MASTER

Distance adaptive picture quality of natural images

van Etten, M.

Award date:
2009

Distance adaptive picture quality of natural images

Final version

Marcel van Etten
m.v.etten@student.tue.nl

Supervisor at the Eindhoven University of Technology,
Human-Technology Interaction Group
Wijnand IJsselsteijn
w.a.ijsselsteijn@tue.nl

Supervisors at Philips Research Laboratories
Ingrid Heynderickx
ingrid.heynderickx@philips.com

Michael Murdoch

Table of Contents

Preface

This report is the result of my graduation project for the completion of the Master Human Technology Interaction at the Eindhoven University of Technology. Over the last 10 months I have worked on this project intensely and with much fun. This is among others accountable to the very interesting subject and the people surrounding me during this project, at Philips Research, the TU/e and at home. Apart from the fact that these people where great to work (and/or to hang out) with, it were all very inspirational, fun and intelligent people. During my graduation I learned a lot from them, not only on scientific levels, but also personally. Of those people surrounding me I would like to thank a number of people more specifically.

At first I would like to thank my main supervisor at Philips Research, Ingrid Heynderickx. Due to her inspirational character, she always made me think twice about the choices that I made. She is one of those people that can be very critical on a very nice and supportive way. I learned a lot from her in setting up experiments, organizing projects, writing skills, etc, etc. I would also like to thank Michael Murdoch, my second supervisor at Philips Research. He always had time for me when I had a question and was very supportive and helpful in answering these questions. He has also spend a lot of time in helping me achieving the current state of this report and setting up the experiments in a proper way. I would also like to thank Wijnand IJsselsteijn, my first supervisor from the University of Technology for being critical in reviewing my report, my decisions and arranging the logistic part of my graduation. Although he was very busy, he did find the time to be there at critical times during my graduation. Other people that I would like to thank more specifically are Arjen Damstra from Philips Consumer Lifestyle, Stefan Swinkels, Dragan Sekulovski, Marco van Boven and another graduation student Robert de Volder. They all helped me out in many different kinds of ways during measurements, experimental setup, evaluations of the results, etc. and made me think thoroughly about decisions and actions that had to be taken during this project. They were not only good support but also very nice sparring partners. Furthermore I would like to thank all the other students and co-workers at Philips for the fun and interesting time during work, in the coffee/lunch breaks and after work

I would also like to thank my wonderful girlfriend Cheryl, for her unconditional support, her interest and her patience. She has spent a lot of evenings on her own downstairs while I was working on my thesis upstairs. She never complained and always stood by my side when I needed her, again unconditionally!

Last but certainly not least, I want to thank my fantastic parents for giving me the opportunity to follow and finish this study, for their interest and for their dedication in helping me out at some tricky times.

Marcel van Etten

Abstract

Optimizing image quality of TVs is a very important aspect of making TVs look more attractive to potential buyers and owners. Image quality is a percept produced by the image quality attributes which are the things people perceive in an image.

In this study, the influence of the most important image quality attributes on image quality, are evaluated with different environmental characteristics. These most important attributes were determined to be brightness, contrast and sharpness. This was found in a field study which is discussed in this report. Two controlled psychophysical experiments were performed to find (1) the effects of environmental characteristics: room illumination, viewing distance and natural image content, on Just Noticeable Differences (JNDs) of the image quality attributes, and (2) the optimal settings of the image quality attributes in the different environmental settings to obtain the optimal perceived image quality for natural images in different environmental conditions.

With use of the first experiment, the JNDs of the image quality attributes brightness, contrast and sharpness were found to depend on viewing distance, content and ambient illumination. The results of the second experiment show that the optimal settings of the image quality attributes depend on the setting of the environmental characteristics.

1 Introduction

Since 1980, marketing research has shown that picture quality is an important aspect in the decision process of consumers for buying a TV and that it can be used as a competitive factor (Roufs & Bouma, 1980). Therefore manufacturers are constantly trying to optimize the picture quality of their TVs. During this optimization process, they have the dilemma of optimizing the quality for specific environments such as the shop or the home environment. These kinds of environments mainly differ in the amount of ambient illumination in which the TV is displayed and in the viewing distance at which the TV is judged. Both aspects are known to affect the perceived image quality (Ardito, 1994; Lund, 1993). Fortunately, with current developments in sensor technology, it would be relatively easy and cheap to implement a sensor system that measures the amount of ambient illumination and the viewing distance towards the TV. Hence, if it is known how to adapt the quality of the TV according to the values of these sensors, it would be possible to optimize the perceived image quality of a TV simultaneously for all kinds of different environments. This research project mainly focuses on how to optimize perceived image quality as a function of viewing distance.

