses on the variation in the quality of different views, and the effectiveness of objective and subjective S3D video evaluation techniques.

The major challenge in this work is the level of stereoscopy achieved with different resolutions in the video. The research in this thesis has two coding approaches as use-cases to be evaluated using both objective and subjective metrics:

1. Assess the “hybrid” S3D video coding scheme proposed for evaluation using objective metrics. With the aim of achieving the highest possible subjective quality through “hybrid” coding schemes for video sequences, this “hybrid” approach is used to reduce the amount of data transmitted over a given period without changing the perceptual quality of the S3D videos.
2. Enhance S3D video through a higher quality of the stereo pair coded.

The “hybrid” S3D video coding scheme proposed as a use-case for applying S3D evaluation is shown in Figure 5.1. The main aim of “hybrid” sequencing is to reduce the level of redundancy and maintain a comparable video quality for different views experienced due to the limited transmission bandwidth. The proposed “hybrid” coding approach mixes one uncompressed (raw) 1080p video and one compressed 1080p video into a hybrid 1080p video in the same channel; the compressed videos used the standardised and widely used AVC/H.264 codec. The “hybrid” compressed versions of the files were then stored in MP4 format. For each video sequence tested, the bitrate of uncompressed, compressed and hybrid video sequences are the same within the sequence but may differ between sequences.

Figure 5.2 shows the proposed time sequencing of the resulting hybrid video. The uncompressed video and compressed video are denoted as ‘U’ and ‘C’, respectively. Due to human visual perception and the computation time of video sequences when viewing the 3D video [7, 253], the time duration for alternative sequencing is in one-second segments.

In objective evaluations, the Peak Signal-to-Noise Ratio (PSNR) is computed between uncompressed and compressed video, and “hybrid sequencing” video is computed as the difference between the left and right views of the S3D video sequence.

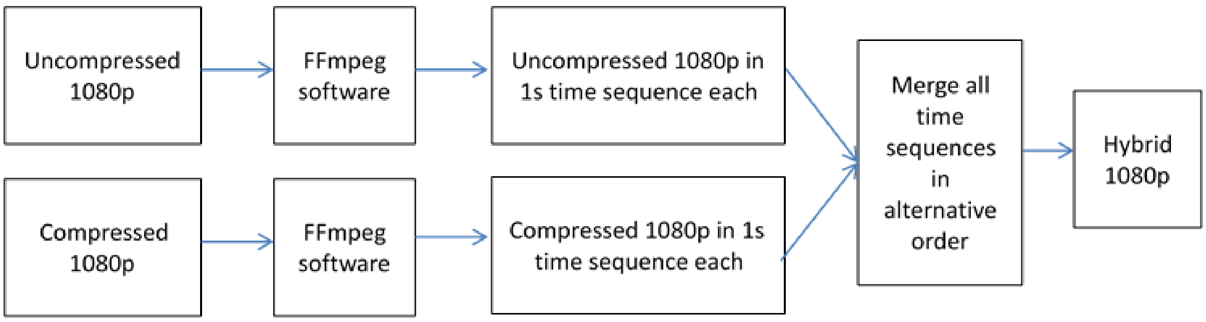


Figure 5.1: The process of creating a hybrid video sequence from different video formats in the same channel.

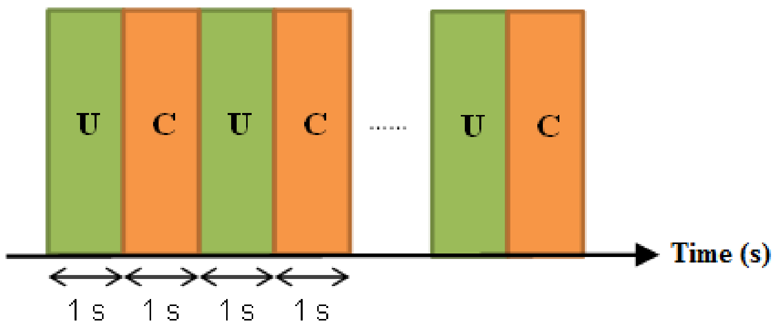


Figure 5.2: The timing of the proposed hybrid video sequence.

In this research, all S3D videos were coded using H.264/AVC coding to handle the Full Resolution (FR) and the Mixed Resolution (MR) formats similarly. This ensures minimising unpredictable effects on the outcome by varying down-sampling ratios. Stereoscopic video encoding with MR is a widely utilised and well-studied form of asymmetry between views. One of the main factors driving many video coding research efforts is to minimise the sophistication of the primary encoder and decoder execution since the spatial precision of a single view decreases the number of pixels used in coding and decoding relative to the usage of FR information [27].

5.3.2 S3D video dataset

To select appropriate video sequences for evaluation, there should also be a variance for depth prediction and disparity, which is the negative and positive parallax that is encoded into one measurement [254]. The depth prediction determines the farthest and nearest objects that our viewer can perceive. A bad depth prediction of images leads to visual degradations of the 3D experience [255], where crosstalk artefacts can also become more pronounced [69]. It is therefore important to include for evaluation various depth perception and different disparities of focal objects, and use a variety of scenes that interact differently with the 3D effect.

Further, the coding "difficulty" posed by the video content should range from easy-to-code scenes to hard-to-code scenes [4]. Different algorithms affect the different ranges of coding difficulty. Video encoders and decoders, video quality metrics and error correction software vary with variance in coding difficulty. To understand the full scope, it is therefore important to have an acceptable range of scenes with various coding difficulties. This allows for full characterization of the coding approach that is being assessed.

A total of three video sequences as shown in Figure 5.3 were selected from the RMIT3DV [75] database to apply the "hybrid" video coding as a use-case for S3D video evaluation. All the video sequences in the RMIT3DV databases were recorded in high resolution, uncompressed (raw), 25 fps and stored in MOV format [75]. For the S3D evaluations in this chapter, different versions of the stereoscopic videos using the "hybrid" coding were generated from these uncompressed videos using the FFmpeg software [241]. Both the uncompressed and compressed files were rendered into two resolutions, 720p and 1080p, to enable evaluation at different video resolutions.

The S3D video sequences selected were the "Water fountain", "Tram stop" and "Wishing well", which provided a range of video content for evaluation. For experimental consistency, the "Water fountain" and "Wishing well" video sequences are as used in Chapters 3 and 4. The "Water fountain" video sequence has the characteristic of being difficult to encode whilst the "Tram stop" video sequence can be encoded easily. The "Water fountain" video sequence is difficult to code due to the changing depth perception due to the moving water and objects. In contrast, the mostly static scene of the "Tram stop" video sequence results in easy encoding. Lastly, the "Wishing well" video sequence features a number of small objects that vary in size.

(a) (b) (c)

Figure 5.3: Video sequences selected from the RMIT3DV database (a) Water fountain, (b) Tram stop, (c) Wishing well.

5.4 Evaluation Results

5.4.1 Objective quality assessment

For objective quantitative quality assessment of the different hybrid video sequences, the PSNR was compared between uncompressed video, compressed video, and the proposed “hybrid” sequence coding approach at the same video resolution (720p or 1080p). The first set of PSNR computations is the comparison between the uncompressed and compressed video at the same resolution using the left view of the video sequences. The second set of PSNR computations is the comparison of the differences (if any) in the disparity of the S3D video sequence between the left and right views. Unlike the first set of PSNR computations previously shown in Equations (2.1) and (2.2), the second set of PSNR computation is modified as Equations (5.1) and (5.2):

$$\mathrm{MSE} = \frac{1}{mn}\sum_{i=0}^{m-1}\sum_{j=0}^{n-1}\left|Left(i, j) - Right(i, j)\right|^2 \tag{5.1}$$

$$\mathrm{PSNR} = 10\log_{10}\left(\frac{\mathrm{MAX}_I^2}{\mathrm{MSE}}\right) \tag{5.2}$$

where MSE is the Mean Square Error to measure the quality, *m* and *n* represents the image pixel, *Left*(i, j) and *Right*(i, j) are the left and right views of the S3D video sequence, respectively, and MAX_I is the maximum value of a pixel.

5.4.2 Subjective quality assessment

The proposed "hybrid" coding approach was previously described in Section 5.3.1. S3D video encoding with MR (Mixed Resolution) is a widely utilised and well-studied form of stereoscopy between two views. One of the main factors driving video coding research is to minimise the sophistication of primary encoder and decoder execution since the spatial precision of a single view decreases the number of pixels used in coding and decoding relative to the usage of full referenced information [27]. Another advantage of MR stereoscopic video coding is limiting bitrate to the reduced amount of encoded pixels compared to the Full Referenced (FR) situation. The bitrate required to encode MR with the same Quantization Parameters (QPs) as FR S3D video is reduced when the left and right views are encoded in simulcast mode with different levels of quality [251]. The magnitude of the bitrate decrease depends on the video frame ratio.

To verify the subjective quality of MR video [249], four different video sequence pairs have been compared using subjective quality assessment, as shown in Table 5.1. For each video sequence tested, the bitrate of uncompressed, compressed and hybrid video sequences are the same within the sequence but may differ between sequences.

Ten participants (8 male and 2 female) aged 20 to 38 years old participated in the experiment. Before the experiment, the participants performed 3D vision assessments according to the ITU-R BT.2021 [25] standard. The S3D videos in the subjective tests were displayed on a 25.5'' Panasonic BT-3D L2550 Full HD LCD 3D monitor, where participants wore passive 3D glasses to watch the S3D content. A 36° viewing angle and 0.9 m viewing distance were implemented using the THX Cinema Qualification criteria [36] in quantitative standard measurements.

When a comparative exercise was done in pairs, participants were required to sit near the 3D display for 10 s to view the first S3D video. After 10 s, the participants rated the subjective quality of the video content using a scale of five Absolute Category Rating (ACR) points (1: bad, 2: poor, 3: fair, 4: good, 5: excellent). Each participant then viewed the second video (of a different quality, as in Table 5.1) for 10 s to reassess the video quality. Each participant was requested to watch a combination of three separate S3D videos. The sequence of the four video experiments shown in Table 5.1 was randomly selected to mitigate sequencing bias. Therefore, a total of 12 video pairs were viewed by participants (i.e. 3 video sequences $\times$ 4 compression sequence pairs). The questionnaire of this experiment can be referenced to Appendix 3.

Table 5.1: Video sequence pairs for subjective assessment.

Video Sequence pairs	**Video 1**	**Video 2**
1	Compressed 1080p	Compressed 720p
2	Compressed 720p	Compressed 1080p
3	Uncompressed 1080p	Hybrid 1080p
4	Uncompressed 720p	Hybrid 1080p

5.5 Results and Discussion

5.5.1 PSNR of hybrid sequencing

In this experiment, the raster file used for uncompressed (raw) video is represented in MOV format, while the use of compressed video is represented by MP4 format. Table 5.2 shows the computations of PSNR for the three S3D video sequences evaluated (Figure 5.3). Table 5.2 shows that the 1080p PSNR values for all video sequences were lower than for 720p resolution. These results reveal that a higher resolution can lead to more errors from uncompressed video compared to compressed video.

Furthermore, the PSNR between the left and right views were compared to study the relationship between objective and subjective video quality evaluation [67]. Generally, all the PSNR between the left and the right views (i.e., Set 2 from Table 5.2) are lower than the PSNR of videos with the same resolution (i.e., Set 1). The results are potentially due to the different content and display between two stereoscopic views. Among the videos tested on the left and right views, the "Tram stop" video sequence obtained the lowest PSNR values, whereas the "Water fountain" video sequence obtained the highest PSNR values. However, the subjective evaluations suggested that the "Tram stop" video sequence showed the least score variation among participants evaluated in response to the varying video quality and resolutions. These findings indicated that S3D video disparity might not be the only factor affecting 3D video quality. Subjective assessments showed that other features of 3D videos, such as scene movements, encoding methods, and image orientation, may also affect the perceived video quality [256]. For the proposed "hybrid" approach, the PSNR for all video sequences are higher than the compressed videos, but lower than uncompressed videos at 1080p resolution. The results are correlated to the video quality that compression causes distortion of the video quality, leading to lower PNSR values.

5.5.2 Subjective quality assessments from the ACR

Figure 5.4 shows the average ACR participant scores recorded for each pair of video formats shown in Table 5.1 for the “Water Fountain”, “Tram Stop” and “Wishing Well” S3D videos tested. The goal of testing video Pair 1 and Pair 2 in Figure 5.4 was to examine how the ACR scores differ from low (high) to high (low) resolution videos. Pair 1 and Pair 2 of these subjective assessments show that the smallest difference between the ACR values for the 1080p and 720p videos was recorded for the “Tram Stop” video. This is most likely because both the people and the trams travelled steadily in this video [137].

Table 5.2: PSNR comparison for the three video sequences.

PSNR(dB) / *Video quality*	**“Water fountain” sequence**	**“Tram stop” sequence**	**“Wishing well” sequence**
Set 1: Comparison between the uncompressed video (reference) and compressed video (distorted) with the same resolution			
1080p	35.78	35.32	35.12
720p	36.00	35.87	35.23
Set 2: Comparison between the left view and the right view of the video sequence, using left view as the reference			
1080p uncompressed	17.52	12.35	16.34
1080p hybrid	17.54	12.36	16.37
1080p compressed	17.57	12.37	16.40
720p uncompressed	17.81	12.44	16.47
720p compressed	17.84	12.45	16.52

Pair 3 and Pair 4 in Table 5.1 were used to compare uncompressed videos at 1080p and 720p with the “hybrid” coding approach. Figure 5.4 shows that the “hybrid” coded video average rating is less than that for the uncompressed video while it is at the same resolution, as in the

Pair 3 test (1080p). Compared to the lower resolution of the uncompressed file (720p) in the Pair 4 test, hybrid sequencing for all three video sequences resulted in higher ACR scores.

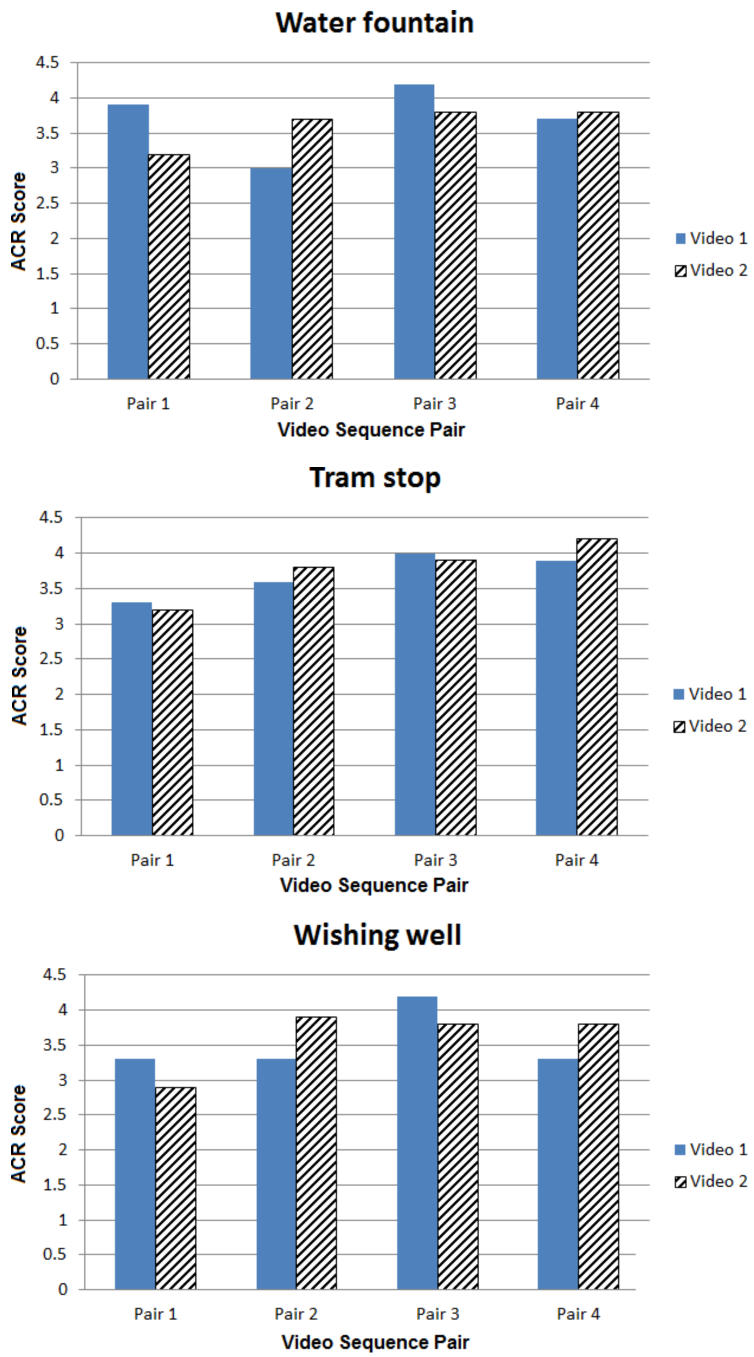


Figure 5.4: Average ACR scores for different pairs of video quality for the three S3D videos.

Furthermore, the ACR scores for the "Tram stop" video sequence shows again the smallest score difference for Pair 3 and Pair 4, similar to those for the Pair 1 and Pair 2 tests. The larger difference in ACR scores for the "Water fountain" video sequence may be caused by the rapid movement of water in the video, which induces a high 3D effect with potentially distorted images [257]. In comparison, the less consistent ACR ratings of the "Wishing well" video sequence could be attributed to the video with multiple tiny artefacts with varying depth perceptions due to the moving water content, which may cause more viewers to focus on those objects.

The results show that the video content with more scene movement has less ACR score than stable scene. Therefore, "hybrid" coded video sequence is more preferable for the content with static scene. Also, the average ACR rating of the "hybrid" coded video is comparable to the same resolution of the uncompressed video content, and the rating of this "hybrid" coded video is better than the lower resolution of the uncompressed file.

5.6 Conclusion

S3D video compression is studied in this chapter as a use-case for applying objective and subjective evaluation techniques. A "hybrid" sequencing video coding approach was thus proposed for evaluation for both uncompressed and compressed S3D video evaluations. The study investigated the effects of hybrid sequencing on the S3D QoE videos viewed on a 3D flat screen. This study found that hybrid sequencing can provide perceptual quality similar to that of uncompressed videos. This was found to be true for videos with scenes that have 3D effects that are moderate, and more stable scenes. The different artefacts that affect the quality of 3D content and which therefore need to be reduced to maintain a comparable video quality were also noted in the evaluation results; therefore, future work can further evaluate the effect of the different types of scenes in an S3D video such as the level of complexity and effect over the proposed "hybrid" sequencing approach.

