MASTER

Effects of noise and blur on depth impression and overall image quality in 3-DTV

van Vugt, F.L.

Award date:
2007

Link to publication
Effects of noise and blur on Depth impression and overall image quality in 3-DTV

F.L. van Vugt
S463057
Effects of noise and blur on depth impression and overall image quality in 3-DTV

F.L. van Vugt
S463057

Master’s thesis
June 2006

Graduation committee:
Dr. W.A. IJsselsteijn  Department of Technology Management
Dr. Ir. L.M.J. Meesters  Department of Technology Management
Prof. Dr. Ir. I. Heynderickx  Philips Research Natlab, TU Delft
Preface

I have worked for twelve years on the Overseas-TV-laboratory at Philips in Eindhoven where we were dealing with product-design of TV-sets. The term ‘picture quality’ or ‘image quality’ frequently popped up during these design activities. Generally it concerned physical image quality problems like colour purity, convergence problems, ringing phenomena and non-linearity. At each development stage of a new TV-set, be it in the prototype stage, the hand-made models stage, or in the pilot series, these aspects always returned. More often than not it was found that these quality problems could be assigned either to problems in the combination picture tube with deflection system, or to the imperfections in the electronic circuits which had to drive the picture tube system. We measured these imperfections always in a physical way, sometimes on our knees in front of the tube with help of small rulers or sometimes with a microscope when the landing of the electron beam on the phosphor dots had to be checked. In all these cases the imperfections were expressed in numbers or proportions of physical quantities and we never used statistical techniques to process the measured data.

The judgments described in this report concerns assessing 2D- and 3D-images. These assessments occurred by means of criteria which can be summarised by the macro term ‘quality’. Seen in the light of the present view concerning image-judgment, this term is a very important, but also a highly controversial judgment criterion.

Engeldrum, when he explains the quality circle in his book ‘Psychometric Scaling’ (Engeldrum, 2000), pleads for using subjective image evaluation as a tool in the development process. We hardly did, at the very most on an individual base. From this point of view this research, sitting on the side of the ‘customer’, was very interesting to me.

The research took place under supervision of the TM-department from the faculty ‘Technology and Management’ of the Technical University in Eindhoven and was executed in one of the rooms in the test room hall in the IPO-building.

I want to thank those people who have accompanied me during this research. These persons are in particular Pieter Seuntiëns, Wijnand IJsselsteijn, Ingrid Heynderickx and Lydia Meesters.

My special thanks go out to René Olthof, editor of the New In Chess Yearbook Series and New In Chess Magazine. His editorial advice was more than welcome to me.

Without the input of these people this report would not have been accomplished the way it did.

Frans van Vugt
Summary

Within two years from now 3D-TV will be reality and people can watch 3D movies and 3D program scenes out of their armchairs. Before this milestone could be reached years of research was needed. Not only important investigations in the area of technical and system possibilities have been made but also in the psychological area. Questions such as what bandwidth or what resolution is acceptable or what camera base distance must be used, found their answers in this area.

The research described in this thesis is about the psychological aspects of 3D. The first theoretical part describes in short how and why we experience depth when we watch the outside world. It also gives a little history about 3D-viewing and some important studies in relation to 3D-quality aspects. Investigations showed that image quality is a controversial and vague criterion for 3D images, especially when the 3D image is impaired with artefacts like blur, noise or blocking.

Then in the second part the three research questions for this thesis are brought up. In short, one question is related to two criteria as possible replacements for the quality criterion. The criteria Viewing Experience and Naturalness were investigated, i.e. do these criteria perform better than the quality criterion for 3D and should the standards ITU-R BT.500 and BT-1438 be adapted because they both consistently use the term image quality?

The second question is to investigate the influences of blur and noise as functions of those criteria with various base distances as parameter. Especially in the area of artefacts, noise is still underexposed in scientific research. Descriptions, or, if possible, definitions of both the criteria and the artefacts used are given.

The third question is to investigate possible relations between the assessments of noise and blur and if these assessments are depending on the content of the images. Then a description about the test set-up is given. It describes how the 2D and 3D-images were prepared and how the images were presented. It describes also the stereoscope equipment used and the instructions the participants received. In fact, other researchers or students can repeat these tests in the same way if they read this chapter well.

From the results with the blur experiments it can be derived that blur rate significantly affects all the four criteria, Perceived Depth, Perceived Image Quality, Perceived Viewing Experience and Perceived Naturalness. The values of these various assessment criteria all decrease when the blur grade increases. Observers anchor their judgments on the most salient features, in this case the blur level. The observers of the Depth assessments clearly recognize the depth information but the observers which used the Quality criterion recognized clearly the blur gradations while they didn’t recognize depth or recognized it unsatisfactorily. The Quality criterion was therefore inferior compared to the other criteria. A comparison between the criteria Viewing Experience and Naturalness is in favour for the criterion Naturalness. Naturalness is more sensitive than Viewing Experience.

The noise impairment experiments show more or less the same results. The noise levels significantly affect all the four criteria. The various assessment values all decrease when the noise grade increases. The other conclusion is that observers with the Depth assessments clearly recognize the depth information. Also with noise impairments the Quality criterion is not a good qualifier. The criterion recognizes
clearly the noise gradations but it doesn’t recognize depth or recognizes it unsatisfactorily. A comparison between the criteria Viewing Experience and Naturalness gives a more or less equal result.

So the general conclusion is that the term ‘Image Quality’ is not sufficient when 3D-images are involved and should be replaced by the term ‘Naturalness’. So in the recommendation BT-1438, where 3D is used, the word ‘Quality’ should be replaced by the word ‘Naturalness’ when it relates to 3D.

With the given picture contents and the chosen blur and noise grades the noise impaired pictures with more detailed information have higher assessments for all criteria than the same pictures with blur impairments. With the exception of the depth criterion, the noise and blur impaired pictures with less detailed information have equal assessments or assessments in favour of blur.
# Table of Contents

PREFACE ......................................................................................................................... II

SUMMARY ........................................................................................................................ III

TABLE OF CONTENTS .......................................................................................................... V

1. INTRODUCTION ............................................................................................................ 1

2. THEORETICAL BACKGROUND .................................................................................... 2
   2.1 INTRODUCTION ........................................................................................................ 2
   2.2 MONOCULAR CUES TO DISTANCE PERCEPTION ..................................................... 2
   2.3 DISTANCE CUES AND BINOCULAR CUES ................................................................. 3
   2.4 HISTORICAL REVIEW CONCERNING 3D-ASPECTS .............................................. 6

3. RECENT RESEARCH CONCERNING 3D-ASPECTS .................................................... 11

4. INTRODUCTION TO THE FOUR ASSESSMENT CRITERIA ......................................... 18
   4.1 PERCEIVED DEPTH .................................................................................................. 18
   4.2 WHAT DOES PERCEIVED QUALITY MEAN? ........................................................... 18
   4.3 WHAT DOES NATURALNESS MEAN? ...................................................................... 21
   4.4 WHAT DOES VIEWING EXPERIENCE MEAN? ........................................................ 22

5. IMAGE ARTEFACTS USED IN THESE EXPERIMENTS ........................................... 24
   5.1 INTRODUCTION ........................................................................................................ 24
   5.2 DESCRIPTION OF BLUR ........................................................................................... 24
   5.3 DESCRIPTION OF NOISE .......................................................................................... 25

6. RESEARCH QUESTIONS .............................................................................................. 27
   QUESTION 1 ....................................................................................................................... 27
   QUESTION 2 ....................................................................................................................... 27
   QUESTION 3 ....................................................................................................................... 27

7. DESIGN AND DESCRIPTION OF THE TEST ............................................................. 28
   7.1 INTRODUCTION ........................................................................................................ 28
   7.2 DESIGN OF THE TEST .............................................................................................. 28
   7.3 PARTICIPANTS TO THE TEST ................................................................................. 28
   7.4 TEST MATERIALS, TEST EQUIPMENT AND TEST ROOM ....................................... 29
      7.4.1 Test materials ....................................................................................................... 29
      7.4.1.1 The stereoscopic images ................................................................................. 30
      7.4.1.2 Design of the image sequence test files .......................................................... 30
      7.4.2 Test equipment ..................................................................................................... 31
      7.4.3 Test room ............................................................................................................... 32
   7.5 MEASURES .................................................................................................................. 32
   7.6 THE TEST PROCEDURE ............................................................................................ 32

8. RESULTS OF THE TESTS ............................................................................................ 34
   8.1 INTRODUCTION TO THE ANALYSES ..................................................................... 34
   8.2 AVERAGE VALUES WITH BLUR DISTORTIONS ..................................................... 34
   8.3 AVERAGE VALUES WITH NOISE DISTORTIONS ................................................. 37
   8.4 RELATIONSHIPS BETWEEN THE VARIOUS CRITERIA WITH BLUR DISTORTIONS 40
   8.5 RELATIONSHIPS BETWEEN THE VARIOUS CRITERIA WITH NOISE DISTORTIONS 42
   8.6 RELATIONSHIPS BETWEEN THE NOISE AND THE BLUR IMPAIRMENTS .................... 44

9 CONCLUSIONS FROM THE EXPERIMENTS .............................................................. 48
   9.1 CONCLUSIONS FROM THE BLUR EXPERIMENTS ............................................... 48
   9.2 CONCLUSIONS FROM THE NOISE EXPERIMENTS .............................................. 48
   9.3 CONCLUSIONS WITH RESPECT TO THE RELATIONS BETWEEN NOISE AND BLUR 49
   9.4 ANSWERS TO THE QUESTIONS .............................................................................. 50
1. Introduction

For thousands of years people try to record the daily images they receive on their retinas, into pictures with help of available techniques. The oldest known human made images are the murals in the cave of Lascaux in France, made during the Palaeolithic period about 18,000 BC. Was the cave used as a religious temple or was it used as a shelter for the people during the cold winter periods? We don’t know. We also don’t know how the reactions of the spectators were to those pictures at the time. Did they like the pictures? Probably yes, because, although the used techniques were very primitive, we still find them very beautiful and amazing.

Well known are the reactions of spectators when Rembrandt van Rijn finished his most famous oil painting ‘The Night Watch’ in 1642. They had no eye for the composition of the painting, no eye for the light effects that make the composition so striking. They only saw the hardly recognizable faces in the dark, the informal posing of the watchers and the playing kids in the middle. How horrible! Nowadays we consider this painting with the painted light effects as one of the best in its sort and one of the best ever made by human beings.

When people watch pictures they judge them. This brings us to the core questions of this report. How do people judge images? In this report we don’t put our attention to the composition of the images but to the effects of the used techniques. The developments of image-techniques are going fast. Starting with the photography in the middle of the nineteenth century and the film at the beginning of the twentieth century there exists a long row of new techniques, new image carriers, new colour techniques, new transmission techniques etc, time after time. Every time when a new image technique was introduced, it only could do so by comparing and judging the new with the old techniques and trying to improve, to come to more spectacular effects.

Now we are standing at the point of the introduction of yet another new technique: stereoscopic images and films. This new technique gives rise to new questions to answer like how to reduce the bandwidth for transmission or what will be the influence of introduced impairments? Can we still talk about the term ‘Quality’? Recent investigations showed that the use of this term can give cause to confusion. This report tries to give answers on some of the questions put above.

Figure 1.1: Famous murals in the cave of Lascaux in France
2. Theoretical Background

2.1 Introduction

Before we come to the definition of our research questions and the description and discussion of the tests, we need to know something about the theory and definitions of monocular and binocular viewing as well as the historical backgrounds and the results of scientific research in the 3D-area, because the tests described in this report are based on and have their origin in those investigations.

In the second and third paragraph of this chapter we will treat the monocular and the binocular cues respectively. This theory will be given in short because most of the theory can be found in reference books. The fourth paragraph features a historical overview of the latest scientific results. Especially the researches concerning the judgments of impaired 2D- and 3D-images are very important for our own investigation.

2.2 Monocular cues to distance perception

Question: Can a soccer keeper with only one eye hold a ball, kicked from the edge of the penalty area?

Answer: probably yes, when the shot is not too hard, but it is very difficult. For the estimation of the speed, the direction of the ball and the estimated time he will need for catching the ball at the goal line, he only avails of monocular cues. The keeper lacks the binocular cues like convergence and binocular parallax. These binocular cues are particularly effective for distance estimates of a nearby object closer than 10 m.

The depth perception theory makes a clear distinction in seeing distance with one eye, the so called monocular vision, and seeing distance or depth by watching with two eyes, the so called binocular vision. A lot of people think that one only can see a sense of depth with two eyes, but this is a misunderstanding. With one eye only (monocular viewing) you can also see depth but only to a certain extent. Strangely enough monocular viewing has even more specific depth cues than binocular viewing. Table 2.1 on next page, taken from the book of Matlin & Foley (1997) ‘Sensation and Perception’, gives an overview of the cues in the several groups (Matlin & Foley, 1997). As can be seen, monocular vision has a large number of factors from which distance information can be derived. They can be divided into factors involving motion and factors not involving motion.

One of the factors not involving motion is accommodation. When you look to a picture you have to focus your eye depending on the distance of the picture to your eye. The eye lens is thinner when the image is further away and becomes thicker when the image is more nearby. The eye muscles control the lens shape during focussing. This changing of the lens is called accommodation. See Figure 2.1.

![Figure 2.1: Accomodation](http://en.wikipedia.org/wiki/Eye#Varieties_of_eyes)
Another group of factors not involving motion are the so called pictorial cues. From the pictures you receive on the retina, a lot of specific pictorial cues can be derived. The table below gives seven cues. In Appendix 1 you can find brief explanations with examples of these pictorial clues.

### Monocular factors

<table>
<thead>
<tr>
<th>No movement</th>
<th>Movement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eye muscles involved</td>
<td>Pictorial cues</td>
</tr>
<tr>
<td>Inter-position</td>
<td>Size</td>
</tr>
<tr>
<td></td>
<td>Texture gradient</td>
</tr>
<tr>
<td></td>
<td>Linear perspective</td>
</tr>
<tr>
<td></td>
<td>Atmospheric perspective</td>
</tr>
<tr>
<td></td>
<td>Shading</td>
</tr>
<tr>
<td></td>
<td>Height</td>
</tr>
</tbody>
</table>

**Table 2.1:** Table taken from ‘Sensation and Perception’, page 166 (Matlin & Foley, 1997)

One of the monocular cues involving motion is parallax. Watching an object and at the same time moving our head give us a lot of information about the position and the distance of that object. Because a change in position is called parallax, this phenomenon is called motion parallax. When the attention is fixed at an object close to the observer, the object seems to move in the other direction as the movement of the head, while another object further away seems to move into the same direction as the movement of the head. It is not possible to show this effect in a 2D-picture. Kinetic depth effects involve the motion of the observed object itself. Objects that look flat when they don’t move appear to have depth once they start moving. Changes in the observed size of the moving object, the observed effects when the object rotates, the influence of gravity on a moving object, they all provide us with cues about depth and distance.

### 2.3 Distance cues and binocular cues

A stereoscopic view is only possible with two eyes. Someone who has only one eye or two strongly different eyes will miss these depth aspects. This ability, also called stereopsis, arises because both eyes are about 6.3 cm situated from each other. When the eyes will be focussed to an certain object, the picture on the fovea of one eye will slightly deviate from the picture on the fovea of the other eye because of this distance. The brain will receive two slightly different images of the subject and will fuse them to one picture. The differences between both images (binocular disparity) will be experienced as depth.

All existing 3D-techniques are based on this principle: create two paired pictures, one for the left and one for the right eye and take care that the retina of the left eye will receive only the left image and the right retina will only receive the right image. The brain will combine the two pictures to one picture and the viewer will experience ‘depth’.

In the book ‘Sensation and Perception’ from M.W.Matlin & H.J.Foley (1997), two binocular factors have been given: Convergence and Binocular parallax. In the same breath the last term is also called ‘Binocular disparity’.