1.1 Image quality

(Perceived) Image quality is a subjective measure for how well an image is rendered. During the optimization of a TV set, it is important to obtain a perceived image quality as high as possible and therefore, to consider what peoples image quality preferences are. Engeldrum (2000) introduced the Image Quality Circle model (see Figure 1) as a method to describe optimal image quality.

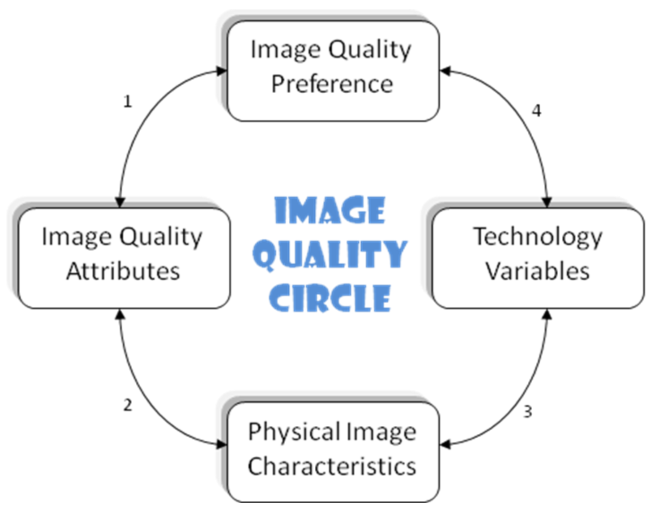


Figure 1 The image quality circle by Engeldrum (2000)

In this model, "Technology variables" refer to the things that can be manipulated in the TV set to vary the "physical image characteristics". These "physical image characteristics" are the things we measure at physical or electronic level, such as luminance, resolution, coding artifacts and color gamut. Combinations of the "physical image characteristics" yield the "image quality attributes". These attributes are what people actually see, such as sharpness and brightness. The right balance between these attributes will result in the highest image quality preference.

When steps 1, 2 and 3 are known, manufacturers are able to see which technology variables they have to vary to increase the perceived image quality.

1.1.1 Quality attributes, physical characteristics and technology variables

From the model of Engeldrum we can see that image quality attributes influence the image quality preference. The most significant attributes that influence this perceived image quality are contrast, brightness, sharpness, motion and color (Young Park, Triantaphillidou, & Jacobson, 2009; Heynderickx, Xia, Qin, Liu, & Yin, 2005). The most important attributes for this study, i.e. contrast, brightness and sharpness, will be explained below, together with the technology variables that can be used to influence these percepts.

1.1.1.1 Contrast and brightness

The contrast of a TV is the ratio between the physical image characteristics light over dark colors. The higher this ratio, the bigger the perceived contrast of an image. The brightness is the amount of perceived luminance of the screen. It is often thought that brightness and contrast are perceptually independent, but as Keelan (2002) already wrote:

'Sometimes the levels of different attributes are correlated because each of the attributes is affected by a common process. For example, blurring an image decreases both sharpness and noisiness, if the noise is suprathreshold. This does not imply that the attributes are perceptually dependent, but rather is an indication that they are physically linked under some circumstances.' (p. 232)

Keelan, 2002

This can be applied to the attributes brightness and contrast. They are physically linked by two underlying physical image characteristics: the white level and black level of the display. Heynderickx et al. (2007) also stated that contrast and brightness are not independent. Therefore, they used the independent physical characteristics black level and white level in their experiments to establish just noticeable differences of image quality attributes.

To understand the impact of changing the black and white level, we will first examine the grey scale histogram of a natural image. Figure 2 shows an image of a polar bear together with its grey level histogram. The grey levels are plotted on the x-axis and the number of pixels that contain a specific grey level is plotted on the y-axis. The zero grey level refers to saturated black, and the 255 grey level refers to saturated white. For the image of the polar bear we can state that it contains much grey and white, but little dark grey levels since it does not contain a lot of pixels with grey levels near the 0 grey level point.