The compression technique used in this research was the AVC/H.264 codec. Another option for future work for a practical use-case is the coding of S3D videos with High Efficiency Video Coding (HEVC). The resolution is higher than HD and the computational efficiency of the approach has been improved. Samelak et al. [258] conducted a study to compare the performance in compression achieved through screen content coding HEVC and multi-view HEVC. The results suggested that the compression time was reduced by almost 10% in comparison to the standard recommended configuration for computer generated content. Hence, the use of HEVC can be one of the potential ways to carry out further research on the basis of the proposed "hybrid" sequencing coding approach.

Chapter 6 Visual Attention of S3D Video in relation to QoE using Eye-tracking Analysis

6.1 Introduction

Previous chapters have investigated subjective evaluation approaches measuring visual and eye fatigue in the viewing of S3D videos through the use of consumer grade EEG headsets. In addition to the measurement of visual and eye fatigue, the existing literature in Section 2.1.4 discussed that visual attention may correlate to the quality of experience when participants view S3D videos. Therefore, it is also important to consider the visual attention of viewers when watching S3D video and how it relates to QoE. Thus, Section 2.5.5 discussed how researchers have adopted eye-tracking methods to assess how the human visual system responds to visual attention when viewers watch 3D video contents. With this aim, this chapter presents eye-tracking experiments conducted to measure human visual attention when viewing S3D videos to see whether there is any correlation to the quality perception and experience from viewers. Some visual attributes such as brightness, colour, texture, direction, motion and depth were selected for the prediction of visual attention when viewing S3D video.

In the experiments detailed in this chapter, eye-tracker equipment was used to detect the eye movements from viewers and a human saliency map (or fixation density map) was developed to show the most important objects or areas in a scene. This model was compared with existing human saliency maps and saliency map prediction models from other researchers for analysis, evaluating whether visual attention measured using the eye-tracking methods contributes to QoE assessments of S3D video.

6.2 Background

6.2.1 Current saliency prediction models

The concept of visual saliency comes from human perception in viewing a scene, and correlates with the ability of a region to attract attention [259]. Section 2.5.2 described how a human saliency map can be obtained with an eye-tracking experiment. However, this approach is a time-consuming and costly method, requiring many viewers to conduct eye-tracking experiments, and processing large amounts of visual data. Another approach is the prediction of a visual attention model by computer stimulation as a perception mechanism of the HVS to view a scene, called the saliency prediction model.

Existing saliency prediction models of 3D videos have been based on two approaches. The first approach uses a weighting factor for the disparity or depth map derived from a current 2D saliency detection model. i.e., the 2D saliency map is generated and a weight based on its disparity value was used to generate a 3D saliency map. Chamaret et al. [260], Zhang et al. [261] and Maki et al. [262] designed their 3D saliency models by adopting the above approach. In these three models, objects that are closer to the observer are considered to be more salient. Therefore, the results of these 3D saliency models are considered to be more salient when closer to the objects. In addition, qualitative evaluations show improvements compared with 2D saliency models. However, this method did not provide a quantitative evaluation of these 3D models. The results are not necessarily more salient when the viewers are closer to the objects. The second approach adopts the prediction of a depth saliency map from the depth information of the image. The 2D saliency map is combined with the resultant depth saliency map to generate a 3D saliency map. Ouerhani and Hugli [263] suggested a model including the gradient of depth features and the surface curvature. The model did not include the quantitative assessment using eye-tracking data, but a qualitative assessment was included. Lang et al. [264]

suggest a depth saliency map for the evaluation of a saliency ratio at different ranges of depth by statistical analysis with the use of a training dataset. This method has been validated by comparing the existing 2D saliency models with the generated depth saliency map by using different combinations of multiplication and summation.

The above two approaches to 3D saliency map prediction methods that can imitate the HVS is still a challenge for researchers. In addition, the key features of the HVS to be examined are still an important task for investigation.

6.2.2 Dataset

Previous research has established fixation prediction datasets and salient object detection datasets for saliency prediction [132, 154, 265, 266]. The following sub-sections describe three recent datasets which are widely adopted in the current research.

6.2.2.1 MIT300 dataset

Judd et al. [267] established the MIT300 dataset in 2012. This dataset contains 300 natural images and 39 sets of observer eye movement data. This database is the most widely used in the field of image human eye focus detection. However, the dataset still has challenges. The ground truth of the human eye's focus is not public and it is generally not used as a training set [268]. Another dataset MIT1003 was also established by Judd et al. [269]. This dataset contains about 1000 images obtained from Flickr [270] and LabelMe [271] websites, of which 779 images are landscapes, 228 images are portraits, and 15 observations are included. At the same time, the recording process of eye movement data also considered human memory mechanisms. Each participant was required to indicate which of the 100 images was previously seen. Thus, the MIT1003 dataset can be used as a supplement to the MIT300 dataset. For example, this dataset could be used to train an attention model based on machine learning on the MIT1003 dataset, and then used on the MIT300 dataset as the test set for performance evaluation [272].

6.2.2.2 PASCAL-S dataset

The PASCAL-S dataset was established by Li et al. [273] in 2014. This dataset used 850 images from the PASCAL VOC 2010 dataset validation set [274], viewed by 8 participants in “free viewing” mode within 2 seconds. Eye movement data was obtained by observing the image. In the gaze prediction experiment, 8 participants were asked to perform a "free viewing" task to explore images. The eye movement data was sampled at 125 Hz using the EyeLink 1000 eye-tracker equipment. In the salient target segmentation experiment, a complete segmentation was manually performed to crop out all objects in the image. It has been determined that the ground-truth of the saliency object segmentation followed these three rules:

1. Unintentionally mark parts of the image such as human faces.
2. Separately identify disconnected areas of the same object.
3. A solid area approximates a hollow object, such as a bicycle wheel.

Twelve participants were asked to mark prominent objects. There was no time limit or restriction on the number of objects a person could choose. The final significance value of each segmentation is the total number of clicks it receives divided by the number of characters. The marking of salient objects was based on the complete segmentation of an image. Each image in the PASCAL-S dataset was marked by multiple markers, and there was no limit to the number of salient targets [275].

6.2.2.3 Eye-tracking dataset for stereoscopic videos

Fang et al. [276] established an eye-tracking dataset for stereoscopic videos in 2014. This dataset contains 41 video sequences selected from the following two main databases: 1) The RMIT3DV database [75] containing 24 video sequences with 10-bit 4:2:2 YUV format at 25 fps; 2) The IVC stereoscopic video database [277] including 12 video sequences with AVI format at 25 fps.

This dataset includes both indoor and outdoor scenes in full HD resolution. This dataset used an SMI RED-60Hz eye-tracker at 60 Hz to record eye movement data. When using eye movement data recorded in different ways as training samples to train a saliency model, the different training samples will have different effects on the final performance of the model. A typical ground truth saliency map is generated by performing Gaussian filter aggregation, which is common practice for generating gaze views from eye-tracking data [137]. However, the saliency map in this dataset was generated in two parts. The first part creates two gaze point maps from the left and right eye respectively. Then, the “left” gaze point maps were directly generated by using the gaze position data of the left eye, whilst the “right” gaze point maps were generated by using each gaze point coordinate of the right eye plus a horizontal and vertical displacement. The final output gaze point map is the sum of the above gaze point maps referenced to the left and right eyes. The final ground truth saliency map is then created using a Gaussian kernel.

Therefore, taking into consideration the findings from previous researchers’ studies using these three key datasets, this dataset for stereoscopic video was adopted for the experimental work in this chapter.

6.2.3 Evaluation indicators of eye focus detection

In human eye focus detection tasks, previous researchers have proposed many evaluation indicators, among which the most commonly used include Earth Movers' Distance (EMD), cross-entropy such as Kullback-Leibler (KL) divergence, standardised scan path significance such as Normalised Scan-path Saliency (NSS), Similarity Metric (SIM), Pearson's Correlation Coefficient (CC) and Area Under Curve (AUC) indicator; that is, the area under the Receiver Operating Characteristic (ROC) curve [267, 278-280]. These indicators follow different design principles: for example, the cross-entropy indicator treats the saliency prediction result and the real human eye attention calibration as a probability distribution. The AUC indicator treats the saliency prediction result as a binary classification to evaluate from the perspective of analysing the classification performance of the classifier. The Linear CC or NSS can be treated to measure two correlations: the saliency prediction result and the real human eye attention calibration as random variables. These evaluation indicators provide different dimensions of evaluation for consistency between the saliency detection result and the attention distribution from the real human eye. A variety of evaluation methods are used to evaluate the model to find the most preferable approach to predict eye focus distribution from experiments [128].

Based on the saliency prediction result: $P = [0, 1]^{W \times H}$, the true binary human eye attention point record $R = [0, 1]^{W \times H}$ and the visual attention true value distribution $Q = [0, 1]^{W \times H}$, Q is obtained by using a smaller Gaussian kernel convolution on the binary human eye attention point distribution map. The parameters of the Gaussian kernel are mainly based on different eye parameters. The eye size and eye movement equipment on the movement dataset are set previously [281].

In this thesis, two saliency prediction models, Itti [133] and GBVS [134] as described in Section 2.5.4, are adopted in the experimental work. These two saliency predictions are still the most commonly used methods by researchers because they are relatively simple and the values of AUC are also relatively high in most scenes tested.

6.3 Experimental Methodology

Figure 6.1 below shows the method used to conduct the eye-tracking experiment in this thesis. The participants were given five sets of S3D video sequences at random, and they watched each video sequence for one minute, taking an average total duration of 15 minutes. In line with the ethical conduct of participant experiments, participants who suffered from dizziness, physical or motion sickness can discontinue the study at any time.

Similar to the experimental work conducted in Chapter 4 , bias was minimised by viewing the S3D videos randomly without any specific order. The random order of the S3D videos was generated using the Stat Trek [238] Random Number Generator.

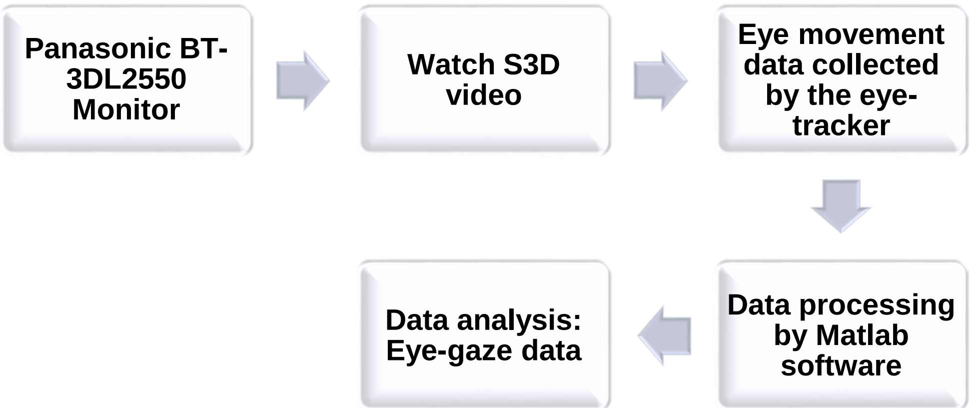


Figure 6.1: Method for the eye-tracking experiment.

Before participants watched the S3D videos, they were required to conduct an eye-tracking calibration to ensure accuracy in collecting their eye movement data. The details of the eye-tracking calibration will be described in Section 6.3.1. Whilst participants were watching an

S3D video, the eye movement of the participant was recorded by the eye-tracker. The eye-tracker obtained the eye position of both left and right eyes: each eye position is shown in *X, Y* and *Z* coordinates with respect to this equipment. The eye-tracker can also obtain the eye-gaze locations on the screen when participants are watching video sequences.

All of the signals obtained from the eye-tracker were processed in Matlab [282] software, where the Matlab scripts control the eye-tracker to start and stop the data logging and analyse the gaze data [283], as shown in Appendix 1. Those gaze points show more visual attention of relative parts whilst the fixation tracking can help to identify which objects the human eyes focus on. For the eye-tracking equipment, a higher-performance eye-tracker can help to build up different models for analysis such as the heat map, fixation map and saliency map. The eye-tracking data thus provides the time spent, fixation and Area of Interest (AOI) information.

6.3.1 Experimental setup and eye-tracker calibration

The flat 3D screen used in the research was a 25.5'' Panasonic BT-3D L2550 Full HD LCD 3D screen. The participants sat 0.9 m in front of the S3D screen with a viewing angle of 36° when viewing 3D videos, based on the THX Cinema Certification [87]. The screen, position of the screen, the viewing conditions and the selection of the five video sequences were the same as for the experiments using the flat 3D screen described in both Chapter 3 and Chapter 4 to maintain the same controlled environment.

To record the eye movement of the participants, a Tobii EyeX eye-tracker [284] was used. The specifications of the Tobii eye-tracker are shown in Table 6.1. The sampling rate of this eye-tracker is 70 Hz, which can trace eye movement up to 70 times per second. The reason for the selection of this eye-tracker is that it was readily available to consumers, it non-intrusively recorded eye-tracking signals, and was non-contact to the participant. Figure 6.2 shows the experimental eye-tracking setup showing how to connect all the required devices for the experiment. The distance between the eye-tracker and the participant is set at the minimum operating distance of 0.5 m to maintain the highest eye-tracking accuracy due to the fact that wearing passive 3D glasses increased the difficulty of eye-gaze tracking.

Table 6.1: Specifications of Tobii EyeX eye-tracker [284].

Operating distance	0.5-0.9 m
Maximum screen size	27”
Frequency	70 Hz
Illustration	Backlight assisted near-infrared: red light (650 nm) and near-infrared red (850 nm)
Tracking population	95%

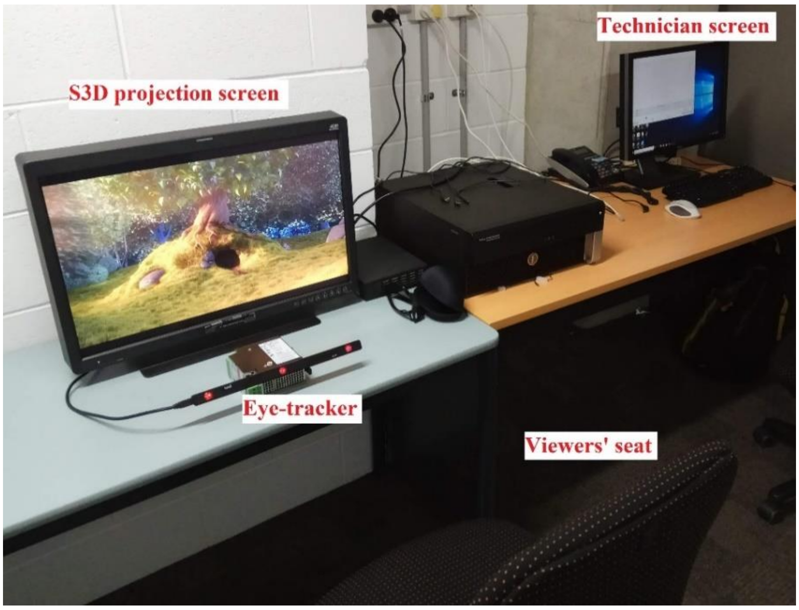


Figure 6.2: Experimental setup of the eye-tracking experiment.

Participants were required to complete an eye-tracking calibration before the experiment, consisting of two parts. The first part was the calibration of the five gaze points provided by the software tools of the Tobii eye-tracker. These five points are illustrated in Figure 6.3 using an active display coordinate system.

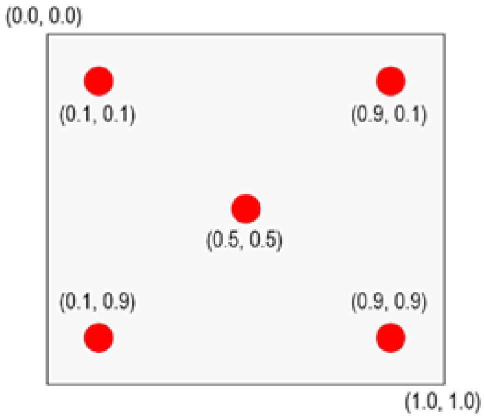


Figure 6.3: Five gaze points for eye-tracker calibration.

The second part of the calibration procedure was based on the study from O'Connor et al. [114], where a 2-second scene including both fixation and pursuit conditions for participants to view and collect data for further calibration [112]. The aim of the design is used to compare the precision of retinal and extra-retinal motion signals between young and old observers [114]. This approach to direction and speed discrimination over a 2-second period is implemented in this work for the calibration. Kaernbach et al. [285] designed the two angle intervals as an adaptive test to collect the response of eye movement. Table 6.2 shows two examples of the speed and orientation of dots in two situations: 1) The dots are moving to 0° direction, 2) The dots are moving to 45° direction, for direction and speed discriminations of pursuit conditions at the time in the 0 second and the 0.8th second, respectively [114].

Table 6.2: Two examples of dot arrangements in pursuit conditions.

Time / *Discrimination* / Orientation	0 s		0.8 s	
	Direction	***Speed***	***Direction***	***Speed***
Direction at 0°				
Speed	2°/s	4.8°/s	2°/s	4.8°/s
***X*,*Y* coordinates at centre point**	(50, 200)	(50, 200)	(63, 200)	(81.2, 200)
Direction at 45°				
Speed	2°/s	-	2°/s	-
***X*,*Y* coordinates at centre point**	(50, 200)	-	(66, 216)	-

6.3.2 Eye-tracking data processing

Figure 6.4 shows an example of the simulation result during the experiment, which collected eye-gaze data from the Tobii eye-tracker, which was processed in Matlab. The eye-gaze data included the x and y coordinates of the eye gaze points, the position of both eyes in front of the S3D screen, and the access time for data logging. The data “L” and “R” correspond to the detection of left and right eyes respectively. For data consistency, the data is only used when the detection of both left and right eyes were valid. The accuracy rate to detect both left and right eyes in this experimental work is 96.89%, which is higher than the reference accuracy rate (95%) from the Tobii specifications.

