De Montfort University

Institute of Creative Technologies

Major Project

An investigation into how binaural sound may add presence in stereoscopic 3D-TV

Major Project

IOCT 5000

Dr Ian Sexton

December 2012

Tove Noorjahaan Dalenius

P10468717

Executive Summary

Limitations formed by the position and performance of the human eye restricts the perceivable depth that can be created through using stereographic display technologies. This report investigates whether adding binaural 3D sound can bring an experience of additional depth and presence to the viewer. The emphasis is on recording sound and image in a live/documentary setting. It was hypothesised that adding further auditory depth clues to stereoscopic video could enhance the viewer's perception of the space in which the subject was captured, and thereby enhance the viewer's perception of presence.

In order to find potential evidence in support of the hypothesis an initial experiment was carried out where participants viewed stereoscopic video recorded with binaural and stereo sound. Participants expressed their experiences of depth and presence in the stimuli they were shown and the results were analysed.

Preface

This study was carried out at the Institute of Creative Technologies, at De Montfort University. The research took place in 2012 and the report was completed at the end of this year. This report forms part of the Major Project module for the MSc in Creative Technologies.

The report is accompanied by a Blu-ray disc containing the stimuli used in the experiment.
These are also available online at http://youtu.be/J5UIJgyLZug

The report consists of eight parts. After the Introduction, which outlines the research, follows a section on the Context, which provides a background and rationale for the study. Hypothesis and Related Studies relates the focus of the research and provides grounding for the ideas put forward.
Analysis and Design follow the production of the stimuli whilst Experiment Design relates the setup and procedure for the investigation. In Results the outcomes from the experiments are described and a discussion ensues. The project as a whole is reviewed and concluded in Discussion and Conclusion.

De Montfort University, Leicester
December 2012

Tove Noorjahaan Dalenius

Acknowledgement

I would like to thank Dr Ian Sexton for his supervision during the project.
Dr Lorenzo Picinali for the use of the binaural dummy head and Christopher A. Feakes for support with binaural recording techniques.
Dr Howell Istance for assistance in treating the data.
Phil Surman for dialogue around stereoscopic imaging issues.
I would also like to thank Sally Hossack, Roger Wilson, Dave Everitt, Chris Conway and Mike Sole for giving their time and agreeing to the use of their musical performances as part of this research.

1 Table of Contents

List of Figures

List of Tables

List of Abbreviations

2DBi	2D video/binaural sound.
3D	Three-Dimensional
3D-TV	Three-Dimensional Television
3DS	Three-Dimension Stereoscopic
AVCHD	Advanced Video Coding High Definition
CES	Consumer Electronics Show
BFI	British Film Institute
GPA	Generalized Procrutes Analysis
HD	High Definition
HDTV	High Definition Television
IID	Interaural Intensity Difference
IOCT	Institute of Creative Technologies
ITD	Interaural Time Difference
ITU	International Telecommunication Union
LCD	Liquid crystal display
LCD glasses	Active liquid crystal shutter glasses
LED	Light-emitting diode
LED-TV	LED backlit LCD television
MOV	File extension used by the QuickTime-wrapped files
MPEG	Moving Picture Expert Group (audio/video formats)
MTS	File extension - MPEG Transport Stream
S3DBi	Stereographic 3D video/binaural sound,
S3DSt	Stereographic 3D video/binaural sound and
UHDTV	Ultra High Definition Television
VBAP	Vector base amplitude panning
WAVE	Waveform Audio File Format (also .WAV)
WFS	Wave field synthesis

2 Introduction

Limitations formed by the position and performance of the human eye restrict the perceivable depth that can be created through using stereographic video and TV. This report investigates whether adding binaural 3D sound can bring an experience of additional depth and presence to the viewer. The emphasis is on recording sound and image in a live/documentary setting.

3 Context

The history of technological innovations and market development is closely interwoven in the history of 3D film and television development. It emerges that 3D-TV, just as 3D cinema has developed in a fluctuating pattern, with peaks of interest and times of relative disregard. Consumer 3D television sets have been available in the market since 2010 but content availability has not increased substantially by 2012. Currently the film industry is keeping the 3D momentum, whilst the trajectory of 3D television industry is tentative. Commentators suggest that the possibility of auto-stereoscopic, glasses free, consumer displays may be the determining factor.[1] This study is focusing on the independent film/video maker with limited budgets, and the possibility of using binaural dummy head recording to increase audience perception of presence and immersivety. The study takes place presently as professional purpose designed stereoscopic video cameras are entering the market for the first time, simultaneously to the increase in consumer 3D displays. As well as investigating the technology and audience perception of one detailed aspect of stereoscopic filmmaking and use of binaural sound the study aims to build a background for further development of the emerging 3D imaging and sound creative language.

3.1 Development of 3D film and television, technology and industry

This chapter will summarise the development of the 3D technologies available to consumers and filmmakers today. It attempts to retain some separation between developments of 3D-TV and 3D cinema, as the specific medium of the research is 3D-TV.

3.1.1 Brief history of 3D-TV in the UK

British television pioneer John Logie Baird demonstrated both colour and stereoscopic television for the first time in 1928 using electro-mechanical and cathode-ray tube techniques. There were developments in 3D cinema and display techniques but the first broadcast of a full length stereoscopic film on British television would not take place until 1982 when ITV franchise Television South broadcast Fort Ti. Anaglyph glasses had been given out with the previous weeks TV Times. This broadcast was seen as an experiment and only occasional 3D broadcasts followed. In 2010 many well-known electronics manufacturers unveiled purpose built 3D-TVs during the Consumer Electronics Show, CES, in Las Vegas. Sales were predicted to rise for the following years and 3D-TV was hoped to ensure growth in the electronics industry. The systems released employed HDTV technique with active glasses. By the end of 2011 the BBC ran the first test broadcasts aimed for viewing on purpose built 3D-TV sets on their HD (high definition) channel.[2] In 2012 the BBC recorded having 113,000 viewers of their 3D broadcast of the Olympics opening, which is 0.5% of the total viewer number for the event. Sky has currently more than 250,000 Sky 3D subscribers in the UK. Although the numbers of 3D-TV viewers are increasing they still form a small part of the overall broadcast audience. The development of technology used in the consumer displays today owes much to the developments in 3D cinema which will be described further on.

3.1.2 Stereopsis/ human 3D vision

To understand the technology used in stereographic imaging it is important to understand human 3D vision. The human perception of a visual three-dimensional space is created in the brain. The eyes of a human are positioned 50–75 mm apart and this interpupillary distance makes it possible to perceive a slight difference between the image captured by the right and the left eye. Through a process of stereopsis, the brain extracts depth information from this difference. Another term for this process is binocular vision. Figure 1 illustrates binocular disparity, the difference in the position of an object between the two eyes; the amount of difference between the images gives the brain visual clues to infer depth.[3] Stereography uses this to create a perception of depth by presenting to different images to each eye. The human form stereopsis from birth and it is fully developed by the age of two, with some variations. Some individuals do not sense stereopsis.

There are also a number of monocular depth cues that form part of human depth perception. These are part of human perception without stereopsis. Occlusion or interposition, relative size, texture gradient, shading, shadows, atmospheric perspective, linear perspective and height cues and cues based on motion such as motion parallax and kinetic depth perception are examples of monocular depth cues.[4]

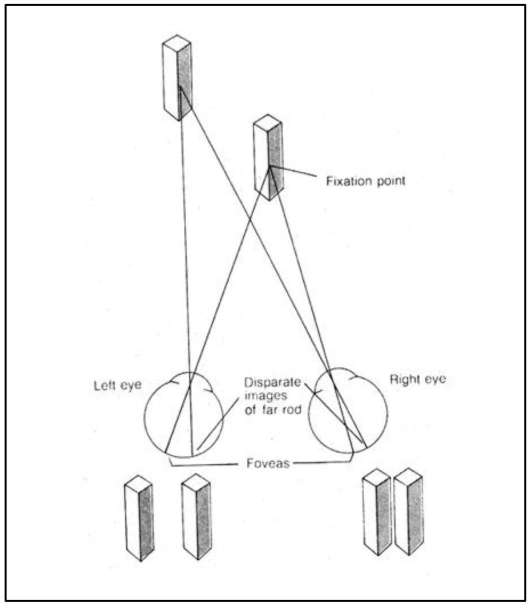

Figure 1 Binocular disparity[5]

Figure 2 Holmes stereoscope[6]

3.1.3 3D cinema - creating a perception of depth through stereographic imaging

Charles Wheatstone is credited with creating the first stereoscope in 1860. It used two images of the same subject taken with intraocular distance. The images were viewed through a stereoscope with lenses. Different types of stereoscopes were developed and they were popular in the Victorian era, creating a 3D industry making 3D cards. Figure 2 shows a Holmes stereoscope display. A well-known toy, which was originally intended for use by all ages, was developed by View-Master in 1939 and is known as its brand name. William Friese-Greene filed a patent for a 3D stereoscopic motion picture process in 1893. [7] Although trials took place 3D cinema did not become popular until the 1950s when the first American feature film to be made in colour and 3D, Bwana Devil, was screened, creating profit for the studio. The audience wearing polarised glasses admitting light polarised at different angles to the different eyes achieved the stereoscopic effect in Bwana Devil.[8] This is defined as a passive process and is used in developed forms in cinemas today. The linear polarised glasses method is illustrated in Figure 3. The 3D film industry grew during the fifties and another passive glasses based method, the anaglyph system was also used, although to a lesser extent. In anaglyph systems the viewer wear glasses with different colour lenses to discriminate between two images projected together, and the two parallax images are differentiated by colour.

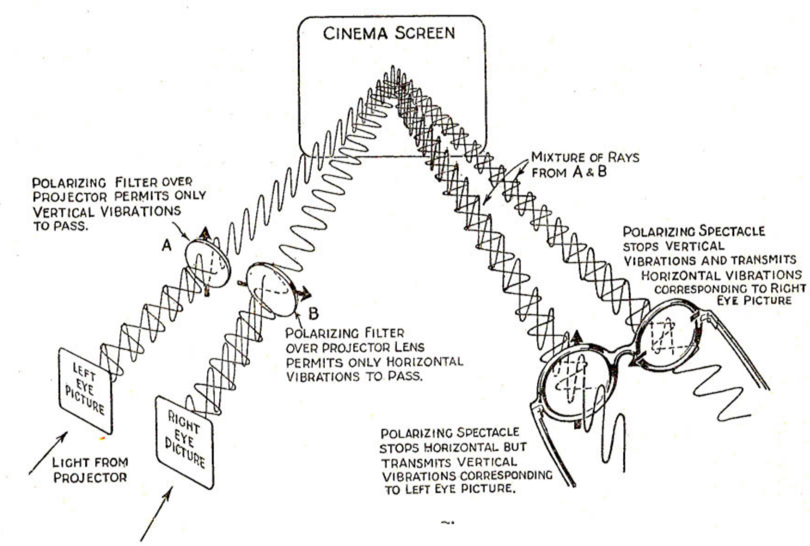

Figure 3 Linear polarised passive glasses system in a cinema setting[9]

Although 3D films drew significant audiences in the fifties the popularity started to wain and the costs of the projection technology led the industry in the direction of 2D production. Technical difficulties included the expense of keeping two matching celluloid film copies for simultaneous projection and of keeping the projectors accurately coordinated. There was a sharp fall in 3D production in the sixties, an increase in the production in the eighties and a slowdown again in the nineties. In the new millennium production has again reached the levels of the fifties. Avatar grossed more than $2.5 billion worldwide in 2009 and became the most profitable 3D motion picture of all time.[10]