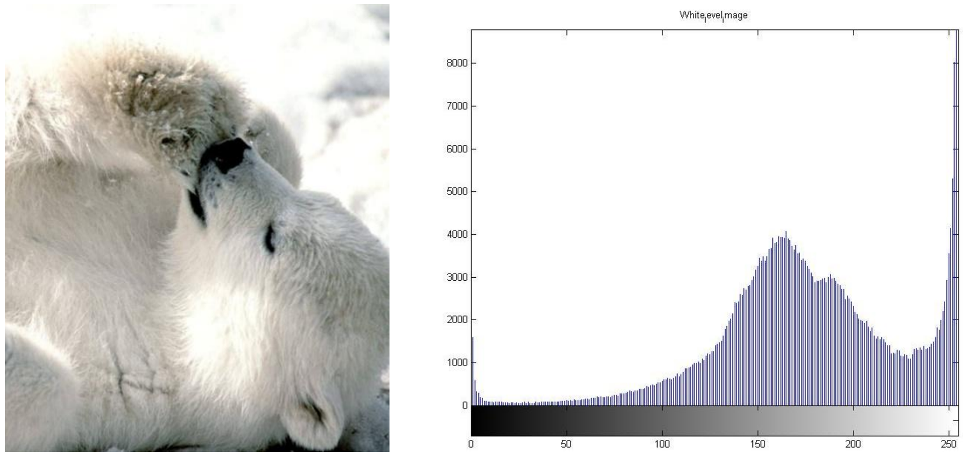


Figure 2 An original image "polar bear" together with its grey level histogram showing that this image contains mostly bright grey levels.

As stated, the 255 grey level of this image is the point that allows for the whitest white, so the image in Figure 2, allows for grey levels of 255 and thus the white level of this image is maximum. But, the perceived contrast is a combination of the black and white level of the display that presents the image, the histogram of the displayed image, and the display function (or gamma) of the display. Of course, if an image does not use all (255) grey levels, its (perceived) contrast will not be maximal. But if an image uses all grey levels, its contrast can still be enlarged by (1) decreasing the display black level, (2) increasing the display white level, and by (3) changing the gamma or display function. When the maximum grey level of the image is changed to, for example, 195, the image contains no more grey levels near the whitest white, as can be seen in Figure 3. The image itself now has less contrast (the difference between the darkest colors and the lightest colors is now smaller) and less brightness (it contains less light colors).

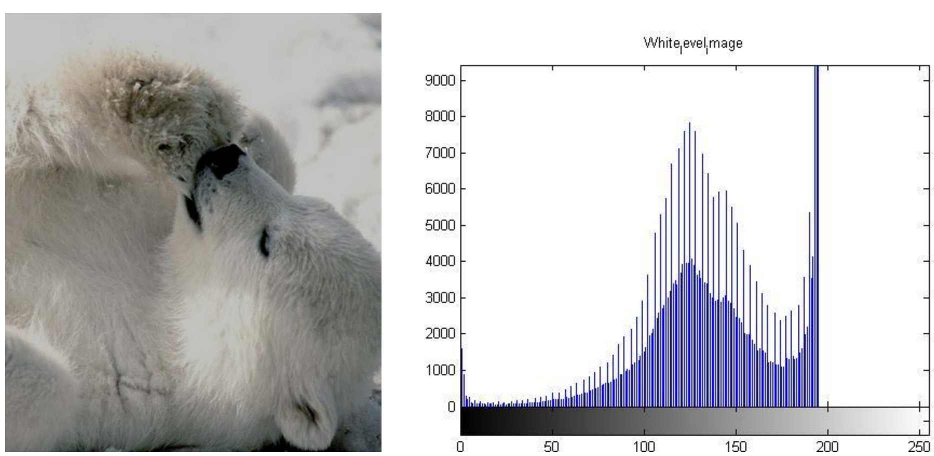


Figure 3 The maximum grey level of the "polar bear" image is decreased by 60 grey levels making the image appear darker and of lesser contrast.

The black level is the point that allows for the blackest black. In the histogram above the minimum grey level is zero, but also this level could be shifted to change the contrast and brightness. Again this also depends on the display black and white level and/or gamma. Hence, the two physical image characteristics, black level and white level, work together to form perceived brightness and contrast. When, for example, the black level is increased, the image appears brighter because all grey levels are increased in luminance. The contrast becomes smaller because the perceived difference between dark and light colors becomes smaller.

Again, it should be noted that it depends on the contrast and brightness characteristics of the screen on which the image is presented how the contrast and brightness of the image appears. Hence adjusting contrast and brightness can be done by image processing and/or adjusting display characteristics.

1.1.1.2 Sharpness

How sharp an image appears is dependent on several physical image characteristics, such as the resolution of the image and the steepness edges in the image. Blurring an image makes the edges less steep, so that they appear less sharp. This is mainly done by using a Gaussian blur filter which filters the high frequency components out of an image.

In current TV set a peaking algorithm is used to make the image appear sharper (or less blurry) by increasing the visibility of the edges. A peaking algorithm increases the amplitude of the high and/or middle frequency components of an image (Haan, 2006). The effect of peaking is illustrated in Figure 4 in which an original image is shown at the top left-hand side and an extremely peaked image is shown at the top right-hand side. The graph at the bottom of Figure 4 presents what actually happens to an edge when a peaking algorithm is applied, i.e. when the amplitude of an edge is increased.