GazeX_px	GazeY_px	GazeTimestamp	L	R	LeyeX_mm	LeyeY_mm	LeyeZ_mm	ReyeX_mm	ReyeY_mm	ReyeZ_mm
2930.60	665.59	185019188.07	1	1	-10.49	117.05	689.40	55.22	109.59	671.64
3027.59	608.61	185019202.86	1	0	-10.74	117.20	689.16	0.00	0.00	0.00
3008.59	614.40	185019217.90	1	0	-10.74	117.20	689.16	0.00	0.00	0.00
2996.43	618.78	185019232.96	1	1	-11.33	117.63	689.27	54.38	109.89	670.91
3015.50	569.54	185019247.58	1	1	-11.33	117.63	689.27	54.38	109.89	670.91
2928.44	635.99	185019292.54	1	1	-12.27	118.49	689.49	53.59	110.55	671.07
2905.55	656.26	185019308.14	1	1	-12.40	118.49	689.31	53.45	110.66	671.06
2913.85	657.89	185019323.13	1	1	-12.58	118.66	689.50	53.32	110.90	671.47
2908.57	653.22	185019337.75	0	1	0.00	0.00	0.00	53.23	111.12	671.87

Figure 6.4: Example of eye-gaze data logging by the Tobii eye-tracker.

The collected gaze point coordinates were the sum of the resolution from the two screens (i.e., technician's screen plus the S3D flat screen, the setup was previously shown in Figure 6.2). The *x*-coordinate (i.e., GazeX_px) is revised to fit the gaze point viewing on the S3D flat screen only. Also, the first two seconds of eye-tracking data were discarded to allow for participants to adjust to the experimental setup and process.

A procedure for the generation of a human saliency map (or fixation density map) is shown in Figure 6.5. All calibrated gaze points recorded by all participants are processed by the binary representation of the location of the gaze points to form a human fixation map. The human fixation map is then processed by a Gaussian Mixture Model (GMM) [129, 286], as shown in Equation (6.1), for the distribution of all gaze points, where (x_f, y_f) is the gaze points obtained by the eye-tracking experiment, σ^2 is the variance of the fixations obtained, f_{num} is the number of fixations, and α_f is the weight of the distribution of fixations.

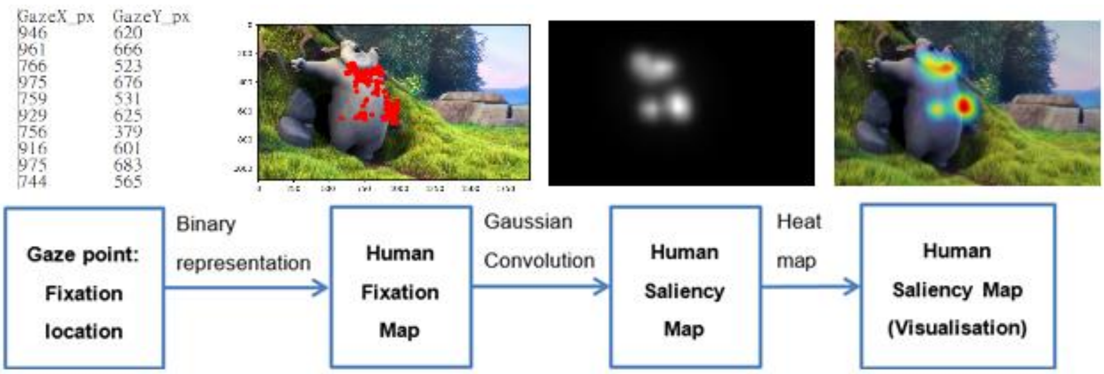


Figure 6.5: Procedure for the generation of a human saliency map for visualisation.

Gaussian mixture model (GMM):

$$f_{XY}(x,y)=\sum_{f=1}^{f_{num}}\alpha_f\frac{1}{2\pi\sigma^2}\exp\left(\frac{(x-x_f)^2+(y-y_f)^2}{2\sigma^2}\right) \tag{6.1}$$

Finally, the human saliency map is further processed to form a heat map visualisation in a “jet” colour map format, as shown in Figure 6.6. The red colour represents the most salient parts whilst the blue colour represents the least salient parts. The human saliency map can thus be used to analyse the eye movement behaviour.

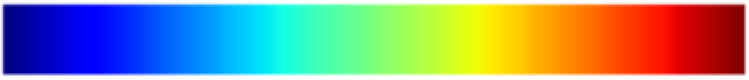


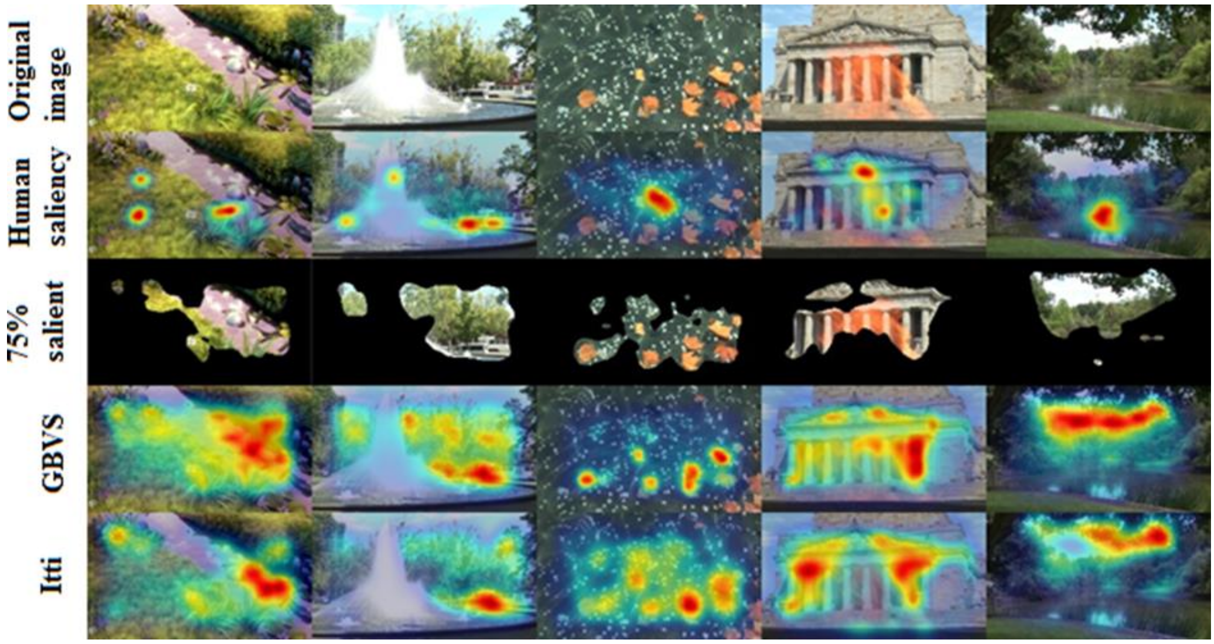
Figure 6.6: A “jet” colour map is presented for a heat map visualisation.

6.4 Results and Discussion

A total of ten participants took part in the viewing experiment, consisting of 3 females and 7 males. A minimum sample size of 10 is suggested based on the recommendations of the Quality of Multimedia Experience (QoMEX) research community under the recommendation of the ITU-T P.913 International standard [5]. The participants’ ages were between 18 and 46 years and the mean and standard deviation of the participants’ ages were 28.3 years and 7.26 years respectively. Before the start of the experiment, each participant was asked about their fatigue level, and each participant completed three kinds of calibration: eye-tracker, eye-gaze point and pursuit and fixation conditions. Such calibration methods were previously described in Section 6.3.1.

6.4.1 Comparison of human saliency maps

The human saliency map for computational eye movement prediction in this work was compared with two widely used saliency prediction models, the Graph-Based Visual Saliency (GBVS) model [134] and the model of Itti et al. [133], which were previously described in both Section 2.5.4 and Section 6.2.1 based on the literature and existing current saliency prediction models. The procedure to formulate the saliency maps is illustrated in Figure 6.5. Figure 6.7 shows the comparison between the proposed saliency maps with these two existing saliency prediction approaches. The five sets of images shown from the top to the bottom are represented as: 1) original image, 2) proposed human saliency map, 3) 75th percentile of most salient parts of the GBVS saliency map, 4) the heat map of the GBVS saliency map, and 5) the heat map of the Itti saliency map, respectively.

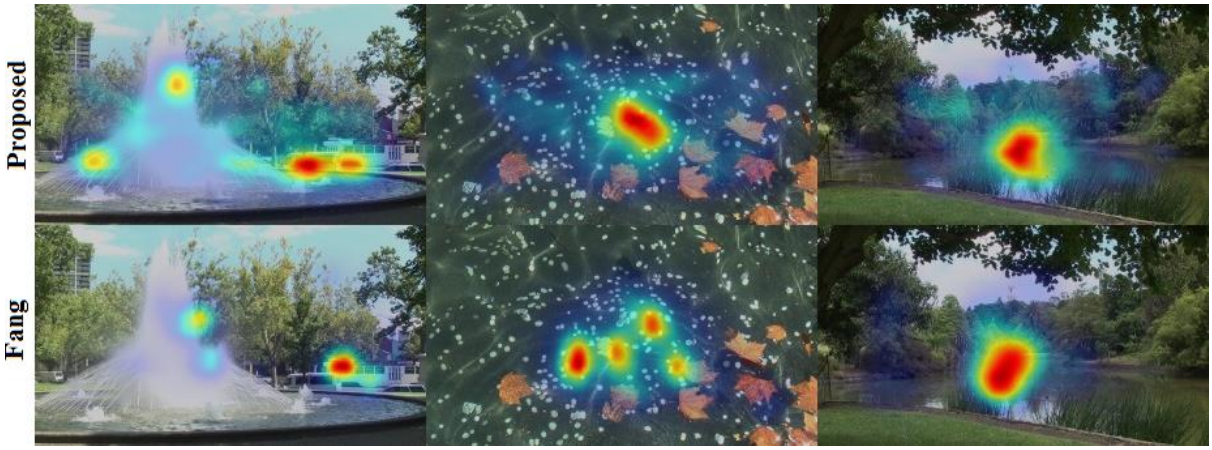


(a) (b) (c) (d) (e)

Figure 6.7: Comparison of the proposed human saliency map to two saliency prediction models, GBVS [134] and Itti [133], from five video sequences (a) BBB, (b) Water fountain, (c) Wishing well, (d) Flame, (e) Garden.

In the “Wishing well”, “Flame” and “Garden” video sequences, participants were more focused on watching the centre of the scene rather than on the main object whilst the existing two saliency map models predicted the most salient parts on the object. In the “BBB” and “Water fountain” video sequences, the prediction of the two saliency models are comparable to the proposed approach. However, the prediction of the fountain from the “Water fountain” video sequence is not accurate, especially for the Itti model [133]. This is probably due to the difficulty in the prediction of chaotic and fast-moving water jet streams from the fountain. In addition, Figure 6.7 shows the 75th percentile outline of the most salient parts from the GBVS model, which reflected mostly on stationary objects in the area of interest. Therefore, the prediction may not be correct for the proposed human saliency map.

Figure 6.8 shows two sets of human saliency maps generated from watching three video sequences from the RMIT3DV database. These videos are: a) Water fountain, b) Wishing well, and c) Garden. The three images on the top show the human saliency maps generated by the method proposed in this thesis, whilst the images at the bottom show the output generated by Fang’s dataset [276] as previously described in Section 6.2.2.3. Comparing the two sets of the human saliency map, they are more consistent for viewing the area of interest and are in line with the viewing pattern from the participants. In addition, the most salient parts are also focused at the centre of the scene in both the “Garden” and “Wishing well” video sequences. Based on the findings of the “Wishing well” video sequence, the participants are more focused on watching the centre of the scene rather than watching the tree cover at the right-hand side. The reason for this is potentially due to the continuity of the objects of interest at the centre of the scene [272], and the chaotic movement of the tree cover. Some researchers consider this situation as “centre-bias” from eye-tracking experiments, due to the tendency of participants to view at the centre [287]. In the “Water fountain” video sequence, most participants viewed the right-hand side where a car is located in the scene. The reason for less focus on the fountain may be due to the chaotic water jet stream of the fountain.



(a) (b) (c)

Figure 6.8: Two sets of human saliency maps generated by the proposed approach (top) and Fang’s dataset [276] (bottom) for three video sequences (a) Water fountain, (b) Wishing well, (c) Garden.

According to the findings presented in this section, eye-tracking experiments are thus essential to an investigation of the visual attention and QoE, especially for 3D videos; however, a saliency prediction model may not contribute reliable results to the research of QoE.

6.4.2 Evaluation indicators of saliency map performance

As described in Section 6.2.3, a number of evaluation indicators are available to measure the performance of the saliency prediction models. Riche et al. [288] compared an eye-tracking database with 12 different saliency prediction models and found that using only one similarity metric is insufficient for evaluation. In addition, the Kullback-Leibler (KL) divergence indicator with complementary interpretation can be compared fairly with the saliency prediction model. With reference to Riche’s consideration, the following three evaluation indicators of a saliency map are used to compare the distribution between human saliency maps and saliency prediction models. Bylinskii et al. [278] has shown the detailed computation of the following three evaluation metrics:

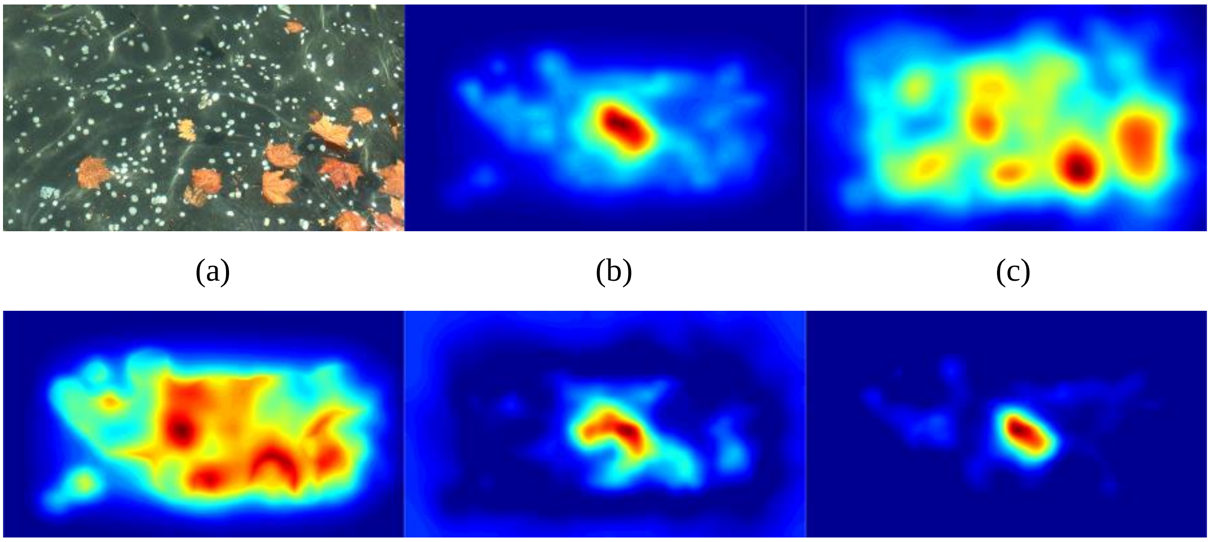
1. Similarity metric (SIM): Measure of similarity between two distributions, viewed as a histogram to compare the intersection between two distributions [278].
2. Pearson's Correlation Coefficient (CC): A similarity metric by statistical methods to measure the linear relationship between two distributions [278].
3. KL divergence: A dissimilarity metric to measure the difference between two probability distributions. A lower score represents a better approximation of the ground truth of the human saliency map by the saliency prediction model [278].

Table 6.3 shows the comparison of two human saliency maps with two saliency prediction models using the three evaluation indicators. Higher scores obtained from the SIM and CC evaluation indicators and a lower score obtained from the KL evaluation indicator represents a more accurate result of the human saliency approximation. In general, the three indicators show that the models of this research have an improved human saliency approximation than Fang's dataset. However, there is no clarity on which saliency prediction model is better for a stable or a moving scene. In addition, the variation of the CC and KL indicators are comparably higher than the SIM indicators when comparing the two saliency prediction models. Figure 6.9 shows the "Wishing well" video sequence with the least variations of the three evaluation indicators when comparing with Itti's saliency maps [133]. From parts (d) to (f) of Figure 6.9, the CC and KL evaluation indicators show a better match than SIM in terms of similarity. Therefore, the findings in this thesis are in line with Riche's recommendations [288].

Table 6.3: Comparison of the two human saliency maps (proposed research and Fang's dataset* [276]) with two different saliency prediction models (GBVS [134] and Itti [133]) using three saliency evaluation indicators (SIM, CC and KL).

Video sequence	*Water fountain*		*Wishing well*		*Garden*		**BBB*		**Flame*	
Model	GBVS [134]	Itti [133]	GBVS [134]	Itti [133]	GBVS [134]	Itti [133]	GBVS [134]	Itti [133]	GBVS [134]	Itti [133]
Proposed approach										
SIM↑	**0.681**	0.595	0.719	**0.734**	**0.546**	0.479	0.403	**0.418**	**0.726**	0.636
CC↑	**0.603**	0.420	0.572	**0.714**	**0.275**	0.080	0.280	**0.311**	**0.741**	0.492
KL↓	**0.336**	0.619	0.259	**0.244**	**0.655**	0.875	1.07	**1.025**	**0.249**	0.454
Fang's Dataset [276]										
SIM↑	**0.568**	0.524	0.562	**0.571**	**0.480**	0.416				
CC↑	**0.473**	0.448	0.466	**0.622**	**0.195**	0.031				
KL↓	**0.592**	0.782	0.598	**0.564**	**0.864**	1.13				

*Note: Fang's dataset does not include the "BBB" and "Flame" video sequences



(a) (b) (c)

(d) (e) (f)

Figure 6.9: An example of the comparison between the proposed approach and the Itti [133] saliency map model (a) Original image of "wishing well" video sequence, (b) Proposed human saliency map, (c) Itti's saliency map, and the corresponding three saliency evaluation indicators for visualisation (d) SIM, (e) CC, (f) KL.