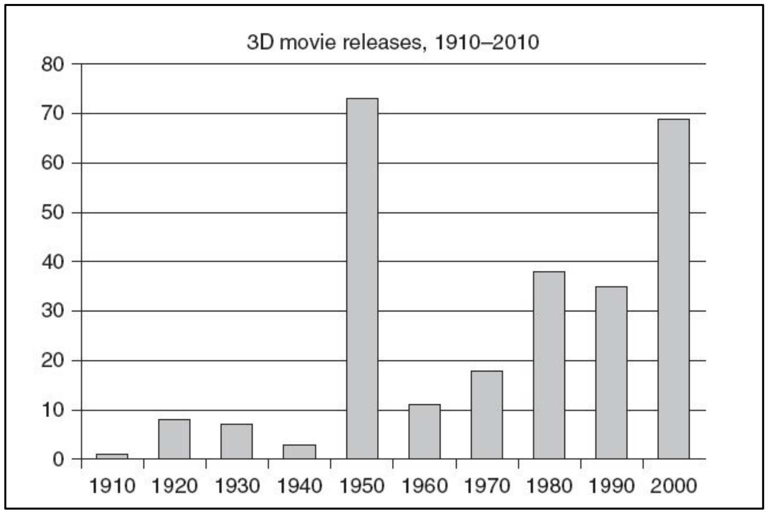

Table 1 Figure 4 3D movie releases 1910 to 2010[11]

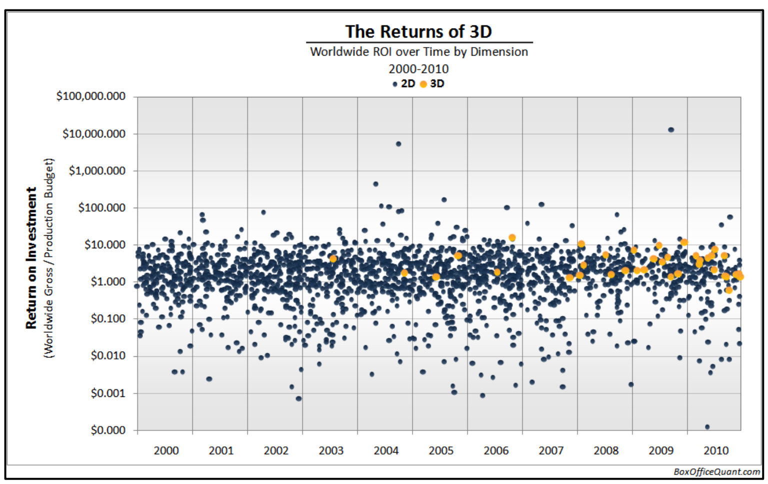

Table 2 Return on investment for 2D and 3D films; note that the scale is logarithmic. [12]

3.1.4 Contemporary development in the UK 3D filmmaking industry

Currently the UK 3D film and television industries are predicted to grow further. BFI (British Film Institute) has identified 3D competence as one of the key strategic skills prioritised.[13] Recognised UK industry training providers Skillset states that “Evidence indicates that there is a core audience and demand for this medium, and exhibitors (cinema operators and broadcasters) are catering for that demand. 3D is at a tipping point and during this time of flux it is vital that UK film makers grow and maintain the skills and knowledge they need to make informed decisions about 3D Stereoscopic (3DS) film making, their own projects and how to support international projects in the UK.”[14]

Film studios have a team of camera, lighting and soundrecordists; and the budget to build complex audio landscapes in postproduction. The independent film/video maker often goes out single handed instead of in a team, and the camera operator also does sound recording on the spot. Postproduction is kept to a minimum and the time needed to record and mix folio sound is inexpedient. In this environment the use of a single binaural recording dummy head could be used to attain 3D sound efficiently.

3.1.5 3D television technology available in the consumer market

The 3D-TV displays currently available consumer markets are of three types; Passive, Active or Autostreoscopic systems. Active shutter systems use synchronised active liquid crystal shutter glasses, which opens and close the view of each eye in time with changes of the displayed image. They have the advantage of often providing a clear image but the disadvantage of the need of more costly headsets. Passive glasses use polarised light technology where only the image matching the polarisation of the glasses is let through to each eye. These have the advantage of less costly headsets, but arguably provide lower picture quality. Autostreoscopic systems use technologies that are glasses-free and there are a few available in the consumer market. Autosterescopic systems build on parallax barrier or lenticular array technologies, and some utilise head or eye tracking of the viewers. The ITU (International Telecommunication Union) has during 2012 recommended a standard for the next generation UHDTV criteria, the 4K system corresponding to about 8 megapixels (3 840 x 2 160 image system), and the next tier 8K system corresponding to about 32 megapixels (7 680 x 4 320 image system).[15] UHDTV (Ultra High Definition Television) 4K high-resolution TVs with autostereoscopic 3D technologies are predicted to be released in 2013.[16]

3.2 Issues and limitation in stereoscopic imaging

Stereographic 3D-TV transmits two slightly horizontally separated images to the individual eyes and with the effect the viewer perceives three-dimensional space, both in front of and behind the TV screen. However this illusion is constricted to the actual area of the screen itself. The image can appear as being three dimensional in front of and behind the screen, but it cannot transcend the screen or reach sideway and out, in the way 3D-TV is often portrayed in adverts. Figure 1 illustrates an unrealistic portrayal of 3D-TV.

Figure 4 Unrealistic portrayal of stereoscopic 3D-TV[17]

The limit of the human vergence forms a restriction in that humans can only cross their eyes to a certain threshold comfortably. This forces the image producer to maintain a limit on the disparity between the images that lies within the viewer comfort zone when producing images that appear in front of the screen, with a negative parallax, forcing the eyes to converge. The same is true for trying to create images that appear to be behind the screen, using positive parallax. The perceivable depth is limited as the human eyes cannot comfortably diverge beyond a parallel position. Figure 5 illustrates human vergence. To avoid excessive positive or negative parallax it is recommended that the filmmaker keeps the parallax to a maximum 3% of the display size.[18]

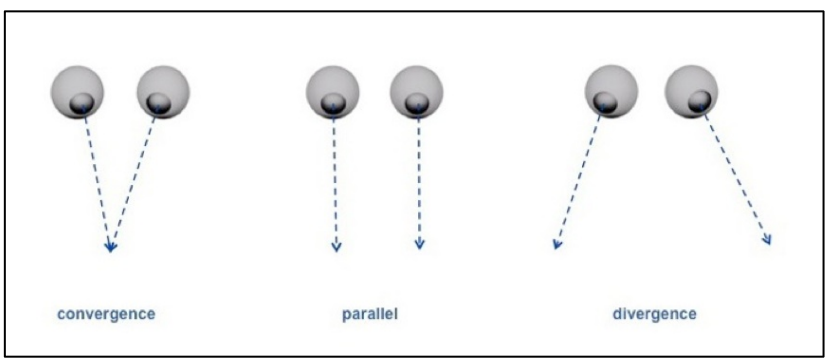

Figure 5 Convergence and divergence of the human eyes[19]

Stereographic video has a number of image composition issues that have different implications than in 2D filmmaking. As two separate images are relayed and joined in the brain it is important that the information in the images match up effortlessly. An example of what may cause discomfort is a frame violation, when the images projected in front of the screen are broken by the edge of the frame. The two images mismatching so that an object appear on the screen in one eye, but is not present in the other eye may also cause discomfort. Light reflecting in object can become visible to one view but not to the other. Crosstalk can be a problem if the contrast in the image is large, creating ghosting which may disrupt the 3D illusion.

Another problem is the conflict of accommodation and convergence. The eye focuses on an object according to the distance of the object. At the same time the two eyes adjust to the convergence point of the object, the angle at which the two eyes need to be to perceive the object clearly. To do this humans slightly cross their eyes when looking at something close up and have parallel positioned eyes when looking far away. In real life the focus and convergence fixes at the same point. When watching stereoscopic 3D-TV the human need to focus on the plain of the screen, whilst the object may appear in front of or behind the screen, forcing a mismatch to what the brain is used to perceive.

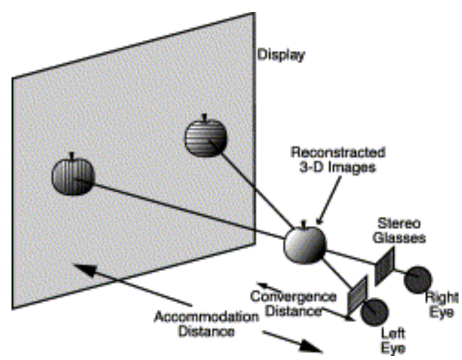

Figure 6 Accommodation and convergence mismatch [20]

These limitations of human vision and stereoscopic display technologies create a number of restrictions for the 3D filmmaker. Image composition must be restricted so that all objects are framed within an acceptable depth budget. This is easier to achieve in a studio setting or in animated image creation where objects can be moved to fit the criteria and create maximum possible depth. However in documentary setting the filmmaker often has a situation where some objects cannot practically be moved to suit image composition. The proposition here is that use of 3D sound could help create additional depth and space clues for the viewer and by doing this contribute an additional perception of presence in the scene/room captured.

3.3 Binaural sound – creating a perception of 3D space through interaural difference

Using two omni-directional microphones fitted inside modelled ears fitted on a dummy head sound reflected in the space and by the ear shape is recorded.[21] The dummy head recreates the human physique and the effect it has on sound before it reaches the microphones, which are placed to approximate the outer ear's tympanic membrane.[22] This closely represents what a human hears and allows the listener to accurately perceive where the source of the sound is situated and the shape and size of the room through intensity and delay clues. Humans can recognise room sizes accurately by listening to binaural recordings.[23] Binaurally recorded sound need to be played back through headphones to have the intended effect as the illusion would be ruined by introducing cross-ear couplings using open speakers.[24]

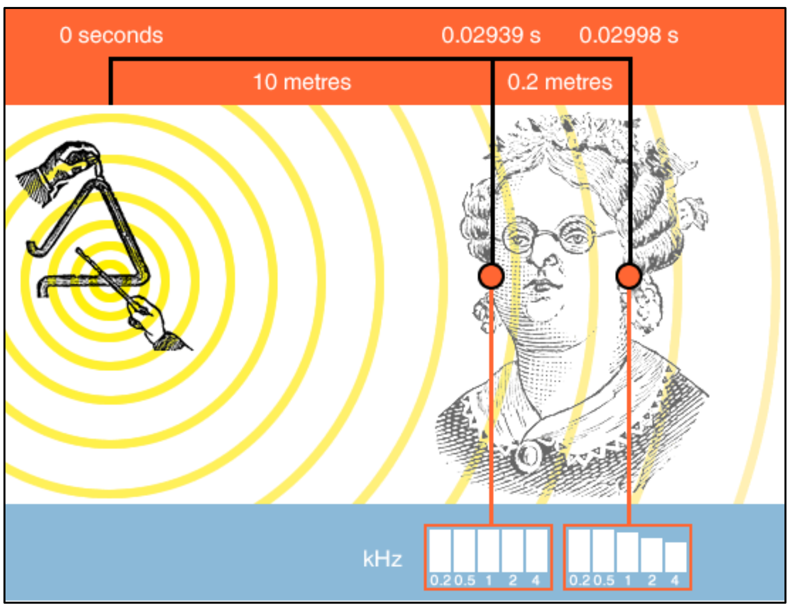

Figure 7 Interaural differences[25]

Three inter-channel differences may be in use when recording and reproducing binaural sound. Humans experience them in everyday life as interaural differences, the difference in sound reaching each ear. The differences are produced by the size and mass of the head, the torso and the shape and size of the individual's ears.[26] The first is the difference in sound intensity in each stereo channel, the second is the very small time delay between the sound from an off-centre source reaching each ear, the third encompasses differences in the spectrum of frequencies which reach each ear. These differences are illustrated in Figure 7.