3
Our eyes point more inward when we look at an object close to us than at an object further away. See Figure 2.2. This phenomenon is called convergence. Depending on the degree of convergence we can make an estimate about the distance to the object. Convergence is only effective when the distance is less than 10 m.

For a better understanding of binocularity the horopter concept needs to be introduced. See Figure 2.3. In the picture the ideal horizontal horopter is shown. The horopter is a perfect circle and every point at the circle will give the same deviation from the reference F on both fovea, with the eye lens centres as crossing points (angle $\alpha_1 = \alpha_2$). That means that, when the eyes are focused on a certain point at the circle all other points on that circle will be seen single and at equal distance if the focusing is kept unchanged. For every focussing point such a circle can be constructed and those circles answer to the name Vieth-Muller circles (Vieth 1818, Muller 1826).

The brain will fuse corresponding points to a dept perception when those points don’t fall on the circle and when the disparity is not to large.

Two remarks have to be made on this concept. First: experiments showed that the horopter is not laying on a perfect circle but is somewhat flattened out to the back (Figure 2.4). This is the so called empirical horopter. The reason is that the shape of the eye and the lenses are not perfect either. Second: although the horopter is a sharp line there is a small band or area around the ideal line in which fusion into a single image still occurs. Spatial points that lay within this band stimulate non-corresponding retinal points but the brain still fuses these points into single images. This band is called Panum’s area or Panum’s zone (Figure 2.5).

Two special cases of disparity occur when a object is situated before or behind the object laying on the horopter. See Figure 2.6. When object Q is in front of the horopter-object A, then the viewer will see object Q with his right eye as object B on the left side of object A and with his left eye as object C on the right side of the horopter-object A. The brain will still fuse the front-object Q to a single object unless the disparity is to large, then the object will be seen double. The same is valid for the back-object P with this difference that the left eye will see the back object as
object B on the left side of object A and the right eye as object C on the right side of A. The first case is called crossed disparity and the second case uncrossed disparity.

As can be seen in Figure 2.6 the disparity depends on the distance between the eyes. The base distance between the fovea of the eyes on average is 6.3 cm. For a true-to-life picture the base distance between the stereo cameras should therefore also be 6.3 cm, see Figure 2.7. The depth experience can be influenced by varying this base distance between the two cameras. In this way the distance and the angular size of objects can be uncoupled. When the base distance increases the depth perception increases too, but at the same time the width of the object decreases. This uncoupling can cause all kinds of typical 3D-distortions such as keystone distortion, puppet theatre effect and card board effect. Pastoor (1993) concluded that when the camera distance is greater than the screen distance, the stereoscopic images look unnaturally small.

Just as there is a horopter in horizontal direction a vertical horopter also exists. This vertical horopter is a straight line and has a backwards tilt such that it passes through the fixation point and a point near the feet of the observer. We all know this effect when we place our laptop on the table in front of us and open it. The LCD-screen will always be placed in a backward tilt position. See Figure 2.8.

Till so far the ‘depth’ theory. We went through this theory with rather big strides, but it’s sufficient to be able to follow the tests described in this report. This overview is not complete. For instance, one factor, which is not mentioned, are the oculomotor cues, they have been described indirectly via topics such as focussing and convergence. The eye muscles are involved in oculomotor cues.
2.4 Historical review concerning 3D-aspects

To understand the investigations of the recent researches better we will first describe in short some interesting historical facts about 3D-techniques and 3D-science.

The first known description of 3D comes from the mathematician Euclid in 280 B.C. He had come to the conclusion that depth perception can only be obtained when each eye simultaneously receives one of two dissimilar images of the same object (Turing Institute, 1996). Around 200 B.C. Galenos, who is considered to be the founder of the physiology, came to the same conclusion that dept perception can only be caused by dissimilarity (Turing Institute, 1996).

An Arabic scientist, Abu Ali Al-Hazan ibn Al-Hazan ibn Al-Haytham, noticed at the end of the tenth century that under normal viewing conditions, the world appears to us as seen from a virtual eye placed in the middle between the left and right eye positions. The geometry of this cyclopean view is illustrated in Figure 2.9

This simple geometrical arrangement has an important consequence: since the outside world appears different from these two viewing points we never perceive an outside world image recorded directly by any sensory array, but our brain has to construct that image! (Henkel, 1994)

![Figure 2.9](http://axon.physik.uni-bremen.de/research/stereo)

A man who needs no introduction, Leonardo da Vinci, was architect, musician, anatomist, inventor, engineer, sculptor, geometer and painter. Nowadays he is described as the archetype of the "Renaissance man".

![Figure 2.10](http://www.dartmouth.edu/~matc/math5.geometry/)
He invented tools and compasses and rules to draw perspective and worked them out in his paintings. See Figure 2.10. His paintings and sketches showed a clear understanding of shading, texture and viewpoint projection (Turing Institute, 1996). He concluded that the two eyes of a viewer receive different images in horizontal direction and that therefore it is not possible to paint ‘depth’, or what he called relief, in his paintings.

Around 1600 Jacopo Copo Chimenti da Empoli painted a double painting based on disparity. See Figure 2.11. It is assumed that he made the double painting in commission of Johann Baptista Porta, a scientist from Naples. In those days it was impossible to view the painting via a stereoscope, but when this painting is now viewed with the help of a modern stereoscope, it gives a real 3D-picture.

In 1677 Père Cherubin of Orleans described stereoscopic three dimensional vision. Obviously his story was neglected because it had no follow up (Thimbleby, 1997).

The modern history of stereoscopy can be traced back to 1838. Sir Charles Wheatstone (1802-1875) was an English physicist who was the inventor of the bridge of Wheatstone, an electric circuit to calculate electric resistance; the concertina, a kind of accordion; the Diaphonicon, a primitive type of microphone or sound conductor; the five needle telegraph and a lot more. Then he presented his classic paper “On some remarkable, and hitherto unobserved, phenomena of binocular vision” to the Royal Society of London, which discussed his theories on stereopsis.
With his stereoscope viewer, an invention from the same time, the viewer was able to see a 3D picture that was build up out of two slightly different views of the same painting. The two pictures were projected to respectively the left and the right eye with help of mirrors and the brain fuses the two pictures to one stereoscopic image. The effect was overwhelming. For a sterilised impression of the Wheatstone stereoscope see Figure 2.12.

When Wheatstone wrote his paper, the only method easily available for preparing a stereo pair was drawing. Soon, however, photography became available. This method was ideally suited to make stereo pairs. For this purpose two cameras, a camera with two lenses and film holders, or one camera moved between exposures were used.

For one of the oldest stereoptic cameras see Figure 2.13. First one photograph was taken, then a new photographic plate was inserted and the second photograph was taken.

Sir David Brewster (Scottish physicist, 1781-1868) further developed the stereoscope by utilizing prismatic lenses to magnify and fuse the stereo images. See Figure 2.14.

His stereoscope, invented around 1844, was much more convenient, and became the standard throughout the age of popularity of stereograms. Viewing stereoscopic still images became popular in the western world, and from 1860 to the 1930’s stereography flourished. Brewster's lenticular stereoscope became a commercial success, selling 250,000 in a short time. See Figure 2.15. Many homes had
stereoscopic viewing devices of various sorts during the nineteenth century. The images were permanently mounted side by side, and only had to be inserted into the instrument to be viewed.

Figure 2.15: Brewster’s lenticular stereoscope

In 1959 Bela Julesz demonstrated with his random dot stereogram the brain is able to perceive 3-D information from stereoscopic disparity information only. Starting with a square random dot picture, he constructed a second image by shifting the first one slightly. Then he surrounded both squares with an identical dot pattern and combined them to a stereo pair. When viewing this stereo image in a stereoscope the central part of the image seems to be floating. In 1971 he wrote "Foundations of Cyclopean Perception", in which he described photographic techniques to produce random dot stereograms. See Figure 2.16

Figure 2.16: Random dot stereogram

Then in 1979, the article "A computational theory of human stereo vision" appeared in which Marr and Poggio described computational models. Based on these models,
Christopher Tyler designed his first single image random dot stereogram (SIRDS). With these SIRDS an unaided eye can view depth. They use the same technique as random dot stereograms but with only one image in which a pattern is repeated several times (Henkel, 1994). See Figure 2.17.

Starting with the introduction of photography, around the middle of the nineteenth century, a lot of new techniques were introduced to record images, such as film, TV, printing, digital images and so on. With the rise of those new techniques, the need to compare the images made by those techniques and to measure their improvements became more and more stringent. One was looking for methods to express the quality in a way that comparison was possible at different times and at different places. Extensive research was devoted to this subject and various new insights became available in accordance to the technical developments in recording and transmitting images. Nowadays 3D techniques are actually implemented in new TV-systems. With the implementation of these techniques, new applications become available. One can think of professional applications like monitoring spots from a distance when precise handling is needed, such as a medical surgery inside the human body, but also applications for pleasure, like 3D-TV in the living rooms of our homes. In next chapter we will discuss recent investigations which are based on typical problems related to transmission and the impairments introduced with that.
3. Recent research concerning 3D-aspects

Within two years from now we can expect high-quality three-dimensional TV-broadcasting (3-D TV). A lot of investigations and effort went in advance into understanding quality aspects of the system and reaching well defined and accepted standards. An important part of these investigations aimed at investigating acceptable restrictions in quality and bandwidth in order to reduce costs.

Transmitting a stereoscopic video signal requires twice the bandwidth of a monocular signal. A method to reduce the bandwidth can be asymmetrical coding of the two video streams. In 1992 Perkins introduced the theory of binocular suppression into stereo coding. He demonstrated that the final perception of a stereoscopic image is dominated by the highest quality component of one of the channels of a stereo-pair. When one view of a stereo image is of a high quality then the quality of the other view can be reduced without visible distortion of the binocular perception (Perkins, 1992).

This theory formed the base for some investigations by a.o Stelmach, James and Meegan (1998) and Stelmach, Tam, Meegan, Vincent and Corriveau (2000). Their first research was done on 10-second stereo video-sequences. Here one view of the stereoscopic video pairs was degraded in two different ways. One way was a low-pass filtering of the video stream (spatial resolution). This was done in horizontal and/or vertical direction to full, ½ and ¼th of the original resolution. This gave nine video-sequences with a different level of spatial resolution. The second method was temporal filtering by either taking 2-Field averages, or by dropping and repeating frames. This temporal filtering was applied on a subset of three spatially filtered video sequences.

![Spatial Filtering](image1)

![Spatio-Temporal Filtering](image2)

Fig. 3.1: Quality, depth and sharpness rating of image sequences where the right eye view was spatially filtered. Error bars show the 95% confidence interval of the mean

Fig. 3.2: Quality, depth and sharpness rating of image sequences where the right eye view was temporarily filtered. Error bars show the 95% confidence interval of the mean

Participants had to judge the stereoscopic video streams on three criteria; Quality, Depth and Sharpness. It was found that spatial filtering produced acceptable results:
The overall sensation of depth was unaffected by low-pass filtering, while ratings of quality and of sharpness were biased towards the eye with the greater spatial resolution. By comparison, temporal filtering produced unacceptable results: these conditions yielded images with poor quality and sharpness, even though perceived depth was relatively unaffected. See also Figure 3.1 and 3.2. The conclusion was that spatial filtering of one channel of a stereo video-sequence may be an effective means of reducing transmission bandwidth (Stelmach, James & Meegan, 1998).

In the year 2000 the same researchers issued another study with the same theme; asymmetrical coding of the two video streams with the goal to reduce bandwidth. The first part of the test was done with three different static stereoscopic images (Stelmach, Tam, Meegan, Vincent & Corriveau, 2000). The applied bandwidth reduction techniques were again low-pass filtering, a DCT-based quantization and a combination of both techniques. DCT stands for ‘Discrete Cosine Transformation’. An example of this technique is for instance jpg-coding. In this part of the test the participants had to tune the still test images until it had the same amount of blur/quantization as the source images. The conclusion of this part of the test was that the human visual system treats low-passed filtered and quantized images differently when computing the binocular image. For low-pass filtering, the high-quality source image dominates the perception. For quantization, the binocular perception is roughly the average of the left- and right-eye input. This formula also applied to the combination of low-pass filtering and quantization (Stelmach et al., 2000). In the second part of the test 10 second stereoscopic movies were used with the same compression techniques. The results were the same as with the still pictures. The final conclusion was that image quality was affected by quantization and to a smaller extent by spatial filtering. The other conclusion was that asymmetric coding is a promising method to reduce bandwidth.

Coding of stereo images uses the same techniques as coding of movie images or still images. With these coding techniques conventional coding distortions will also be introduced. The effects of those distortions, also called artefacts, have been studied extensively. The most important investigations for the purposes of the experiment discussed in this report, will now be presented.

Fig. 3.3: Images taken from Tam et al. (1998)
In 1997 Stelmach investigated the effect of blocking, an artefact that arises with JPEG-coding or MPEG-coding. He did his tests with help of three short movies: Flower Pot, Trapeze and Street Organ. His participants had to assess the movies on three criteria: depth sensation, perceived sharpness and quality.

Stelmach came to the conclusion that stereo depth in natural scenes did not improve image quality ratings compared with the same 2D-images (see Figure 3.3, graph on the right). (Tam, Stelmach & Corriveau, 1998) (Stelmach, Tam & Corriveau, 1997). Annika Berthold, a scientist working in Kyoto, Japan, at the time took a stand against this result. She expressed it as follows in her report: ‘this result was surprising since intuitively it might be expected that both the extra depth and the extra sharpness of stereo images would boost image quality ratings’ (Berthold, 1997). She started a new experiment with still blur-impaired pictures and with the same criteria for the participants: depth, sharpness and quality. Blur is a widespread and important degrading factor which can degrade the appreciation of quality considerably.

![Fig. 3.4: Images taken from Berthold (1997)](image-url)

For the results see Figure 3.4. Her findings for the depth perception are equal to those by Stelmach et al. Depth perception increases with stereo pictures. The difference between the 2D and the 3D-pictures is more or less constant, also when the blur factor increases. But the perception for the other two criteria was different. Berthold et al. explained this as follows: the improved image quality ratings for stereo may be the improved image sharpness. This is due to the fact that the brain receives information about the scene from two sources (left and right eye image), and since the noise between the two images is uncorrelated, the stereo images are usually perceived as more crisp than the non-stereo images (Berthold, 1997).

She also gave some potential reasons why her results deviated from the results of Stelmach et al. The main reason was some negative aspects of stereo viewing which causes the expected sharpness to be reduced. It is important that when adding stereo to a picture there should be no image deteriorations or factors unpleasant to the viewers present in the pictures. With a simple preference study she proved that this was the case with the Stelmach tests. Furthermore she indicated two secondary effects that could influence the results between her tests and Stelmach’s: firstly the
different impairments, blur in the test of Berhold and blockiness in the test of Stelmach and secondly the content of the images.

Depth perception of an image can be influenced by varying the camera base distance when taking the photograph. The average distance of the eyes is 6.3 cm. If the base distance of the cameras gets larger, the perceived depth will increase and vice versa (when the base distance is 0 then we have a 2D-image). IJsselsteijn, de Ridder & Vliegen (2000) investigated the effects of stereoscopic film parameters and display duration on observer’s appreciation of stereoscopic images. In that research he manipulated the camera base distance, the convergence distance, the focal length and the display duration. The observers had to assess the images on the criteria Quality and Naturalness. To have an impression of what happens if one of the parameters changes, Table 1 is taken from the report.