The results show that the saliency prediction model could be considered to be influenced by the centre bias from participants. To reduce the influence of centre bias, the centre map could be first computed by taking the average of all human saliency maps obtained from eye-gaze data [287]. Then, the centre bias can be measured by the distance between the image at the centre and the centroid of the centre map. Then, the performance of the centre bias can be analysed by statistical analysis such as a one-way ANOVA [287]. In addition, the selection of saliency evaluation indicators is also important for a comparison of the existing saliency prediction models with the human saliency map. Furthermore, similar to Fang's dataset [276], choosing more video samples including both indoor and outdoor scenes are also key factors for better estimation of saliency prediction models [289].

6.5 Conclusion

Measuring the human vision experience is the most direct way to assess the visual attention of viewers when watching S3D videos and how it relates to QoE evaluation. Eye-tracking experiments have been detailed in this chapter to assess human visual attention, which compared human saliency maps based on eye-tracking gaze data with existing human saliency maps and saliency prediction models when watching S3D video sequences. Three evaluation indicators were used to evaluate the similarity between the human saliency map proposed in this research and other saliency models. The results found that eye-tracking is an important approach to identify the visual attention of viewers for QoE evaluation. However, the existing saliency prediction models are not yet accurate enough to predict a visual attention model, especially for S3D video sequences consisting of depth information, viewing preference and other factors, such as the chaotic movement of objects. However, the results can reflect some characteristics of the viewing preference by viewers and thus can be beneficial for visual attention and QoE assessments.

Further studies can further elaborate on S3D videos of both indoor and outdoor scenes to compare with other saliency prediction models. Also, the study in this thesis was conducted by examining a 3D passive flat screen only. The study has not included any VR device or other viewing environment; however, this can be included in future research work. For example, FOVE VR [224] and HTC VIVE Pro Eye VR [290] headsets can conduct eye-tracking experiments in a VR environment as they have embedded eye-tracking functionality to capture eye-gaze data for further analysis.

In future studies, the research focused on salient object detection can investigate deep learning technology to explore more effective network structures that can retain more spatial details [132, 268]. For example, Lin et al. [126] used different scale inputs to obtain depth information and interconnected the deep neural network features of each layer. In addition, Qin et al. [291] developed a BASNet model that detected visually salient objects and developed saliency object detection maps with clear boundaries. This model used visual attention as a high-level understanding of the entire scene, learning through higher-level neural network layers and salient object detection tasks.

Chapter 7 Conclusions and Future Work

7.1 Introduction

The research conducted in this thesis addresses the video quality assessments of stereoscopic 3D (S3D) video technology, in order to develop novel objective and subjective quality assessments of S3D videos. All of the work and experiments conducted in this thesis were performed with four specific objectives: 1) an investigation of both objective and subjective tests to evaluate visual fatigue measured through Electroencephalography (EEG) through a comparison of viewing 2D and S3D videos; 2) an investigation of both objective and subjective assessments using Quality of Experience (QoE) methods to evaluate the visual fatigue of S3D videos with three different viewing environments; 3) an application of the objective quality metrics and subjective evaluation approaches to video coding applications as a practical use-case; 4) an investigation of eye-tracking analysis when viewing S3D video to develop human saliency maps and compare with existing human saliency maps and saliency prediction models for visual attention whilst watching S3D video.

Firstly, this thesis conducted objective and subjective assessments to investigate differences in the viewers' QoE between 2D and S3D videos. This work proposed and investigated the use of EEG biosignals to measure visual and eye fatigue to augment Visually Induced Motion Sickness (VIMS) measurement through participants' self-reporting using the Simulator Sickness Questionnaire (SSQ). Experiments were conducted to compare fatigue caused by 2D and S3D video sequences of the same video content. Experimental results indicate that the 3D video sequences caused more fatigue for the participants than the 2D video sequences. Further, participants exhibited more eye blinking movement for 2D videos, indicating less eye and visual fatigue. Congruently, the brain wave power ratio results from the EEG biosignals showed that 3D videos caused higher power ratio values than 2D videos, and larger than the fatigue threshold value of 0.05. Previous researchers have shown that 3D videos cause the most VIMS for viewers, and the results in this work confirm these same outcomes. Most VIMS and fatigue were seen whilst watching the video sequence with the greatest image distortion through water movement in the content, depth perception change, time-varying and rapid higher frequency motion.

Secondly, this thesis extended to the QoE assessment of S3D videos in three viewing environments: a flat 3D screen, a panoramic screen and a Virtual Reality (VR) headset. This experiment was performed to identify the effects of VIMS and to analyse 3D fatigue in the three viewing environments. The evaluation of the viewer experience whilst watching S3D video sequences used five metrics. The metrics were evaluated with a variety of content and the results revealed that the viewers who used a VR device to view the S3D video sequences resulted in higher visual fatigue compared to the other two screen displays. In addition, the experimental results indicated that participants who watched S3D video sequences were more likely to exhibit higher VIMS. For the panoramic screen, the participants reported the lowest SSQ scores and the highest enjoyment ratings when viewing S3D video sequences. The difference in results reveal that the screen and viewing environment are an essential factor that influences the level of visual fatigue, QoE and VIMS. The results also revealed that the chaotic movement in some video sequences has an effect on the SSQ, level of visual fatigue and enjoyment ratings. Hence, such findings show that the content of the video sequence and the viewing screen used are key factors that affect the enjoyment rating of the S3D videos. The findings of this chapter are essential for the advancement of the identification of quality factors in the assessment of stereoscopic visual fatigue. In addition, these factor are key to enhancing the development of better quality experiences (and measurement thereof) for 3D screens in various viewing environments, and this study contributes to the adoption of the design of 3D video sequences to suit particular viewing environments. For instance, a better user experience can be achieved by designing a 3D video optimal for a particular viewing environment. In this regard, a user preference of typical S3D video content could be adapted to the 360° video content in a panoramic screen, or the omnidirectional visual content in a VR environment. In this research, the University of Technology Sydney (UTS) Data Arena provided a suitable yet unique immersive surround, panoramic multimedia facility on-campus.

Thirdly, eye-tracking experiments were conducted for the investigation of human visual attention to determine whether visual attention obtained from eye-tracking methods contribute to the QoE assessments of S3D videos. Experiments have been conducted to compare human saliency maps based on eye-tracking gaze data with existing human saliency maps and saliency prediction models when watching S3D video sequences. Three evaluation indicators were used to evaluate the similarity between the human saliency map proposed in this research and other saliency models. The results found that eye-tracking is a key method to identify the visual attention of viewers for QoE evaluation. The existing saliency prediction models are not yet accurate enough to predict a visual attention model, especially for S3D video sequences consisting of depth information, viewing preference and other factors, such as the chaotic movement of objects. However, the results can reflect some characteristics of the viewing preference by viewers and thus can be beneficial for visual attention in relation to QoE assessments.

Finally, the application of the objective quality metrics and subjective evaluation approaches were applied as a practical use-case to video coding evaluation. A “hybrid” sequencing video coding approach was thus proposed for evaluation for both compressed and compressed S3D videos. The study investigated the effects that hybrid sequencing has on the viewer QoE of the S3D finding that hybrid sequencing can provide perceptual quality similar to that of uncompressed videos. This was found to be true for videos with scenes that have 3D effects that are moderate and video whose scenes are stable. The different artefacts that affect the quality of 3D content and which therefore need to be reduced to maintain a comparable video quality were also noted in the evaluation results.

7.2 Research Question Outcomes

Research Question 1: How to assess the visual fatigue of S3D video in both objective and subjective quality assessments?

Following experimental analysis of various factors, two methods are used to measure visual fatigue of S3D video in this thesis: 1) the self-reported Simulator Sickness Questionnaire (SSQ) to measure Visually Induced Motion Sickness (VIMS); and 2) the analysis of the correlation between biosignals and 3D visual fatigue, in which Electroencephalography (EEG) signals record the bioelectrical activities in the brain to quantify the subjective emotional engagement of viewers when watching S3D videos, and the EEG indices are used to assess S3D visual fatigue. Chapter 3 and Chapter 4 discussed in detail the research conducted to assess 3D fatigue with experimental evaluations.

The SSQ results show that none of the participants felt any VIMS after the experiment in flat 3D screen. However, the majority of the participants exhibited eye fatigue when participants watched video sequences for a prolonged period of time. Thus, the brain wave power ratio results from the measurement EEG biosignals showed that 3D videos caused higher power ratio values than 2D videos, and larger than the fatigue threshold value of 0.05.

- **Research question 1(a):** How to assess the viewers' perceptual experiences and visual attention when watching S3D video?

The measure of eye blinking movement rate, attention and meditation levels, and the self-reported enjoyment level by Absolute Category Rating (ACR) method were used to measure the perceptual experiences and the visual attention of viewers. Chapter 3 and Chapter 4 discussed these methods for evaluations and presented relevant results.

The eye blink movement rate indicates that the visual fatigue is likely to be higher for 3D video sequences compared to 2D video sequences. In addition, both video sequence and the viewing device are not significant factors affecting attention and meditation levels. Furthermore, chaotic movement of video sequences may be correlated with lower enjoyment rating.

- **Research question 1(b):** How to evaluate the perceptual experience and visual attention of viewers watching S3D video in different viewing environments?

Three different viewing environments that are commonly used as the viewing environments of S3D videos were evaluated: a VR headset device, a panoramic screen, and a flat 3D screen. Evaluation methods investigated including the SSQ, EEG biosignals, eye blinking, attention and meditation levels, enjoyment rating by ACR and statistical analysis by Analysis Of Variance (ANOVA). Chapter 4 discussed in detail this research design and experimental methods in three different viewing environments with evaluations and results presented.

The experimental results show that VR device to view the stereoscopic video sequences resulted in higher visual fatigue compared to the other two screen display types. In addition, the participants reported the lowest enjoyment ratings for the VR device. For the panoramic screen, the participants reported the lowest SSQ scores and the highest enjoyment ratings when viewing S3D video sequences. The difference in results experienced when viewing the same S3D video sequence on different screens reveal that the screen and viewing environment are an essential factor that influences the level of visual fatigue, QoE and VIMS. The results also revealed that the chaotic movement in video sequences had an effect on the SSQ, level of visual fatigue and enjoyment ratings. Hence, such findings show that the content of the video sequence and the viewing screen used are key factors that affect the enjoyment rating of the S3D videos.

Research Question 2: How to assess the visual attention of viewers watching S3D video in relation to the quality of experience?

Measuring the human vision experience is the most direct way to understand the visual attention of viewers when watching S3D videos. Eye-tracking experiments were conducted to measure human visual attention when viewing S3D videos. Some visual attributes were selected for the prediction of visual attention when viewing S3D video. Human saliency maps based on eye-tracking gaze data were compared with existing human saliency maps and saliency prediction models when watching S3D video sequences. Three evaluation indicators were used to evaluate the similarity between the human saliency map and other saliency models. Chapter 6 discussed eye-tracking techniques, comparing with existing human saliency map saliency prediction models and presented experimental results to identify if there is any correlation to the quality of experience of viewers.

The results show that the saliency prediction model could be influenced by the centre bias from participants. In addition, the selection of saliency evaluation indicators is also important for a comparison of the existing saliency prediction models with the human saliency map. Furthermore, choosing more video samples including both indoor and outdoor scenes are also key factors for better estimation of saliency prediction models.

Research Question 3: How are objective metrics and the quality of experiences of viewers affected when implementing a video sequencing model for video coding as a practical application?

The commonly utilised PSNR and evaluation by ACR were used to evaluate the objective metrics and the subjective QoE, respectively. A practical use-case of “hybrid” sequence video coding approaches were proposed and adopted for the practical application of these evaluation methods. Evaluation results found the proposed “hybrid” S3D video coding approach to provide perceptual quality similar to that of uncompressed videos with scenes where the 3D effects are moderate and where the scenes are stable. Chapter 5 discussed in detail the application of video coding evaluation as a practical use-case to measure and evaluate objective metrics and the viewer quality of experience, presenting relevant results.

The results show that hybrid sequencing can provide perceptual quality comparable to that of uncompressed videos, with the 3D effects that are moderate and stable scenes. The different artefacts that affect the quality of 3D content and which therefore need to be reduced to maintain a comparable video quality were also noted in the evaluation results.

7.3 Summary of Contributions

The list below summarises the main technical contributions of the work in this thesis:

- Chapter 3 : Comparison of Objective and Subjective S3D Evaluation using EEG Biosignals (published in [36])
  - Employed Simulator Sickness Questionnaire (SSQ) scores to evaluate the Visually Induced Motion Sickness (VIMS) of 2D and S3D videos.
  - Presented the effect of the eye blinking frequency of participants whilst viewing 2D and S3D videos.

- Adopted Electroencephalography (EEG) signals and brain wave power algorithms to evaluate visual fatigue of participants whilst viewing 2D and S3D videos.
- Conducted objective and subjective tests to evaluate the visual fatigue of participants whilst viewing both 2D and S3D videos.

- Chapter 4 : Comparison of S3D Video QoE in Different Viewing Environments (published in [37])
  - Employed SSQ scores and its sub-scales to evaluate the VIMS of S3D videos in different viewing environments.
  - Presented the effect of eye blinking frequency and neuro signal conditions (EEG) of participants viewing S3D videos in different viewing environments.
  - Adopted the use of EEG signals and its brain wave power algorithms to evaluate visual fatigue.
  - Suggested the major factors that affect visual fatigue and evaluated the QoE of viewers whilst watching S3D videos in different viewing environments with three different projection devices.
- Chapter 5 : S3D Objective Quality Metrics and Subjective Evaluation Methods applied to Video Compression Evaluation (published in [38])
  - Compared and evaluated PSNR between the uncompressed and compressed S3D video at the same resolution.
  - Proposed a practical use-case, a "hybrid" coding approach, to compare and evaluate PSNR between the two stereoscopic views.
  - Proposed the "hybrid" coding approach to compare and evaluate the QoE of viewers whilst watching the different video pairs of S3D video.

- Suggested the major factors that affect the objective metrics and subjective evaluation of the QoE of viewers in the practical S3D video coding use-case for evaluation.

- Chapter 6 : Visual Attention of S3D Video in relation to QoE using Eye-tracking Analysis
  - Contributed a dataset to obtain the eye-gaze data of participants when viewing S3D videos and developed an approach to estimate the fixation map.
  - Developed a human saliency map for the visual attention of S3D video to correlate with eye-gaze data.
  - Compared the human saliency map with existing human saliency maps and saliency map prediction models from other researchers for analysis.
  - Evaluated if visual attention from the eye-tracking analysis contributed to QoE assessments of S3D video.

7.4 Future Work

For the future development of objective and subjective quality assessments of S3D video, the following research directions are suggested for extending the work in this thesis:

1. **QoE of different video formats in different viewing environments**

Further studies should include QoE assessments of 360° 2D videos, S3D videos and S3D videos with omnidirectional content in panoramic screening environments; and include the SMPTE and THX specifications for the 3D fatigue assessment of UHD-1 video.

2. **Larger sample size of participants and more video sequences**

Regarding the impact of the viewing device, the sample size of video sequences and the number of participants, it is recommended for future work to use a larger sample size of video sequences and more participants to further assess the correlation of ACR enjoyment rating with both the brain wave power ratios and SSQ scores.

3. Eye-tracking in VR technology

The work in this thesis may inspire more S3D videos, including both indoor and outdoor scenes. The viewing device, especially a VR headset, can be used in future experiments. Furthermore, the FOVE VR [224] and HTC VIVE Pro Eye VR [290] headsets are potential device options to conduct the eye-tracking experiments in a VR environment because they have embedded eye-tracking functionality to capture eye-gaze data for further analysis.

4. Multi-view S3D video technology

The challenge of multi-view S3D video technology includes the design of a quality metric for assessments and the adoption of a QoE method to assess the visual attention of viewers when viewing multi-view S3D video. Future work should expand to develop evaluation metrics for multi-view 3D technology and the assessment of 3D fatigue with the enhancement of video quality, such as multi-view High Efficiency Video Coding (HEVC). Also, future research work should expand the objective metrics and subjective evaluations proposed in this thesis to assess higher-resolution stereoscopic video, such as Full-HD and 4K, which are available on the commercial market in commercial movies and 3D-enabled consumer televisions [23, 26-28, 73].

References

[1] A. Vetro, A. M. Tourapis, K. Muller and T. Chen, "3D-TV content storage and transmission," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 384-394, 2011.

[2] F. Albiol, A. Corbi and A. Albiol. "3D measurements from X-ray images and dense surface mappings," In *2016 Global Medical Engineering Physics Exchanges/Pan American Health Care Exchanges (GMEPE/PAHCE)*, 2016, pp. 1-5.

[3] G. Van Wallendael, S. Van Leuven, J. De Cock, F. Bruls and R. Van De Walle, "3D video compression based on high efficiency video coding," *IEEE Transactions on Consumer Electronics*, vol. 58, no. 1, pp. 137-145, 2012.

[4] M. H. Pinson and M. Barkowsky, "Selecting scenes for 2D and 3D subjective video quality tests," *EURASIP Journal on Image and Video Processing*, vol. 50, 2013.

[5] ITU-T P.913, "Methods for the subjective assessment of video quality, audio quality and audiovisual quality of Internet video and distribution quality television in any environment," International Telecommuncation Union, Mar., 2016.

[6] L. Chiariglione, J. Ostermann and M. Tanimoto, "*The MPEG Representation of Digital Media*". New York: Springer, 2016.

[7] H. G. Koumaras, "Objective and subjective assessment methods of video quality in multimedia broadcasting," in *Mobile Multimedia Broadcasting Standards*: Springer, Boston, MA, 2009, pp- 417-434.

[8] O. Marques, "*Practical Image and Video Processing Using MATLAB*". New Jersey: John Wiley & Sons, Inc, 2011.

[9] MPEG. "Call for proposals on multi-view video coding," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. N7327*, Poznan, Poland, 2005.

[10] M. Tanimoto, T. Fujii and N. Fukushima. "1D parallel test sequences for MPEG-FTV," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. M15378*, Archamps, France, 2008.

[11] M. P. Tehrani, S. Shimizu, G. Lafruit, T. Senoh, T. Fujii, A. Vetro and M. Tanimoto. "Use cases and requirements on free-viewpoint television (FTV)," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. N14014*, Geneva, Switzerland, 2013.

[12] K. Wegner and O. Stankiewicz. "Depth estimation reference software (DERS) 6.1 software manual," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. MPEG2014/M34302*,

Sapporo, Japan, 2014.