Figure 8 Purpose built binaural recording microphone[27]

Figure 9 In-ear binaural microphones & earphones[28]

4 Hypothesis and Related Studies

This study makes use of stereographic video in combination binaural sound recording to investigate whether the use of binaural sound can add depth and presence to the viewer experience. It is an initial experimental and descriptive study where the viewer perception of the 3D space sensed when viewing a 3D-TV is contrasted between using stereo and binaural sound. The study is set against the practical background of an independent production situation and the workflow of for example a single cameraoperator /soundrecordist. It is hypothesised that adding the further auditory depth clues to the stereoscopic video could enhance the viewer's perception of the space in which the subject was captured, and thereby enhance the viewer's perception of presence.

4.1 Improving immersivety by combining stereographic video and binaural sound.

That perception of video and audio quality is influenced by the quality of each mode is well established. "When subjects are asked to judge audio quality of an audiovisual stimulus, the video quality will contribute significantly to the perceived audio quality" states Berrends and De Caluwe (1999) [29]
It is not a long step to the question whether 3D audio would influence the perception of the 3D image. Stereopsis is a multi-dimensional psychological construct that reflects several image attributes and has been described as by the CEO of RealD, Josh Greer as "more like a feeling than a perception".[30]
Current immersive media such as virtual reality technologies has been leading the development of interactive environments, however the particular situation of the filmmaker is slightly different as immersion rather than interactivity is the goal. During the research the question of what the stereoscopic filmmaker is trying to achieve arose. Is the aim to make the subject, here the artist, appear as being in the room where the audience is based, or to move the audience in to the room where the subject was

captured? The implications for sound reproduction are pressing, if the former is the case the acoustics would aim to recreate the performance as if in the room where the audience is based. If the latter is the case sound reproduction would aim to move the audience into the space of the performance. In this study the intention is to allow the audience to experience the performance as where it was performed.

4.2 Related Work

There are few studies relating directly to binaural sound and stereoscopic video being used in conjunction, and a study looking at depth perception using physical loudspeakers and stereoscopic video will be detailed under this heading. The conference proceedings 'Adding 3D Sound to 3D Cinema: Identification and Evaluation of Different reproduction Techniques' investigates the use of audio in a movie theatre setting.[31] Vector base amplitude panning (VBAP), Binaural techniques, Transaural techniques, Wave field synthesis (WFS) and Ambisonics are compared as to ascertain their effectiveness in a cinema setting. Difficulties in efficient reproduction of sound with a multitude of sweet spots prevent most of the described methods from being deployed successfully, only binaural techniques using headphones had the possibility of tackling all the challenges of 3D-stereo movies. Many of the particulars can be translated from a cinema to a TV audience and the findings substantiate the expediency of the use of the binaural technique.

Studies concerning methods of describing 3D imaging quality were examined, with a view to identify appropriate investigational procedures. Three studies were selected as being of particular interest, the first due to the direct correlation to this study, investigating the perception of depth in stereoscopic images and 3D sound. The second paper discusses the development of a 3D Quality Model, adding naturalness, presence and viewing experience are suggested as higher level evaluation concepts to the concept of image quality, which is described as more suitable to the evaluation of 2D images. The third study suggests Free-choice profiling and Generalized Procrutes Analysis methods to understand audiovisual quality. The paper 'Descriptive Quality of Experience for Mobile 3D video' also aims to identify general descriptive characteristics if the experience of 3D video on mobile devices and uses the Open Profiling of Quality and the Grounded Theory framework approaches to analyse participants contributions, but will not be described in further detail here.

When carrying out the experimental face of this study a quantitative methodology, using quality measures proposed by Lambooij and IJsselsteijn (2011)[32] was employed as the Free-choice profiling approaches investigated by Strohmeier and Jumisko-Pyykkö (2010) [33], although interesting and valuable appeared complex beyond the scope of the study at hand.

Methodology used in the three studies examined

Study: Can the perception of depth in stereoscopic images be influenced by 3D sound
Hypotheses → Method → Creation of stimuli → Quantitative Testing → Statistical Analyses→ Discussion → Conclusion

Study: Evaluation of Stereoscopic Images: Beyond 2D Quality
Hypotheses → Discussion of connotations of terminology → Method → Creation of stimuli → Establishing relation of terminology to predicted pattern through Quantitative Testing → Statistical Analyses → Further Quantitative testing to determine if terminology and measurement relate to predicted pattern on a wider scale → Statistical Analyses → Discussion → Conclusion

Profiling experienced quality factors of audiovisual 3D perception
Hypotheses → Method → Creation of stimuli → Quantitative Testing → Statistical Analyses (not proven effective) → Participants create their own terminology and scales → Statistical Analyses through Generalized Procrustes analysis (analysing participants' scale usage and comparing the data in shapes) → Discussion → Conclusion

4.2.1 Can the perception of depth in stereoscopic images be influenced by 3D sound

The study focuses on whether visual perception of depth can be influenced by audio information.[34] It proposes that the restricted scope of image depth budget, constricted by the limits of image disparity, forms a challenge for stereoscopic image creators. Auditory localisation cues include the intensity of sound in each ear and the time it takes for the sound to reach each ear. Inaural intensity difference (IID) refers to the sound intensity and inaural time difference (ITD) to the time difference. This study makes use of the term inaural rather than the more common interaural. Apart from these binaural sound depth cues humans also perceives distance by sound pressure, amount of reverberation and the spectral shape of the sound signal.

Method

Using openGl (an application programming interface for rendering 2D and 3D computer graphics) an image of a mobile telephone was projected on a 3D monitor using horizontal line interleaving and polarised glasses. Two hidden speakers were placed in front of the monitor at a distance of 0.25m. The disparity was adjusted so that the telephone appeared to be positioned over speaker A. The image was shown twice to each participant; once with a ring tone played on speaker A, and a second time on speaker B. The participants were asked to identify in which image display the image was nearest to them. This was repeated 20 times.

Conclusion

The participants did report that they saw the stimulus closer when the sound source was closer, than when the sound source was in line with the projected image. This study suggests that "depth in a stereoscopic image can be made to appear greater if a nearer 3D sound is played with the image."[35]
As stereoscopic image creators must work within the stringent parameters of the depth budget, this could be of tangible applied value. The study recommends further study into the methods of relay of 3D sound appropriate to commercial implementation.

4.2.2 Evaluation of Stereoscopic Images: Beyond 2D Quality

The study 'Evaluation of Stereoscopic Images: Beyond 2D Quality'[36] discusses the value of perceived image quality, which is a standard evaluation model for 2D imaging technology. It is found that it does not include the added value of stereoscopic depth appropriate to 3D imaging. The analyses in the study points to that the concepts 'viewing experience' and 'naturalness', contain variation in image quality to a similar extent, whilst 'naturalness' features the value of stereoscopic depth to a larger extent.
The psychovisual perception of stereoscopic images on audiences with regards to perceived image quality, sharpness and perceived depth was assessed by Tam et al. (1998). They reported a low correlation linking perceived image quality and perceived depth, and a high correlation between perceived sharpness and image quality. [37] Seuntiëns et al. (2006) relates that perceived image quality tracked variations in JPEG-compression, but remained unaltered by various quantities of stereoscopic depth.[38] This indicated that stereoscopic depth in the image is not described well by the term image quality. The study sites research describing how "non-expert observers described sensations of presence or of 'being there' while viewing stereoscopic image material, and related the concept of presence to involvement, realism and naturalness." Presence and depth, and presence and naturalness are described as being associated in additional research. The study proposes that when evaluating moving images the term presence is more relevant. The study proposes the use of the described 3D Quality Model as an extension of Engeldrum's Image Quality Circle in current use.

Method

The study compares evaluation by participants of stereographic still images where the image quality was altered by adding white Gaussian noise, Gaussian blur and CBD (camera based depth) using a [bad]-[poor]-[fair]-[good]-[excellent] scale Likert scale questionnaire. Previous to the experiment the participants undertook a training session to anchor the evaluation concepts. The match of the answers was then compared to the predicated outcome to quantify the suggested 3D Quality Model.

Conclusion

The study suggests the 3D Quality Model as an extension of Engeldrum's Image Quality Circle. 'Naturalness', 'presence' and 'viewing experience' are suggested as higher level evaluation concepts suitable for the 3D environment as these constructs are sensitive to image quality as well as stereoscopic depth. It proposes that the 3D Quality Model can contribute to a more effective design circle for 3D-TV.

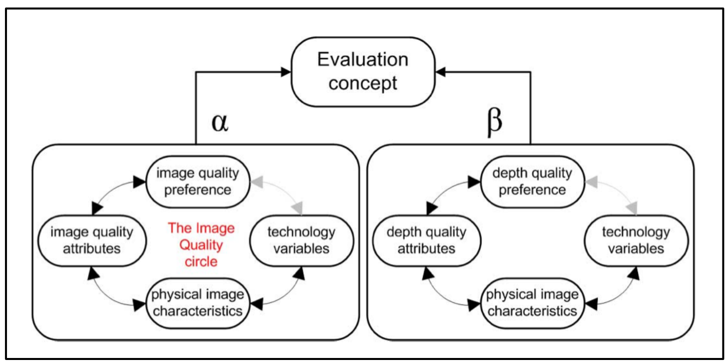

Figure 10 Three-dimensional Quality Model.
The higher value of perceived depth quality is added to the perceived image quality in Engeldrum's Image Quality Circle.[39]

4.2.3 Profiling experienced quality factors of audiovisual 3D perception

This paper explores experienced video and audio multimodal quality.[40] It focuses on a Free-choice Profiling approach to understand audiovisual quality. It also explores differences in perceptual styles of participants and the identification of those. The paper suggest the application of sensory profiling for the evaluation of audio and video quality, rather than the current quantitative semi-structured interview commonly used to gather qualitative data. The sensory profiling focus on direct elicitation of individual quality factors and allows participants to develop their own personal quality attributes.

Method

Two rooms were modelled on Maya software and the room acoustics were modelled using the Perceptual Approach. One large room modelled as a classroom and a small room modelled as a student living room. The acoustics for the large room was a male speaker and for the small room drum and bass music. The stimuli were presented on a 15" auto stereoscopic parallax barrier display and the sound played back on a four channel surround set up. Participants were asked to evaluate overall quality on an 11-point unlabelled scale to apply an Absolute Category rating quantitatively. Open Profiling of Quality was applied by asking participants to write down all their individual quality attributes. The participants were not limited in what they could describe as quality attributes as the researchers were interested in every aspect of a holistic understanding of the quality rational. The identified attributes were then linked a line with a maximum and a minimum rating at each end. Participant used these scales to evaluate stimuli which were shown in 2D and 3D variations. The qualitative data did not show any significant influence of the presentation modes on overall quality perception. The Free Choice profiling tasks were analysed by applying Generalized Procrutes Analysis (GPA). The GPA fits all participants' configurations to a common consensus. Figure 11 shows correlation plots for individual quality attributes in the GPA model. Participants individual quality attributes were analysed to establish a correlation to the model. The content dimension was described with attributes like 'tidy' or 'comic like', additionally affective attributes like 'monotone' or 'sympathy' were be observed. The video quality dimension was described in polarities such as the monoscopic and stereoscopic. 3D quality factors were described in terms such as 'three-dimensional' or 'spacious'. This demonstrates that test participants were able to perceive the depth not picked up with useful effect in the quantitative study. Some attributes indicated that 3D did not deliver added value for the participants. Stereoscopic videos were associated with attributes like 'blurry', 'smeary', and 'interlaced lines' and 'exhausting' whilst monoscopic videos were associated with 'sharp', 'colourful', and 'clear'. Artefacts are deemed to be the determining factor in 3D video quality. Although depth was perceived added value attributes such as presence or involvement were not presented by participants. Individuals presented tendencies towards different human information processing styles such as visual and verbal.