<table>
<thead>
<tr>
<th>Parameter</th>
<th>Effect of increase</th>
<th>Effect of decrease</th>
</tr>
</thead>
<tbody>
<tr>
<td>Camera separation</td>
<td>Increase in disparity values</td>
<td>Decrease in disparity values</td>
</tr>
<tr>
<td>(values used in the experiment: 0, 4, 8, 12, 16 and 24 cm)</td>
<td>Increase in perceived depth</td>
<td>Decrease in perceived depth</td>
</tr>
<tr>
<td></td>
<td>Constant size of objects</td>
<td>Constant size of objects</td>
</tr>
<tr>
<td></td>
<td>Constant field of view</td>
<td>Constant field of view</td>
</tr>
<tr>
<td>Camera focal length</td>
<td>Increase in disparity values</td>
<td>Decrease in disparity values</td>
</tr>
<tr>
<td>(values used: 10 and 20 mm)</td>
<td>Increase in perceived depth</td>
<td>Decrease in perceived depth</td>
</tr>
<tr>
<td></td>
<td>Increase in size of objects</td>
<td>Decrease in size of objects</td>
</tr>
<tr>
<td></td>
<td>Decrease in field of view</td>
<td>Increase in field of view</td>
</tr>
<tr>
<td>Convergence distance</td>
<td>Decrease in disparity values</td>
<td>Increase in disparity values</td>
</tr>
<tr>
<td>(values used: 1.30 and 2.60 m)</td>
<td>Shift in perceived depth</td>
<td>Decrease in disparity values</td>
</tr>
<tr>
<td></td>
<td>Constant size of objects</td>
<td>Constant size of objects</td>
</tr>
<tr>
<td></td>
<td>Constant field of view</td>
<td>Constant field of view</td>
</tr>
</tbody>
</table>

Table 1: Taken from IJsselsteijn et al. (2000)

A more detailed explanation of the criterion Naturalness will be given in the Chapter 4.3. Below the most important graphs are shown.

Fig. 3.5: Quality and naturalness scores plotted against camera base distance (in cm). The graphs in the top panel show the results of the 5s display duration condition, the graphs in the bottom panel show the results for the 10s condition.

Fig. 3.6: Quality and naturalness scores plotted against camera base distance (in cm). The graphs in the top panel show the results for the scene ‘Playmobiles’, the graphs in the bottom panel show the results for the scene ‘Bureau’.

The most important conclusions from this experiment are: Firstly, the effect of display duration on the subjective evaluation of stereoscopic images is relatively small. Secondly, naturalness is judged higher than quality as long as disparity values are
within natural limits. When the base distance becomes too high, naturalness judgments drop below the quality judgments. This effect seems to be content-dependent. Thirdly, the results of this experiment indicate that stereoscopic camera toe-in should be avoided if possible. A toed in camera formation can give rise to a so called puppet theatre effect.

One important remark has to be made regarding this experiment. IJsselsteijn et al. didn’t add artefacts to the pictures!

Elaborating on these findings, Seuntiëns, Meesters & IJsselsteijn (2003) started a research in 2003, in which they investigated the influence of various camera base distances, and various grades of JPEG compression (blockiness) with still pictures. The three levels of JPEG compression used were 1:1 (Original, no compression), 1:30 (Q30), 1:40 (Q20) and 1:60 (Q10), while the three camera base distances used were 0 (is 2D), 8 and 12 cm. The JPEG pictures were fed to the participant’s left respectively right eye in the following, mostly asymmetrical, combinations: 10_10, 10_20, 20_20, 10_30, 20_30, 30_30, 10_original, 20_original, 30_original and original_original.

The conclusions were: Firstly, blockiness had a negative influence on quality, sharpness and eye-strain but not on perceived depth. Secondly, with increasing camera base distance, the perceived depth and eye-strain increased too, but the increased base-distance, within the limits used in this experiment, had no influence on perceived sharpness. Thirdly, both sharpness and eye-strain were highly correlated with perceived quality.
A short experiment was done and issued by the same authors in 2005. The motivation for this experiment was the confusion with respect to the term ‘Image Quality’, which made no clear distinction between assessments of a same picture when it was produced in 2D or 3D, assuming that 3D always should be assessed as ‘higher quality’ as long as the 3D dimensions are within natural boundaries. Seuntiëns et al. (2006) were looking for a replacement for the word ‘Quality’ that should be more distinguishing when assessing the same images in a 2D or 3D-version (Seuntiëns et al., 2006).
For this experiment the artefact ‘Noise’ was implemented at six different levels and the new criteria the participants had to apply were ‘Naturalness’ and ‘Viewing Experience’. Seuntiëns et al. (2006) used a 20” Philips multi-view auto-stereoscopic display for this experiment. The most important graphs are shown in Figure 3.11 and Figure 3.12. A more detailed explanation of the criterion Viewing Experience will be given in Chapter 4.4. The conclusions from this experiment were: both the level of noise and dimension (2D or 3D) significantly affect viewing experience and naturalness and there were no interaction effects. Secondly, the added value of depth is taken into account both in viewing experience and naturalness, with the latter criterion more than with the former. According to Seuntiëns further investigations in combination with different 2D and 3D artefacts are needed to get more insight in the behaviour of these criteria (Seuntiëns et al., 2006).

The overview of recent investigations described above is far from complete, but suffices to understand the experiment described in this report.

The most important findings for our experiment were:

Stelmach et al. (1997) came to the conclusion that stereo depth in natural scenes did not improve image quality ratings compared with the same 2D-images, while Berthold (1997), under slightly different conditions, came to the opposite conclusion.

Naturalness is judged higher than quality as long as disparity values are within natural limits (IJsselsteijn et al., 2000). These conclusions came from a study where images without impairments were used.

JPEG-compression has a negative influence on quality, sharpness and eye-strain, but not on perceived depth (Seuntiëns et al., 2003).
With increasing camera base distance, the perceived depth and eye-strain increases too, but increased base-distance has no influence on perceived sharpness (Seuntiëns et al., 2003).

Sharpness and eye-strain are highly correlated with perceived Quality (Seuntiëns et al., 2003).
Both the level of noise and dimension (2D or 3D) significantly affect viewing experience and naturalness (Seuntiëns et al., 2006).

The added value of depth is taken into account both in viewing experience and naturalness, with the latter criterion more than with the former (Seuntiëns et al., 2006).

We can now almost define the research questions but in the following chapter we will first explain the criteria and artefacts used in our experiment.
4. Introduction to the four assessment criteria

In the experiments, described later, participants had to assess 2D and 3D-pictures on four criteria. A quarter of the number of participants had to use the criterion ‘Perceived Depth’, another quarter had to use the criterion ‘Perceived Image Quality’, another part ‘Viewing Experience’ and the last quarter ‘Perceived Naturalness’. This chapter will not explain what the various criteria do, but what they are. An impression of the effects of these criteria could already be read in Chapter 3 about recent researches. This chapter will shed more light on the nature of those criteria, and the question why observers may give one reaction to a criterion and a different reaction to another. The reasons why these four criteria will be used is dealt with in Chapter 6 where the Research questions are discussed.

4.1 Perceived depth

Why human beings experience depth and the psychological basis of that perception, the monocular cues and binocular cues, is already explained in the theoretical chapters above. As with all human senses the perception of the viewing sensation for any individual human being is not always the same. So a certain 3D-image with a particular depth can be perceived differently from one person to another. A reason for this difference in perception is the individual difference in acuity between the observers.

Nevertheless there are still differences in perception of depth and they have a psychological background. To quote Seuntiëns et al. (2006) and also IJsselsteijn (2004): ‘According to a number of studies, the addition of the third dimension (is depth) influences the psychological impact of viewers. This influence expresses itself in terms of experienced realism, power, and presence.’

In the experiment described in this report only three camera base distances will be used, and thus three different depth impressions of the test images, viz. for 2D-images (base is 0 cm.), and for the 3D-images 4 cm and 8 cm. Because these distances are close to the average base distance of 6.3 cm, it is expected that the measuring results will not be influenced by eye-strain.

4.2 What does Perceived Quality mean?

There is no unambiguous definition of image quality. Some investigations showed that quality assessments are especially influenced by the image technique used. So it can by all means happen that a black-white picture is judged by observers with the same quality rate as a colour picture. Obviously the quality assessments occur within the limits of the possibilities of the used technique. Nowadays most researchers don’t consider perceived image quality as a one-dimensional quantity but rather as a multi-dimensional one. Underlying aspects like sharpness, contrast, blur etc, form the elements of this concept.

In her experiment with blur impairments on 2D and 3D images, Berthold (1997) defined image quality as loosely and being based on a combination of sharpness, depth and colour. She considered image quality as a multidimensional parameter and obviously, looking at the results of her experiment which showed a large
variation, each participant defined it in his/her own way. According to her, even participants with different cultural backgrounds and other spoken languages played a role in the wide variation of the results. This showed how important it is to define the quality aspects accurately.

Based on the concept of multi-dimensionality Kayargadde & Martens (1995) did experiments in an attempt to define the underlying dimensions of the attributes in a multidimensional image quality space. Images often contain several kinds of impairments of degradations such as blur or noise. Those impairments influence the overall quality impression. In their experiment a multidimensional (MD) approach to image quality was used. They considered image quality to be mediated by perceptual attributes like sharpness and noisiness. The image is regarded as having a position in a MD perceptual space and in this space the attributes can be correlated with directions or vectors. Via two multidimensional scaling (MDS) techniques the experiment showed that those images can be placed in a two-dimensional perceptual space with their underlying vectors for sharpness and noisiness. The attribute vectors for blur (unsharpness) and noise are approximately orthogonal. This indicates that both attributes have almost no correlation and so there is only little interaction between the two attributes. The impairment vector is opposite to that of the quality vector and lies at an angle of about 30 degrees between unsharpness and impairment.

Engeldrum made a constructive contribution in his book about psychometric scaling by describing mathematical models for calculating direct and indirect interval scales (Engeldrum, 2000). He defines quality as "the integrated set of perceptions of the overall degree of excellence of the image." In the first part of his book he introduces the broader subject of image quality to establish a reference for psychometric scaling and highlights how psychometric scaling fits within the overall theme of image quality. He called this larger framework of image quality the Image Quality Circle. Image quality, according to him, can not be ignored. Market studies show that image quality, along with purchase factors like price, is one of the customer considerations for buying a product. Achieving good quality still requires much effort. He gives several reasons why image quality remains a difficult target to achieve. Firstly, the study of image quality has always been an objective image evaluation which involves physical measurement. This requires costly investments in measurement equipment and a connection or relationship between image quality and technology variables. But simply taking a lot of measurements does not bring us any closer to revealing this relationship. This is because the efforts generally fail to account for the human observer characteristics. Secondly, until recently it didn’t get attention of academies, although this is slowly changing now (Engeldrum, 2000, p. 7). Here at the University of Eindhoven we have been working on IQ for over two decades now. The study of image quality is a multidimensional and multidisciplinary subject, driven mostly by equipment and product development. For a lot of producers, image quality doesn’t form a part of the management policy. Thirdly, a lot of myths and mysteries surround image quality judgments by human beings. In the photographic industry the process of collecting this evaluation data is called subjective image evaluation. The prevailing myth about this data is that humans can’t be meters. Finally, there is a lack of a unified view on image quality and this has lead to confusion and chaos (Engeldrum, 2000, p. 8).
The image quality circle serves as a process mode and can simplify and focus product development activities and planning. See Figure 4.1 below.

![Image Quality Circle Diagram](image)

Fig 4.1: Engeldrum’s quality circle

The element ‘Technology variables’ contains the variables under control and describes the imaging product. Intended are variables such as dots per pixel, paper thickness, water fastness etc.

The element ‘Physical Image Parameters’ contains the so called objective measures of image quality. One can think of physical quantities such as optical density, spectral reflectance factor etc.

The ‘Customer perceptions’, mostly visible, are the attributes to be judged. With the “nesses” are meant colourfulness, brightness, sharpness etc. A large number of “nesses” are applicable to any set of images (Engeldrum, 2000, p. 12).

The element ‘Customer Quality Preference’ represents the assessment of the customer or the customer’s quality opinion of the image under evaluation.

The connection link ‘System models’ also called ‘Image models’ can be formulas, physical models, algorithms etc.

The function of the link ‘Visual Algorithms’ is to predict a value on a ‘ness’ scale from a suitable set of Physical Image parameters.

The input of the link ‘Image Quality Models’ are the values of the ‘nesses’ while the output is the image quality value.

All links between the elements are bi-directional, so can be passed through in both directions. Engeldrum gives as a general rule to take no more than one single link in a development process and not to use shortcuts by going directly from ‘Technology variables’ to ‘Customer perceptions’ or from ‘Physical Image Parameters’ to ‘Customer perceptions’ (or vice versa). Although this rule has its exceptions the reason not to take short cuts is that it is intuitive and unorganized.

The thin two-directional link between ‘Technology variables’ and ‘Customer Quality Preference’ presents a simple process, i.e. changing technology variables and assessing the outcome by observers. This way of working, for the same reason as with the short cuts, is not recommended because it doesn’t yield understanding of the quality components.
This circle only has meaning to people working in the image industry on e.g. cameras, printers, LCD-screens etc. Because the tests described in this report are stand alone tests and not part of a development process, only the last two steps ‘Customer Perceptions’ and ‘Customer Quality Preference’ are done by participants.

IJsselsteijn, Seuntiëns & Meesters (2002) make a distinction between three categories of technical variables in their ATTEST report, describing the state of the art of 3D broadcasting techniques: 1) content generating variables such as camera separation, focal length and convergence distance, 2) coding algorithm variables such as Q-parameter, bit-rate reduction by quantization and 3) display variables such as picture size, viewing distance, viewing angle, exposure and contrast. For customer perceptions they give attributes such as sharpness, depth and eye strain (IJsselsteijn et al. 2002)

With perceived quality as criterion the added value of depth in 3D pictures can clearly be recognized when observers have to assess artefact-free 3D as well as 2D pictures. However when the pictures are impaired with artefacts such as noise or blur etc, then the impairments have a great influence on the adjustments and the quality assessments are mainly determined by the impairment artefacts. If we consider depth, thus 3D, as an extension of the picture value, then we would like to find that extra value in our assessments, even when impairments are involved. This makes the perceived quality criterion less suitable for 3D-assessments. See also Seuntiëns (Seuntiëns et al., 2006).

In their survey of perceptual evaluations and requirements of 3D TV, Meesters, IJsselsteijn & Seuntiëns (2004) reached the conclusion that a multidimensional 3D-image quality model is required which has incorporated perceptual factors related to reproduced depth, 3D image impairments and visual comfort.

4.3 What does naturalness mean?

As Seuntiëns stated in one of his experiments: Image Quality may not be the most appropriate term to capture the evaluation processes associated with 3D images (Seuntiëns et al., 2006). The added value of depth in 3D images if compared to 2D images can clearly be identified as long as the pictures are not impaired with artefacts such as noise, blockiness or blur. When the image is impaired, the quality assessment becomes more dependent on the grade of the impairment and much less on the depth dimension. Therefore, the more general concept of Naturalness could be a good replacement instead of the Quality criterion in the assessment of 3D images.

Originally the term naturalness was introduced as a criterion for determining the perceived quality of colour reproductions where, according to Yendrikhovskij (1998) it was assumed that images of high quality should also be perceived as natural. This implies that quality and naturalness should have a very strong correlation.

Strothotte proposed a relationship between realism and naturalism. Naturalism is a term closely related to Naturalness and is often seen as photo realism, the ultimate goal for a pictorial representation of reality (Strothotte, 1998). According to Strothotte naturalism refers to the following two aspects: firstly, the quality of a pictorial representation evokes a visual impression as close as possible to that of the scene depicted. Secondly, realism is a property of a representation that gives an impression of a configuration of spatial objects that exists or could exist in the real world. From
the visual aspects, he lists the following most important components that contributes to the naturalism of a realistic picture:

1) Colour or colour realism: a picture can be coloured, uncoloured or unnaturally coloured.
2) Texture: a representational element may show the ordinary distributed texture, no texture or the wrong texture.
3) The shape of the element: the dimensions of the element can range from a photo realistic 3D projection to a sketch.
4) The configuration of the elements: this can be considered as natural or as unnatural (Strothotte, 1998).