[13] T. Senoh, K. Yamamoto, N. Tetsutani, H. Yasuda and K. Wegner. "View Synthesis Reference Software (VSRS) 4.2 with improved inpainting and hole filing," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. MPEG2017/M40657*, Hobart, Australia, 2017.

[14] Y.-S. Ho and Y. Song. "MPEG activities for 3D video coding," In *Signal and Information Processing Association Annual Summit and Conference (APSIPA), 2014 Asia-Pacific*, 2014, pp. 1-4.

[15] ITU-T Rec. BT.500-13, "Methodology for the subjective assessment of the quality of television pictures," International Telecommuncation Union, Jan., 2012.

[16] MPEG. "3D video subjective quality assessment test plan v2," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. N13920*, Geneva, Switzerland, 2013.

[17] G. Lafruit, K. Wegner, T. Grajek, T. Senoh, K. P. Tamás, P. Goorts, L. Jorissen, B. Ceulemans, P. C. Lopez, S. G. Lobo, Q. Wang, J. Jung and M. Tanimoto. "FTV software framework," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. MPEG2014/N15349*, Warsaw, Poland, 2015.

[18] K. Wegner, O. Stankiewicz, T. Grajek and M. Domański. "Depth estimation from circular projection of 360 degree 3D video," In *ISO/IEC JTC1/SC29/WG11 MPEG Doc. MPEG2017/M40596*, Hobart, Australia, 2017.

[19] H. Duan, G. Zhai, X. Min, Y. Zhu, W. Sun and X. Yang, "Assessment of visually induced motion sickness in immersive videos," In *Advances in Multimedia Information Processing – PCM 2017*, 2018, pp. 662-672.

[20] K. Terzic and M. Hansard, "Causes of discomfort in stereoscopic content: A review," *arXiv*, 2017.

[21] K. Terzić and M. Hansard, "Methods for reducing visual discomfort in stereoscopic 3D: A review," *Signal Processing: Image Communication*, vol. 47, pp. 402-416, 2016.

[22] T. Bando, A. Iijima and S. Yano, "Visual fatigue caused by stereoscopic images and the search for the requirement to prevent them: A review," *Displays*, vol. 33, no. 2, pp. 76-83, 2012.

[23] S. A. A. Naqvi, N. Badruddin, A. S. Malik, W. Hazabbah and B. Abdullah. "Does 3D produce more symptoms of visually induced motion sickness?," In *Engineering in Medicine and Biology Society (EMBC), 2013 35th Annual International Conference of the IEEE*, 2013, pp. 6405-6408.

[24] R. S. Kennedy, J. Drexler and R. C. Kennedy, "Research in visually induced motion sickness," *Applied Ergonomics*, vol. 41, no. 4, pp. 494-503, 2010.

[25] ITU-T Rec. BT2021.1, "Subjective methods for the assessment of stereoscopic 3DTV systems," International Telecommuncation Union, Feb., 2015.

[26] F. Qi, T. Jiang, X. Fan, S. Ma and D. Zhao. "Stereoscopic video quality assessment based on stereo just-noticeable difference model," In *2013 20th IEEE International Conference on Image Processing (ICIP)*, 2013, pp. 34-38.

[27] A. S. Umar, R. M. Swash and A. H. Sadka. "Subjective quality assessment of 3D videos," In *IEEE AFRICON 2011*, Victoria Falls, Livingstone, 2011.

[28] T. Kawano, K. Yamagishi and T. Hayashi. "Performance comparison of subjective assessment methods for 3D video quality," In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop on*, 2012, pp. 218-223.

[29] Z. Chen, W. Lin and K. N. Ngan. "Perceptual video coding: Challenges and approaches," In *2010 IEEE International Conference on Multimedia and Expo*, 2010, pp. 784-789.

[30] S. Wilkinson, "Quad HD to UHDTV: Making the difference," *SMPTE Webinar*, 2014. [Online]. Available: www.avsforum.com/forum/286-latest-industry-news/1609674-smpte-webinar-quad-hd-uhdtv.html. [Accessed: Mar. 8, 2016].

[31] H. Sohn, Y. J. Jung and Y. M. Ro, "Crosstalk reduction in stereoscopic 3D displays: Disparity adjustment using crosstalk visibility index for crosstalk cancellation," *Opt Express*, vol. 22, no. 3, pp. 3375-3392, 2014.

[32] A. J. Woods, "Crosstalk in stereoscopic displays: a review," *Journal of Electronic Imaging*, vol. 21, no. 4, 2012.

[33] A. Saxena, M. Sun and A. Y. Ng, "Make3D: depth perception from a single still image," In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 3*, Chicago, Illinois, 2008, pp. 1571-1576.

[34] K. Karsch, C. Liu and S. Kang. "Depth extraction from video using non-parametric sampling," In *12th European Conference on Computer Vision*, Florence, Italy, 2012, pp. 775-788.

[35] B. Oztas, M. T. Pourazad, P. Nasiopoulos, I. Sodagar and V. C. M. Leung. "A rate adaptation approach for streaming multiview plus depth content," In *2014 International Conference on Computing, Networking and Communications (ICCNC)*, 2014, pp.

1006-1010.

[36] S.-M. Choy, K.-H. Chiu, E. Cheng and I. Burnett. "3D fatigue from stereoscopic 3D video displays: Comparing objective and subjective tests using electroencephalography," In *TENCON IEEE Region 10 Conference*, 2015, pp. 1-4.

[37] S.-M. Choy, E. Cheng, R. H. Wilkinson, I. Burnett and M. W. Austin, "Quality of experience comparison of stereoscopic 3D videos in different projection devices: Flat screen, panoramic screen and virtual reality headset," *IEEE Access*, vol. 9, pp. 9584-9594, 2021.

[38] S.-M. Choy, E. Cheng and I. Burnett, "Hybrid sequencing of uncompressed and compressed 3D stereoscopic video: A preliminary quality evaluation," In *The International Conference on Electrical Engineering (ICEE)*, Okinawa, Japan, 2016, pp. 1-5.

[39] R. S. Kennedy, N. E. Lane, K. S. Berbaum and M. G. Lilienthal, "Simulator sickness questionnaire: An enhanced method for quantifying simulator sickness," *The International Journal of Aviation Psychology*, vol. 3, no. 3, pp. 203-220, 1993.

[40] M. Karimi, N. Soltanian, S. Samavi, K. Najarian, N. Karimi and S. M. Reza Soroushmehr, "Blind stereo image quality assessment inspired by brain sensory-motor fusion," *Digital Signal Processing*, vol. 91, pp. 91-104, 2019.

[41] D. De Silva, G. Nur Yilmaz, E. Ekmekcioglu and A. Kondoz, "QoE of 3D media delivery systems," in *Media Networks: Architectures, Applications, and Standards*. United Kingdom: Taylor & Francis Group, 2012, pp- 275-292.

[42] P.-A. Blanche, "Toward the Ultimate 3-D Display," *Information Display*, vol. 28, no. 2-3, pp. 32-37, 2012.

[43] D. Page, T. Thomas, S. Kelley, P. Jones and D. Miller. "Vergence and accommodation in simulation and training with 3D displays," In *Interservice/Industry Training, Simulation, and Education Conference*, 2014, pp. 1-12.

[44] E. Lueder, "*3D Displays*". West Sussex: A John Wiley & Sons Ltd., 2012.

[45] Y. Kim, J. Kim, K. Hong, H. K. Yang, J.-H. Jung, H. Choi, S.-W. Min, J.-M. Seo, J.-M. Hwang and B. Lee, "Accommodative response of integral imaging in near distance," *Journal of Display Technology*, vol. 8, no. 2, pp. 70-78, 2012.

[46] A. J. Wilkins, "*Visual Stress*". Oxford: Oxford University Press, 1995.

[47] D. M. Hoffman, A. R. Girshick, K. Akeley and M. S. Banks, "Vergence–

accommodation conflicts hinder visual performance and cause visual fatigue," *Journal of Vision*, vol. 8, no. 3:33, pp. 1-30, 2008.

[48] GoodFirms, "3D game development companies," 2021. [Online]. Available: www.goodfirms.co/game-development-companies/3d. [Accessed: Jun. 13, 2021].

[49] Facebook, "Facebook Oculus," 2020. [Online]. Available: https://developers.facebook.com/products/oculus/. [Accessed: Jun. 6, 2020].

[50] HTC Corporation, "HTC VIVE," 2020. [Online]. Available: www.vive.com/us/product/#vive. [Accessed: Feb. 4, 2020].

[51] Sony, "Playstation VR," 2021. [Online]. Available: https://www.playstation.com/en-hk/ps-vr/. [Accessed: May. 8, 2021].

[52] S. A. E. M. Nasri, "Multiview coding and compression for 3D video." Ph. D. thesis, Department of Electronics, Annaba University, Annaba, 2018.

[53] M. Takada, M. Miyao, F. Kinoshita, A. Sugiura, K. Fujikake and H. Takada. "Changes in brain blood flow during stereoscopic video clips," In *2017 12th International Conference on Computer Science and Education (ICCSE)*, 2017, pp. 65-68.

[54] R. R. Rubio, R. D. Bonaventura, I. Kournoutas, D. Barakat, V. Vigo, I. El-Sayed and A. A. Abla, "Stereoscopy in surgical neuroanatomy: Past, present, and future," *Operative Neurosurgery*, 2019.

[55] J. Gil and M. Kim, "Motion depth generation using MHI for 2D-to-3D video conversion," *Electronics Letters*, vol. 53, no. 23, pp. 1520-1522, 2017.

[56] P. Seuntiëns, I. Heynderickx, W. Ijsselsteijn, P. Avoort, J. Berentsen, I. Dalm, M. Lambooij and W. Oosting. "Viewing experience and naturalness of 3D images," In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2005.

[57] M. M. Walsh, G. Gunzelmann and H. P. A. Van Dongen, "Computational cognitive modeling of the temporal dynamics of fatigue from sleep loss," *Psychonomic Bulletin & Review*, vol. 24, no. 6, pp. 1785-1807, 2017.

[58] J. Li, M. Barkowsky and P. Le Callet, "Visual Discomfort in 3DTV: Definitions, Causes, Measurement, and Modeling," in *Novel 3D Media Technologies*. New York: Springer, 2015, pp- 185-209.

[59] M. Z. Ramadan, M. H. Alhaag and M. H. Abidi, "Effects of Viewing Displays from Different Distances on Human Visual System," *Applied Sciences*, vol. 7, no. 11, pp. 1153, 2017.

[60] N. Manshouri and T. Kayikcioglu, "A Comprehensive Analysis of 2D&3D Video Watching of EEG Signals by Increasing PLSR and SVM Classification Results," *ArXiv*, vol. abs/1903.05636, 2020.

[61] M.-K. Kang, H. Cho, H.-M. Park, S. C. Jun and K.-J. Yoon, "A wellness platform for stereoscopic 3D video systems using EEG-based visual discomfort evaluation technology," *Applied Ergonomics*, vol. 62, pp. 158-167, 2017.

[62] M. Urvoy, M. Barkowsky and P. L. Callet, "How visual fatigue and discomfort impact 3D-TV quality of experience: a comprehensive review of technological, psychophysical, and psychological factors," *Annals of Telecommunications*, vol. 68, pp. 641-655, 2013.

[63] M. T. M. Lambooij, W. A. IJsselsteijn and I. Heynderickx. "Visual discomfort in stereoscopic displays: a review," In *Electronic Imaging 2007*, 2007.

[64] H. Sohn, Y. J. Jung, S.-I. Lee, H. W. Park and Y. M. Ro. "Attention model-based visual comfort assessment for stereoscopic depth perception," In *2011 17th International Conference on Digital Signal Processing (DSP)*, 2011, pp. 1-6.

[65] S.-I. Lee, Y. J. Jung, H. Sohn, Y. M. Ro and H. W. Park. "Visual discomfort induced by fast salient object motion in stereoscopic video," In *IS&T/SPIE Electronic Imaging*, 2011.

[66] S. Tian, L. Zhang, L. Morin and O. Deforges, "A full-reference image quality assessment metric for 3D synthesized views," *Electronic Imaging*, vol. 2018, no. 12, pp. 366-1-366-5, 2018.

[67] J. Seo, X. Liu, D. Kim and K. Sohn, "An objective video quality metric for compressed stereoscopic video," *Circuits, Systems, and Signal Processing*, vol. 31, no. 3, pp. 1089-1107, 2012.

[68] Q. Huynh-Thu and M. Ghanbari, "The accuracy of PSNR in predicting video quality for different video scenes and frame rates," *Telecommunication Systems*, vol. 49, no. 1, pp. 35-48, 2012.

[69] U. Sara, M. Akter and M. S. Uddin, "Image quality assessment through FSIM, SSIM, MSE and PSNR - A comparative study," *Journal of Computer and Communications*, vol. 7, pp. 8-18, 2019.

[70] A. S. Malik, "*Depth Map and 3D Imaging Applications: Algorithms and Technologies*". Portland, United States, Portland: Book News, Inc., 2012.

[71] T. Tominaga, T. Hayashi, J. Okamoto and A. Takahashi. "Performance comparisons of subjective quality assessment methods for mobile video," In *2010 Second International Workshop on Quality of Multimedia Experience (QoMEX)*, 2010, pp. 82-87.

[72] Y. Liu and O. Sourina. "EEG-based dominance level recognition for emotion-enabled interaction," In *2012 IEEE International Conference on Multimedia and Expo (ICME)*, 2012, pp. 1039-1044.

[73] K. Gu, G. Zhai, X. Yang and W. Zhang, "No-reference stereoscopic IQA approach: From nonlinear effect to parallax compensation," *Journal of Electrical and Computer Engineering*, 2012.

[74] H.-T. Quan, M. Barkowsky and P. Le Callet, "The importance of visual attention in improving the 3D-TV viewing experience: Overview and new perspectives," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 421-431, 2011.

[75] E. Cheng, P. Burton, J. Burton, A. Joseski and I. Burnett. "RMIT3DV: Pre-announcement of a creative commons uncompressed HD 3D video database," In *Quality of Multimedia Experience (QoMEX), 2012 Fourth International Workshop*, 2012, pp. 212-217.

[76] F. Qi, T. Jiang, S. Ma and D. Zhao. "Quality of experience assessment for stereoscopic images," In *2012 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2012, pp. 1712-1715.

[77] A. Chammem, "Robust watermarking techniques for stereoscopic video protection." PhD thesis, Institut National des Télécommunications, 2013.

[78] S. A. Mahmood and R. F. Ghani. "Objective quality assessment of 3D stereoscopic video based on motion vectors and depth map features," In *2015 7th Computer Science and Electronic Engineering Conference (CEEC)*, 2015, pp. 179-183.

[79] Z. Wang and A. C. Bovik, "A universal image quality index," *IEEE Signal Processing Letters*, vol. 9, no. 3, pp. 81-84, 2002.

[80] W. J. Tam, F. Speranza, S. Yano, K. Shimono and H. Ono, "Stereoscopic 3D-TV: Visual comfort," *IEEE Transactions on Broadcasting*, vol. 57, no. 2, pp. 335-346, 2011.

[81] ITU-T Rec. P.910, "Subjective video quality assessment methods for multimedia applications," International Telecommuncation Union, Apr., 2008.

[82] K. Wang, M. Barkowsky, K. Brunnström, M. Sjöström, R. Cousseau and P. Le Callet, "Perceived 3D TV transmission quality assessment: Multi-laboratory results using

absolute category rating on quality of experience scale," *IEEE Transactions on Broadcasting*, vol. 58, no. 4, pp. 544-557, 2012.

[83] SMPTE, "Society of motion picture and television engineers," 2021. [Online]. Available: https://www.smpte.org/. [Accessed: Jun. 20, 2021].

[84] THX Ltd., "Tomlinson Holman experiment," 2021. [Online]. Available: https://www.thx.com/. [Accessed: Jun. 20, 2021].

[85] F. Zeri and S. Livi, "Visual discomfort while watching stereoscopic three-dimensional movies at the cinema," *Ophthalmic and Physiological Optics*, vol. 35, 2015.

[86] E. Garci, "Pixels per degree calculator," 2014. [Online]. Available: https://res18h39.netlify.com/calculator. [Accessed: Apr. 7, 2019].

[87] P. Kauff, N. Atzpadin, C. Fehn, M. Müller, O. Schreer, A. Smolic and R. Tanger, "Depth map creation and image-based rendering for advanced 3DTV services providing interoperability and scalability," *Signal Processing: Image Communication*, vol. 22, no. 2, pp. 217-234, 2007.

[88] P. Rotter, "Why did the 3D revolution fail? The present and future of stereoscopy," *IEEE Technology and Society Magazine*, vol. 36, no. 1, pp. 81-85, 2017.

[89] M. Y.-C. Yim, Y. Abdourazakou, P. L. Sauer and S.-Y. Park, "Modelling the dimensionality effects on brand placement effectiveness in stereoscopic 3-D versus 2-D sports games," *International Journal of Advertising*, vol. 37, no. 6, pp. 958-983, 2018.

[90] B. Zou, Y. Liu, M. Guo and Y. Wang, "EEG-based assessment of stereoscopic 3D visual fatigue caused by vergence-accommodation conflict," *Journal of Display Technology*, vol. 11, no. 12, pp. 1076-1083, 2015.

[91] R. S. Kennedy, M. G. Lilienthal and K. S. Berbaum, "Simulator sickness in U.S. Navy flight simulators," *Aviation, space, and environmental medicine*, vol. 60, no. 1, pp. 10-16, 1989.

[92] A. Solimini, G. La Torre, A. Mannocci, D. Di Thiene and A. Boccia, "Prevalence of symptoms of visually induced motion sickness and visual stress during and after viewing a 3D movie," *European Journal of Public Health*, vol. 20, 2010.

[93] R. Cassani, M.-A. Moinnereau and T. Falk. "A neurophysiological sensor-equipped head-mounted display for instrumental QoE assessment of immersive multimedia," In *2018 Tenth International Conference on Quality of Multimedia Experience (QoMEX)*, 2018, pp. 1-6.

[94] J. Beyer, R. Varbelow, J.-N. Antons and S. Möller. "Using electroencephalography and subjective self-assessment to measure the influence of quality variations in cloud gaming," In *2015 Seventh International Workshop on Quality of Multimedia Experience (QoMEX)*, 2015, pp. 1-6.