Conclusion

The qualitative aspect of the study did not show any difference in the experienced quality between the variables. The sensory profiling showed that the participants did experience differences as they could qualitatively differentiate them. Perceived depth was emphasised in both modalities contributing to the global audiovisual perception. Suggestions for further work include the development of a well-validated device for recognition of groups with different information processing styles for the audiovisual quality practitioner.

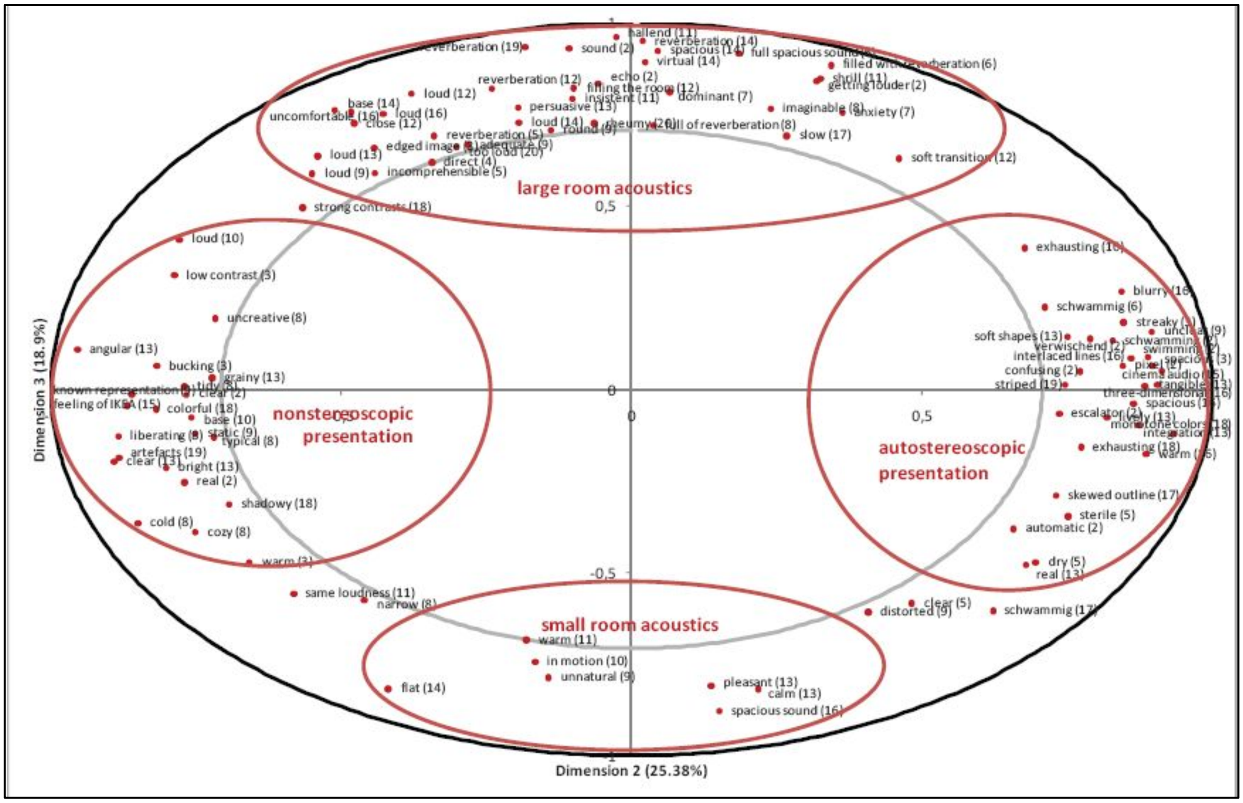

Figure 11 Correlation plot of inividual quality attributes and GPA model[41]

5 Analysis and Design

To test whether using binaural sound could make any difference in the viewers' experience, and whether the viewer would experience the space and presence differently, three similar video clips where recorded. The integral Panasonic AG-3DA1E stereographic video camera was used. Sound was recorded both with a binaural dummy head and with the on-board stereo microphones of the camera. The videos were edited so that each clip was available in three versions; version one had stereographic video with binaural sound, version two stereographic video with stereo sound, in version three the stereographic video was muxed from two left views so that a the video had no depth and binaural sound.

To allow viewer interest and realism the subjects chosen were three local acoustic musicians. Each was filmed doing a short song, about three and a half minute long. The musicians were filmed in their working environments, two in their homes and one in a music studio to create a genuine documentary live feel and to build the challenge of filming in a real life situation.

Initially filming in larger spaces with distinct audio-scapes, such as three places of worship, a church, a mosque and a temple was considered. However after consultation with senior researchers in the field it was established that smaller rooms were better for a clearer three-dimensional definition as room reverberation is more direct and sound localisation cues easier to isolate.

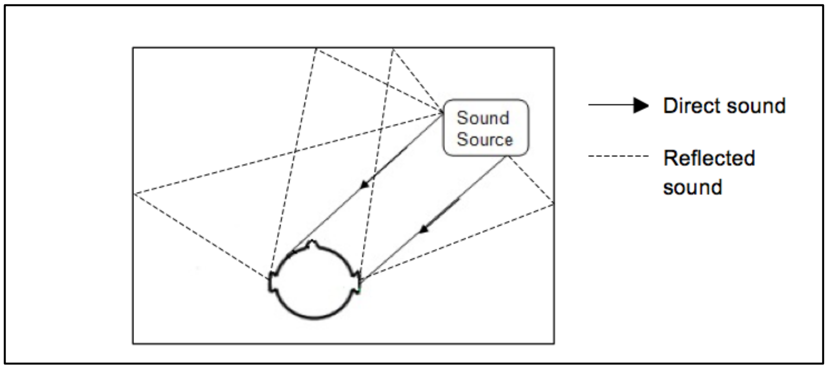

Figure 12 Direct and reflected sound propagating to a listener[42]

5.1 Production

Recording sound and video in the working environments of musicians live provided the challenges foreseen. Initial test recordings were made with a fourth musician and after editing and analysing some changes were made at a second shoot. Changes made included a stricter image composition and changes in lighting. The AG-3DA1E requires manual setting of focus, iris and parallax and the shooting is involved in comparison to a similar 2D video shoot. Due to the out of studio setting constant challenges included frame violations, keeping the composition within a conservative depth budget, lighting issues, and sound distractions such as humming fridges and traffic. These technical challenges had to be considered and counteracted by extemporisation, as it was important to work within the available time, and the interest attention span of the musicians, to capture a vivid performance. The artist performed their songs about three to four times to achieve this. Each scene was recorded in rooms of comparable size in a single take, with a one fixed camera angle. The artist walked in to the room where the performance was taking place. The sound of the artist walking in was intended to spacialise the recording. Each scene was filmed at a similar distance to the subject. The dummy head with the binaural microphones was placed near to and at the same height as the camera. Initially there was a slight concern that the audio from the binaural microphones (DPA 40-60 omnidirectional microphones) and the on-board camera stereo microphone was not of a comparable quality. However after playback to an experienced sound-engineer who preferred the stereo recording and a binaural sound-recordist who preferred the binaural recording it could be concluded that the preference could diverge between individuals with different experience, and that both stimuli modes were of adequate quality.

 

Figure 13 Binaural dummy head with Zoom H4n digital recorder[43]

Figure 14 Panasonic AG-3DA1E stereographic video camera[44]

5.2 Post Production

The recorded material, which consisted of AVCHD (Advanced Video Coding High Definition) video in .MTS (MPEG Transport Stream) files in two separate folders, was muxed using the Cineform and FirstLight software. Muxing is the process of combining two separate full-size AVCHD video stream into a side-by-side video format which is compatible with purpose built 3D televisions, projectors and and Blu-ray players. FirstLight produces muxed .MOV files. Cineform is compatible with the video formats produced by the Panasonic AG-3DA1E and the files do generally not need pre-muxing adjustments. The muxing of the two left eye views, to create a stimulus with no depth, produced a clip that caused some eye strain, but it was chosen as a better option than making the viewer remove their active glasses during the test sequence.

The muxed files where then edited in Premier CS5 using a 37" 3D LED TV screen with active glasses as an editing monitor. Monitoring editing with glasses based systems is inelegant as the editor will need to view both a computer screen, which frequency is blacked out by the active glasses, and the 3D display repeatedly. This makes the workflow somewhat more cumbersome than 2D editing. After becoming familiarised with the content editing is often done watching the muxed material in vertically stretched side-by-side format in the Premier monitor window as illustrated in Figure 15. Sound was edited in Premier CS5. Binaural WAVE files where imported separately whilst the audio from the camera microphone was included in the native .MOV files. Each scene with the three different musicians was filmed at a similar distance to the subject and no editing between different shots was made to ensure that the clips where comparable and that the viewer's perception was not influenced by the editing, rather than the sound and image. As the negative parallax was often excessive as the musician walked into the shot and past the camera sound was allowed to start earlier in the edits whilst the video was black with the video fading in after a few seconds. The sound volume was adjusted so that the viewer did not perceived one clip as being quieter or louder than another, and to make the experience comfortable, as the viewer would not be able to adjust the sound during the tests. A short 3D clip of a piano was inserted at the start of the test reels to ensure that the display converted side-by-side video into a 3D viewing mode correctly at each showing. Many 3D televisions adjust automatically when detecting side-by-side video content but video with side-by-side text does not appear to trigger this function reliably. On reviewing the development it was noticed that this clip could be perceived as having a problem with negative parallax (in front of screen) frame violation due to the relative close up of the clip, and could be substituted with a similar clip of all positive parallax (behind screen).