A disadvantage of this peaking algorithm is that it also increases unwanted edges that exist in compression artifacts and noise. In Figure 4 for example the artifact of mosquito noise (i.e. noise around edges, such as around the women or the parrot) becomes clearly visible. Compression artifacts, such as blockiness and ringing, also known as mosquito artifacts (Meesters & Martens, 1999; Ridder, & Willemsen, 2000; Kirenco, Muijs, & Shao, 2006; Subedar, & Caviedes, 2007) therefore become more visible when a peaking algorithm is applied on an image. For still images which are extracted from a movie, another type of artifact exists, namely capture artifacts (Bailey & Punchihewa, 2002), which are the result of imperfections of the optical capture device. In Figure 5 this type of artifact can be seen in the hands of the lady in front and in the white shirt of the man in the middle.

Another disadvantage of peaking is ‘electronic ageing’ (Haan, 2006). When applying the peaking algorithm on human skin, the skin may show more wrinkles. This might, for example, occur when noise is present in the image on the skin. The peaking algorithm boosts high frequencies, and also the high frequencies present in noise. This makes the skin look wrinkly.

For a more elaborate explanation of all artifacts we refer to the literature above.

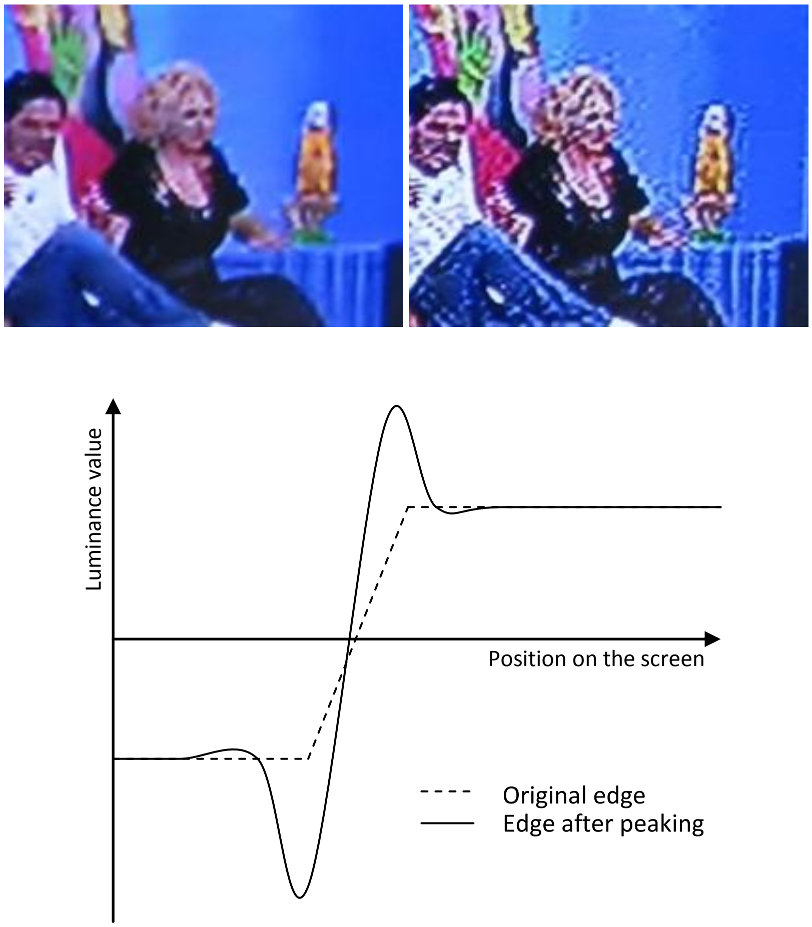


Figure 4 Applying the peaking algorithm makes edges appear sharper by increasing its amplitude. The image in the upper left corner shows a blurred image while the image in the upper right corner shows a peaked image making the edges more visible

Figure 5 An example of capture artifacts in a still image. These can especially be seen in the hands of the woman in front who is clapping.

1.2 Earlier research on image quality optimization

People and manufacturers used to have the impression that the larger the brightness, contrast and other image quality attributes, such as color gamut, sharpness, etc. the higher the perceived image quality. From scientific studies, however, it appears that this is not always the case. Chronologically, people first tried to find the overall preferred viewing distance to obtain the optimal perceived image quality. From these studies it appeared that this preferred distance was very dependent on the image quality attributes and therefore no single preferred viewing distance could be found.