IJsselsteijn et al. (2000) underlined the interesting relation between Quality judgments and Naturalness judgments. According to them the main difference between the two criteria as a subjective evaluation concept is that Naturalness refers to what participants perceive as a truthful representation of reality or perceptual realism, whereas Perceived Quality refers to a subjective preference scale. Research in the colour domain of Image Quality showed that observers managed to discriminate the two concepts and the experiment showed there was a slight but systematic difference between the two concepts. De Ridder et al. explained this difference stating that the observers had a preference for more colourful, and at the same time more unnatural pictures (de Ridder, Blommaert & Fedorovskya, 1995) & (de Ridder, 1996). The same effect between perceived quality and naturalness could also be noted in the area of stereoscopic depth perception. Recent research revealed that observers preferred a reproduction with more disparity, and thus more depth, but judged those 3D-images at the same time as slightly unnatural (IJsselsteijn et al. 2000).

In a study to assess perceived depth, naturalness and presence (another judgment criterion but not used in this thesis), IJsselsteijn, de Ridder, Hamberg, Bouwhuis, & Freeman (1998) showed that depth and naturalness were related but could vary independently, depending on scene content and image parameters. According to IJsselsteijn et al. (2002) these image parameters, i.e. stereo disparity and motion parallax, can give typical 3D-distortions, thereby influencing quality and naturalness. Conventional stereoscopic systems can cause unnatural visual effects because of the fixed pair of views which causes the absence of motion parallax information and whereby the perceived shape of stereoscopic objects is deformed if the viewing position of the object doesn’t coincide with the fovea of the eyes.

In the experiments of this thesis the observers were asked to assess images according to their perceived naturalness without giving a definition or description of the criterion.

4.4 What does Viewing Experience mean?

Viewing Experience is a new and recent shoot on the tree of evaluation criteria. This criterion is meant as another alternative for the criterion “perceived image quality” applied to 3D-image assessments. This is because research showed that the Quality criterion is not the most appropriate term. With exception of one study by Berthold (1997), research outcomes demonstrated that Image Quality could not account for
the added value of 3D compared to 2D when the picture was impaired with noise or blur. It is expected that the criterion “Viewing Experience” can give a better discrimination between evaluations of 2D and 3D images (Seuntiëns et al., 2003). To date only a few studies are known which have used this criterion for image perception. Seuntiëns et al. (2003) used it in one of their experiments as a generally known concept without giving a definition of the term or an explanation how the observers were instructed. They assumed that viewing experience accounts for image quality, but also incorporates depth perception, experienced realism, power and presence. Lambooij (2005) too, used this concept in his thesis and described it as a complex, multidimensional expression that reflects the users' general perception taking all possible characteristics of the application into account and summarizes a rather high amount of aspects such as image quality, presence and naturalness i.e. a degree of realism, naturalism and usefulness, visual discomfort, spatial inconsistencies and spatial distortions, sharpness, involvement and control, content and depth.

In the experiment described in this report the term is used as a generally known concept and participants were asked to judge images according to this concept without giving them a definition or a description.
5. Image artefacts used in these experiments

5.1 Introduction

Image artefacts, or image degradations, contribute to a large extent to the quality of an image. In Chapter 3 about recent research concerning 3D-aspects, effects of artefacts such as blur, noise and blockiness were clarified briefly. In this thesis we use two artefacts, namely blur and noise. The next paragraph describes the characteristics of blur, some models and their applications. In paragraph 5.3 noise and its characteristics will be treated.

5.2 Description of blur

Blur is defined with help of characteristics. Kayargadde & Martens (1994b) considers the following edge parameters: position, orientation, amplitude, mean value, and edge slope. Most mathematical models make use of a Gaussian blurring kernel situated at the blurred point, edge or line. An edge is modelled as a step function while the mathematical representation of the kernel is as follows:

\[ g(x, y, \sigma_b) = \frac{1}{2\pi\sigma_b^2} e^{-\frac{(x^2+y^2)}{2\sigma_b^2}} \]

Here \( \sigma \) is the (mostly unknown) spread of the blurring kernel while \( x \) and \( y \) are the spatial dimensions around the central point. See Figure 5.1

[Fig 5.1: Ideal blurred step. The thick arrows represent the gradient estimates while the thin arrows represent second derivative estimates.]

Blur can be introduced in several ways. Kayargadde & Martens (1994b) give a number of causes: camera defocus, low-pass filtering performed to obtain reduced data rates while coding, pre-filtering before sampling, etc. Elder & Zucker (1998) defines different types of blur. See Figure 5.2. They distinguish between focal blur (due to finite depth of field), penumbral blur (mostly the edge of a shadow) and shading (blur at a smoothed object edge).

In order to find a relation between physical derivation characteristics and perceived blur, Kayargadde & Martens (1994b) describe a method for estimating blurred edge...
parameters using polynomial transforms. In another research they quantify blur with the help of a blur index using these edge parameters (Kayargadde & Martens, 1996). To find accurate measures Elder et al. describe a mathematical model for edge detection and blur modelling where they involve local scales. That is to say, the local estimation depends upon the local structure of the edge at each point of the image (Elder & Zucker, 1998).

![Fig. 5.2: Elder's three types of blur](image)

Blur, or unsharpness, degrades the quality of natural scenes on television or in print and it affects the performance of medical or astronomical images when diagnostic analyses are involved. Blur can not be avoided and is inherent to image systems. In most cases it is considered as an undesirable component, although we sometimes use it, for example in the reduction of aliasing effects in images. Blur harms sharp transitions as edges and lines in an image. These components undergo extensive changes while the uniform regions almost remain unaltered.

Most of the applications concerning blur are with help of software programs. Image deblurring is one of the applications that needs an estimate of the blurring kernel. Software is used to estimate depth in images for tasks such as object recognition and scene interpretation. By measuring the size of the blurring kernel (caused by camera defocus) at different positions in the image, a depth map can be drawn for the entire scene. Applications for image quality prediction also exist. They require quantitative approaches for the blur parameters.

For the experiments of this thesis pictures were blurred with help of the software program MatLab. For a brief explanation of the used routine see Appendix 2.

### 5.3 Description of noise

Every physical system shows spontaneously fluctuations and is therefore stochastical by nature. In general these fluctuations are called noise. Because of the stochastic nature of noise, expressions and terms related to noise come from the field of statistics, although some noise parameters have their origin in the field of frequency analyses. The type and amount of noise is mostly described in colours. The colours describe the way in which noise energy is scattered over the image, or similar to that, the time domain. When the noise signal has a flat power spectral density (PSD) then this type of noise is so called white noise (see Figure 5.3). For example the frequency range from 40 Hz to 60 Hz contains the same amount of power as the
range between 4000 and 4020 Hz. In practice, a signal can be "white" with a flat spectrum over a defined frequency band. One speaks about pink noise if the frequency spectrum is flat in a logarithmic space. That means that pink noise will have equal power in the range from 40 to 60 Hz as in the band from 4000 to 6000 Hz (see Figure 5.4). With blue noise the power density increases with 3 Db per octave with increasing frequency (see Figure 5.5). More types of noise exist and have other colour names like brown, purple etc, all with their own definitions. (Wikipedia ,2006a & 2006b).

As blur, noise decreases the quality of natural scenes on television or in print and it affects the performance of images when diagnostic analyses are involved. In contrast to blur, noise doesn’t affect the sharp edges or lines to the same extent as blur, but it harms the uniform regions of an image more. Noise is a natural phenomenon and can arise in several ways. Most known are weakened received signals when the distance between transmitter and receiver is too long causing a low signal to noise ratio (SNR). But noise can also arise from electronic systems such as amplifiers, mostly thermic noise, or from camera systems, so called photo emission noise. Tape, the signal carrier for the band recorder or the video recorder is also a famous source for noise.

Applications of noise hardly exist, sometimes it is used in the graphic industry. Just as with blur, methods exist to quantify noise, for example for automatic quality determination. Kayargadde & Martens (1994) describes a method for estimating the standard deviation on the basis of local gradient energy.

The images for the experiments were provided with noise with help of the MatLab program. For a description of the used algorithm see Appendix 3.
6. Research questions

A number of issues arise from the enumerated results in paragraph 4. Firstly, in the tests and findings of IJsselsteijn no image impairments were introduced (IJsselsteijn et al., 2000). So the question rises what influence this could have on his results. Secondly, in all investigations mentioned the impact of noise distortion in images has hardly been examined. Two experiments are known so far and both have been discussed in paragraph 4. A test by Kayargadde & Martens (1995) is described where they used these image degradations for determining the values and directions of vectors in a multi-dimensional space representing ‘Quality’. The second experiment is a very recent investigation by Seuntes where he used the noise impairments to investigate the influence of noise on ‘Viewing Experience’ and ‘Naturalness’ (Seuntes et al., 2006). Thirdly, the standards ITU-R BT.500 (2002) and BT-1438 (2000) only use the term ‘Quality’. This term is very relative and is often used within the same applied technique. So 2D-images can have a very high quality just as 3D-images have. The assessment results are mainly determined by the impairments implemented and the ‘depth’ factor hardly plays a role within assessments using the ‘Quality’-term.

In this graduate study these questions will be investigated and the three investigation questions are defined as follows:

**Question 1**

*To what extent do existing image quality assessment methods, as described in the recommendations of ITU-R BT.500 and BT-1438, take into account the added value of depth? Are they sufficient for reliably judging stereoscopic television quality? Can other concepts, such as naturalness, value better the total assessments?*

**Question 2**

*From the literature it appears that two artefacts, which can influence the image quality in 2D-pictures, are underexposed or give conflicting results in stereoscopic television research. These aspects are:*

- Pictures provided with a high level of blur. Blur often can be seen in an image taken from a close distance where the background is vague or out of focus.
- Stereoscopic vision signals mixed with noise. This can occur when the 3DTV-receiver gets a weak antennae signal and the signal is analogy processed.

*Both noise and blur will be used as artefacts. What are their relative contributions to the overall appreciation of a 3D image or image sequence?*

**Question 3**

*From these tests, assessment data become available both from the noise and blur impairments. Are there relations between the perception of blur and the perception of noise within the various criteria and are these dependent on the image content?*
7. Design and description of the test

7.1. Introduction
In the next paragraph the test design will be described. Then the selection and the composition of the group participants for the four tests will be described. Then the design of the images with their artificial impairments will be explained and also the display equipment and the test room. In the fourth paragraph the images with their impairments are explained and at the end the test procedure with the appraisal forms will be presented.

7.2 Design of the test
A mixed design was employed where four groups of subjects was asked to rate one particular assignment attribute. Totally four dependent attributes or variables had to be tested. These four variables were ‘Perceived Depth’, ‘Naturalness’, ‘Image Quality’ and ‘Viewing Experience’. Every group tested only one dependent variable; this is the ‘in between-subject design’ part of the test. See also Figure 7.1.

![Figure 7.1: Test set-up for the stimuli](image_base64)

For every participant the test was divided in two independent parts. One half of the test was made up of rating images with blur artefacts and the other part of rating images with noise artefacts. This is the so-called within-subject design testing. See Figure 7.1.

The whole range of pictures was tested twice, resulting in 48 images per artefact. For every group with the same assignment attribute half of the participants started with the blur session, while the other half started with the noise session. Every set of 48 images was presented randomly to avoid order effects. This guaranteed a maximal counter balanced testing, free from possibly systematic sequence failures.

7.3 Participants to the test
In total 44 participants were selected to do the test. These 44 persons were split up in four groups of eleven participants. Each group had to assess one of the four dependent variables – Quality, Depth, Naturalness, or Viewing Experience--. See also Figure 7.1.
From the test persons twelve participants were female and thirty-two participants were male. The major part of the participants was aged between 18 and 30 years. Only two male participants were respectively 32 and 58 years old. The persons were selected on the basis of their eye acuity. For the selection procedure see the description in Chapter 7.6. All the persons had none or hardly any experience in assessing 3D-images and therefore can be considered as non-experts.

7.4 Test materials, test equipment and test room

7.4.1 Test materials

7.4.1.1 The stereoscopic images

The stereoscopic images used in this experiment were kindly provided by CCETT in France. Every picture had a resolution of 720 * 576 pixels. Figure 7.2 shows the ‘Bureau’ image consisting of a tailor’s dummy sitting behind a desk with some office material and a curtain in the background. The other picture is named ‘Playmobiles’ and shows a mountain landscape with some stone quarry workers. This picture is much more detailed than the ‘Bureau’ image and is shown in Figure 7.3.

![Figure 7.2: Stereoscopic picture with the tailor’s dummy](image1)

![Figure 7.3: Stereoscopic picture with the playmobiles figures](image2)

The pictures were taken with a professional stereoscopic camera in a toed-in configuration and various camera-base distances. The focal length was 20 mm and the convergence distance was 1.30 m respectively. For this test, three camera-base distances of 0 cm (monoscopic view), 4 cm, and 8 cm were chosen. The camera base distances of 4 and 8 cm resulted in 3D-images containing a pleasant amount of depth.

In total, six different pictures were thus available. Besides the original images also two types of artefacts, blur and white noise, both at three different levels, were...
introduced to the pictures. For the blurring session the two scenes had both three different depth levels of 0, 4 and 8 cm base distance and every image had one original plus three blurred images at different levels with standard deviations of 1, 1.5 and 5 pixels. See also the equation in figure 5.1, page 24. This gives a total of $2 \times 3 \times 4 = 24$ different images. For examples of blurred images see Figures 7.4 and 7.5.

![Figure 7.4: Stereoscopic picture ‘Bureau’ provided with blur](image1)

![Figure 7.5: Stereoscopic picture ‘Playmobiles’ provided with blur](image2)

In the same way 24 images became available for the noise session. Here the noise mean used was always 0 and the variances for the noise were 0.00125, 0.005 and 0.01 pixels respectively. For examples of pictures impaired with noise see Figures 7.6 and 7.7.

For both blur and nose the artefact levels were selected on the face and the artefacts were added to the pictures with help of the program MathLab. The blur routine was based on a Gaussian low-pass filter and applied separately to the three colours red, green and blue of the image. The noise routine makes use of a Gaussian white noise generator and was applied to the total image. For the program routines see Appendices 2 and 3.

![Figure 7.6: Stereoscopic picture ‘Bureau’ provided with white noise](image3)

![Figure 7.7: Stereoscopic picture ‘Playmobiles’ provided with white noise](image4)

7.4.1.2 Design of the image sequence test files

Every image had to be tested by the observer without reference image, the so-called single stimulus method. Between every image a mid-grey image with white text was shown in the test. This image had two functions: first the eyes could be restored from the impressions of the preceding image, this is the adaptation function. Secondly, the
number of the next stimulus was shown by the white text. See Figure 7.8. This minimized the possibility for inaccuracy errors on the ranking form.

For every participant a test file was formed with a unique random sequence of the images. Between every test image a grey image with the matching sequence number was placed. For 22 participants the slide show started with 48 blurred images followed by 48 pictures with the noise impairments. For the other 22 participants the test file started with the images with noise impairments. This was divided over the assignment groups as follows:

<table>
<thead>
<tr>
<th>Assignment Attribute</th>
<th>Participants started with Blur assignments:</th>
<th>Participants started with Noise assignments:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Dept</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Naturalness</td>
<td>5</td>
<td>6</td>
</tr>
<tr>
<td>Quality</td>
<td>6</td>
<td>5</td>
</tr>
<tr>
<td>Viewing Experience</td>
<td>5</td>
<td>6</td>
</tr>
</tbody>
</table>

Table 7.1 Numbers of participants starting with blur or noise impairments.

### 7.4.2 Test equipment

For the tests use was made of a PC with a high-resolution monitor with a size 17 inches and 1600 by 1200 pixels. On this monitor the left and right image of each stimulus (each 720 by 576 pixels) were placed side by side. The remaining pixels on the screen were black. See left picture of Figure 7.9
A stereoscope was attached to the monitor itself. The stereoscope was adjustable to a limited extent to the height of the observer’s head. The base distance of the stereoscope could be adjusted to the distance between the observer’s eyes with help of two screws. The advantage of the use of this type of stereoscope was the zero crossover between the two pictures. The distance from viewer to screen was about 40 cm. See also figure 7.9.