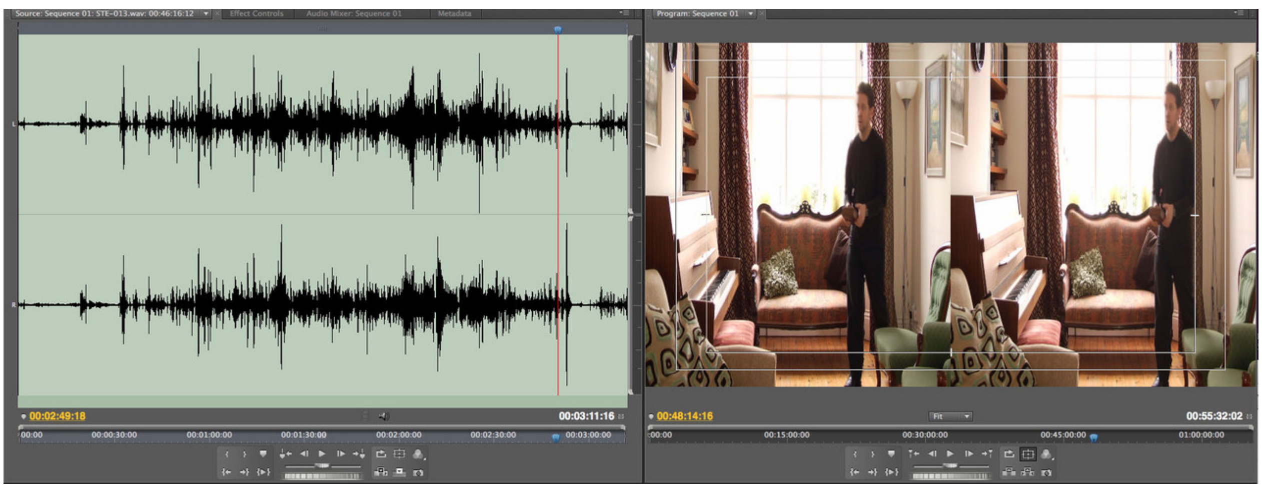

Figure 15 Editing binaural sound and stereoscopic video[45]

On screen text instructions with negative and positive parallax on different titles were inserted to ensure that the viewer had the glasses switched on and functioning and instructions included about wearing the headphones the right way around, as this is essential to the binaural sound experience. A fourth clip of a musician walking in a figure eight in front of and behind the camera playing a portable instrument was inserted at the start of all the versions to enable the viewer to acclimatise to the 3D set up. This clip used binaural sound. This clip can have problems with ghosting caused by image contrast in inferior viewing conditions; however several participants referred to it as the most engaging and remarkable of the clips. After editing the clips where imported to Adobe Encore for Blu-Ray encoding. Blu-ray disc was chosen rather than playback from a computer to ensure uniform replay. Three different discs where assembled, each holding one 3D/stereographic video/binaural sound, one 3D/stereographic video/stereo sound and one 2D/video/binaural sound clip. The order of the video and sound combinations was changed on each version as well as which artist performed on each clip, to ensure that the experiment was not biased by the order of the clips.

To simplify this report artists will be designated the names Artist X, Artist Y and Artist Z. The mode of the clips will be designated as S3DBi = stereographic 3D video/binaural sound, S3DSt = stereographic 3D video/binaural sound and 2DBi = 2D video/binaural sound.

Figure 16 Artist X   Figure 17 Artist Y   Figure 18 Artist Z

6 Experiment Design

Built on the evidence described in the preceding chapters an experiment was designed to test the hypotheses that additional depth and presences could be added to stereoscopic video by adding binaural sound. The methodology employed, and the values investigated builds on the strategies described in the works reviewed under the heading Related works.

Planned methodology for this study
Hypotheses → Method → Creation of stimuli → Quantitative and Qualitative Testing → Statistical Analyses → Qualitative analyses → Discussion → Conclusion

6.1 Construction of questionnaire

All participants were asked to complete one questionnaire. The first page of the questionnaire acquired information about gender and age. Age and gender data was gathered to make the research analogous to other research in the field, but names and identities of participants where anonymised. The following three pages contained six questions about the participants' perception of quality and presence in the stimuli. This included questions about perceived 'naturalness', 'image quality', 'depth' and 'eye strain' as well as about 'perceived presence in the room' and 'presence with the artist'. These values were derived from the studies described under the heading Related works, with the addition of the specification of 'presence in the room' and 'presence with the artist' to elucidate any differences between the two concepts. Each page with the same six questions was completed after watching the relevant clip. The answers were completed on a nine point Likert scale. The design of these questions again followed the outline of work by Lambooijd and IJsselsteijn (2011) and a [bad]-[poor]-[fair]-[good]-[excellent] scale was utilised corresponding to the ITU recommendations for methodology for the subjective assessment of quality of television pictures.[46] Figure 19 is an example of the Likert scales used in the questionnaire. Questions only concerned the visual and viewer experience and did deliberately not focus the participants' attention on the sound. As the research emphasis was on whether the sound influenced the perception of the image, not directly on the participants' view of sound recording method or quality, these where not brought in to direct focus. The full questionnaire can be seen in the appendix. On completing watching all three clips the participants were then asked to compare the overall the viewer experience and rank the three clips in the order. As well as quantitative questions there were two qualitative questions included where the participants were asked to detail why they rated the overall viewing experience the best or the least good for the associated clip. The qualitative questions were included to give more depth to the understanding of the viewer experience, to pick up on any unpredicted viewer reactions and to ensure that a clear and multifaceted understanding was developed about why certain clips were chosen.

Perceived Depth

Bad		Poor		Fair		Good		Excellent

Figure 19 An example of the 9-point Likert scale utilised in the questionnaire

6.2 Experiment setup

Testing took place at the IOCT lab where an optimal viewing set up was constructed. Optimal viewing conditions include low light, facing the display from a horizontal position and a correct distance. The optimal viewing distance was calculated from manufacturer's instructions and relevant literature and display screen size, and was established as 2 meters. [47] [48] The participants sat straight in front of the TV and within a 40% angle of the bottom of the screen as illustrated in figure 20.
All participants wore active LCD glasses and viewed a 27" active stereoscopic display. All participants wore headphones with a minimum of distortion such as enhanced base. It was insured that each participant wore the headphones the right way around and that the glasses were switched on and in correct 3D-mode.

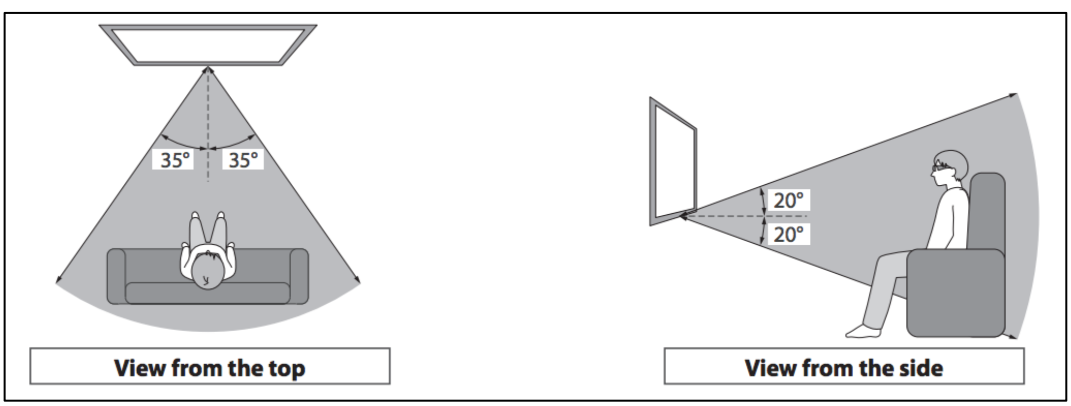

Figure 20 Viewing angles for 3D display used in experiment[49]

6.2.1 Participants

Six females and nine males participated in the experiment. Participants where aged 20 to 43 years old. All were naive participants[50]. All had a stereo acuity of <70 seconds of arc, tested with the Randot stereo test.

6.2.2 Stimuli

The stimuli were made up from three video clips. Each has a version with stereographic video and binaural sound, stereographic video with stereo sound and two-dimensional video with binaural sound. A clip of each artist was arranged on each test disc with a different audiovisual mode, so that each participant would view three clips in total. Rather than randomising the order, which would have been technically cumbersome, the stimuli were designed in order to counter balance any effect of the order of the clips. It was ensured that the same artist in combination with the same audiovisual mode was not repeated twice and that the order of the clips was varied. Each clip was approximately 3.30 min long and the whole test sequence lasted 15 minutes.

Disc A	2DBi X	S3DSt Y	S3DBi Z
Disc B	S3DSt Z	2DBi Y	S3DBi X
Disc C	S3DBi Y	S3DSt X	2DBi Z

6.2.3 Procedure

A participant permission form was designed stating research purpose and procedure to ensure informed consent. Participants underwent a polarised Randot stereo test to ensure that all participants had stereovision. They were familiarised with the questionnaire and the researcher insured that all the equipment was working and worn correctly. Participants watched the first stimulus, paused the Blu-ray player, and completed the relevant page of the questionnaire and so on. After watching all three films participants answered a question about overall best viewing experience, two questions about the rational for their choices.

6.2.4 Sampling methods

The investigation used a convenience sampling approach. University students at DMU were approached and asked to join the experiment, the rational for this being that the experiment location was situated within the university and the participants present at the locality. The disadvantages of this approach include that the participants come from a similar background and are of similar age. Therefore it is difficult to generalise results. The participants can be described as naïve in the sense that they were not directly involved with stereoscopic or binaural 3D research, as established in conversation.

6.2.5 Statistics methods

For quantitative data a non-parametric statistical test was made. The median of each the three sets of data was calculated. The absolute difference between the means was established. The probability of getting a score higher than or the same as the absolute difference by chance by resampling the data 10 000 times was established and is described as a p value. The significance level was set to $p = 0.05$. The tables are visualising individual participant responses to each question and the error bar indicates standard error.

7 Results

Results for the quantitative and the qualitative study will be presented and discussed.

7.1 Results for quantitative elements study

Question 1. Perceived Naturalness

Participants were asked to indicate perceived naturalness on a nine-point scale.

The mean of group one was 7, group two 7 and group three 7. There was no absolute difference.

The non-parametric analyses found no significant difference between the groups.

None of the clips where perceived as bad, and participants largely perceived the naturalness of clips presented in all three modes as Good.

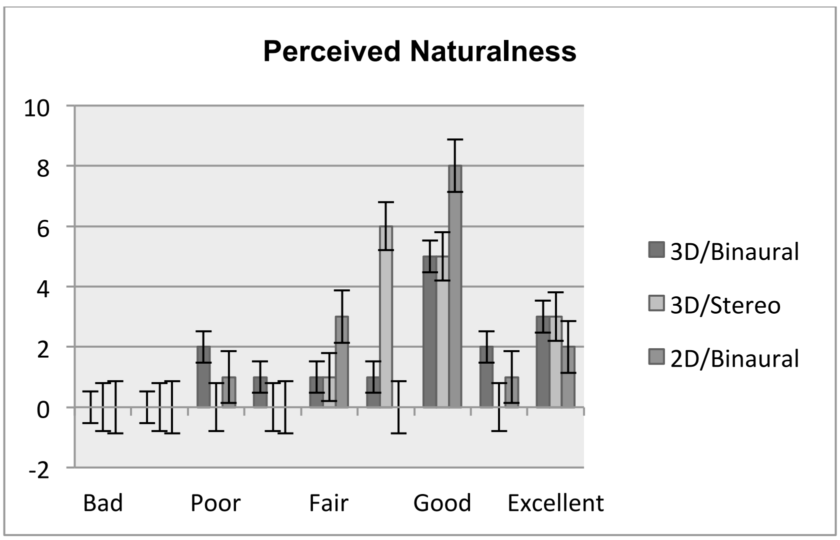

Table 3 Perceived Naturalness

Question 2. Perceived Depth

Participants were asked to indicate perceived depth on a nine-point scale.

The mean of group one was 5, group two 8 and group three 8. The observed value was 6.

The non-parametric analyses found highly significant difference between the groups ($p = 0.036$).

Participants perceived more 3D depth in the two 3D clips than in the 2D clips. More participants reported the 3D depth in the binaural stimuli as top of the scale, excellent, compared to the stereo stimuli. 3 participants scored the depth in the 2D video as good.