1.2.1 Viewing distance

Previous literature showed that the preferred viewing distance was linked to the screen height with a constant factor of about 7 to 8 for NTSC video material (Fujio, 1985; Fish, 1991). Other studies also found this ratio to be ranging from 5 to 8 times the screen height. One of those studies by Jesty found this ratio to be 7.8, but he mentioned that attributes affecting image quality can have an effect on the viewing-distance preference, and therefore, this ratio might not be constant for varying image quality attributes such as resolution (Jesty, 1958). In a study done by Lund the effects of image size and image resolution on the preference for viewing distance were further evaluated to see if this indeed has an effect on the ratio (Lund, 1993). In his study, Lund performed five psychophysical experiments in which he asked participants to position their chair at the distance at which the image appeared best. He found that the relation between preferred viewing distance and screen height was not linear, but that the ratio became smaller for larger image heights, approaching 3 or 4 times the image height for larger image sizes. Furthermore he was unable to find an effect of screen resolution on the optimal viewing distance. This finding disagreed with the earlier note from Jesty that resolution would have an effect on the optimal viewing distance (Jesty, 1958). It also disagreed with basic visual acuity theory stating that the farther the viewing distance, the harder it is to distinguish between lines (Blake & Sekuler, 2006). Lund however noted that the lack of an effect of image resolution might stem from uncontrolled variables in the experimental design and that it would need further examination. He also indicated that earlier research, for example by Neuman (1990), did find strong effects of resolution on quality judgments.

That resolution does have an effect on the subjective image quality as found in an earlier study by Westerink & Roufs (1989). In one experiment they influenced image 'resolution' (it was actually not the resolution that was changed, but the focus (sharpness) of the image and the image height). People were then asked to give their subjective image quality rating on a scale from 0.1 to 10 at a constant viewing distance and ambient illumination for five different types of content. From this experiment Westerink and Roufs found an effect of both the 'resolution' and the picture height on image quality. For a higher 'resolution' and for larger image widths, the subjective image quality increased. In a second experiment also the viewing distance was varied. Here, an effect of the visual angle covered by the image (related to the viewing distance, since at a larger distance this angle becomes smaller) on the subjective image quality was found. They concluded that a larger viewing distance had a negative effect on the subjective image quality since the visual angle decreased, but that, on the other hand, the quality improved because the angular resolution became larger.

In another study by Ardito the preferred viewing distance for HDTV programs was evaluated (Ardito, 1994). This was done by asking people to choose the viewing distance they preferred most. In this study it was found that the viewing distance expressed in image heights decreased with larger dimensions of the screen. This was in agreement with the findings of Lund. Furthermore the shape of the curve matched with the curve found by Lund. Ardito also found that for increased brightness, the preferred viewing distance increased as well. This, together with the note from Jesty was one of the first indications that image quality attributes may have an influence on the preferred viewing distance (Jesty, 1958). Hence, from this it can be stated that the settings for these attributes need to be changes depending on the viewing distance in order to obtain the optimal perceived image quality.

Findings and indications like the ones discussed above, made people more aware of the influence of environmental factors such as the viewing distance and ambient illumination on the image quality attributes such as sharpness and brightness. As a result, more recent studies were done on the effect of variables like distance, content and ambient illumination on the perception of the image quality attributes.

1.2.2 Environmental factors and image quality attributes

Although the literature in this area is quite scarce, there are some researchers that have done investigations in this area. In a study by Johnson and Montag (2005) for example a psychophysical study was performed to find the influence of viewing distance on the perceived spatial frequency and contrast. This was done with two images of noise patterns and by asking participants to adjust the contrast and spatial frequency of one image until it matched those of another image located at a different distance. The size of the farthest image was adjusted to match the perceived image size of the image closest by. From this study Johnson found that participants perceived the two images equal when the spatial frequency and contrast of the image farthest away were higher than those of the image closest by. This again indicates that viewing distance has an effect on how image attributes like sharpness (resolution) and contrast are perceived.

In another study by Radun et al. (2007) the main goal was to find if sharpness was dependent on content and if so, if these differences could be explained by interpretations related to the changes which were perceived. From a free description experiment, they found that for image quality perception, the effect of sharpness variations was dependent on the type of content. For some images low sharpness received a more positive rating than for other images. They also found that people used different terms of interpretation for these two separate cases. If people required more sharpness they used terms like: irritating, dirty, amateurish, etc. When people required a more blurry image however, they used terms like: artistic, soft, light colors, etc.

In another study, Damstra (2009) presented participants with two images varying in sharpness settings on two TVs and asked them to indicate the TV with the best image quality. This was done on two different distances for a collection of fourteen images. Damstra found a strong interaction between the sharpness setting and the preferred viewing distance. At a close distance (1.3 meters) the less sharp image was preferred over the sharper image and at a far distance (4.5 meters) the sharper image was preferred.