[95] S. Arndt, J.-N. Voigt-Antons, R. Schleicher, S. Moller and G. Curio. "Perception of low-quality videos analyzed by means of electroencephalography," In *2012 Fourth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2012, pp. 284-289.

[96] QoMEX, "Quality of multimedia experience community," 2021. [Online]. Available: http://qomex.org/. [Accessed: Jun. 20, 2021].

[97] NeuroSky, "MindWave mobile: User guide," *NeuroSky, Inc.*, 2015. [Online]. Available: http://download.neurosky.com/support_page_files/MindWaveMobile/docs/mindwave_mobile_user_guide.pdf. [Accessed: Jan. 7, 2020].

[98] Biofeedback and neurofeedback therapy, "NeuroSky home use device for neurofeedback," 2019. [Online]. Available: https://biofeedback-neurofeedback-therapy.com/neurosky/. [Accessed: Jun. 5, 2021].

[99] J. Lagopoulos, J. Xu, I. Rasmussen, A. Vik, G. S. Malhi, C. F. Eliassen, I. E. Arntsen, J. G. Sæther, S. Hollup, A. Holen, S. Davanger and O. Ellingsen, "Increased theta and alpha EEG activity during nondirective meditation," *Journal of Alternative and Complementary Medicine*, vol. 15, pp. 1187-1192, 2009.

[100] NeuroSky Inc., "NeuroSky," 2021. [Online]. Available: http://neurosky.com/. [Accessed: Jun. 15, 2021].

[101] InteraXon Inc., "Muse," 2021. [Online]. Available: https://choosemuse.com/. [Accessed: Jun. 15, 2021].

[102] EMOTIV Inc., "EMOTIV," 2021. [Online]. Available: www.emotiv.com/. [Accessed: Jun. 15, 2021].

[103] OpenBCI, "OpenBCI," 2021. [Online]. Available: https://openbci.com/. [Accessed: Jun. 15, 2021].

[104] A. D. Bateson, H. A. Baseler, K. S. Paulson, F. Ahmed and A. U. R. Asghar, "Categorisation of mobile EEG: A researcher's perspective," *BioMed Research International*, vol. 2017, no. 5496196, pp. 1-15, 2017.

[105] H. Yuan, J. Calic, A. Fernando and A. Kondoz. "User requirements elicitation of

stereoscopic 3D video interaction," In *2012 IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, 2012, pp. 31-36.

[106] P. C. Petrantonakis and L. J. Hadjileontiadis. "EEG-based emotion recognition using hybrid filtering and higher order crossings," In *2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops*, 2009, pp. 1-6.

[107] J. Chen, W. Cranton and M. Fihn, "*Handbook of Visual Display Technology*". London: Springer-Verlag Berlin Heidelberg, 2012.

[108] C. Klein and U. Ettinger, "*Eye Movement Research: An Introduction to its Scientific Foundations and Applications*". Switzerland: Springer International Publishing, 2019.

[109] M. F. Land and S. Furneaux, "The knowledge base of the oculomotor system," *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, vol. 352, no. 1358, pp. 1231-1239, 1997.

[110] D. W. Hansen and Q. Ji, "In the eye of the beholder: A survey of models for eyes and gaze," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 3, pp. 478-500, 2010.

[111] R. P. G. Van Gompel, M. H. Fischer, W. S. Murray and R. L. Hill, "*Eye Movements: A Window on Mind and Brain*". Oxford: Elsevier, 2007.

[112] K. R. Gegenfurtner, D. Xing, B. H. Scott and M. J. Hawken, "A comparison of pursuit eye movement and perceptual performance in speed discrimination," *Journal of Vision*, vol. 3, no. 11, pp. 865-876, 2003.

[113] K. A. Turano and S. M. Heidenreich, "Eye movements affect the perceived speed of visual motion," *Vision Research*, vol. 39, no. 6, pp. 1177-1187, 1999.

[114] E. O’Connor, T. H. Margrain and T. C. A. Freeman, "Age, eye movement and motion discrimination," *Vision Research*, vol. 50, no. 23, pp. 2588-2599, 2010.

[115] P. J. Bennett, R. Sekuler and A. B. Sekuler, "The effects of aging on motion detection and direction identification," *Vision Research*, vol. 47, no. 6, pp. 799-809, 2007.

[116] O. Ferhat and F. Vilariño, "Low cost eye tracking: The current panorama," *Computational Intelligence and Neuroscience*, vol. 2016, pp. 1-14, 2016.

[117] P. Hanhart and T. Ebrahimi. "Subjective evaluation of two stereoscopic imaging systems exploiting visual attention to improve 3D quality of experience," In *Electronic Imaging*, 2014.

[118] K. Dierkes, M. Kassner and A. Bulling. "A novel approach to single camera, glint-free 3D eye model fitting including corneal refraction," In *2018 ACM Symposium on Eye Tracking Research & Applications*, 2018, pp. 1-9.

[119] F. B. Narcizo, J. E. Rangel de Queiroz and H. M. Gomes. "Remote eye tracking systems: technologies and applications," In *26th Conference on Graphics, Patterns and Images Tutorials*, 2013, pp. 15-22.

[120] C. Hennessey, B. Noureddin and P. Lawrence, "A single camera eye-gaze tracking system with free head motion," In *Proceedings of the Eye Tracking Research & Application Symposium*, San Diego, California, USA, 2006, pp. 87-94.

[121] F. Battisti, M. Carli, P. Le Callet and P. Paudyal, "Toward the assessment of quality of experience for asymmetric encoding in immersive media," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 392-406, 2018.

[122] M. Cognolato, M. Atzori and H. Müller, "Head-mounted eye gaze tracking devices: An overview of modern devices and recent advances," *Journal of Rehabilitation and Assistive Technologies Engineering*, vol. 5, pp. 2055668318773991, 2018.

[123] A. Ciancio, J. F. L. de Oliveira, F. M. L. Ribeiro, E. A. B. da Silva and A. Said. "Quality perception in 3D interactive environments," In *2013 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2013, pp. 9-12.

[124] A. Kar and P. Corcoran, "A review and analysis of eye-gaze estimation systems, algorithms and performance evaluation methods in consumer platforms," *IEEE Access*, vol. 5, pp. 16495-16519, 2017.

[125] Tobii Technology, "Accuracy and precision test method for remote eye trackers, test specification version 2.1.1," 2011. [Online]. Available: https://stemedhub.org/resources/3311/download/Tobii_Test_Specifications_Accuracy_and_PrecisionTestMethod_version_2_1_1_.pdf. [Accessed: Mar. 8, 2021].

[126] C. J. Lin, Y. T. Prasetyo and R. Widyaningrum, "Eye movement parameters for performance evaluation in projection-based stereoscopic display," *Journal of Eye Movement Research*, vol. 11, no. 6, 2018.

[127] D. Xu, "Capturing and post-processing of stereoscopic 3D content for improved quality of experience." Ph. D. thesis, Faculty of Graduate Studies (Electrical & Computer Engineering), University of British Columbia, Vancouver, Canada, 2013.

[128] C. T. E. R. Hewage, J. Wang, M. G. Martini and P. Le Callet. "Visual saliency driven error protection for 3D video," In *2013 IEEE International Conference on Multimedia*

and Expo Workshops (ICMEW), 2013, pp. 1-6.

[129] S. Ahn, J. Kim, H. Kim and S. Lee. "Visual attention analysis on stereoscopic images for subjective discomfort evaluation," In *2016 IEEE International Conference on Multimedia and Expo (ICME)*, 2016, pp. 1-6.

[130] Y. Liu, S. Guo, L. Yu and E. D. Reichle, "Word predictability affects saccade length in Chinese reading: An evaluation of the dynamic-adjustment model," *Psychonomic Bulletin & Review*, vol. 25, no. 5, pp. 1891-1899, 2018.

[131] H. G. Debarba, M. Montagud, S. Chague, J. Lajara, I. Lacosta, S. F. Langa and C. Charbonnier, "Content format and quality of experience in virtual reality," *ArXiv*, 2020.

[132] S. Winkler and R. Subramanian. "Overview of eye tracking datasets," In *2013 Fifth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2013, pp. 212-217.

[133] L. Itti, C. Koch and E. Niebur, "A model of saliency-based visual attention for rapid scene analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 20, no. 11, pp. 1254-1259, 1998.

[134] J. Harel, C. Koch and P. Perona, "Graph-based visual saliency," In *Proceedings of the 19th International Conference on Neural Information Processing Systems*, Canada, 2006, pp. 545-552.

[135] Q. Jiang, F. Shao, G. Jiang, M. Yu, Z. Peng and C. Yu, "A depth perception and visual comfort guided computational model for stereoscopic 3D visual saliency," *Signal Processing: Image Communication*, vol. 38, pp. 57-69, 2015.

[136] T. Liu, J. Sun, N.-N. Zheng, X. Tang and H.-Y. Shum. "Learning to detect a salient object," In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, 2007, pp. 1-8.

[137] S. Wang, K. S. Ong, P. Surman, J. Yuan, Y. Zheng and X. Sun, "Quality of experience measurement for light field 3D displays on multilayer LCDs," *Journal of The Society for Information Display*, vol. 24, pp. 726-740, 2016.

[138] S. A. Fezza and M.-C. Larabi. "Stereoscopic 3D image quality assessment based on cyclopean view and depth map," In *2014 IEEE Fourth International Conference on Consumer Electronics Berlin (ICCE-Berlin)*, 2014, pp. 335-339.

[139] C. He, Z. Chen and C. Liu, "Salient object detection via images frequency domain analyzing," *Signal, Image and Video Processing*, vol. 10, no. 7, pp. 1295-1302, 2016.

[140] I. Iatsun, M.-C. Larabi and C. Fernandez-Maloigne, "Investigation and modeling of visual fatigue caused by S3D content using eye-tracking," *Displays*, vol. 39, pp. 11-25, 2015.

[141] P. Hanhart and T. Ebrahimi. "EYEC3D: 3D video eye tracking dataset," In *2014 Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014.

[142] Y. Fang, J. Wang, M. Narwaria, P. L. Le Callet and W. Lin, "Saliency Detection for Stereoscopic Images," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2625-2636, 2014.

[143] A. Banitalebi-Dehkordi, E. Nasiopoulos, M. T. Pourazad and P. Nasiopoulos, "Benchmark three-dimensional eye-tracking dataset for visual saliency prediction on stereoscopic three-dimensional video," *Journal of Electronic Imaging*, vol. 25, no. 1, 2016.

[144] Y.-S. Ho and K.-J. Oh. "Overview of multi-view video coding," In *2007 14th International Workshop on Systems, Signals and Image Processing and 6th EURASIP Conference focused on Speech and Image Processing, Multimedia Communications and Services*, 2007, pp. 5-12.

[145] C. Fehn. "Depth-image-based rendering (DIBR), compression and transmission for a new approach on 3D-TV," In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2004.

[146] Y. Park, K. Jung, Y. Oh, S. Lee, J. Kim, G. Lee, H. Lee, K. Yun, N. Hur and J. Kim, "Depth-image-based rendering for 3DTV service over T-DMB," *Signal Processing: Image Communication*, vol. 24, pp. 122-136, 2009.

[147] Y.-C. Fan, P.-C. Chi and Z.-S. Li. "Implementation of multi-view generation system based on DIBR," In *2012 International Symposium on Computer, Consumer and Control (IS3C)*, 2012, pp. 325-328.

[148] Y. Mori, N. Fukushima, T. Fujii and M. Tanimoto. "View generation with 3D warping using depth information for FTV," In *2008 3DTV Conference: The True Vision - Capture, Transmission and Display of 3D Video*, 2008, pp. 229-232.

[149] K. Mueller, H. Schwarz, D. Marpe, C. Bartnik, S. Bosse, H. Brust, T. Hinz, H. Lakshman, P. Merkle, H. Rhee, G. Tech, M. Winken and T. Wiegand, "3D high-efficiency video coding for multi-view video and depth data," *IEEE Transactions on Image Processing*, vol. 22, 2013.

[150] C.-S. Chen, H.-M. Cheng, H.-J. Chen, S.-Y. Tsai, C.-H. Kao, H.-J. Lin, L. Wan and T.-

Y. Yang, "Dry eye syndrome and the subsequent risk of chronic fatigue syndrome-a prospective population-based study in Taiwan," *Oncotarget*, vol. 9, no. 55, pp. 30694-30703, 2018.

[151] J. J. J. LaViola, E. Kruijff, R. P. Mcmahan, D. A. Bowman and I. Poupyrev, "*3D User Interfaces: Theory and Practice*". Addison-Wesley, 2017.

[152] B.-M. Choi, "Characteristics of electroencephalogram signatures in sedated patients induced by various anesthetic agents," *Journal of Dental Anesthesia and Pain Medicine*, vol. 17, no. 4, pp. 241-251, 2017.

[153] M. Remmele, E. Schmidt, M. Lingenfelder and A. Martens, "The impact of stereoscopic imagery and motion on anatomical structure recognition and visual attention performance," *Anatomical Sciences Education*, vol. 11, no. 1, pp. 15-24, 2018.

[154] R. Yang, W. Wang, Q. Lai, H. Fu, J. Shen and H. Ling, "Salient object detection in the deep learning era: An in-depth survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2019.

[155] J. S. Young and C. S. Cashwell, "*Clinical Mental Health Counseling: Elements of Effective Practice*". SAGE Publications, 2016.

[156] L. Trejo, R. Kochavi, K. Kubitz, L. Montgomery, R. Rosipal and B. Matthews. "EEG-based estimation of cognitive fatigue," In *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2005, pp. 105-115.

[157] R. Yuvaraj, M. Murugappan, N. M. Ibrahim, M. I. Omar, K. Sundaraj, K. Mohamad, R. Palaniappan and M. Satiyan, "Emotion classification in Parkinson's disease by higher-order spectra and power spectrum features using EEG signals: A comparative study," *Journal of Integrative Neuroscience*, vol. 13, no. 01, pp. 89-120, 2014.

[158] X. Hu, F. Deng, K. Li, T. Zhang and H. Chen, "Bridging low-level features and high-level semantics via fMRI brain imaging for video classification," In *Proceedings of the 18th ACM international conference on Multimedia*, Firenze, Italy, 2010, pp. 451-460.

[159] V. Vinhas, L. P. Reis and E. Oliveira, "Emotion-based multimedia retrieval and delivery through online user biosignals multichannel online biosignals towards adaptive GUI and content delivery," In *1st International Conference on Agents and Artificial Intelligence*, Porto, Portugal, 2008, pp. 232-237.

[160] R. M. Mehmood, R. Du and H. J. Lee, "Optimal feature selection and deep learning ensembles method for emotion recognition from human brain EEG sensors," *IEEE Access*, vol. 5, pp. 14797-14806, 2017.

[161] Y. Wang, G. Zhai, S. Zhou, S. Chen, X. Min, Z. Gao and M. Hu, "Eye fatigue assessment using unobtrusive eye tracker," *IEEE Access*, vol. 6, pp. 55948-55962, 2018.

[162] V. Ronca, A. Giorgi, D. Rossi, A. Di Florio, G. Di Flumeri, P. Aricò, N. Sciaraffa, A. Vozzi, L. Tamborra, I. Simonetti and G. Borghini, "A video-based technique for heart rate and eye blinks rate estimation: A potential solution for telemonitoring and remote healthcare," *Sensors*, vol. 21, no. 1607, 2021.

[163] Y. Yamada and M. Kobayashi, "Detecting mental fatigue from eye-tracking data gathered while watching video: Evaluation in younger and older adults," *Artificial Intelligence in Medicine*, vol. 91, pp. 39-48, 2018.

[164] I. NeuroSky, "EEG of NeuroSky: Comparison of Neurosky vs Biopac systems," 2009. [Online]. Available: http://sleepshepherd.com/wpcontent/uploads/2016/02/neurosky-vs-medical-eeg.pdf. [Accessed: Mar. 4, 2016].

[165] P. A. Abhang, B. W. Gawali and S. C. Mehrotra, "*Introduction to EEG- and Speech-Based Emotion Recognition*". Elsevier Inc., 2016.

[166] K. Sharma, R. Daniel, S. Chandra and A. G. Ramakrishnan, "Post meditation changes in EEG spectrum of long-term Brahmakumaris Rajayoga meditators," In *59th Annual Meeting of the Society for Psychophysiological Research*, Washington DC, USA, 2019.

[167] C. Andreu-Sánchez, M. Á. Martín-Pascual, A. Gruart and J. M. Delgado-García, "Chaotic and Fast Audiovisuals Increase Attentional Scope but Decrease Conscious Processing," *Neuroscience*, vol. 394, pp. 83-97, 2018.

[168] M. Mouzé-Amady and F. Horwat, "Evaluation of Hjorth parameters in forearm surface EMG analysis during an occupational repetitive task," *Electroencephalography and Clinical Neurophysiology/Electromyography and Motor Control*, vol. 101, no. 2, pp. 181-183, 1996.

[169] A. Pérez, M. Carreiras and J. A. Duñabeitia, "Brain-to-brain entrainment: EEG interbrain synchronization while speaking and listening," *Scientific Reports*, vol. 7, no. 1, pp. 4190, 2017.

[170] Blender Foundation, "Big buck bunny," 2015. [Online]. Available: http://bbb3d.renderfarming.net/download.html. [Accessed: Jun. 13, 2018].

[171] L. Barclay and M. Chu, "The impact of visual fatigue for students with visual impairment: Issues and solutions," *Paths to literacy - A joint project between Perkins School for the Blind and Texas School for the Blind and Visually Impaired*, 2020. [Online]. Available: www.oberlo.com/blog/video-marketing-statistics. [Accessed: Jun.

15, 2021].

[172] Z. Zhang and L. Schomaker, "DTGAN: Dual attention generative adversarial networks for text-to-image generation," *ArXiv*, vol. abs/2011.02709, 2020.

[173] C. Kinsey, "Blind windows," *History of the Human Sciences*, vol. 31, no. 5, pp. 154-182, 2018.

[174] B. Kan, Y. Zhao and S. Wang, "Objective visual comfort evaluation method based on disparity information and motion for stereoscopic video," *Optics Express*, vol. 26, no. 9, pp. 11418-11437, 2018.

[175] K. Hirvonen, S. Puttonen, K. Gould, J. Korpela, V. F. Koefoed and K. Müller, "Improving the saccade peak velocity measurement for detecting fatigue," *Journal of Neuroscience Methods*, vol. 187, no. 2, pp. 199-206, 2010.