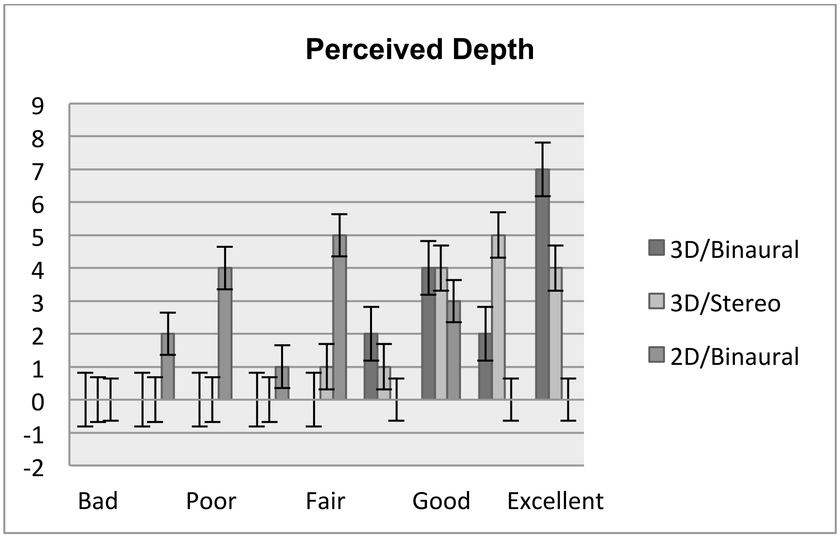

Table 4 Perceived Depth

Question 3. Perceived Image Quality

Participants were asked to indicate perceived image quality on a nine-point scale.

The mean of group one was 7, group two 7 and group three 7. There was no absolute difference.

The non-parametric analyses found no significant difference between the groups.

The perceived image quality was largely perceived as Fair to Excellent; with one participant scoring one clip as Bad in the range Bad to Poor.

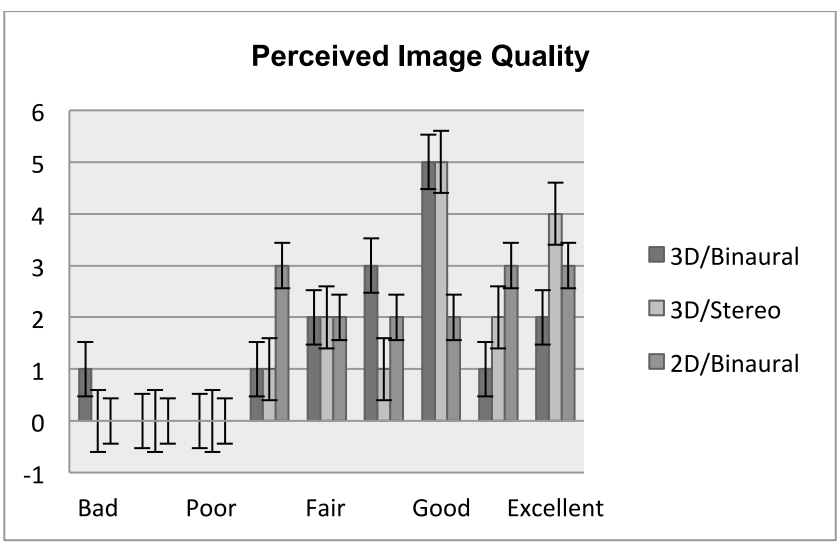

Table 5 Perceived Image Quality

Question 4. Perceived Eye Strain

Participants were asked to indicate perceived eye strain on a nine-point scale.

The mean of group one was 3, group two 4 and group three 3. The observed value was 2.

The non-parametric analyses found no significant difference between the groups (p =0.928).

The participants largely described eye strain below the middle value. In no stimuli was the eye strain perceived to be Very High.

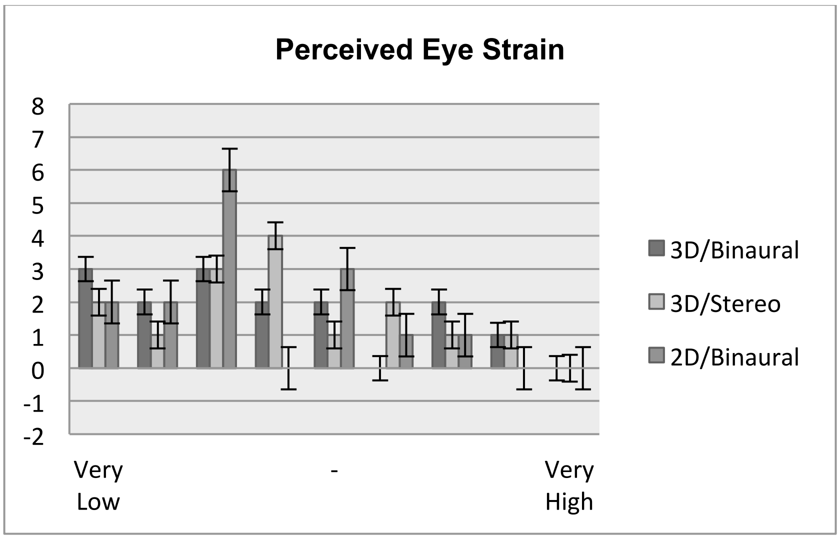

Table 6 Perceived eye strain

Question 5. Perceived presence in the room

Participants were asked to indicate perceived in the room on a nine-point scale.

The mean of group one was 5, group two 7 and group three 7. The observed value was 4.

The non-parametric analyses found no significant difference between the groups (p =0.505).

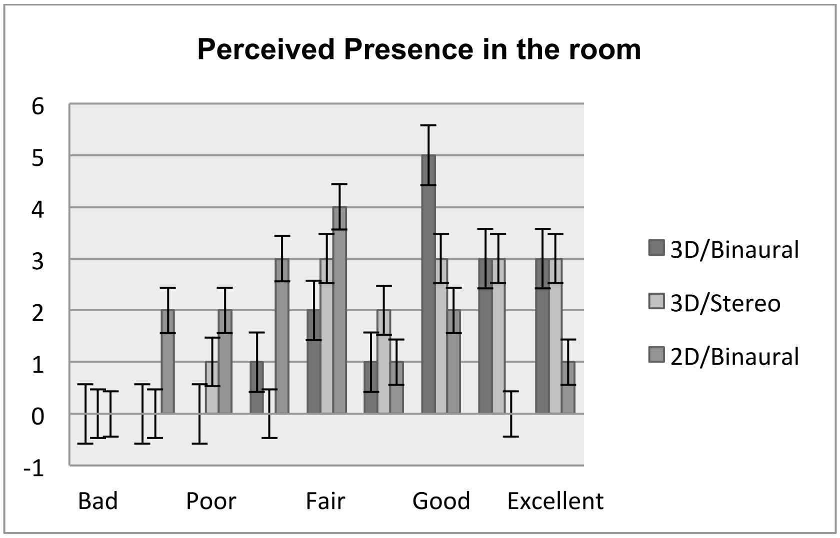

Table 7 Perceived Presence in the room

Question 6. Perceived Presence with the artist

Participants were asked to indicate perceived in the room on a nine-point scale.

The mean of group one was 5, group two 6 and group three 7. The observed value was 4.

The non-parametric analyses found no significant difference between the groups (p =0.263).

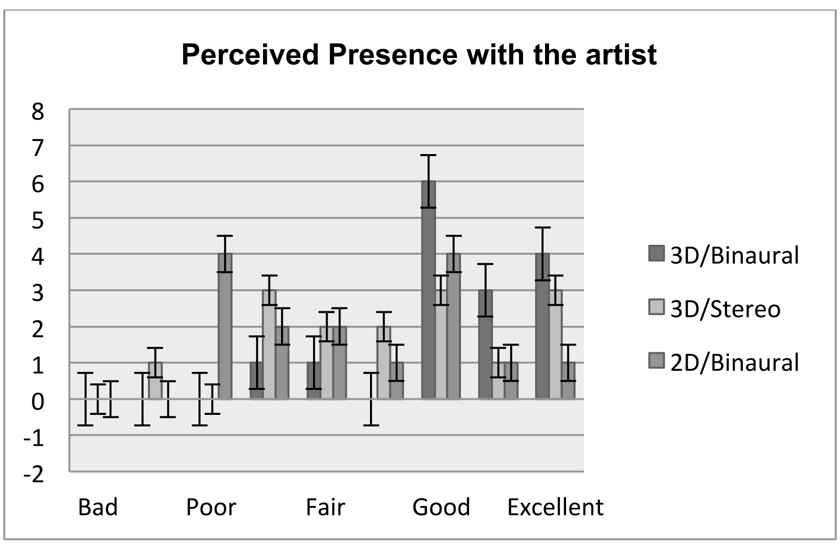

Table 8 Perceived Presence with the artist

Follow up question. Overall viewing experience.

After completing individual questioner pages for all clip the participants were asked to rate the film they thought provided the best, second best and lest good viewing experience. The mean of group one was 2, group two 2 and group three 1. The observed value was 2. The non-parametric analyses found no significant difference between the groups (p =0.270).

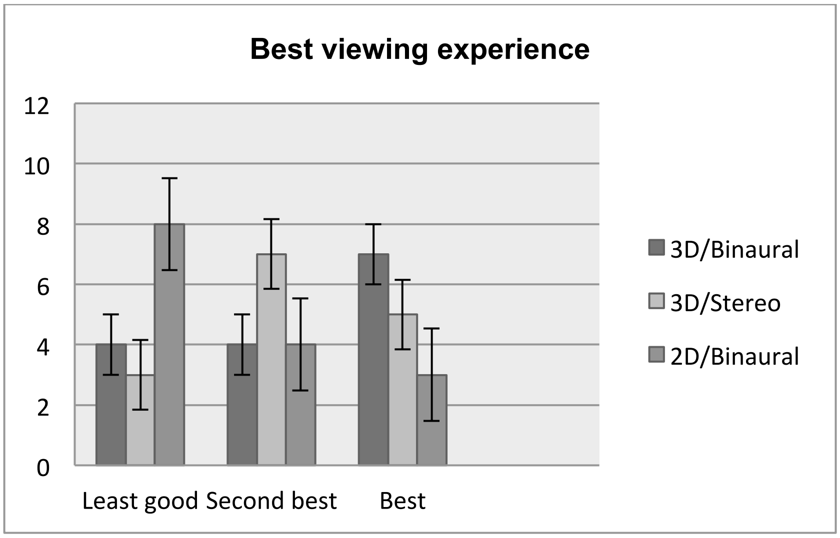

Table 9 Best viewing experience

7.2 Discussion of results of quantitative study

Several explanations to the lack of an effect can be offered. There could be a problem of the differences in the stimuli being too subtle. To avoid influencing the participants by factors such as movement by actors (in this case the musicians) and cameras, and to keep the binaural effect subtle rather than obvious, the stimuli was created without camera movement and without actors moving, apart from when entering the room. Stimuli with more distinct audiovisual clues may have produced a different result. It is possible that the scales are not refined enough to pick up and describe fine differences in perception adequately. The question that did relay statistically significant data, where participant scored perceived depth has a very clear difference between the perceived depth in 2D film with binaural sound and 3D film with stereo or binaural sound. The scales utilised could successfully differentiate between these greater differences. Stimuli could be also be having some differences outside the parameters investigated that influenced participants disproportionately. For example one of the clips featured a frame violation that some participants mentioned as distracting, however out of the four participants who descried the stereoscopic binaural clip as the least good, with references to the image quality, two described a clip with artist X, one a clip with artist Y and one a clip with artist Z. It cannot be verified that specific image artefacts were of consequence. Another possibility is that participants did not experience any difference in the stimulis' audiovisual modes It is not possible to support the hypotheses that binaural sound ads depth or immersivity on the bases of the quantitative study alone.