A disadvantage was that people with glasses couldn’t participate in the test. Another disadvantage was the vertical line that became visible by looking through the stereoscope, which was obviously caused by the lens and mirror system itself.

7.4.3 Test room

The test room is situated in Hall 1 of the IPO-building. In the test room the monitor was placed on a table with some backlight behind it. During the test the room was dark with exception of the backlight. The environment requirements of the test room fulfilled the requirements of ITU-R BT.500-11 (2002).

7.5 Measures

The appraisal method was based on the adjective categorical method as described in Recommendation ITU-R BT.500-11 (2002). A scale is divided in five ordinal values: excellent, good, fair, poor, bad. These five values are represented in a grading scale according to Figure 7.10.

In the appraisal form one grading scale was assigned to every stimulus. Because in the test every picture was shown twice, total 98 grading scales were taken up in the appraisal form (2 * 24 scales for the blur testing and 2* 24 scales for the noise setting). For an example of one sheet of an appraisal form see Appendix 7.

7.6 The test procedure

Before taking part in the test every participant was screened on visual acuity. This eye screen test existed of two parts: the first test was the traditional eye acuity test with help of the Landolt C-eye test Card. The minimum requirement here was that both eyes had an acuity of minimal 1.0. For an explanation of the eye acuity test see Appendix 4.

The second test was the Randot Stereotest. Here the ability of the participant to binocularly discern was tested. The ability for stereoscopic viewing had to be 30 arcsec or smaller at a viewing distance of about 40 cm. For various reasons 4 candidates had to be ejected because of eye problems (two of those had had a laser correction procedure to their eyes!). For an explanation of the Randot Stereotest see Appendix 5.

After succeeding the eye test, the test procedure was explained to the observer. Then the observer had to do a trial session of 6 images. These images were carefully selected and contained the whole range of blur and noise deviations of the image.
set, so the results of the trial were predictable to some extent. After the trial a short check was done to ensure that the participant had understood his/her task, possibly with an additional explanation afterwards.

After successfully fulfilling the trial session, the real test could be started. The test consisted of two parts, a part with 48 blurred images and a part with 48 images containing noise. For counter balancing the observers were numbered from 1 to 44 and the odd numbers started with the blur part while the even numbers started with the noise part. For an explanation of the Instruction Form see Appendix 6.

The images were represented one by one with every time a mid-grey text image in between. The time to assess the image could be determined by the participant himself, but after the next image was chosen, the previous image was not accessible anymore. On average, each participant needed about 35 minutes to do the test.

During the test the observer could indicate his/her adjustment by putting a pencil mark on only one of the horizontal/vertical crossings of the scale (Figure 7.10).

Later on, these ordinal test results were translated into numeric values. This ‘translation’ was as follows:

bad = 1; poor = 2; fair = 3; good = 4 and excellent = 5
8. Results of the tests
8.1. Introduction to the analyses

Overall ANOVA analyses were conducted per criterion (Depth, Quality, Viewing Experience and Naturalness) with help of the SPSS12 statistics program. In these analyses the following factors were used:

- Dependent Variable: Average values per scene assignment per subject
- Fixed factors: stimulus (‘Bureau’ and ‘Playmobiles’), the artefact rate and the base distance
- Random factor: subjects

In all ANOVA tests the SPSS calculation model was customized involving all fixed factors and all possible combinations of two fixed factors. Two way interactions were included.

The SPSS12 program doesn’t perform post hoc tests for less than three groups. For this reason it didn’t perform post hoc tests for the stimulus factor because it has only the two groups ‘Bureau’ and ‘Playmobiles’. To give more information about possible content influences of the images separate ANOVA tests were performed with only the Playmobiles images or the Bureau images involved. In the following paragraphs the results of the ANOVA tests will only be mentioned when the mean difference is significant, i.e. when \( p \leq 0.05 \).

For checking the model fit the average merged values per artefact (Blur and Noise), per criterion (Depth, Quality, Viewing Experience and Naturalness) and per stimulus were calculated with help of the ThurCatD program. The ThurCatD calculations gave a model fit for all the ordinal values in the figures: \( P > 0.05 \). For the explanation of the ThurCatD program see Appendix 8. For an example of a ThurCatD calculation see Appendix 9.

8.2. Average values with Blur distortions

The Figures 8.1 to 8.4 show the average ratings for the various test criteria. The values for the criteria (y-axis) are plotted against the grade of blur distortion (x-axis). The x-value 0 represents no distortion and 1 to 3 represent the three increasing distortion grades, where 1 is the lowest and 3 the highest grade. Distortion grade 1 stands for a blur deviation of 1.0, distortion grade 2 for a blur deviation of 1.5 and distortion grade 3 for a blur deviation of 5.0, see for more explanation Chapter 5.2 and appendix 2. For every base distance, measured in cm, a separate graphic line is presented. The error bar on every measuring point reflects the standard error of the mean of each data point.

Figure 8.1 (see next page) shows the average ratings for the Depth criterion: The figures show a clear effect of the various base distances, where with every base distance the depth perception increases. Also the effect of the blur impairment is clear, with increasing grade the perceived depth becomes smaller. The ‘between subjects’ overall analysis of variance (ANOVA) we performed showed a significant mean difference of camera base distance \( (F[2,246] = 154; p = 0) \) and of the impairment artefact \( (F[3,246] = 45.96; p = 0) \). See also Appendix 10, Figure 1 and Figure 2.
Figure 8.1

Figure 8.2 shows the average ratings for the Quality test criterion. The figures show no clear effect for the various base distances. For every grade of blur impairment the quality perception for the various base distances is almost equal. But the effect of the blur impairment is clear, with increasing grade the perceived quality becomes smaller. The performed overall ‘between subject’ ANOVA showed a significant main effect for the impairment artefact (F[3,246] = 221; p = 0), for the two stimuli Bureau and Playmobiles (F[1,246] = 24.075; p = 0) and a borderline case for the interaction ‘impairment artefact’ * ‘stimuli Bureau or Playmobiles’ (F[6,246] = 2.633; p = 0.051). There is no significant ‘between subject’ main effect for the camera base distances (F[2,246] = 0.417; p = 0.66).

Figure 8.2

For the average ratings of the Bureau images see Figure 1, Appendix 11 and for the Playmobiles images see Figure 2, Appendix 11. As can be seen in the graphs, the Image Quality judgments largely coincide with and without blur impairments.

Figure 8.3 (see next page) shows the average ratings for the ‘Viewing Experience’ test criterion. The figure shows a discriminating effect between the 2D-images and the 3D-images when there is no blur impairment. The performed overall ‘between subject’ ANOVA showed significant main effects for the impairment artefact (F[3,246] = 446; p = 0), the base distance (F[2,246] = 3.347; p = 0.037), the two images bureau and playmobiles (F[1,246] = 35.164; p = 0) and a significant interaction for the combination camera base distance and blur degree (F[6,246] = 2.567; p = 0.02).
Post hoc tests indicated here a significant difference between the 2D-images and the 3D images (p = 0.026 for base distances 0cm/4cm and p = 0.026 for base distances 0cm/8cm). Between the 4 cm and 8 cm base distance there was no significant mean difference at all (p = 1).

**Figure 8.3**

For the separate Bureau images between 0 and 8 cm there was no significance (p = 0.131), while between 4 and 8 cm there was also no significance (p = 0.312). See appendix 12, Figure 1. The main cause for the low overall mean difference between the base distances 4 cm and 8 cm was caused by the Playmobiles image assessments.

The performed ‘between subject’ ANOVA for the playmobiles showed no significant main effect for the base distance (F[2,120] = 1.416; p = 0.247). The post hoc tests didn’t indicate significant differences for all combinations of base distances (all p’s ≥ 0.10). See also Appendix 12, Figure 2.

As can be seen in the graphs of Figure 8.3, without impairment there is a clear distinction between 2D and 3D images. Within 3D (4 and 8 cm base) there is no distinction.

Figure 8.4 shows the average ratings for the Naturalness test criterion. The figures show a small but discriminating effect between the 2D-images and the 3D-images, especially when there is no blur impairment.

**Figure 8.4**
The ANOVA showed a significant main effect for the impairment artefact (F[3,246] = 226; p = 0), the base distance (F[2,246] = 13.037; p = 0), and the two images bureau and playmobiles (F[1,246] = 24.079; p = 0). Post hoc tests indicated a significant difference between the 2D-images and the 3D images (p = 0 for the base distances 0cm/4cm and 0cm/8cm). Between the 4 cm and 8 cm base distance there was no significant mean difference (p = 0.402). This effect could be found back in the separate Bureau and Playmobiles images.

For the separate Bureau images between 0 and 8 cm there was no significance (p = 0.131), while between 4 and 8 cm there was also no significance (p = 0.312). See appendix 13, Figure 1. The main cause for the small overall mean difference between the base distances 4 cm and 8 cm was caused by the Playmobiles image assessments.

The performed ‘between subject’ ANOVA for the playmobiles showed no significant main effect for the base distance (F[2,246] = 1.416; p = 0.247) and the post hoc tests didn’t indicate significant differences for all combinations of base distances (all p’s ≥ 0.10). See also Appendix 13, Figure 2.

The distinction for depth is better with the Naturalness criterion than with the Viewing Experience criterion. There is even a distinction when the pictures are impaired with blur rate 3. Then the significant mean difference between 0 cm and 4 cm base distance is p = 0.012 and the significant mean difference between 0 cm and 8 cm base distance is p = 0.01.

For the various figures of the separate images ‘Bureau’ and ‘Playmobiles’ with blur impairments see the following appendices:
Depth criterion: Appendix 10, Figures 1 and 2; Quality criterion: Appendix 11, Figures 1 and 2; Viewing Experience criterion: Appendix 12, Figures 1 and 2; Naturalness criterion: Appendix 13, Figures 1 and 2.
For a one page-overview of the four criteria assessments to a specific image with blur impairment see following appendices: Appendix 14: Bureau images and Appendix 15: Playmobiles.

8.3. Average values with Noise distortions

The Figures 8.5 to 8.8 show the average ratings for the various test criteria. Here also the values for the criteria (y-axis) are plotted against the grade of the impairment, in this case noise distortion (x-axis). The x-value 0 represents no distortion and 1 to 3 represent the three increasing distortion grades. Distortion grade 1 stands for a noise variance of 0.00125, distortion grade 2 for a noise variance of 0.005 and distortion grade 3 for a noise variance of 0.01, see for more explanation Chapter 5.3 and appendix 3. For every base distance, measured in cm, a separate graphic line is presented.

Figure 8.5 (see next page) shows the average ratings for the Depth criterion: Here the effect of the noise impairment is the same as the effect of the Blur artefact. With increasing base distance, the depth perception increases. Also the effect of the noise impairment is clear, with increasing grade the perceived depth becomes smaller. The performed overall ‘between subject’ ANOVA showed a significant main effect of
camera base distance \( (F[2, 246] = 232; p = 0) \) and the impairment artefact \( (F[3, 246] = 12.041; p = 0) \).

Fig. 8.5

The separately ‘between subjects’ analysis of variance (ANOVA) of the Bureau and Playmobiles images showed a significant mean difference of camera base distance and of the impairment artefact. See also Appendix 10, Figure 3 and Figure 4.

Figure 8.6 shows the average ratings for the Quality test criterion. The figures show no clear effect of the various base distances although the 2D assessments are always lower than the other assessments. For every grade of Noise impairment the quality perception for the various base distances is almost equal. But the effect of the noise impairment is clear, with increasing grade the perceived quality becomes smaller. The performed ‘between subjects’ overall ANOVA showed a significant main effect for the impairment artefact \( (F[3, 246] = 330; p = 0) \), for the camera base distances \( (F[2, 246] = 4.589; p = 0.011) \) and the two stimuli Bureau and Playmobiles \( (F[1, 246] = 22.767; p = 0) \). Post hoc tests indicated a significant difference between 0 cm/4cm \( (p = 0.03) \), a borderline case for 0cm/8cm \( (p = 0.055) \) and no significant mean difference for 4cm/8cm \( (p = 0.289) \).

Fig. 8.6

The separately calculated ‘between subjects’ ANOVA analyses of the Bureau images showed both a significant mean difference for the camera base distance and for the impairment artefact. The Playmobiles images showed a significant mean difference for the impairment artefact but no significant mean difference for the base distances
(F[2,220] = 1.025; p = 0.362). The post hoc tests didn’t indicate significant differences for all combinations of base distances (p ≥ 0.15). See appendix 11, Figure 3 and 4.

Figure 8.7 shows the average ratings for the ‘Viewing Experience’ test criterion. The performed ‘between subjects’ overall ANOVA showed a significant main effect for the impairment artefact (F[3,246] = 283; p = 0), for the camera base distances (F[2, 246] = 18.236; p = 0), the two stimuli bureau and playmobiles (F[1,246] = 13.892; p = 0) and a significant interaction for the combination ‘impairment artefact’ * ‘stimuli Bureau and Playmobile’ (F[3,246] = 5.616; p = 0.01). The figures show a discriminating effect between the 2D-images and the 3D-images. Post hoc tests indicated a significant difference between 0cm/4cm and 0cm/8cm (p = 0), and no significant mean difference for 4cm/8cm (p = 0.437).

Fig. 8.7

Those post hoc test results could be found back in the separate image ANOVA analyses. Both the Bureau images and the Playmobiles images showed no significant difference between 4 and 8 cm base (respectively p = 1 and p = 0.472). See appendix 12, Figure 3 and 4.

Fig. 8.8

Figure 8.8 shows the average ratings for the Naturalness test criterion. The figures show a discriminating effect between the 2D-images and the 3D-images. This is in agreement with the ‘between subjects’ overall ANOVA which showed a significant
main effect for the impairment artefact \( (F[3,246] = 181; p = 0) \), a significant main effect for the camera base distances \( (F[2, 246] = 8.543; p = 0) \) and a border line case for the interaction ‘impairment artefact * stimuli Bureau and Playmoble’ \( (F[3,246] = 2.616; p = 0.52) \). Post hoc tests indicated a significant difference between 0cm/4cm and 0cm/8cm \( (p = 0) \), and no significant difference for 4cm/8cm \( (p = 0.498) \).

This was both valid for the overall average values and for the separate values for the Bureau and the Playmobiles images. See appendix 13, Figure 3 and 4.

For the various figures of the separate images ‘Bureau’ and ‘Playmobiles’ with noise impairments see the following appendices:
Depth criterion: Appendix 10, Figures 3 and 4; Quality criterion: Appendix 11, Figures 3 and 4; Viewing Experience criterion: Appendix 12, Figures 3 and 4; Naturalness criterion: Appendix 13, Figures 3 and 4.

For a one page-overview of the four criteria assessments to a specific image with noise impairment see following appendices: Appendix 16: Bureau images and Appendix 17: Playmobiles.

8.4. Relationships between the various criteria with Blur distortions

Just as with the previous paragraphs, all considerations are based on mean values of the Bureau images and the Playmobiles pictures together.

Now we know the relations between the four assessment criteria and the impairments blur and noise, we are interested in the criteria relationships amongst themselves. With exception of the Depth criterion every possible combination of two criteria will be investigated. Two phenomena will be investigated more closely: which of the two criteria has the most discriminating effect with increasing grade of impairment. This conclusion can be derived from the angle of the regression line between the two criteria, where one criterion is placed along the x-axis and the other criterion along the y-axis. The slope of the angle is expressed as \( \beta_1 \), and the Y-intercept as \( \beta_0 \). Secondly, which of the criteria will be influenced more by the various base distances? These conclusions can be derived by comparing the assessment values as a function of the base distances.