[176] Y. Wang, G. Zhai, S. Chen, X. Min, Z. Gao and X. Song, "Assessment of eye fatigue caused by head-mounted displays using eye-tracking," *Biomedical engineering online*, vol. 18, no. 111, pp. 1-19, 2019.

[177] K. Duval, H. Grover, L.-H. Han, Y. Mou, A. F. Pegoraro, J. Fredberg and Z. Chen, "Modeling physiological events in 2D vs. 3D cell culture," *Physiology*, vol. 32, no. 4, pp. 266-277, 2017.

[178] E. G. González, M. S. Mandelcorn, E. D. Mandelcorn and L. Tarita-Nistor, "Effect of visual feedback on the eye position stability of patients with AMD," *Vision*, vol. 3, no. 4, 2019.

[179] F. Yang, H. Gu, M. Li, J. Cai, Q. Zhou and W.-K. Lam, "The impact on human visual performance when viewing 2-D and 3-D movies," *Technology and Health Care*, vol. 26, pp. 79-86, 2018.

[180] N.-T. Tran, M.-Q. Nghiem, N. T. H. Nguyen, N. Nguyen, N. V. Chi and D. Dien, "ViMs: A high-quality Vietnamese dataset for abstractive multi-document summarization," *Language Resources and Evaluation*, pp. 1-28, 2020.

[181] L. Salvati, M. d'Amore, A. Fiorentino, A. Pellegrino, P. Sena and F. Villecco, "On-road detection of driver fatigue and drowsiness during medium-distance journeys," *Entropy*, vol. 23, no. 2, 2021.

[182] B. Pursel, "*Information, People, and Technology*". Pennsylvania, U.S.: Wikipedia, 2021.

[183] E. Habibi, H. Rajabi and M. Arbabi, "An examination of the relationship between visual fatigue symptoms with flicker value variations in video display terminal users," *Iranian*

journal of health, safety and environment, vol. 2, pp. 211-216, 2015.

[184] A. L. Sheppard and J. S. Wolffsohn, "Digital eye strain: prevalence, measurement and amelioration," *BMJ Open Ophthalmology*, vol. 3, no. 1, 2018.

[185] H. D. Wiyor and C. A. Ntuen, "Empirical evaluation of visual fatigue from display alignment errors using cerebral hemodynamic responses," *Journal of Medical Engineering*, vol. 2013, no. 521579, 2013.

[186] H. J. Lee and S.-J. Kim, "Factors associated with visual fatigue from curved monitor use: A prospective study of healthy subjects," *PloS one*, vol. 11, no. 10, 2016.

[187] C. Liu, Z. Li and X. Du, "The effect of musical stimulation in sports on sports fatigue of college students," *Journal of Internet Technology*, vol. 22, no. 1, pp. 187-195, 2021.

[188] H. T. T. Tran, N. P. Ngoc, C. T. Pham, Y. J. Jung and T. C. Thang. "A subjective study on QoE of 360 video for VR communication," In *2017 IEEE 19th International Workshop on Multimedia Signal Processing (MMSP)*, 2017, pp. 1-6.

[189] Y. Nehmé, J.-P. Farrugia, F. Dupont, P. Le Callet and G. Lavoué, "Comparison of subjective methods for quality assessment of 3D graphics in virtual reality," *ACM Transactions on Applied Perception*, vol. 18, no. 1, pp. 1-23, 2020.

[190] A. Singla, S. Fremerey, W. Robitza and A. Raake. "Measuring and comparing QoE and simulator sickness of omnidirectional videos in different head mounted displays," In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*, 2017, pp. 1-6.

[191] B. Simons, "UTS Data Arena," 2019. [Online]. Available: https://dataarena.net/. [Accessed: May. 8, 2019].

[192] E. Upenik, M. Řeřábek and T. Ebrahimi. "Testbed for subjective evaluation of omnidirectional visual content," In *2016 Picture Coding Symposium (PCS)*, 2016, pp. 1-5.

[193] W. Sun, X. Min, G. Zhai, K. Gu, H. Duan and S. Ma, "MC360IQA: A multi-channel cnn for blind 360-degree image quality assessment," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 1, pp. 64-77, 2020.

[194] D. Narciso, M. Bessa, M. Melo, A. Coelho and J. Vasconcelos-Raposo, "Immersive 360° video user experience: impact of diferent variables in the sense of presence and cybersickness," *Universal Access in the Information Society*, journal article vol. 18, pp. 77-87, 2017.

[195] D. Narciso, M. Bessa, M. Melo and J. Vasconcelos-Raposo. "Virtual reality for training - The impact of smell on presence, cybersickness, fatigue, stress and knowledge transfer," In *2019 International Conference on Graphics and Interaction (ICGI)*, 2019, pp. 115-121.

[196] M. S. Anwar, J. Wang, S. Ahmad, A. Ullah, W. Khan and Z. Fei, "Evaluating the factors affecting QoE of 360-degree videos and cybersickness levels predictions in virtual reality," *Electronics*, vol. 9, no. 1530, pp. 1-20, 2020.

[197] H. Duan, G. Zhai, X. Min, Y. Zhu, Y. Fang and X. Yang. "Perceptual quality assessment of omnidirectional images," In *2018 IEEE International Symposium on Circuits and Systems (ISCAS)*, 2018, pp. 1-5.

[198] M. S. Anwar, J. Wang, W. Khan, A. Ullah, S. Ahmad and Z. Fei, "Subjective QoE of 360-degree virtual reality videos and machine learning predictions," *IEEE Access*, vol. 8, pp. 148084-148099, 2020.

[199] S. Johnson, I. Rae, B. Mutlu and L. Takayama. "Can you see me now? How field of view affects collaboration in robotic telepresence," In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, Seoul, Republic of Korea, 2015, pp. 2397-2406.

[200] Z. Li, S. Miyafuji, T. Sato, H. Koike, N. Yamashita and H. Kuzuoka, "How display shapes affect 360-degree panoramic video communication," In *Proceedings of the Designing Interactive Systems Conference*, Hong Kong, China, 2018, pp. 845-856.

[201] B. Zhang, J. Zhao, S. Yang, Y. Zhang, J. Wang and Z. Fei. "Subjective and objective quality assessment of panoramic videos in virtual reality environments," In *2017 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, 2017, pp. 163-168.

[202] B. Appina, S. V. R. Dendi, K. Manasa, S. S. Channappayya and A. C. Bovik, "Study of subjective quality and objective blind quality prediction of stereoscopic videos," *IEEE Transactions on Image Processing*, vol. 28, no. 10, pp. 5027-5040, 2019.

[203] L. Tychsen and P. Foeller, "Effects of immersive virtual reality headset viewing on young children: Visuomotor function, postural stability, and motion sickness," *American Journal of Ophthalmology*, vol. 209, pp. 151-159, 2020.

[204] N. Li, Y. K. Wong, W. L. Chan and K. M. Tsang. "Analysis of EEG signals during acupuncture using spectral analysis techniques," In *2010 IEEE International Conference on Automation and Logistics*, 2010, pp. 391-395.

[205] A. Salami, F. Ghassemi and M. H. Moradi. "A criterion to evaluate feature vectors based on ANOVA statistical analysis," In *2017 24th National and 2nd International Iranian Conference on Biomedical Engineering (ICBME)*, 2017.

[206] A. S. Malik, R. N. H. R. Khairuddin, H. U. Amin, M. L. Smith, N. Kamel, J. M. Abdullah, S. M. Fawzy and S. Shim, "EEG based evaluation of stereoscopic 3D displays for viewer discomfort," *BioMedical Engineering OnLine*, vol. 14, pp. 1-21, 2015.

[207] A. Banitalebi-Dehkordi, M. T. Pourazad and P. Nasiopoulos, "The effect of frame rate on 3D video quality and bitrate," *3D Research*, vol. 6, no. 1, pp. 1-13, 2015.

[208] T. Meitzler, D. Bednarz, K. Lane, E. Sohn, D. Bryk, E. Bankowski, R. Jozwiak and R. Andrews. "Comparison of 2D and 3D displays and sensor fusion for threat detection, surveillance and telepresence," In *AeroSense 2003*, 2003.

[209] J. Urakami, H. Matulis, S. Miyafuji, Z. Li, H. Koike and M. Chignell, "Comparing immersiveness and perceptibility of spherical and curved displays," *Applied Ergonomics*, vol. 90, no. 103271, 2021.

[210] G. Kyung and S. Park, "Curved versus flat monitors: Interactive effects of display curvature radius and display size on visual search performance and visual fatigue," *The Journal of the Human Factors and Ergonomics Society*, 2020.

[211] J. Bizzocchi, "Video as ambience: Reception and aesthetics of flat-screen video display," *Journal of Moving Image Studies*, 2004.

[212] Y. Zhang, Y. Wang, F. Liu, Z. Liu, Y. Li, D. Yang and Z. Chen, "Subjective panoramic video quality assessment database for coding applications," *IEEE Transactions on Broadcasting*, vol. 64, no. 2, pp. 461-473, 2018.

[213] S. N. M. Yacoub, "Quality evaluation for stitched panoramic videos." Ph. D. thesis, Computer Vision and Pattern Recognition, Université Grenoble Alpes, Saint-Martin-d'Hères, France, 2018.

[214] S. Leorin, L. Lucchese and R. G. Cutler. "Quality assessment of panorama video for videoconferencing applications," In *2005 IEEE 7th Workshop on Multimedia Signal Processing*, 2005, pp. 1-4.

[215] A. Philpot, M. Glancy, P. J. Passmore, A. Wood and B. Fields, "User experience of panoramic video in cave-like and head mounted display viewing conditions," In *Proceedings of the ACM International Conference on Interactive Experiences for TV and Online Video*, Hilversum, The Netherlands, 2017, pp. 65-75.

[216] Monash University, "Monash CAVE2," 2021. [Online]. Available: www.monash.edu/researchinfrastructure/mivp/access/facilities/cave2. [Accessed: Jun. 16, 2021].

[217] R. Folen, S. Miyahira and M. Stetz, "Immersive panoramic video display compared to flat screen display: psychological and physiological reactions to anger stimuli," *Frontiers in Neuroengineering*, vol. 2, 2009.

[218] S. Wu, Z. Chen, N. Liao and X. Chen. "Study on user experience of panoramic images on different immersive devices," In *19th Pacific-Rim Conference on Multimedia*, Hefei, China, 2018, pp. 519-527.

[219] A. Sabarudin and Y. J. Tiau, "Image quality assessment in panoramic dental radiography: A comparative study between conventional and digital systems," *Quantitative Imaging in Medicine Surgery*, vol. 3, no. 1, pp. 43-48, 2013.

[220] P. R. Desai, P. N. Desai, K. D. Ajmera and K. Mehta, "A review paper on oculus rift- A virtual reality headset," *International Journal of Engineering Trends and Technology (IJETT)*, vol. 13, 2014.

[221] J. A. Juanes, J. J. Gómez, P. D. Peguero, J. G. Lagándara and P. Ruisoto, "Analysis of the oculus rift device as a technological resource in medical training through clinical practice," In *Proceedings of the 3rd International Conference on Technological Ecosystems for Enhancing Multiculturality*, Porto, Portugal, 2015, pp. 19-23.

[222] T. J. Buker, D. A. Vincenzi and J. E. Deaton, "The effect of apparent latency on simulator sickness while using a see-through helmet-mounted display," *Human Factors: The Journal of the Human Factors and Ergonomics Society*, vol. 54, no. 2, pp. 235-249, 2012.

[223] A. B. Craig, W. R. Sherman and J. D. Will, "*Developing Virtual Reality Applications: Foundations of Effective Design*". Boston: Morgan Kaufmann, 2009.

[224] FOVE Inc., "FOVE," 2020. [Online]. Available: https://fove-inc.com/. [Accessed: Jan. 8, 2020].

[225] Google, "Google cardboard," 2021. [Online]. Available: https://arvr.google.com/intl/en_uk/cardboard/. [Accessed: Feb. 8, 2021].

[226] UL, "VRMark headset comparison March 2021," 2021. [Online]. Available: https://benchmarks.ul.com/compare/best-vr-headsets. [Accessed: Mar. 31, 2021].

[227] W. Greenwald, "Oculus Rift vs. HTC Vive: Which virtual reality headset is best?,"

PCMag, 2018. [Online]. Available: www.pcmag.com/comparisons/oculus-rift-vs-htc-vive-which-virtual-reality-headset-is-best. [Accessed: Jan. 3, 2021].

[228] T. Parisi, "*Learning Virtual Reality: Developing Immersive Experiences and Applications for Desktop, Web, And Mobile*". U.S.: O'Reilly Media, Inc., 2015.

[229] D. Lee, K. Baek, J. Lee and H. Lim, "A development of virtual reality game utilizing Kinect, Oculus Rift and smartphone," *International Journal of Applied Engineering Research*, vol. 11, pp. 829-833, 2016.

[230] J. Bailenson, "*Experience on Demand: What Virtual Reality is, how it works, and what it can do*". New York: W. W. Norton & Company, 2018.

[231] C. T. Tan, T. W. Leong, S. Shen, C. Dubravs and C. Si, "Exploring gameplay experiences on the Oculus Rift," In *Proceedings of the 2015 Annual Symposium on Computer-Human Interaction in Play*, London, United Kingdom, 2015, pp. 253-263.

[232] D. Egan, S. Brennan, J. Barrett, Y. Qiao, C. Timmerer and N. Murray, "An evaluation of heart rate and electrodermal activity as an objective QoE evaluation method for immersive virtual reality environments," In *2016 Eighth International Conference on Quality of Multimedia Experience (QoMEX)*, 2016, pp. 1-6.

[233] D. C. Niehorster, L. Li and M. Lappe, "The accuracy and precision of position and orientation tracking in the HTC Vive virtual reality system for scientific research," *i-Perception*, vol. 8, no. 3, 2017.

[234] A. Borrego, J. Latorre, M. Alcañiz and R. Llorens, "Comparison of Oculus Rift and HTC Vive: Feasibility for virtual reality-based exploration, navigation, exergaming, and rehabilitation," *Games for Health Journal*, vol. 7, no. 3, pp. 151-156, 2018.

[235] J. Radianti, T. A. Majchrzak, J. Fromm and I. Wohlgenannt, "A systematic review of immersive virtual reality applications for higher education: Design elements, lessons learned, and research agenda," *Computers & Education*, vol. 147, no. 103778, 2020.

[236] L. T. De Paolis and V. De Luca, "The impact of the input interface in a virtual environment: the Vive controller and the Myo armband," *Virtual Reality*, vol. 24, no. 3, pp. 483-502, 2020.

[237] The Lab, "Myo Armhand," 2021. [Online]. Available: https://developerblog.myo.com/get-hands-game-myo-armband-oculus-rift/. [Accessed: Jun. 15, 2021].

[238] Stat Trek, "Random number generator," 2011. [Online]. Available:

https://stattrek.com/statistics/random-number-generator.aspx. [Accessed: Jul. 6, 2020].

[239] SideFX, "Houdini," 2021. [Online]. Available: www.sidefx.com/products/houdini/. [Accessed: Jun. 16, 2021].

[240] Equalizer Graphics, "Equalizer," 2012. [Online]. Available: https://www.equalizergraphics.com/. [Accessed: Jun. 16, 2021].

[241] F. Korbel, "FFmpeg," 2019. [Online]. Available: www.ffmpeg.org. [Accessed: Mar. 5, 2019].

[242] R. Schleicher, N. Galley, S. Briest and L. Galley, "Blinks and saccades as indicators of fatigue in sleepiness warnings: looking tired?," *Ergonomics*, vol. 51, no. 7, pp. 982-1010, 2008.

[243] B. Ülker, M. B. Tabakcıoğlu, H. Çizmeci and D. Ayberkin, "Relations of attention and meditation level with learning in engineering education," In *2017 9th International Conference on Electronics, Computers and Artificial Intelligence (ECAI)*, 2017, pp. 1-4.

[244] M. V. Navarro-Haro, Y. López-del-Hoyo, D. Campos, M. M. Linehan, H. G. Hoffman, A. García-Palacios, M. Modrego-Alarcón, L. Borao and J. García-Campayo, "Meditation experts try virtual reality mindfulness: A pilot study evaluation of the feasibility and acceptability of virtual reality to facilitate mindfulness practice in people attending a mindfulness conference," *PLOS ONE*, vol. 12, no. 11, 2017.

[245] I. Kosunen, M. Salminen, S. Järvelä, A. Ruonala, N. Ravaja and G. Jacucci, "Relaworld: Neuroadaptive and immersive virtual reality meditation system," In *Proceedings of the 21st International Conference on Intelligent User Interfaces*, Sonoma, California, U.S., 2016, pp. 208-217.

[246] D. M. Shafer, C. P. Carbonara and M. F. Korpi, "Factors affecting enjoyment of virtual reality games: A comparison involving consumer-grade virtual reality technology," *Games for Health Journal*, vol. 8, no. 1, pp. 15-23, 2019.

[247] M. M. Nasralla, M. Razaak, I. U. Rehman and M. G. Martini, "Content-aware packet scheduling strategy for medical ultrasound videos over LTE wireless networks," *Computer Networks*, vol. 140, pp. 126-137, 2018.

[248] A. Asan, W. Robitza, I.-H. Mkwawa, L. Sun, E. Ifeachor and A. Raake. "Impact of video resolution changes on QoE for adaptive video streaming," In *2017 IEEE International Conference on Multimedia and Expo (ICME)*, 2017.

[249] B. Mallik, A. Sheikh-Akbari and A.-L. Kor, "Mixed-resolution HEVC based multiview video codec for low bitrate transmission," *Multimedia Tools and Applications*, vol. 78, no. 6, pp. 6701-6720, 2019.

[250] M. S. Banks, D. M. Hoffman and J. Kim, "3D displays," *Annual Review of Vision Science*, vol. 2, no. 1, pp. 397-435, 2016.

[251] B. Du, M. Yu, G. Jiang, Y. Zhang and T. Zhu, "A novel texture-based asymmetric visibility threshold model for stereoscopic video coding," *EURASIP Journal on Image and Video Processing*, vol. 1, no. 27, 2018.

[252] P. Aflaki, M. M. Hannuksela, M. Homayouni and M. Gabbouj. "Cross-asymmetric mixed-resolution 3D video compression," In *2012 3DTV-Conference*, 2012, pp. 1-4.