7.3 Results for the qualitative element of the study

The participants were asked to describe why they selected the films they did as providing the best and the least good overall viewing experience. The questions were: Why did you feel that the film you scored the best provided a better experience for you? Why did you feel that the film you scored the least good provided a poorer experience for you? The replies were sorted in groups according to which stimuli mode they were referring to. The text was analysed looking for common attributes and concepts in the participants' evaluation.

7.3.1 Attributes of the clips that provided the best overall viewing experience

3D video with binaural sound - 7 participants

Stimuli S3DBi, artist X, Y and Z

The most commonly used attributes for the stereoscopic clips with binaural sound were 'depth' and 'quality'. Participants used expressions such as 'The image quality was better, there seemed to be a real depth to image', 'both the video and the audio elements were equal in depth', 'The depth of the picture and the quality', 'there was a natural level if depth on the clip as if I was watching the person in real life'. Another participant referred to depth without using the actual attribute 'overall better level of quality and 3D perception'. Several participants referred to the attributes of 'audio' or 'sound' in contexts such as 'the added sound made it feel very realistic', 'the way sound worked made me feel like I was in the room'. Immersivity was mentioned 'more of an illusion of being in the film'. Three participants in this group mentioned the artist or the performance 'the performance in the third film was also more intimate, personal and engaging', 'the artist had presence' and 'I liked the environment the musician was in'.

3D video with stereo sound - 5 participants

Stimuli S3DSt, artist X and Y

The most commonly used attribute for the stereoscopic clips with stereo sound was 'room'. Participants expressed notions like 'looked really natural, I thought I was in the room' and 'you felt like you were in the room with the artist'. Depth was mentioned once and other attributes were used to describe depth: 'film provided a good sense of depth', 'impressive 3D reality' and 'images appear to be popping out of the screen and also far into the screen'. The attribute natural was used 'looked really natural' and 'it was more natural feeling'. Immersivity was mentioned with different attributes 'I felt most immersed', 'made me feel like I was actually there'. One participant mentioned that there was no eye strain 'It did not cause any eye strain and felt most comfortable watching'.

2D video with binaural sound - 3 participants

2DBi = artist Y and Z

The most commonly used attribute in connection with the 2D film was 'sound'; 'the sound quality' and 'the use of sound'. Other attributes were connected with the musician and the performance, 'the music and the artist were the most moving'. Participants mentioned presence 'despite there being less depth to the room the artist appeared to be sitting right across from me' and 'the feeling of presence'. The attribute depth was used once to describe the lack of it 'despite there being less depth to the room'.

7.3.2 Attributes of the clips that provided the least good overall viewing experience

2D video with binaural sound - 7 participants

2DBi = artist X, Y and Z

The most commonly used attribute when describing the lower quality in the 2D clips was 'depth'. 'The depth was poor' and 'Poorer depth sensation'. Depth was also referred to in other terms 'flatter image- not much 3D effect' and 'more resemblance of normal 2D.' Four participants mentioned eye strain 'seems to hurt my eyes watching it' and ' greater eye strain'. Two participants refer to immersivity 'I didn't [feel] like I was part of the scene' and 'the experience of being there wasn't as strong'. One participant sighted the artist/ performance 'I didn't feel engaged with the artist'. One participant experienced that the sound, which was binaural, was flatter than in the other clips 'sound felt flat on the piece'.

3D video with stereo sound – 4 participants

Stimuli S3DSt, artist X and Y

Audio was the most frequently used attribute when describing the lower quality in this clip. 'I could not resolve the visual space to the audio' and 'audio did not seem consistent with visuals'. Audio was also described with other attributes 'the recording seems rough and noisy' and' I felt like I was listening to record whilst watching a hologram'. Depth was mentioned by two participants, one in negative and in positive terms ‘the depth of the film was poor’ and ‘there was a nice depth to the room’. One participant referred to poor ‘image quality’. Two participants referred to the artist/ performance ‘the performer also seemed to be going through the motions’.

3D video with binaural sound – 4 participants

2DBi = artist X, Y and Z

The most commonly used attribute for these clips were referring to eye strain ‘slightly had to strain my eyes’ and ‘some details … looked blurry’. Depth was referred to once ‘there was too much depth’. Sound was referred to once ‘the sound was great’. Immersivity was denoted once ‘seemed less real’.

7.4 Discussion of results of quantitative element of this study

The most commonly used attribute when commenting on both the 3D binaural stimuli, most often seen as the best overall viewing experience, and the 2D binaural stimuli, most often seen as the least good overall viewing experience is ‘depth’. In the 3D binaural grouping this is frequently in in combination with ‘quality’ whilst in the 2D binaural group with the lack of depth. The 2D binaural group also relate eye strain. The 3D stereo stimuli are related to as a best viewing experience with attributes referring to ‘room’ and ‘depth’, and as least good viewing experience with ‘audio’. The 2D binaural best overall viewing experience group refers to ‘performance’ and ‘sound’, attributes not directly to do with positive experience of image depth. The 3D binaural least good experience group refers to eye strain. The groups are illustrated in table 10 and 11.

This could be seen to indicate that participants are experiencing differences, and are referring to different attributes when assessing viewing experience for the different groups. These types of differences are applied as evidence in the quality assessment models suggested by Strohmeier et al. (2010). In their study they indicated similar observations in regards to qualitative and quantitative data: “Firstly, the nonsignificant difference [in quantitative data] was not caused by the non-detectable differences between stimuli, as participants qualitatively differentiated them. Secondly, perceived depth was underlined in both audio and visual modalities, thus contributing to the overall audiovisual perception”.[51] Strohmeier et al. (2010) proposes using a Free-choice profiling and Generalized Procrustes Analysis method of study to take these observations into account. The current study has not been based on this method, so cannot fully conclude the significance of the qualitative evidence at hand, however it will allow returning to the quantitative data for analyses with the qualitative results in mind.

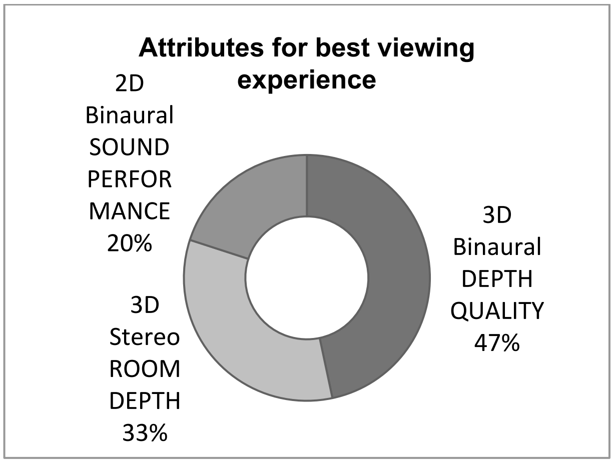

Table 10 Attributes for best viewing experience

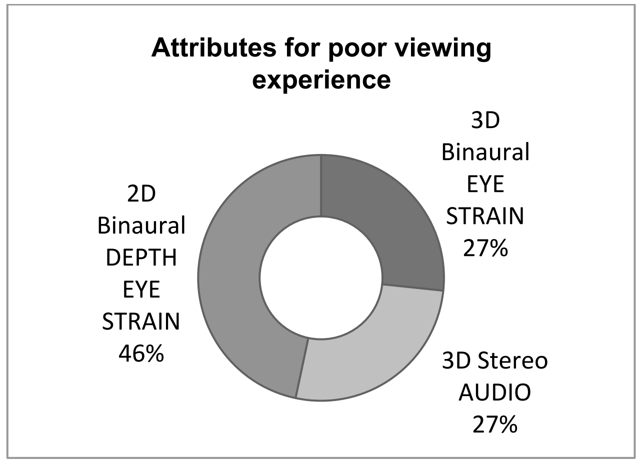

Table 11 Attributes for least good viewing experience

8 Discussion and Conclusion

8.1 Discussion

The results of the quantitative parts of this study alone cannot support the hypotheses that binaural sound can add presence to 3D-TV, as most of the results did not show statistically significant differences. Participants did record data that had significant difference with regards to perceived depth. Looking at other data sets that did show difference but not at a significant level they report a similar narrative to that of the depth data set, illustrated in tables 12, 13 ,14 and 15. The perceived depth, the perceived presence in the room and with the artist follow the same pattern. The null hypothesis for the qualitative data is that there is no difference between the data sets and the statistical analyses calculates the likelihood of the observed differences occurring by chance. Although the statistical analyses showed the same results could have occurred by chance, it is possible those participants did perceive differences and that a different method of analyses would have been more expedient. It is also possible that more contrasting, less subtle in difference, stimuli would have provided data with more effect. The data sets that did not follow this trend were perceived image quality, which as discussed earlier does more often refer to image sharpness, perceived eye strain, which was not predicted to be influenced and naturalness. The naturalness was most often scored as good for all the modes, 2D and 3D and the participants in this study do not appear to use it to describe depth. Perhaps this occurred as there was no in-depth training session to indicate the meaning of this term, as took place in the study by Lambooij and IJsselsteijn (2011).
The decision was made in the beginning of the process to base this study on quantitative methods constructed around examples from studies such as Evaluation of Stereoscopic Images: Beyond 2D Quality

(Lambooijd and IJsselsteijn 2011) and Can the perception of depth in stereoscopic images be influenced by 3D sound (Turner et al. 2011) but carried out in a much simplified form. However when reviewing the process it can be advocated that methods originally rejected due to their more complicated and time consuming processes could be of value when analysing perceived quality issues in 3D-TV. Building a model of analyses based on structuring participants own descriptions of their experience as suggested in the work of Strohmeier et al. (2010) may be an useful tool for further study.

Actual Methodology used in this study

Hypotheses → Method → Creation of stimuli → Quantitative and Qualitative Testing → Statistical Analyses (quantitative study proven partly successful) → Qualitative analyses → Discussion → Conclusion

Table 12 Perceived Depth, highly statistically significant (p = 0.035)

Table 13 Perceived Presence in the room, not statistically significant (p = 0.505)

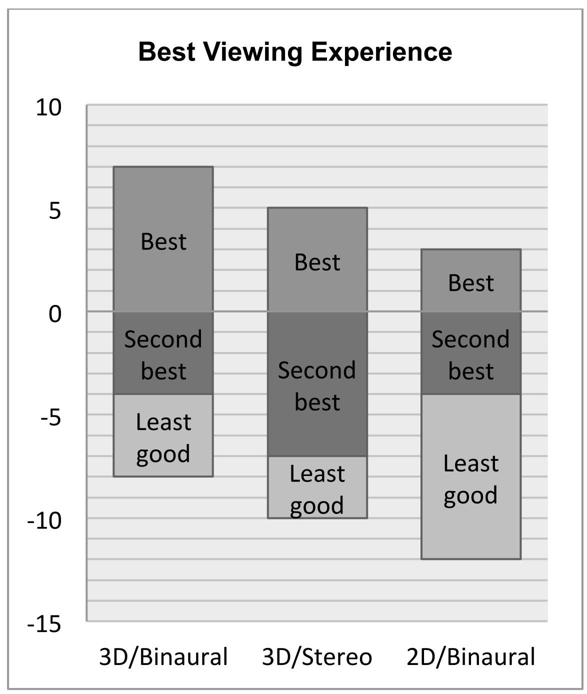

Table 14 Perceived presence with artist, not statistically significant (p = 0.263)

Table 15 Best Viewing Experience, not statistically significant (p =0.270)

Note that the tables on this page add a positive score above zero for participants answers above the description ‘Fair'. Fair has been used as a neutral as it stands in the center of the 9 point Likert scales used. Scores above fair has been evaluated as an expression of overt choice in these tables.