---

![Fig. 8.9: Quality / Viewing Experience Scatterplot.](image)

![Fig. 8.10: Viewing Experience and Quality as function of base distances](image)
First the relationship between Quality and Viewing Experience will be discussed. See Figure 8.9, 8.10 (previous page) and Table 8.1: The correlation (determination coefficient $R^2 = 0.990$) between the Quality criterion and the Viewing Experience criterion is very high. Also the points of the three blur impairments groups are concentrated closely together. This confirms an earlier conclusion that the discrimination of Viewing Experience for impairments is practically equal to the Quality criterion. The discrimination for impairments is slightly higher for Image Quality than for Viewing Experience ($\beta_1 = 1.046$). For Viewing Experience there is only discrimination between 2D and 3D without blur impairment. In all the other unimpaired or impaired situations there is no discrimination within 3D (4 and 8 cm base).

<table>
<thead>
<tr>
<th>Table 8.1: Quality / Viewing Experience relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation factor: $R$</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>0.995</td>
</tr>
</tbody>
</table>

After what is said about the relationship between Image Quality and Viewing Experience we can be short about the relationship between Image Quality and Naturalness. See Figure 8.11, 8.12 and Table 8.2.

<table>
<thead>
<tr>
<th>Table 8.2: Quality / Naturalness relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation factor: $R$</td>
</tr>
<tr>
<td>-------------------------</td>
</tr>
<tr>
<td>0.980</td>
</tr>
</tbody>
</table>

Here, too, the correlation between the Image Quality and the Naturalness is very high (determination coefficient $R^2 = 0.961$) The discriminating factor for the impairments is somewhat higher for Image Quality than for Naturalness ($\beta_1 = 1.189$). The depth discrimination is somewhat better for Naturalness than for Quality, although this is not true for all impairment values according to Figure 8.12.

The last possible relationships is that between Viewing Experience and Naturalness. See Figure 8.13, 8.14 and Table 8.3 on next page. Here the correlation between the two criteria is also very high (determination coefficient $R^2 = 0.978$). The discriminating factor for the impairments is somewhat higher for Viewing Experience than for
Naturalness ($\beta_1 = 0.875$) and the discrimination for depth is better for Naturalness than for Viewing Experience with higher blur impairments (see Figure 8.14).

Table 8.3: Naturalness / Viewing Experience relation

<table>
<thead>
<tr>
<th>Correlation factor: R</th>
<th>R Square</th>
<th>$\beta_0$ (constant)</th>
<th>$\beta_1$ View.Exp.:</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.989</td>
<td>0.978</td>
<td>0.500</td>
<td>0.875</td>
</tr>
</tbody>
</table>

8.5. Relationships between the various criteria with Noise distortions

After what is said about the relations with blur impairments, the explanations regarding the relations with Noise impairments can be short.

First the relationship between Quality and Viewing Experience will be discussed (see Figure 8.15, 8.16 and Table 8.4 on see next page): The correlation between the two criteria is very high (determination coefficient $R^2 = 0.974$). The discriminating factor for the impairments is somewhat higher for Quality than for Viewing Experience ($\beta_1 = 1.112$). The discrimination of depth for the various Noise impairments gives slightly
better results for the Viewing Experience assessments except for the pictures with the highest grade of Noise impairment. In general, according to Figure 8.16 the assessments for the Viewing Experience are slightly better than the Quality assessments.

<table>
<thead>
<tr>
<th>Table 8.4: Quality / Viewing Experience relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation factor: R</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>0.987</td>
</tr>
</tbody>
</table>

For the relationship between Quality and Naturalness see Figure 8.17, 8.18 and Table 8.5: Here, too, the correlation is very high (determination coefficient $R^2 = 0.957$). The discriminating factor for the impairments is somewhat higher for Quality than for Naturalness ($\beta_1 = 1.182$). The discrimination of depth gives better results for the Naturalness assessments than for the Quality assessments, see Figure 8.18.

![Fig. 8.17: Quality / Naturalness Scatterplot](image1)

![Fig. 8.18: Naturalness and Quality as function of base distances](image2)

<table>
<thead>
<tr>
<th>Table 8.5: Quality / Naturalness</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation factor: R</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>0.978</td>
</tr>
</tbody>
</table>

For the relationship between Viewing Experience and Naturalness see Figure 8.19, 8.20 (on next page) and Table 8.6. Again the correlation is very high (determination coefficient $R^2 = 0.986$). The discriminating factor for the impairments is somewhat higher for Naturalness than for Viewing Experience ($\beta_1 = 0.935$). The discrimination of depth gives slightly better results for the Naturalness assessments.

<table>
<thead>
<tr>
<th>Table 8.6: Naturalness / Viewing Experience relation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Correlation factor: R</td>
</tr>
<tr>
<td>----------------------</td>
</tr>
<tr>
<td>0.993</td>
</tr>
</tbody>
</table>

When the images are not impaired, the values for both criteria are exactly equal but when the Noise grades increase some of the Viewing Experience assessments values drop more than the values of the Naturalness assessments, see fig. 8.20.
8.6. Relationships between the noise and the blur impairments

First of all, the three grades of the artefacts noise and blur were determined at face value using trial and error. In this way the noise variance and the blur spread were quantified. No method exists to compare those not related quantified numbers between the blur and noise grades, possible relations can only be made with the restriction that these relations don’t have a collective statistical base and are only valid for the two pictures with the given depths and the given impairment rates. The results cannot be generalized to other pictures.

Figure 8.21 (Bureau) and Figure 8.22 (playmobiles) show graphs from the assessments of the “Perceived Naturalness” data.

To discover the trends in blur and noise, trend lines of the mean values for blur and noise are added to the figures. The mean values are values calculated for every impairment grade over the combined three depth values. The trend lines can be expressed in a slope of the trend line angle ($\beta_i$) and the mean value for grade 0 (Y-intercept).
Comparing the slope values it shows that the decline as function of the impairments, although somewhat more for the noise, can be considered as almost equal (respectively -0.775 and -0.876).

Doing the same with the “Playmobiles” data gives Figure 8.22 and Table 8.8.

Here the picture is different. The decline for the blurred Playmobiles image is much higher than for the noise impaired Playmobiles image (respectively -0.897 and -0.688). Crossed comparisons give the following results: if we compare the blur impaired Bureau images (Table 8.7: -0.775) with the blur impaired Playmobiles images (Table 8.8: -0.897) and if we compare the noise impaired Bureau images (Table 8.7: -0.876) with the noise impaired Playmobiles images (Table 8.8: -0.688) then we can conclude the following trends for the criterion Naturalness: because the Playmobiles images are much more detailed than the Bureau images the impact of blurring for the more detailed Playmobile pictures is higher than for less detailed Bureau images with relatively bigger surfaces in the image, and vice versa, the information of the more detailed Playmobiles picture will be better maintained with noisy detailed pictures than with blurred detailed pictures.

These conclusions are also valid for the criteria “Quality” and “Viewing Experience”. See for the graphics and the figures respectively Appendix 18 and Appendix 19.
The results of the “Depth” criterion are respectively presented in Figure 8.23 and Table 8.9 for the Bureau images and in Table 8.10 and Figure 8.24 (see next page) for the Playmobiles images.

![Depth Average scores Bureau](image)

**Figure 8.23**

<table>
<thead>
<tr>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blur</strong>:</td>
<td>3.838</td>
</tr>
<tr>
<td><strong>Noise</strong>:</td>
<td>3.811</td>
</tr>
</tbody>
</table>

The participants who had to assess the test images on the criterion “Depth”, only focussed on depth and had less eyes for other quality aspects. If we analyse these data in the same way as for the “Naturalness” criterion then we come to the following results:

The decline for the blurred Bureau images is almost equal to the noise impaired Bureau images (slopes respectively -0.465 and -0.455) and the decline for the blurred Playmobiles images is slightly higher than the noise impaired Playmobiles images (slopes respectively -0.267 and -0.188).

Crossed comparisons give the following results: the decline of the blur impaired Bureau images (Table 8.9: -0.465) is higher than the blur impaired Playmobiles images (Table 8.10: -0.269). This is also valid for the noise impaired Bureau and Playmobiles pictures (respectively -0.455 and -0.188).

<table>
<thead>
<tr>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Blur</strong>:</td>
<td>3.522</td>
</tr>
<tr>
<td><strong>Noise</strong>:</td>
<td>3.479</td>
</tr>
</tbody>
</table>

If we compare the slopes for the blur impairments we see that the slope is much smaller for the Depth criterion (Tables 8.9 and 8.10) than with the Naturalness criterion (Tables 8.7 and 8.8). This is valid for both the Bureau and the Playmobiles images. With respect to the noise impairments the result is even more surprising. The slope for the noise impairments is much smaller than with the “Naturalness” criterion.
It can be concluded that the depth information is better maintained for the impairments with the “Depth” criterion than for the impairments with other criteria.

Figure 8.24
9 Conclusions from the experiments

In the first part of this chapter the experiments with the blur impairments will be evaluated and in the second part the experiments with the noise impairments. At the end the general conclusions and, following from these conclusions, recommendations and answers to the research questions will be given.

9.1 Conclusions from the Blur experiments

The results show that blur significantly affects all four criteria, Perceived Depth, Perceived Image Quality, Viewing Experience and Perceived Naturalness. The values of these various assessment criteria all decrease when the blur grade becomes higher. This conclusion is more or less similar to a remark by Seuntiëns who stated that observers anchor their judgments on the most salient features (Seuntiëns et al., 2006). In this test the most salient feature is the blur level.

The other conclusion is that the observers of the Depth assessments clearly recognize the depth information. The graphs show three clearly distinct lines for the three depth values (base distance 0, 4 and 8 cm). The difference of the assessment values between 2D images and the 3D images with the lowest depth information (base distance 4 cm) is greater than the distance between the images with 4 and 8 cm base distance. This is valid for all blur gradations (Appendix 10).

In this experiment it became clear that the Quality criterion is not a good qualifier for the images with the various depths. In the test the Quality criterion recognizes clearly the blur gradations but, obviously depending on image content, it doesn’t recognize or rather recognize unsatisfactorily depth (see Appendix 11, Figure 1 and 2). This conclusion is similar to the investigations of Stelmach et al (1997, 1998), but contrary to the experiments of Berthold (1997). In the discussions in the previous chapter (Figure 8.9, 8.10, 8.11 and 8.12) about the relationship of Quality and Viewing Experience and the relationship of Quality and Naturalness, the Quality criterion was also less satisfactory compared to the other criteria. This suggests that the Image Quality criterion still is and still remains a controversial criterion and can better not be used when 3D-images are involved if the aim is to be sensitive to the potential added value of stereo.

A comparison between the criteria Viewing Experience and Naturalness is in favour for the criterion Naturalness. See Appendix 14 Figure 3 and Figure 4 and Appendix 15 Figure 3 and Figure 4. The reason is that the Naturalness criterion recognizes the depth clues better than the Viewing Experience criterion when the blur impairment increases in the images. This difference is larger for the Bureau images than for the Playmobiles images, so the criterion is depending on the image content. Seuntiëns et al. came to the same conclusion that Naturalness is preferred above Viewing Experience (Seuntiëns et al., 2006).

9.2 Conclusions from the Noise experiments

Here, too, the results show that the noise levels significantly affect all four criteria. The various assessment values all decrease when the noise grade increases. This is in line with the remark of Seuntiëns who stated that observers anchor their judgments on the most salient features, in this case noise (Seuntiëns et al., 2006).
The other conclusion is also valid here: the observers of the Depth assessments clearly recognize the depth information and the difference of the values between 2D images and the 3D images with the lowest depth information (base distance 4 cm) is greater than the distance between the images with 4 and 8 cm base distance. This is valid for all noise gradations (Appendix 10).

Also, with noise impairments the Quality criterion is not a good qualifier for the same reasons as with the blur impairments. The criterion recognizes clearly the noise gradations but, and again obviously depending on image content, it doesn't recognize or rather recognize unsatisfactorily depth (see Appendix 11, Figures 3 and 4). In the discussions in the previous chapter (Figures 8.9 and 8.10) about the relationship of Quality and Viewing Experience and the relationship of Quality and Naturalness (Figures 8.11 and 8.12), the Quality criterion was also inferior in both cases to the other criteria because the other criteria better recognize the depth information.

A comparison between the criteria Viewing Experience and Naturalness gives a more or less equal result. The distinction is that the difference between 2D and 3D for Naturalness is bigger than for Viewing Experience but the 3D perception for Naturalness is rather vague. For some values of Noise impairment images with 4 cm base were assessed higher than images with 8 cm base. See Appendix 16 Figure 3 and Figure 4 and Appendix 17 Figure 3 and Figure 4.

9.3 Conclusions with respect to the relations between Noise and Blur

A lot has already been said about this during the analysis of the relations between the two impairments, so we can repeat briefly the conclusions we made there. For the criteria “Naturalness”, “Viewing Experience” and “Quality” it showed that for every criterion the judgments decline is more or less equal with increasing impairments for the blurred and noised images with relatively large surfaces (Bureau), that is to say the less detailed images. When there is little difference between the judgments of the two impairments than it is in favour of the blurred impairments. The more detailed images (Playmobiles) give an opposite picture. The judgment decline for the blurred images is higher than for the noise impaired images. So the impact of blurring is worse than the impact of noise for a detailed image when the impairments are equally grated.

The results with the criterion “Depth” is different compared with the other three criteria. The “Depth” assessments were higher for the more impaired images than for the corresponding assessments for the three other criteria, this is valid for both the Bureau and the Playmobiles images. Compared with the assessments of the blur impairments, the assessments for the noise impaired images were higher for both the Bureau images and the Playmobiles images. So it can be concluded that the depth information is better maintained with the “Depth” criterion and the noise impaired images have higher assessments than the blur impaired images. Obviously the noise impairment is more independent for detailing than the blur impairments when “Depth” is concerned. This effect can be caused because the noise component is not correlated to specific parts of the image and gives therefore more info of sharp lines and edges. On the other hand, blur is just concentrated on sharp edges and lines and influence the depth information in a negative way.
9.4 Answers to the questions

Combining the results from the previous chapters we come to the following answers:

**Answer to Question 1; Quality concept:** The results here are opposite to the findings of Berthold (1997) and confirm the results of Seuntiëns (2003) and Stelmach (1998) that the “Quality” criterion is not sufficient for assessment of 3D-images. As can be seen in the results and the conclusions the term ‘Image Quality’ is not sufficient when 3D-images are involved and the depth information is important for the concerning application. In that case, the term “Quality” should be replaced by the term “Naturalness”. That means that in the recommendation BT-1438 in which 3D is treated, the word ‘Quality’ should be replaced by the word ‘Naturalness’ at those places where the depth information, related to 3D, is important.

**Answer to Question 2; Influence of artefacts Blur and Noise on assessments:** The experiments show that the artefacts Blur and Noise both influence the assessments in a negative way. The assessment values for the criterion ‘Depth’ decreases with increasing impairment but not that dramatically as with the other three criteria. The assessment values with the criterion ‘Quality’ almost didn’t react to depth information but it recognized the various impairment grades clearly. The assessment values for ‘Viewing Experience’ and ‘Naturalness’ gave clear information about the blur and noise impairment grades. The depth perception for both criteria was much better than with the ‘Image Quality’, from which ‘Naturalness’ gave the best results as function of the impairment grade.

**Answer to Question 3; relation between the artefacts Blur and Noise:** with the given blur and noise rates and the given images ‘Bureau’ and ‘Playmobiles’ the decline of the judgments with increasing impairments is more or less equal for the blurred and noised images with relatively large surfaces. For the more detailed images the judgment decline for the blurred images is higher than for the noise impaired images. So the impact of blurring is bigger than the impact of noise for a detailed image. This is Valid for all criteria except the criterion Depth. With the “Depth” criterion the noisy images better maintain the depth component than the blurred images. The depth information is more independent from the detail grade of the images than its blurred counterpart.
10. Remarks and points for discussion

Criterion of Perceived Naturalness: concerning the tested criteria; a lot is already said about the criterion ‘Image quality’ in relation with 3-D. Because of the dubious results with this criterion, researchers are looking for a replacement and in this thesis the replacements ‘Viewing Experience’ and ‘Naturalness’ were investigated with as outcome that ‘Naturalness’ has the best credentials. The question arises if this criterion can depend on the content of the image. For example in these experiments use was made of an image with a shop-window dummy sitting behind a desk and that is not really natural. What would be the outcome if this rather artificial image was compared with a similar scene where a real man was sitting behind the desk? Would the outcome of both images be the same when the same parameters of depth and impairments are used? This question can be a nice follow up of this research.