[253] W. Yao, L.-P. Chau and S. Rahardja, "Joint Rate Allocation of Stereoscopic 3D Videos in Next-Generation Broadcast Applications," *IEEE Transactions on Broadcasting*, vol. 59, no. 3, pp. 445-454, 2013.

[254] S. Sakata, P. M. Grove, A. Hill, M. O. Watson and A. R. Stevenson, "The viewpoint-specific failure of modern 3D displays in laparoscopic surgery," *Langenbeck's Archives of Surgery*, vol. 401, no. 7, pp. 1007-1018, 2016.

[255] A. Jain and J. Konrad. "Crosstalk in automultiscopic 3-D displays: Blessing in disguise?," In *Electronic Imaging 2007*, 2007.

[256] K.-J. Yun, S.-H. Cho, Y. Choi, J. Lee and C. Ahn, "3D stereoscopic/multiview video processing system and its method," U.S. Patent 8,111,758 B2, Feb. 7, 2012.

[257] W. Wang, C. Ren, X. He, H. Chen and L. Qing, "Video super-resolution via residual learning," *IEEE Access*, vol. 6, pp. 23767-23777, 2018.

[258] J. Samelak, J. Stankowski and M. Domański. "Efficient frame-compatible stereoscopic video coding using HEVC screen content coding," In *2017 International Conference on Systems, Signals and Image Processing (IWSSIP)*, 2017, pp. 1-5.

[259] H. Pashler, "*The Psychology of Attention*". Cambridge: MIT Press, 1999.

[260] C. Chamaret, S. Godeffroy, P. Lopez and O. Le Meur, "Adaptive 3D Rendering based on Region-of-interest," *Proceedings of Society of Photo-Optical Instrumentation Engineers (SPIE)*, 2010.

[261] Y. Zhang, G. Jiang, M. Yu and K. Chen. "Stereoscopic visual attention model for 3D video," In *International Conference on Multimedia Modeling*, 2010, pp. 314-324.

[262] A. Maki, P. Nordlund and J. Eklundh. "A computational model of depth-based attention," In *Proceedings of 13th International Conference on Pattern Recognition*, 1996, pp. 734-739.

[263] N. Ouerhani and H. Hugli. "Computing visual attention from scene depth," In *Proceedings 15th International Conference on Pattern Recognition (ICPR 2000)*, 2000, pp. 375-378.

[264] C. Lang, T. V. Nguyen, H. Katti, K. Yadati, M. Kankanhalli and S. Yan. "Depth Matters: Influence of Depth Cues on Visual Saliency," Berlin, Heidelberg, 2012, pp. 101-115.

[265] T. V. Nguyen and J. Sepulveda, "Salient object detection via augmented hypotheses," In *Proceedings of the 24th International Conference on Artificial Intelligence*, Buenos Aires, Argentina, 2015, pp. 2176–2182.

[266] T. Alshawi, Z. Long and G. AlRegib, "Understanding spatial correlation in eye-fixation maps for visual attention in videos," *arXiv*, 2019.

[267] T. Judd, F. Durand and A. Torralba, "A benchmark of computational models of saliency to predict human fixations," Massachusetts Institute of Technology, Cambridge, United State13 Jan. 2012 2012.

[268] M. Barreda-Á ngeles, F. Battisti, G. Boato, M. Carli, E. Dumic, M. Gelautz, C. Hewage, D. Kukolj, P. Le-Callet and A. Liotta, "Quality of experience and quality of service metrics for 3D content," in *3D Visual Content Creation, Coding and Delivery*: Springer, 2019, pp- 267-297.

[269] T. Judd, K. Ehinger, F. Durand and A. Torralba, "Learning to predict where humans look," In *IEEE 12th International Conference on Computer Vision*, Kyoto, Japan, 2009.

[270] SmugMug Inc., "Flickr," 2021. [Online]. Available: www.flickr.com. [Accessed: Feb. 6, 2021].

[271] B. C. Russell, A. Torralba, K. P. Murphy and W. T. Freeman, "LabelMe: A database and web-based tool for image annotation," *International Journal of Computer Vision*, vol. 77, no. 1-3, pp. 157-173, 2008.

[272] W. Zellinger, B. A. Moser, A. Chouikhi, F. Seitner, M. Nezveda and M. Gelautz, "Linear optimization approach for depth range adaption of stereoscopic videos," *Electronic Imaging*, vol. 2016, no. 5, pp. 1-6, 2016.

[273] Y. Li, X. Hou, C. Koch, J. M. Rehg and A. L. Yuille, "The secrets of salient object segmentation," In *IEEE Conference on Computer Vision and Pattern Recognition*, 2014.

[274] M. Everingham, L. Van Gool, C. K. I. Williams, J. Winn and A. Zisserman, "The PASCAL visual object classes (VOC) challenge," *International Journal of Computer Vision*, vol. 88, no. 2, pp. 303-338, 2010.

[275] A. Antsiferova and D. Vatolin. "The influence of 3D video artifacts on discomfort of 302 viewers," In *2017 International Conference on 3D Immersion (IC3D)*, 2017, pp. 1-8.

[276] Y. Fang, J. Wang, J. Li, R. Pépion and P. Le Callet. "An eye tracking database for stereoscopic video," In *Sixth International Workshop on Quality of Multimedia Experience (QoMEX)*, 2014, pp. 51-52.

[277] M. Urvoy, M. Barkowsky, R. Cousseau, Y. Koudota, V. Ricorde, P. Le Callet, J. Gutiérrez and N. García. "NAMA3DS1-COSPAD1: Subjective video quality assessment database on coding conditions introducing freely available high quality 3D stereoscopic sequences," In *2012 Fourth International Workshop on Quality of Multimedia Experience*, 2012, pp. 109-114.

[278] Z. Bylinskii, T. Judd, A. Oliva, A. Torralba and F. Durand, "What do different evaluation metrics tell us about saliency models?," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 41, no. 3, pp. 740-757, 2019.

[279] C.-Y. Ma and H.-M. Hang, "Learning-based saliency model with depth information," *Journal of Vision*, vol. 15, no. 6, pp. 19-19, 2015.

[280] O. Le Meur and T. Baccino, "Methods for comparing scanpaths and saliency maps: strengths and weaknesses," *Behavior Research Methods*, vol. 45, no. 1, pp. 251-266, 2013.

[281] L. Xing, J. You, T. Ebrahimi and A. Perkis. "Stereoscopic quality datasets under various test conditions," In *Quality of Multimedia Experience (QoMEX), 2013 Fifth International Workshop on*, 2013, pp. 136-141.

[282] MathWorks, "Matlab," 2020. [Online]. Available: www.mathworks.com/products/matlab.html. [Accessed: Feb. 8, 2020].

[283] A. Gibaldi, M. Vanegas, P. J. Bex and G. Maiello, "Evaluation of the Tobii EyeX Eye tracking controller and Matlab toolkit for research," *Behavior Research Methods*, Journal article vol. 49, no. 3, pp. 923-946, 2017.

[284] Tobii Technology, "Tobii eye-tracker," 2020. [Online]. Available: www.tobii.com. [Accessed: Apr. 6, 2020].

[285] C. Kaernbach, "Simple adaptive testing with the weighted up-down method Perception & Psychophysics," *Perception & Psychophysics*, vol. 49, no. 3, pp. 227-229, 1991.

[286] T.-H. Tsai, C.-S. Fan and C.-C. Huang. "Semi-automatic depth map extraction method for stereo video conversion," In *Genetic and Evolutionary Computing (ICGEC), 2012 Sixth International Conference*, 2012, pp. 340-343.

[287] O. Krishna, A. Helo, P. Rämä and K. Aizawa, "Gaze distribution analysis and saliency prediction across age groups," *PLOS ONE*, vol. 13, no. 2, 2018.

[288] N. Riche, M. Duvinage, M. Mancas, B. Gosselin and T. Dutoit. "Saliency and human fixations: State-of-the-art and study of comparison metrics," In *2013 IEEE International Conference on Computer Vision*, 2013, pp. 1153-1160.

[289] Y. Fang, C. Zhang, J. Li, J. Lei, M. P. Da Silva and P. Le Callet, "Visual Attention Modeling for Stereoscopic Video: A Benchmark and Computational Model," *IEEE Transactions on Image Processing*, vol. 26, no. 10, pp. 4684-4696, 2017.

[290] HTC Corporation, "HTC VIVE pro eye," 2021. [Online]. Available: https://enterprise.vive.com/us/product/vive-pro-eye-office/. [Accessed: Feb. 8, 2021].

[291] X. Qin, Z. Zhang, C. Huang, C. Gao, M. Dehghan and M. Jagersand. "BASNet: Boundary-aware salient object detection," In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019, pp. 7471-7481.

Appendices

Appendix 1: Matlab Toolbox for the Data Collection from Tobii EyeX Eye-tracker

Appendix 2: Questionnaire for QoE Experiment in different Viewing Environments

Appendix 3: Questionnaire for S3D Subjective Evaluation Methods applied to Video Compression Evaluation

Appendix 4: Ethics Approval Letter for QoE Experiment in different Viewing Environments

Appendix 1: Matlab Toolbox for the Data Collection from Tobii EyeX Eye-tracker

This Matlab script is modified from the publication of Gibaldi et al. [283].

Main program:

```
%%% MATLAB TOOLBOX for TOBII EYEX EYE-TRACKER %%%

%% CLEAR ALL
clear, close all, clc

%% ADDPATH
addpath ../../tobii_matlab

%% CHECK SOFTWARE
chk = chk_software('../../matlab_server');

%% COLLECT SOME INFO
prompt = {'Subject ID:'};
dlg_title = 'EyeX Vergence';
num_lines = 1;
def = {'SBJ1'};
answer = inputdlg(prompt,dlg_title,num_lines,def);
sName=char(answer(1,1));

%% CREATE DIRECTORIES
save_dir='DATA/';
subject=[sName '/'];

if isempty(dir(save_dir))
   mkdir(save_dir)
end

if isempty(dir([save_dir subject]))
   mkdir([save_dir subject])
```

```
end

if isempty(dir([save_dir subject '/traj']))
   mkdir([save_dir subject '/traj'])
end

%% TOBII SETUP
% START SERVER AND OPEN UDP PORT
%if the Matlab server is in a different folder with the original one, write
the FULL PATH
server_path=fullfile(pwd,'../../matlab_server/');
tobii  =  tobii_connect(server_path);
% INITIALIZE EYE TRACKER
[msg DATA] =  tobii_command(tobii,'init');
% START EYE TRACKER
[msg DATA tobii] =  tobii_command(tobii,'start',[save_dir subject
'traj/']);

%% PSYCHOTOOLBOX SETUP
PsychDefaultSetup(2);
Screen('Preference', 'SkipSyncTests', 1);
% Get the screen numbers
screens  =  Screen('Screens');
% Draw to the external screen if avaliable
screenNumber  =  max(screens);
% Define grey color
grey = (WhiteIndex(screenNumber)+BlackIndex(screenNumber))/2;

% Open an on screen window
[win, winRect]  =  PsychImaging('OpenWindow', screenNumber, grey);
%% CALIBRATION
MAX = .95;  MIN = 0.05; MD = .5;
TargetCalib = [MD, MAX, MIN, MIN, MAX, MD, MIN, MAX, MD;...
              MD, MAX, MAX, MIN, MIN, MIN, MD, MD, MAX];
% POSITION GUIDE
PositionGuide(tobii,win,winRect)
% BINOCULAR CALIB
[CalibL CalibR] = CalibrationProcedure(tobii,TargetCalib,[save_dir subject
```

```
'/DATA_CB'],win,winRect,'B')
% CHECK CALIBRATION
[Lpos Rpos] =
CalibrationCheck(tobii,win,winRect,TargetCalib,CalibL,CalibR);

pause(1)

%% EXPERIMENT PARAMETERS
time_per_scene = 10; %sec
%Read and show image
IM = double(imread('IMAGES/kitchen1.png'))/255;
Screen('PutImage', win, IM, winRect);
Screen('Flip', win);

% ACQUIRE DATA
Time=0;
count=1;
[L(:,count), R(:,count)] = tobii_getGPN(tobii);

tic
stop=false;
while Time(count)<time_per_scene && ~stop
    count = count +1;
    [L(:,count), R(:,count)] = tobii_getGPNcalib(tobii, CalibL, CalibR);
    Time(count) = toc;

    if KbCheck
        stop=true;
    end
end

% Clear the screen
sca;

% STOP EYE TRACKER
[msg, DATA] =  tobii_command(tobii,'stop');
% CLOSE SERVER AND UDP PORT
tobii_close(tobii)
```

```
%% SAVE DATA
save([save_dir subject
'DATA_FIXATION.mat'],'L','R','Time','Lpos','Rpos','time_per_scene','IM')
```

Appendix 2: Questionnaire for QoE Experiment in different Viewing Environments

Quality of multimedia experience in 3D video environments

Online Questionnaire Survey

Please put ☑ in each the following item:

Personal Profile

Gender:	Male ☐		Female ☐
Age:			
Vision:	Myopia (short sighted) ☐	Hyperopia (long sighted) ☐	Normal ☐
Psychic Condition:	Good ☐	Tired ☐	Tired-out ☐

1. Please rate your enjoyment of the 3D video experience on a scale of 1-5 (1: not enjoyable at all; 5: very enjoyable):

1 2 3 4 5

2. Briefly describe your video experience here:

3. Please indicate if you experience any of the following symptoms:

Symptom	None	Slight	Moderate	Severe
General discomfort	☐	☐	☐	☐
Fatigue	☐	☐	☐	☐
Headache	☐	☐	☐	☐
Eyestrain	☐	☐	☐	☐
Difficulty focusing	☐	☐	☐	☐
Increased salivation	☐	☐	☐	☐
Sweating	☐	☐	☐	☐
Nausea	☐	☐	☐	☐
Difficulty concentrating	☐	☐	☐	☐
Fullness of head	☐	☐	☐	☐
Blurred vision	☐	☐	☐	☐
Dizzy(eyes open)	☐	☐	☐	☐
Dizzy(eyes closed)	☐	☐	☐	☐
Vertigo	☐	☐	☐	☐
Stomach awareness	☐	☐	☐	☐
Burping	☐	☐	☐	☐

For Staff Only	
Participant no.:	Date of Test:
Device:	Video sequence:

Appendix 3: Questionnaire for S3D Subjective Evaluation Methods applied to Video Compression Evaluation

S3D Subjective Evaluation Methods applied to Video Compression Evaluation

Online Questionnaire Survey

Please put ☑ in each the following item:

Personal Profile

Gender:	Male ☐		Female ☐
Age:			
Vision:	Myopia (short sighted) ☐	Hyperopia (long sighted) ☐	Normal ☐
Psychic Condition:	Good ☐	Tired ☐	Tired-out ☐

1. Please rate your score of the quality of the video content on a scale of 1-5 (1: bad, 2: poor, 3: fair, 4: good, 5: excellent):

1 2 3 4 5

2. Briefly describe your video experience here:

For Staff Only	
Participant no.:	Date of Test:
Video sequence pair:	
Video 1:	Video 2:

Appendix 4: Ethics Approval Letter for QoE Experiment in different Viewing Environments

Reply all | Delete Junk | ...

HREC Approval Granted - ETH18-2690

Research.Ethics@uts.edu.au

Wed 01/09/2021 07:52

To: Research.Ethics@uts.edu.au; Eva.Cheng@uts.edu.au; Stuart.Perry@uts.edu.au; CHOY SIU MING

Reply all |

Inbox

Ethics Application.pdf

259 KB

Show all 1 attachments (259 KB) Download

Dear Applicant

Re: ETH18-2690 - "Quality of multimedia experience in 3D video environments"

Thank you for your response to the Committee's comments for your project. The Committee agreed that this application now meets the requirements of the National Statement on Ethical Conduct in Human Research (2007) and has been approved on that basis. You are therefore authorised to commence activities as outlined in your application.

You are reminded that this letter constitutes ethics approval only. This research project must also be undertaken in accordance with all UTS policies and guidelines including the Research Management Policy.

Your approval number is UTS HREC REF NO. ETH18-2690.

Approval will be for a period of five (5) years from the date of this correspondence subject to the submission of annual progress reports.

The following standard conditions apply to your approval:

- Your approval number must be included in all participant material and advertisements. Any advertisements on Staff Connect without an approval number will be removed.
- The Principal Investigator will immediately report anything that might warrant review of ethical approval of the project to the Ethics Secretariat.
- The Principal Investigator will notify the Committee of any event that requires a modification to the protocol or other project documents, and submit any required amendments prior to implementation. Instructions on how to submit an amendment application can be found here.
- The Principal Investigator will promptly report adverse events to the Ethics Secretariat. An adverse event is any event (anticipated or otherwise) that has a negative impact on participants, researchers or the reputation of the University. Adverse events can also include privacy breaches, loss of data and damage to property.
- The Principal Investigator will report to the UTS HREC or UTS MREC annually and notify the Committee when the project is completed at all sites. The Principal Investigator will notify the Committee of any plan to extend the duration of the project past the approval period listed above.

Reply all | Delete Junk | •••

- The Principal Investigator will notify the Committee of his or her inability to continue as Principal Investigator including the name of and contact information for a replacement.

This research must be
undertaken in compliance with the Australian Code for the Responsible Conduct of Research and National Statement on Ethical Conduct in Human Research.

You should consider this your official letter of approval. If you require a hardcopy please contact the Ethics Secretariat.

If you have any queries about your ethics approval, or require any amendments to your research in the future, please don't hesitate to contact the Ethics Secretariat and quote the ethics application number (e.g. ETH20-xxxx) in all correspondence.

Yours sincerely,
The Research Ethics Secretariat

On behalf of the UTS Human Research Ethics Committees
C/- Research Office
University of Technology Sydney
E: Research.Ethics@uts.edu.au

Ref: E38

UTS CRICOS Provider Code: 00099F DISCLAIMER: This email message and any accompanying attachments may contain confidential information. If you are not the intended recipient, do not read, use, disseminate, distribute or copy this message or attachments. If you have received this message in error, please notify the sender immediately and delete this message. Any views expressed in this message are those of the individual sender, except where the sender expressly, and with authority, states them to be the views of the University of Technology Sydney. Before opening any attachments, please check them for viruses and defects. Think. Green. Do. Please consider the environment before printing this email.