8.2 Conclusion

As stereoscopic 3D-TV and cinema is developing, and the media industry engage in marketing of the involved technologies across the world, technological, scientific and artistic aspects of the medium are continually being explored. In the technology communities methods of investigating quality aspects of stereoscopic 3D imaging are being explored, as traditional methods of evaluating 2D media do not adequately explore additional values such as depth and presence. A language for describing technical quality values is being developed. This is taking place simultaneously to the development of a creative language and practitioners and theorists are in continuous dialogue about what constitutes good stereoscopic 3D. This preliminary study has not proven conclusively that using binaural sound can add presence to stereoscopic TV. Some evidence has been put forward that it may do, and further research could build on these findings. Testing the hypothesis again with more distinct stimuli, including movement of the subject and the camera, could clarify the issues further. This report has provided a discussion on some methods and processes, which form part of the contemporary discourse, and has suggested that a methodology that builds on Free-choice profiling could provide a good way forward. For the independent filmmaker recording binaural sound is genuinely possible, and during this research audience feedback on the outcomes has been very positive. Exploration of the creative possibilities is open to the practitioner, and there is abundant scope for innovation and discovery.

Additional thought-provoking aspects were brought up during the process, which would be relevant for further investigation. Firstly the relation between placement of the camera, the focal length, and the binaural recording device could be optimised. What would be the most effective configuration, also taking the viewer placement into regard? Secondly can binaural audio clues, preparing the viewer for the appearance of an object from the side of the screen, shorten time it takes for fusion of the two images to occur in the brain, thus effect viewer fatigue?

9 Bibliography

COOLICAN. H. (1994) *Research Methods and Statistics in Psychology*. London. Hodder & Stoughton

DSOUZA. C. (2012) *Think in 3D.* Amazon Digital Services

HOWARD. D. M. (2009) *Acoustics and Psychoacoustics*. Oxford. Focal Press

LIPTON.L (1982) *Foundations of the Stereoscopic Cinema.* Van Nostrand Reinhold

MENDIBURU. B. (2009) *3D Movie Making*. Oxford. Focal Press

MENDIBURU. B. (2012) *3DTV and 3D Cinema.* Oxford. Focal Press

SAXBY. G. (2011) *The Science of Imaging.* Boca Raton. Taylor & Francis

ZONE. R. (2012) *3-D Revolution: The History of Modern Stereoscopic Cinema*. Lexington . University Press of Kentucky

10 References

---

[1] GLUCKMAN. P. 2012) *Glasses-Free 3D 'Really, Really Key' for CE Industry, RealD Says.* Consumer Electronic Daily [Online]. http://consumerelectronicsdaily.com/Content/Glasses-free-3D-big-key-for-CE-industry-RealD-says.aspx [Accessed: 8/12/12]

[2] 3DFOCUS (2012) *BBC HD Channel Screening 24 Minute 3D Promo* [Online]. http://www.3dfocus.co.uk/3d-news-2/3d-broadcasting/bbc-hd-channel-screening-24-minute-3d-promo/6413[Accessed: 8/12/12]

[3] SAXBY. G. (2011) The Science of Imaging. Boca Raton. Taylor & Francis

[4] ibid

[5] HEEGER. D. (2006), *Perception Lecture Notes: Depth, Size, and Shape*. Department of Psychology, New York University

[6] CURTIN. D. P. (2011) *Stereo Photography 3D in the Digital Era.* Short Courses

[7] LIPTON.L (1982) *Foundations of the Stereoscopic Cinema*. Van Nostrand Reinhold

[8] ibid

[9] ZONE. R. (2012) *The 3D Zone: Its Past & Its Future* [Online]. Available from: http://magazine.creativecow.net/article/a-creative-cow-magazine-extra-the-3d-zone [Accessed: 8/12/12]

[10] DSOUZA. C. (2012) *Think in 3D.* Amazon Digital Services

[11] MENDIBURU. B. (2009) *3D Movie Making Stereoscopic Digital Cinema from Script to Screen*. Oxford. Focal Press

[12] EDMUND (2011) *The Rise of 3D.* Box Office Quant. [Online]. Available from: http://boxofficequant.com/ [Accessed: 8/11/12]

[13] BFI. (2012) *New Horizons for UK Film*, Consultation. May 2012

[14] SKILLSET (2012) *3D Stereoscopic 2 3D MOT.* The Skillset Craft and Technical Skills Academy

[15] ITU NEWSROOM. (2012) *Ultra High Definition Television: Threshold of a new age* [Online]. Available from: http://www.itu.int/net/pressoffice/press_releases/2012/31.aspx#.UMkJYuTWjd420120830 [Accessed: 8/11/12]

[16] RODGRIGUEZ. S. (2012) *First glasses-free 3-D TV with 4K resolution to be unveiled.* Los Angeles Times [Online]. Available from: http://articles.latimes.com/2012/sep/01/business/la-fi-tn-3d-tv-without-glasses-20120830 [Accessed: 8/11/12]

[17] WORLDTVPC. (2012) *Samsung To Stick With 3D Glasses TV For Now*. [Online]. Available from: http://www.worldtvpc.com/blog/samsung-stick-3d-glasses-time/ [Accessed: 8/11/12]

[18] PANASONIC. (2012) *Shooting with Panasonic for comfortable viewing.* [Online]. Available from: http://eww.pass.panasonic.co.jp/pro-av/support/content/guide/EN/3dchints_en.html [Accessed: 28/11/12]

[19] BEATO. A. (2012) Understanding Comfortable Stereography. [Online]. Available from: http://64.17.134.112/Affonso_Beato/Understanding_Comfortable_Stereography.html [Accessed: 28/11/12]

[20] SHIBATA. T. (2002) *Head mounted display Displays.* Displays, Volume 23 (1–2)

[21] HOWARD. D. M. (2009) *Acoustics and Psychoacoustics*. Oxford. Focal Press

[22] FEAKES. C. A. (2011) *Binaural for the Blind: A Comparative Perceptual Study.* Thesis (PhD), De Montfort University

[23] Ibid

[24] HOWARD. D. M. (2009) *Acoustics and Psychoacoustics*. Oxford. Focal Press

[25] THE LONDON SOUND SURVEY. (2012) *Building a field recording system 4: introduction to binaural sound recording* [Online]. Available from: http://www.soundsurvey.org.uk/index.php/survey/blog_comments/1617/ [Accessed: 28/11/12]

[26] ibid

[27] MILLER. P.(2008) *Otokinoko Binaural Mic creeps us out in stereo.* [Online]. Available from: http://www.engadget.com/2008/06/15/otokinoko-binaural-mic-creeps-us-out-in-stereo/ [Accessed: 28/11/12]

[28] ROLAND (2012) *http://www.roland.com/products/en/CS-10EM/* [Online]. Available from: http://www.engadget.com/2008/06/15/otokinoko-binaural-mic-creeps-us-out-in-stereo/ [Accessed: 28/11/12]

[29] BEERENDS. J. G and DE CALUWE. F. E (1999) The influence of Video Quality on Perceived Audio Quality and Vive Versa. J. Audio Eng. Soc., 47 (5) pp. 355-362

[30] MENDIBURU. B. (2009) *3D Movie Making*. Oxford. Focal Press

[31] ANDRE. C, EMBRECHTS. J, VERLY J. G. (2010) Adding 3D Sound to 3D cinema. *International Conference on Audio Language and Image Processing (ICALIP)*, Liège, November, 2010. pp. 130 - 137

[32]M. LAMBOOIJD, IJSSELSTEIJN. W. A et al. (2011) *Evaluation of Stereoscopic Images: Beyond 2D Quality.* IEEE transactions on Broadcasting, 57 (2)

[33] STROHMEIER. D, JUMINSKO-PYYKKÖ. S, REITER. U (2010) Profiling experienced Quality factors of Audiovisual 3D perception. In: *International Workshop on Quality of Multimedia Experience (QoMEX),* Ilmenau, June 2010. pp. 70-75

[34] TURNER. A, BERRY. J, HOLLIMAN. N. (2011) Can the perception of depth in stereoscopic images be influenced by 3D sound. In: *Proc. SPIE 7863 Display and Applications XXll.* Vol7863. San Francisco, January 2011.

[35] Ibid

[36] M. LAMBOOIJD, IJSSELSTEIJN. W. A et al. (2011) *Evaluation of Stereoscopic Images: Beyond 2D Quality.* IEEE transactions on Broadcasting, 57 (2)

[37] TAM. W, STELMACH. L and CORRIVEAU. P (1998) *Psychovisual aspects of viewing stereoscopic video sequences*, Proc. SPIE, vol. 3295, pp. 226–235

[38] SEUNTIËNS. P. J. H, HEYNDERICKX. I and IJSSELSTEIJN. W. A. (2008) *Capturing the added value of 3D-TV: Viewing experience and naturalness of stereoscopic images.* J. Imaging Sci. Technol., vol. 52, no. 2, pp. 020504–020504-5

[39] M. LAMBOOIJD, IJSSELSTEIJN. W. A et al. (2011) *Evaluation of Stereoscopic Images: Beyond 2D Quality.* IEEE transactions on Broadcasting, 57 (2)

[40] STROHMEIER. D, JUMINSKO-PYYKKÖ. S, REITER. U (2010) Profiling experienced Quality factors of Audiovisual 3D perception. In: *International Workshop on Quality of Multimedia Experience (QoMEX),* Ilmenau, June 2010. pp. 70-75

[41] Ibid

[42] FEAKES. C. A. (2011) *Binaural for the Blind: A Comparative Perceptual Study.* Thesis (PhD), De Montfort University

[43] DALENIUS. T (2012) *Explore* [Online]. Available from: http:// http://ioctexploratory.tumblr.com/ [Accessed: 28/11/12]

[44] Ibid

[45] Ibid

[46] THE INTERNATIONAL TELECOMMUNICATION UNION (2002) Methodology for the Subjective Assessment of the Quality of Television Pictures, Recommendation ITU-R BT.500-11, ITU-R.

[47] PANASONIC (2011) Operating Instruction 3D Eyewear. China. Panasonic Corporation

[48] BEST3DTVS (2012) *Best 3D TV Screen Size and Viewing Distance.* Available from: http://www.best-3dtvs.com/guides/best-screen-size-viewing-distance/ [Accessed: 29/11/12]

[49] PANASONIC (2011) *Operating Instructions LCD TV*. Berkshire. Panasonic UK Ltd

[50] COLMAN. A.M. (2008) A *Dictionary of Psychology (3 ed.).* Oxford. Oxford University press

[51] STROHMEIER. D, JUMINSKO-PYYKKÖ. S, REITER. U (2010) Profiling experienced Quality factors of Audiovisual 3D perception. In: *International Workshop on Quality of Multimedia Experience (QoMEX),* Ilmenau, June 2010. pp. 70-75