The studies discussed above, describe the influence of some specific environmental conditions on a small subset of image quality attributes. The attributes chosen in these studies seem to be selected rather ad hoc, and not on the basis of their importance to image quality. In the first study described in this report we therefore strive to find the attributes most influencing perceived image quality before performing controlled experiments on the effects. This is done in a field study which is discussed in chapter 2. Some other literature already reports on the importance of attributes. Young Park et al. (2009) investigated which image quality attributes were most affected by image size (which is equivalent to viewing distance, since the image size also decreases visually when the viewing distance increased). This was done by asking people with an imaging or design background which attribute was most different for two images with different image size. The three attributes that were mentioned most for both chromatic and achromatic images were (from high to low) sharpness, contrast and brightness. In another study by Radun, Virtanen and Nyman (2006) participants had to give free descriptions of the images and the most important characteristics were reported. It was found that the most important attributes mentioned by observers for indicating image quality were: graininess, darkness and sharpness. Heynderickx et al. (2005) found that the most important image quality attributes for high-end TVs showing SD and HD frequencies were brightness, contrast, color rendering and sharpness.

The main research question of this study is **if** and **how** viewing distance influences the most important image quality attributes and with that, overall perceived image quality. Although this study mainly investigates the influence of viewing distance, also the influence of ambient illumination and content are included. This main research question can be divided into two sub questions. **Whether** environmental aspects influence the image quality attributes can be answered via the question: ‘is the just noticeable difference (JND) of the image quality attributes dependent on viewing distance?’ **How** environmental aspects influence the image quality attributes can answered via the question: ‘Are the settings for the image quality attributes belonging to an optimal perceived image quality dependent on viewing distance?’

These two sub questions are investigated for the most important image quality attributes in two experiments which are discussed in chapters 3 and 4.

1.2.3 Hypotheses

Based on the literature discussed above, hypotheses can be made for the two sub questions of the main research question. These hypotheses are:

Research question	Hypotheses H_1
Is the level of JND for a specific image quality attribute dependent on viewing distance?	The JND of brightness is larger at a large distance than at a small distance, since the area of the eye perceiving the change is larger at small viewing distances. More cones and rods in the human eye respond to the signal (Blake, 2006).
	The JND of contrast is larger at a large distance than at a small distance, since the area of the eye perceiving the change is larger at small viewing distances. More cones and rods in the human eye respond to the signal (Blake, 2006).
	The JND of sharpness is larger at a large distance than at a small distance, since the visual angle is smaller at large viewing distances, making it harder to discriminate between two points (Blake, 2006).
Are the settings belonging to an optimal perceived image quality dependent on viewing distance?	The settings belonging to an optimal perceived image quality at a large viewing distance are different from those at a small viewing distance (small distance: lower brightness wanted , lower contrast required (Montag, 2005), less sharpness required (Damstra, 2009)).

It was already mentioned that, although it is not the main question of this study, also the effect of ambient illumination and content on the perception of the image quality attributes is investigated in controlled experiments. Therefore for ambient illumination and content, hypotheses are formulated as well. The hypotheses of ambient illumination are based on the theory of visual acuity discussed before by Blake (2006) and by Graham (1965). They stated that at high luminance, the cones are active in the human visual system which allow for a higher visual acuity because the cones have a higher density. This theory applies to the hypotheses of sharpness. However, since the cones are already active at 10^-2 cd/m^2, no effect of room illumination is expected for the visibility of sharpness changes. To formulate the hypotheses for brightness and contrast the DICOM curve can be used (Samei, 2002). This curve specifies a standard display function taking the non-linear nature of human perception into account (for greyscale images), as measured by a side-by-side comparison task for homogeneous images. This curve shows that at higher luminance values, it is harder to discriminate between different light intensities (i.e. the JND of brightness and contrast changes is larger). Therefore, we can hypothesize that the JND of contrast and brightness are bigger with larger room illumination.

Research question	Hypotheses H_1
Is the level of JND for a specific image quality attribute dependent on ambient illumination?	The JND of brightness is larger with a larger room illumination (Samei, 2002)
	The JND of contrast is larger with a larger room illumination (Samei, 2002)
	The JND of sharpness is not dependent on room illumination (Graham, 1965)
Are the settings belonging to an optimal perceived image quality dependent on room illumination?	The settings belonging to an optimal perceived image quality with a large room illumination are different from those with a small room illumination, except for sharpness (Samei, 2002; Graham, 1965).

The hypotheses for content are extracted from an earlier study by Radun et al. (2007) in which they found that the perceived sharpness was dependent on content. They found that for some images a less sharp image was preferred and for some a more sharp image. A study by Heynderickx et. al (2007) was used to formulate the hypotheses of the JND of brightness and contrast with respect to content. In this study, Heynderickx et al. found that for images that contained a lot of intermediate grey levels the JND for black and white level became substantially larger.