Stirring artefacts: still images were used in these experiments, also the artefact noise in the pictures was stationary. Therefore one wonders what results the artefact ‘not-stationary noise’ will produce. Maybe the question can be put in a more general way: what can be the results of stirring artefacts like moving noise or moiré on the assessments? Moiré mostly arises when the dots of the display unit and the content of the (moving) image interfere with each other.

Frame of reference: the ‘Single-stimulus method’ used gave rise to some complaints. Because the sequence of the images was determined randomly, some participants lost their frame of reference when too many pictures with the same grade of blur or noise were displayed in a row.

Learning effect: one complaint of a learning effect was noticed. The observer told me that, after she was halfway through the test and had seen half of the amount of pictures, she exactly knew that there were only four different depths in the pictures, independent of the grade of blur or noise.

Equipment used: the advantage of the stereoscope used was the excellent division between the two stereoscopic pictures, cross-over was not visible. A disadvantage of this stereoscope was that people who wear spectacle could not participate in the test. Also the mirror system of the stereoscope was not optimal. The picture was somewhat trapezium shaped, not perfectly rectangular. Another disadvantage was the very small, but noticeable vertical line that became visible by looking through the stereoscope, obviously caused by the imperfect lens and mirror system itself. All observers could do the tests while ignoring this phenomenon. No complaints about this phenomenon were heard.
11 References


12 Consulted Internet sites

Figure 2.1: http://en.wikipedia.org/wiki/Eye#Varieties_of_eyes
Figure 2.2: http://ccrs.nrcan.gc.ca/resource/tutor/stereo
Figure 2.3: http://webvision.med.utah.edu/
Figure 2.4: http://www.acs.appstate.edu/
Figure 2.5: http://www.2dcurves.com/3d/3dh.html
Figure 2.6: http://www.science.mcmaster.ca/Psychology/psych3j03.old/lectures/depth1/sld009.htm
Figure 2.8: http://www.ankrumassociates.com/articles/
Figure 2.9: http://axon.physik.uni-bremen.de/research/stereo/index.html
Figure 2.10: http://www.dartmouth.edu/~matc/math5.geometry/unit14/unit14.html#mathematics
Figure 2.11: http://www.tomshardware.com/2004/03/04/eye_candy/page3.html
Figure 2.12: http://www.du.edu/~jcalvert/optics/stereops.htm
Figure 2.13: http://www.ise.stanford.edu/class/psych221/projects/03/jgin/history.htm
Figure 2.17: http://axon.physik.uni-bremen.de/research/stereo/index.html
Figure 5.3: http://en.wikipedia.org/wiki/White_noise
Figure 5.4: http://en.wikipedia.org/wiki/Colors_of_noise
Figure 5.5: http://en.wikipedia.org/wiki/Colors_of_noise

Appendix 4, Figure 1: http://www.yorku.ca/eye/thejoy, Kaiser P. K. 'The joy of visual perception',
Appendix 4, Figure 2: http://en.wikipedia.org/wiki/Eye_examination
Appendix 4, Figure 4: http://en.wikipedia.org/wiki/Landolt_C
Appendix 5, Figure 4: http://haag-streit-uk.com/ophthalmic/catalogue/ HAAG-STREIT UK, Ophthalmic Product Catalogue.
Appendix 20, Figure 1: http://www.funsci.com
### Appendix 1: Summary of monocular distance cues

<table>
<thead>
<tr>
<th><strong>Interposition</strong></th>
<th>The blue circle is closer to the observer because it covers partly the red circle.</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Size</strong></td>
<td>When two objects have the same size, we perceive the object that casts a smaller retinal image as farther away than the object that casts a larger retinal image. This cue is known as relative size.</td>
</tr>
<tr>
<td><strong>Texture gradient</strong></td>
<td>The cue is based on the fact that objects seen at greater distances, appear to be smoother and less textured</td>
</tr>
<tr>
<td><strong>Linear perspective</strong></td>
<td>When objects of known distance subtend a smaller and smaller angle, it is interpreted as being further away. Parallel lines converge with increasing distance.</td>
</tr>
<tr>
<td><strong>Atmospheric perspective</strong></td>
<td>Due to the scattering of blue light in the atmosphere distant objects appear bluer. Thus distant mountains appear blue.</td>
</tr>
<tr>
<td><strong>Shading</strong></td>
<td>Light and shadow provide information about an object's dimensions and depth. Our visual system assumes the light comes from above, and a different perception is obtained if the image is viewed upside down.</td>
</tr>
<tr>
<td><strong>Height</strong></td>
<td>Height is a monocular cue to distance and depth and based on the fact that the closer an object to the horizontal plane is, the farther away it appears. Cloud B is farther away than cloud A because the distance of B to the horizon is smaller. For the same reason object C seems to be farther away than object D.</td>
</tr>
</tbody>
</table>

Note: The explanations of these monocular cues are brief and in some cases need supplementation.
Appendix 2: Example of the used MatLab routine for blurring the images

Use is made of the toolkit for the Video and Image processing possibilities of MatLab. The most important routine is fspecial('gaussian',[7 7],x). This function is a special filter (fspecial) that uses a Gaussian low-pass filter. Fspecial('gaussian', [7, 7], x) builds a Gaussian filter matrix of 7 rows and 7 columns, with a standard deviation of x. The filter is a 7x7 matrix, which is applied to input pixels, one at a time, resulting in an output pixel in the filtered image. The filter is used to the three colours of the image, red, green and blue.

In the experiments of these thesis three different standard deviations of 1, 1.5 and 5 pixels are used.

```matlab
% introducing gaussian blur in images.
imfnam = 'Bureau07l.ppm';
RGB = imagread(imfnam);

% Blur factor 1.0
[fdir,fnam,fext] = fileparts(imfnam)
h = fspecial('gaussian',[7 7],1)
RGB1(:,:,1) = filter2(h,RGB(:,:,1));
RGB1(:,:,2) = filter2(h,RGB(:,:,2));
RGB1(:,:,3) = filter2(h,RGB(:,:,3));
%save a blurred file
outputfile = [fdir fnam '_blur_1-0l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);

% Blur factor 1.5
[fdir,fnam,fext] = fileparts(imfnam)
h = fspecial('gaussian',[7 7],1.5)
RGB1(:,:,1) = filter2(h,RGB(:,:,1));
RGB1(:,:,2) = filter2(h,RGB(:,:,2));
RGB1(:,:,3) = filter2(h,RGB(:,:,3));
%save a blurred file
outputfile = [fdir fnam '_blur_1-5l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);

% Blur factor 5.0
[fdir,fnam,fext] = fileparts(imfnam)
h = fspecial('gaussian',[7 7],5)
RGB1(:,:,1) = filter2(h,RGB(:,:,1));
RGB1(:,:,2) = filter2(h,RGB(:,:,2));
RGB1(:,:,3) = filter2(h,RGB(:,:,3));
%save a blurred file
outputfile = [fdir fnam '_blur_5-0l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);
```
Appendix 3: Example of the used MatLab routine for adding noise to an image

The tool kit for Video and Image processing possibilities of MatLab was used to add noise to the images. The most important routine here is RGB1 = imnoise(RGB,'gaussian',x,y). This function adds noise to the picture with the name 'RGB', and RGB1 becomes equal to RGB although with the added noise. The subroutine uses the noise generator for Gaussian white noise 'gaussian' and it uses two variables x and y which values have to be defined. The variable x is the mean of the noise, mostly zero, while y represents the variance. With the size of the variance the grade of noise can be defined. MatLab has also other noise generators like the 'salt & pepper' or the 'speckle' noise generator.

In the experiments of these thesis the mean of the noise added to the images is always zero, while the three variance sizes are 0.00125, 0.005 and 0.01.

```matlab
% Inlezen plaatje
imfnam = 'Bureau08l.ppm';
RGB = imagread(imfnam);

% introducing white gaussian noise 0.00125
[fdir,fnam,fext] = fileparts(imfnam);
RGB1 = imnoise(RGB,'gaussian',0,0.00125);
%save a blurred or noise file
outputfile = [fdir fnam '_Noise-0125l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);

% introducing white gaussian noise 0.005
[fdir,fnam,fext] = fileparts(imfnam);
RGB1 = imnoise(RGB,'gaussian',0,0.005);
%save a blurred or noise file
outputfile = [fdir fnam '_Noise-0500l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);

% introducing white gaussian noise 0.01
[fdir,fnam,fext] = fileparts(imfnam);
RGB1 = imnoise(RGB,'gaussian',0,0.01);
%save a blurred or noise file
outputfile = [fdir fnam '_Noise-1000l' fext];
fprintf('Write file: %s
',outputfile);
imagwrite(RGB1,outputfile);
```
Appendix 4: Measuring methods of visual Acuity

Visual acuity refers to the clarity or clearness of the vision of a person, or in a scientific way; refers to the person’s acceptable spatial resolution. There are different types of eye acuity from which two types are important: Firstly, the Recognition acuity: this acuity requires the viewer to name the target stimuli. This is the most well-known acuity test, as most people will have seen the eye test card in the doctor’s office. Secondly, the Resolution acuity: this refers to the ability to perceive a separation between the elements of a pattern. The most well-known form of this type is the Landolt ring, also used with the eye acuity test cards.

The visibility of a target varies not only with its size but also with its distance to the observer. The visual angle is formed by distance and target size and is expressed by following formula: \( \tan \frac{\theta}{2} = \frac{S}{2D} \) (See Figure.1).

![Diagram](http://www.yorku.ca/thejoy)

Fig. 1: Figure taken from site [http://www.yorku.ca/thejoy](http://www.yorku.ca/thejoy)

The advantage of the visual angle is that it can be expressed in only one singular quantity. The optimal values for the five various acuities for human vision are not equal but for both Recognition and Resolution acuity they are; 30 sec of arc (Schiffman, 2001).

The Snellen eye chart

The Snellen eye card is the most used chart and many people will have seen this card in the doctor’s office. This card is a recognition acuity type card because the letters have to be recognized. With the largest letters at the top, line by line the letters become gradually smaller. Dr. H. Snellen was a 19\(^{th}\) century ophthalmologist who designed the card. Each letter is scientifically designed in a way that at the appropriate distance the total letter encloses a visual angle of 5 degrees and each component of the letter encloses an angle of 1 minute. The letters must be read with one eye open and the other closed.

The Snellen fractions 20/20, 20/30 etc., are measures of sharpness of sight. If a person sees 20/20 at 20 feet distance from the chart, it means that this person can read letters at 20 feet what another person with average sight can also read at 20 feet distance (is...
100 % efficiency). When a person can only read 20/40 it means that he/she can read letters at 20 feet distance what another person can read at 40 feet distance (is 85 % efficiency), etc. There are animals, like the hawk, which may have a vision of 20/2! The relation between the Anglo-American values and the European values are linear (20 feet $\approx 6$ m., or $20/20 = 6/6$)

When the vision of a person is equal or less than 20/200, and there is no possibility of correction with help of surgery, laser or glasses, then the person is legally blind in Australia.

**The Landolt C eye test chart**

The Landolt C ring or Landolt broken ring is a standardized symbol for testing vision. This type of testing is a resolution acuity type of testing. The ring consists of a ring with a gap. The thickness of the ring is, just as the gap, $\frac{1}{5}$th of the outside diameter. The gap can be in the top, right, left and down position and in the $45^\circ$ position. The test person has to decide on which side the gap is.

![Fig. 3: A standard Landolt C optotype with the proportions marked by the grid.](Chen et al., 2005)

![Fig. 4: Landolt C optotypes in various sizes and orientations](Chen et al., 2005)

The read distance according to this card was 5 m and for the range of the lines with the rings see the table. The participants on the tests described in this report, had to do their eye acuity test with this type of chart.
Appendix 5: RANDOT Stereo test

Testing the ability to pure binocularly discernment gives many difficulties. Not only size, contrast and distance between the subjects influence the judgment but also the monocular clues may influence the outcome of the test.

Two groups to measure stereopsis exist, the contour stereo test and the random dot stereo test.

For the contour test is it sufficient to make use of local stereopsis to evaluate the horizontally disparate stimuli. An example of this test is the Titmus Fly Stereo test. In this test the horizontal disparity is presented via the vectographic technique (Fricke & Siderov, 1997).

Random dot stereograms require global stereopsis because for evaluation the correlation of corresponding points over a large retinal area is necessary. The random dot stereograms were invented and used by Bela Julesz (1960) to eliminate the monocular cues. Without contours, depth perception can only be observed when binocular fusion occurs.

Examples used in the clinic are the Lang Stereo test, the Frisby Stereo test, the Randot E Stereo test, and the Randot Stereo test.

The Lang Stereo test uses a panographic technique to present disparity and therefore needs no filters. Patients have to identify pictures when they make this test. The Lang II Stereo test has a monocularly visible shape on it (See Figure 2).

The Frisby Stereo test uses real depth. In the test tool three Perspex plates of different thickness are used and geometric shapes are painted on both sides of the Perspex (See Figure 3; Fricke & Siderov, 1997).
The Random E Dot and the Random Dot uses crossed polarised filters and the disparity is also constructed vectographically.

The Random Dot E Stereo Test consists of three cards and a pair of polarised spectacles. One card is a specimen and shows what the patient should see when using the spectacles and presented with the stereoscopical card. The third card is a blank to expose malingerers (See Figure 4).
Figure 5 of the Randot Stereo test shows two variations to test individuals. Firstly, the cartoon animals. These cartoons must attract the interest of young children and are arranged at three gross levels of disparity. In each of the three tests only one of the animals should appear from the other animals. In row A the cat has 400 seconds of arc, in row B the rabbit 200 seconds and in row C the monkey 100 seconds of arc, all at 16 inch distance.

Secondly, the contoured rings with random dot ground at ten levels of disparity. These rings provide a finely graded sequence for critical testing. For the disparities see the table.

<table>
<thead>
<tr>
<th>Rings</th>
<th>Seconds of arc at 16 In.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>400</td>
</tr>
<tr>
<td>2</td>
<td>200</td>
</tr>
<tr>
<td>3</td>
<td>140</td>
</tr>
<tr>
<td>4</td>
<td>100</td>
</tr>
<tr>
<td>5</td>
<td>70</td>
</tr>
<tr>
<td>6</td>
<td>50</td>
</tr>
<tr>
<td>7</td>
<td>40</td>
</tr>
<tr>
<td>8</td>
<td>30</td>
</tr>
<tr>
<td>9</td>
<td>25</td>
</tr>
<tr>
<td>10</td>
<td>20</td>
</tr>
</tbody>
</table>

Table 1: Randot ring disparities

In this experiment the last test was used to test the stereo acuity of the participants. Potential participants with disturbed binocular vision or a different refractive error in one eye were not allowed to do the test.
Appendix 6: Instructions belonging to the questionnaires

The questionnaires for the group who had to assess the ‘Quality’ criterion were provided with the following instructions:

Instructions for perceived quality judgments

In this experiment you will assess and evaluate 2D and 3D stereoscopic images. You have to give your judgments about the overall perceived image quality for 2 still images in 3 different viewing positions by looking through the stereoscope. If you have formed your judgment about the degree of perceived image quality of an image you indicate your rating on a graphical rating scale by drawing a line at the representative position. The rating scale is labeled by the words “excellent”--“good”--“fair”--“poor”--“bad” for reference. You may mark your judgment on one of these crossing points on the scale (not between the crossing points). An example is given below.

Before the experiment starts you participate in a trial session containing 2 different still stereoscopic images. This is a sample of the stereoscopic images shown during the experiment. Try to calibrate your judgment scale on this subset of images such that the entire scale is used.