Research question	Hypotheses H_1
Is the level of JND for a specific image quality attribute dependent on content?	The JND of brightness is dependent on content (Heynderickx, 2007)
	The JND of contrast is dependent on content (Heynderickx, 2007)
	The JND of sharpness is dependent on content (Radun, 2007)
Are the settings belonging to an optimal perceived image quality dependent on content?	The settings belonging to an optimal perceived image quality are dependent on content (Heynderickx, 2007;Radun, 2007)

2 Field study

In order to find the most important image quality attributes influencing perceived image quality and to select those attributes that were influenced by the viewing distance a field study was performed.

2.1 Methodology

In this field study, people were asked to indicate the TV with the best and the worst perceived image quality out of a number of TVs that were displayed in a shop. They were also asked why they selected this specific TV to be the best or the worst in order to find the image quality attributes that mostly influenced their perceived image quality. The study was performed at two different shops in the Netherlands. Below we elaborate more on the methodology of the field study.

2.1.1 Participants

The participants of this study consisted of a random sample of 58 naïve shop customers (33 male and 25 female). The mean age of the participants was 41.5 (S.D. 14.7) with a minimum of 13 and a maximum of 70 years old. In the first shop (Makro Breda) 40 participants were interviewed and in the second shop (Media Markt Breda, electronics shop) 18 participants.

2.1.2 Setup

Figure 6 and Figure 7 show the setup of the Makro and the Media Markt respectively. Especially the Makro had a very good setup for the sake of this study and therefore, most participants were interviewed here. It included 14 TVs of various brands and sizes that were placed in three rows (the TVs within the yellow square were used for this study). The Media Markt had the disadvantage that other TVs blocked the view on the selected batch of TVs that had to be judged. Here the setup included 17 TVs of various brands and sizes that were placed in two rows.

Figure 6 TV setup at the Makro Breda. The TVs within the dotted yellow line were used for the field study.

Figure 7 Commercial setup of the Media Markt Breda

2.1.3 Procedure

Random customers of the shops were approached and asked if they were willing to answer 4 short questions which would take them about five minutes. If they were willing to participate they were asked if they could walk with the researcher to the TV setup of the shop. First they walked to a distance of 6 meters from the TVs. Here the participant was presented with the first two questions, namely:

- 'According to you, which TV has the best image quality and why?' and
- 'According to you, which TV has the worst image quality and why?'

Next the participant was taken to a distance of 1-2 meters from the TVs. Then, again the questions above were asked.

The researcher noted the number of the TV (as counted from left to right 1-5 and top to bottom 1-3) and the reason why this TV had the best/worst image quality as mentioned by the participant. These reasons were classified along the image quality attributes also mentioned in the introduction (i.e. brightness, contrast, color, sharpness and movement). In Table 1 some examples of this classification are shown. It should be noted that this classification was based on the researcher's expert interpretation of what the participant mentioned, since some reasons were a bit unclear, there is no guarantee that this classification is done correctly in every case. Participants never mentioned reasons that could not be classified along the five image quality attributes.

Table 1 examples of researcher's interpretation

Participant's reason	Factor
I think the image looks grayish	Contrast
I think the image looks dull/matt	Contrast
I think the black is black and the white is white	Contrast
I think the image is (too) bright	Brightness
I think the image looks restless	Movement
I think the image has nice colors	Color
I think the image looks sharp	Sharpness

Sometimes it appeared to the researcher that participants named the design of the TV as determinant for best/worst image quality. When this happened the researcher asked the participants if they could try to ignore the design and to look only at the quality of the image. In the case that the researcher had the feeling that the participant did not correctly name what he/she thought, the researcher asked another question to make sure the participant was sure about his or her reason. For example, sometimes the participant mentioned sharpness while they actually tried to explain brightness. In Dutch sharp (scherp) and bright (fel/helder) are often misused. Dutch people sometimes say: 'Oh, that is sharp for the eyes' while they mean that the brightness is too high. This problem also works the other way around; hence when people indicated an image to be sharp it had to be considered if they actually meant sharpness or brightness. For contrast, color, and movement this problem was not present. By including these measures, the chance of misinterpreting the reasons by the researcher became smaller.

Furthermore, the participants were asked for their age and the average number of hours they watched TV per day. This was also noted.

The image material shown on the TVs in both shops consisted of advertisements concerning products sold in the shop including regular promotional videos altered with a promotion video of the shops consisting of an animation with mainly yellow, blue and red colors.

2.1.4 Measures

To measure people's opinions about the worst/best TV image quality and their reasons, the four questions mentioned in chapter 2.1.3 were used. The 'counts of mentioned TVs' were used to measure best or worst TV. This was analyzed with use of the descriptive statistics in SPSS.