For the groups who had to assess the ‘Depth’ criterion, the ‘Naturalness’ criterion and the ‘Viewing Experience’ criterion, the underlined words were replaced by the respective criterion words.
Appendix 7: Example of one sheet of an inquiry form

In the inquiry form every stimulus of the test has its own grading scale. The number of the grading scale corresponds with the number of the stimulus (1 to 48).
Appendix 8: Rating of ordinal categorical values; The ThurCatD program

Judgment of images mostly happens in categorical ordinal terms. In the tests involved in this report the participants had to judge according to the ordinal categorical terms ‘bad, poor, fair, good and excellent’ (ITU-R BT.500-7, 2002). The question arises if the intervals between those terms are big enough, without any confusion, to separate the judgments in a sufficient way.

Leon Lewis Thurstone (USA 1887 – 1955) who was one of the first and most productive scaling theorists, developed a scaling theory for judgments of ordinal values. In his classical paper from 1927, ‘A law of comparative judgment’ (Thurstone, 1927) he came up with a model existing out of five different cases and identified the assumptions belonging to each of these cases. His basic assumption was that the discriminating factor was a probability density function that followed a Gaussian function on a psychological continuum of a “ness”-scale (colourness, graininess, etc). Further he observed that the proportion of times that a stimulus $S_1$ was judged greater than stimulus $S_2$ ($S_1 > S_2$) was an indirect measure of the distance of the ‘ness’. The average difference between the two values is the statistical difference.

Gaussian probability distribution of the response of four stimuli over the psychological continuum $\Psi$. (Boschman, 2000)

Thurstone’s complete Law of Comparative Judgment is defined by the scale differences, the variances and the correlation coefficient and is expressed in the following formula:

$$S_A - S_B = z_{A-B} \sqrt{\frac{\sigma_A^2 + \sigma_B^2 - 2\rho\sigma_A\sigma_B}{2}}$$

This formula has no general solution and some simplifications concerning the correlation and the variances had to be made. Thurstone designed five assumptions which are summarized in the table on the next page. The most widely applied case V uses the assumptions of case IV plus the supposition that the two variances $\sigma_A$ and $\sigma_B$ are equal. In 1951 Frederik Mosteller (1916, West Virginia) showed that with a slight adaptation of the formula the correlation not necessarily needs to be zero but can also be constant (Case Va). In 1959 Stanley Smith Stevens (USA, 1906 – 1973) came up with case VI, were the standard deviations were proportional to the scale values.

The choice of one of the cases I to VI for calculating the scaling and the deviations depends on the characteristics of the stimuli or samples to be scaled. (Engeldrum, 2000).
Thurstone's six cases of the law of judgment and its conditions

<table>
<thead>
<tr>
<th>Case</th>
<th>Correlation Coefficient</th>
<th>Discriminal Dispersions</th>
<th>Model Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>I</td>
<td>$0 \leq p \leq 1$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} \sqrt{\sigma_i^{2} + \sigma_r^{2}}$</td>
</tr>
<tr>
<td>II</td>
<td>$0 \leq p \leq 1$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} \sqrt{\sigma_i^{2} + \sigma_r^{2}}$</td>
</tr>
<tr>
<td>III</td>
<td>$p = 0$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} \sqrt{\sigma_i^{2} + \sigma_r^{2}}$</td>
</tr>
<tr>
<td>IV</td>
<td>$p = 0$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} \sqrt{\sigma_i^{2} + \sigma_r^{2}}$</td>
</tr>
<tr>
<td>V</td>
<td>$p = 0$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} \sqrt{\sigma_i^{2} + \sigma_r^{2}}$</td>
</tr>
<tr>
<td>Va</td>
<td>$p = k$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$S_x - S_r = z_{\alpha, z} / \sqrt{1-k}$</td>
</tr>
<tr>
<td>VI</td>
<td>$p = 0$</td>
<td>$\sigma_i^{2} = \sigma_r^{2}$</td>
<td>$\ln(S_x) - \ln(S_r) = z_{\alpha, z}$</td>
</tr>
</tbody>
</table>

Thurstone’s six cases of the law of judgment and its conditions

When observers categorise samples they usually construct equal-appearing intervals. The categories are supposed to have equal distances with exception of the begin and end category. For data analyzing purposes weights is assigned to the categories. The first category mostly is given the value 1, the next category value 2, etc.

Based on these considerations Torgerson’s (1958) described a framework, called the ‘Law of Categorical Judgment’, with the following assumptions:

1) The psychological continuum of the observer can be divided into a specified number of ordered categories or steps.

2) It is assumed that the category boundary follows a normal distribution of positions on the continuum and may have different locations and different deviations.

3) The observer judges a given stimulus to be below a given category boundary whenever the value of the stimulus on the continuum is less than that of the category boundary.

Thorgerson’s four conditions of the law of categorical judgment

<table>
<thead>
<tr>
<th>Condition</th>
<th>Correlation Coefficient</th>
<th>Variances</th>
<th>Model Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>$\rho_{xy} = C1$</td>
<td>$\sigma_x^{2} \neq \sigma_y^{2}$</td>
<td>$t_x - t_y = z_{\alpha} \sqrt{\sigma_x^{2} + \sigma_y^{2}} - 2\rho_{xy}\sigma_x\sigma_y$</td>
</tr>
<tr>
<td>B</td>
<td>$\rho_{xy} = 0$</td>
<td>$\sigma_x^{2} = C2$</td>
<td>$t_x - t_y = z_{\alpha} \sqrt{\sigma_x^{2} + \sigma_y^{2}} + C2$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_x^{2} = 0$</td>
<td>$t_x - t_y = z_{\alpha} \sigma_y$</td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>$\rho_{xy} = 0$</td>
<td>$\sigma_x^{2} = C3$</td>
<td>$t_x - t_y = z_{\alpha} \sqrt{\sigma_x^{2} + \sigma_y^{2}} + C3$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_x^{2} = 0$</td>
<td>$t_x - t_y = z_{\alpha} \sigma_y$</td>
<td></td>
</tr>
<tr>
<td>D</td>
<td>$\rho_{xy} = r$</td>
<td>$\sigma_x = k\sigma_y = k2$</td>
<td>$t_x - t_y = z_{\alpha} \sqrt{k1^{2} + k2^{2}} - 2\rho_{xy}k1k2$</td>
</tr>
<tr>
<td></td>
<td>$\sigma_x = \sigma_y = 0$</td>
<td>$t_x - t_y = z_{\alpha}$</td>
<td></td>
</tr>
</tbody>
</table>

Thorgerson’s four conditions of the law of categorical judgment

While Thurstone’s Law of Comparative Judgment is formulated in terms of the scale difference between the samples, the Torgerson’s Law of Categorical Judgment is based on the difference between the sample scale value and the category boundary.
The formal model of Torgerson’s Law of Categorical Judgment is given by the following equation

\[ t_g - S_j = z_{jg} \sqrt{\frac{\sigma_j^2 + \sigma_g^2 - 2 \rho_{jg} \sigma_j \sigma_g}{\sigma_j^2 + \sigma_g^2}} \]  

and \( g = 1, 2, \ldots, m+1; j = 1, 2, \ldots, n \)

Here \( t_g \) = mean location of the upper \( g \)th category boundary, \( S_j \) = scale value of the \( j \)th sample, \( z_{jg} \) = unit normal deviate corresponding to the proportion of times stimulus \( j \) is sorted below category boundary \( g \), \( m+1 \) = number of categories, and \( n \) = number of samples.

The least number of assumptions is given by Condition D and similar to Thurstone’s solution. This makes Condition D suitable for simple calculation techniques.

M. Boschman wrote a computer program based on Condition D, with the name ‘ThurCatD’ (Boschman, 2000). In the tests described in this report this program has been used for checking the goodness of fit of the data.
Appendix 9: Calculations with the ThurCatD programme

In this appendix an example is given of the input files needed by the program ‘ThurcatD’ and the output files generated by that program (Boschman, 2000).

In listing 1 we see the input file of the test with the Bureau images impaired with blur. The listing shows 13 rows. The number 12 of the first row represents the number of stimuli (three different depths and four levels of blur impairments gives 12 stimuli). Starting from row 2, every row represents one stimulus. The number 5 of the first row stands for the number of choices subjects can make with assigning an image. These possibilities can be found in the five columns of row 2 to 13. Column 1 stands for the ‘bad’ choice, column 2 for the ‘poor’ choice, column 3 for the ‘fair choice’, column 4 for the ‘good’ choice and column 5 for the ‘excellent’ choice. So the second row gives the stimulus ‘Bureau’ with a base distance of 0 cm and without blur impairment. Two subjects judged this image as bad, nine subjects judged the image as poor and so on.

Listing 1: the input file ‘Depth Blur Bureau.tcin’

Listing 2 shows the history of the calculation process. It shows respectively:
- The maximum permitted number of iterations (5000)
- The type of output (simple).
- The used input file (‘Depth Blur Bureau.tcin’).
- The desired likelihood level.
- The number of iterations before the stop creation had been reached (110).
- The generated output file (‘Depth Blur Bureau.tcout’).

Listing 2: the history output file ‘Depth Blur Bureau.tchis’

Listing 3 shows the calculated output files. It presents the average values of the twelve stimuli with their deviations. Then it shows the boundary values of the five assignment choices also with their deviations and the intervals of these choices. At the end of the listing something is said about the model fit. If the model fit should not be OK, then one can state
that the intervals between the boundaries are too small so that the bandwidths of the stimuli cover each other too much.

Output for F:\AFSTUDEREN\Data uit Test\ThurCatD22 & TCD-Datafiles\Depth Blur Bureau.tcin generated by ThurCatD 2.2 on 24-2-2006 at 15:07:25.

Configuration after 110 iterations:
Estimated parameters (Noise spread parameter arbitrary set to 1):
Scale value parameters:
stimulus: 1 scale value: -0.4445; S_estimate: 0.2190
stimulus: 2 scale value: -0.6750; S_estimate: 0.2206
stimulus: 3 scale value: -1.7337; S_estimate: 0.2424
stimulus: 4 scale value: -1.9801; S_estimate: 0.2520
stimulus: 5 scale value: 1.2833; S_estimate: 0.2341
stimulus: 6 scale value: 0.6931; S_estimate: 0.2226
stimulus: 7 scale value: -0.0124; S_estimate: 0.2182
stimulus: 8 scale value: -0.6066; S_estimate: 0.2200
stimulus: 9 scale value: 1.6720; S_estimate: 0.2473
stimulus: 10 scale value: 0.6455; S_estimate: 0.2221
stimulus: 11 scale value: 1.1124; S_estimate: 0.2299
stimulus: 12 scale value: 0.0459; S_estimate: 0.2183

Interval bound parameters:
Lower bound of interval 2: -1.9772; S_estimate: 0.1407
Lower bound of interval 3: -0.7271; S_estimate: 0.0906
Lower bound of interval 4: 0.3894; S_estimate: 0.0848
Lower bound of interval 5: 1.6148; S_estimate: 0.1168

Intervals:
interval 1: -infinity ... -1.9772
interval 2: -1.9772 ... -0.7271
interval 3: -0.7271 ... 0.3894
interval 4: 0.3894 ... 1.6148
interval 5: 1.6148 ... +infinity

log likelihood kernel=-3.13346905909971E+0002

Model fit:
Probability stress=0.031630 -> stress<0.031980
Model fit is OK according to rule of thumb for probability stress.

Mosteller Chi-square= 30.4417, 33 Degrees of freedom.
Upper tail P-value=0.5951  Model fit is OK: (P>0.05).

Listing 3: the output file with the results 'Depth Blur Bureau.tcout'

Four criteria, namely depth, quality, naturalness and viewing experience were tested. Every test had two impairments namely blur and noise. So, totally eight calculations were done with the help of this program. The calculations were mainly done to check the model fits.
Appendix 10: Assessments of Depth-criterion

Figure 1

Figure 2

Figure 3

Figure 4
Appendix 11: Assessments of Quality-criterion

Figure 1

Quality scores Bureau Average values

Figure 2

Quality scores Playmobiles Average values

Figure 3

Quality scores Bureau Average values

Figure 4

Quality scores Playmobiles Average values
Appendix 12: Assessments of Viewing Experience-criterion

Figure 1

Viewing Experience scores Bureau Average values

Figure 2

Viewing Experience scores Playmobiles Average values

Figure 3

Viewing Experience scores Bureau Average values

Figure 4

Viewing Experience scores Playmobiles Average values
Appendix 13: Assessments of Naturalness-criterion

Figure 1
Naturalness scores Bureau Average values

Figure 2
Naturalness scores Playmobiles Average values

Figure 3
Naturalness scores Bureau Average values

Figure 4
Naturalness scores Playmobiles Average values
Appendix 14: Assessments of all criteria to Bureau with blur impairment

Figure 1

Depth scores Bureau Average values

Figure 2

Quality scores Bureau Average values

Figure 3

Viewing Experience scores Bureau Average values

Figure 4

Naturalness scores Bureau Average values
Appendix 15: Assessments of all criteria to Playmobiles with blur impairment

**Figure 1**

![Depth scores Playmobiles Average values](chart1)

**Figure 2**

![Quality scores Playmobiles Average values](chart2)

**Figure 3**

![Viewing Experience scores Playmobiles Average values](chart3)

**Figure 4**

![Naturalness scores Playmobiles Average values](chart4)
Appendix 16: Assessments of all criteria to Bureau with Noise impairment

Figure 1
Depth scores Bureau Average values

Figure 2
Quality scores Bureau Average values

Figure 3
Viewing Experience scores Bureau Average values

Figure 4
Naturalness scores Bureau Average values
Appendix 17: Assessments of all criteria to Playmobiles with Noise impairment

Figure 1

Depth scores Playmobiles Average values

Figure 2

Quality scores Playmobiles Average values

Figure 3

Viewing Experience scores Playmobiles Average values

Figure 4

Naturalness scores Playmobiles Average values
Appendix 18: Comparing Noise and blur Assessments of the “Quality” criterion data.

Table A18.1: Quality: Blur and Noise trends for “Bureau” images

<table>
<thead>
<tr>
<th>Value Rate 0</th>
<th>Slope $\beta_1$</th>
<th>Value Rate 0</th>
<th>Slope $\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blur:</td>
<td>4.795</td>
<td>-0.977</td>
<td>Noise:</td>
</tr>
</tbody>
</table>

Figure 1

Table A18.2: Quality: Blur and Noise trends for “Playmobiles” images

<table>
<thead>
<tr>
<th>Value Rate 0</th>
<th>Slope $\beta_1$</th>
<th>Value Rate 0</th>
<th>Slope $\beta_1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blur:</td>
<td>4.553</td>
<td>-1.144</td>
<td>Noise:</td>
</tr>
</tbody>
</table>

Figure 2
**Appendix 19: Comparing Noise and blur Assessments of the “Viewing Experience” criterion data.**

**Table A19.1: Viewing Experience: Blur and Noise trends for “Playmobiles” images**

<table>
<thead>
<tr>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>-0.945</td>
<td></td>
<td>-1.018</td>
</tr>
</tbody>
</table>

**Figure 1**

**Table A19.2: Viewing Experience: Blur and Noise trends for “Playmobiles” images**

<table>
<thead>
<tr>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
<th>Value Rate 0:</th>
<th>Slope $\beta_1$:</th>
</tr>
</thead>
<tbody>
<tr>
<td>Blur:</td>
<td>4.167</td>
<td>Noise:</td>
<td>4.095</td>
</tr>
<tr>
<td></td>
<td>-0.947</td>
<td></td>
<td>-0.738</td>
</tr>
</tbody>
</table>

**Figure 2**
Appendix 20: Just for fun ....make your own stereoscope

Those people who are skilful and interested in the 3D-subject and also want to do some experimenting they can make their own stereoscope. Below you find a construction drawing as can be found on the site http://www.funsci.com.

Fig. 2 - Stereoscope for 6x9 images.