To evaluate the extent to which the factors participants mentioned were consistent with actual physical image characteristics of the TVs, several optical measurements were done on the most selected TVs (best and worst). These measurements included the xy, L, XYZ and u'v' measures for a white (for luminance), black (ratio between black and white for contrast), red, blue and green screen. The images were provided to the displays by use of a HDMI DVD player. Using these measurements the luminance [cd/m^2], static contrast (L_{white}/L_{black}=contrast ratio) and chromaticity (u' v' chromaticity coordinates of Red, Green and blue) of the TVs were determined. The measurement device was a Topcon BM7 meter which was used with a measurement aperture of 1°. The accuracy of the Topcon for Luminance was +/-2% and for chromaticity 0.002.

A picture of the set up is shown in Figure 8.

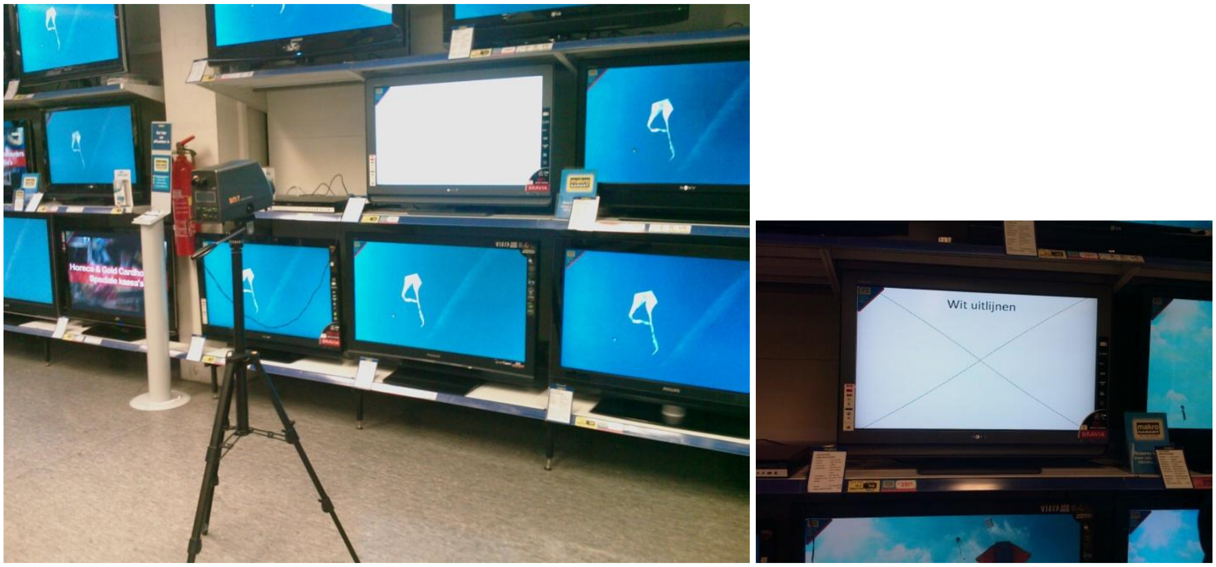


Figure 8 Measurement setup to measure the xy, L, XYZ and u'v' of these TVs in the Marko Breda

From these measures, it is possible to evaluate the extent to which people are able to correctly identify factors as reasons for good or bad image quality. This is done by showing the relation between the measurement values of a specific factor and the times this specific factor was mentioned by participants as being a determinant for high or low image quality.

The Physical measurements were only done at the Makro. This shop was most relevant because of the higher number of participants, but most importantly, at the Media Markt it was not allowed to do measurements.

2.2 Results

The data from participants 1 and 27 were removed from the dataset, since they did not answer all four questions completely. The remaining dataset was used to obtain the results discussed in this paragraph.

These data were used to obtain the graphs presented in Figure 9. These graphs show the number of times a specific attribute was mentioned (i.e. score) when a TV was selected as best or worst in image quality. The blue lines are obtained from the data for the high quality question ('according to you, which TV has the **best** image quality and why?') at the small and large distance, respectively. The red lines represent the answers to the low quality question ('according to you, which TV has the **worst** image quality and why?'). Each attribute is given in a separate graph. From these graphs it can be seen that brightness seems to be a better determinant for high quality, while contrast is a better determinant for low quality. The graphs also show that brightness is used more often as a determinant for image quality at a large than at a small distance. This is the case for both high quality (mentioned 38 times at a large distance in contrast to 32 times at a small distance) and low quality rating (mentioned 24 times on a large distance in contrast to 14 times on a small distance). Sharpness, on the contrary, is a determinant more often used at small than at large viewing distance, again independent of the quality level of the TV. Color is important as a determinant for image quality, but rather independently of the viewing distance and the quality level of the TV. Movement was mentioned by at most four participants to be a determinant for perceived image quality.

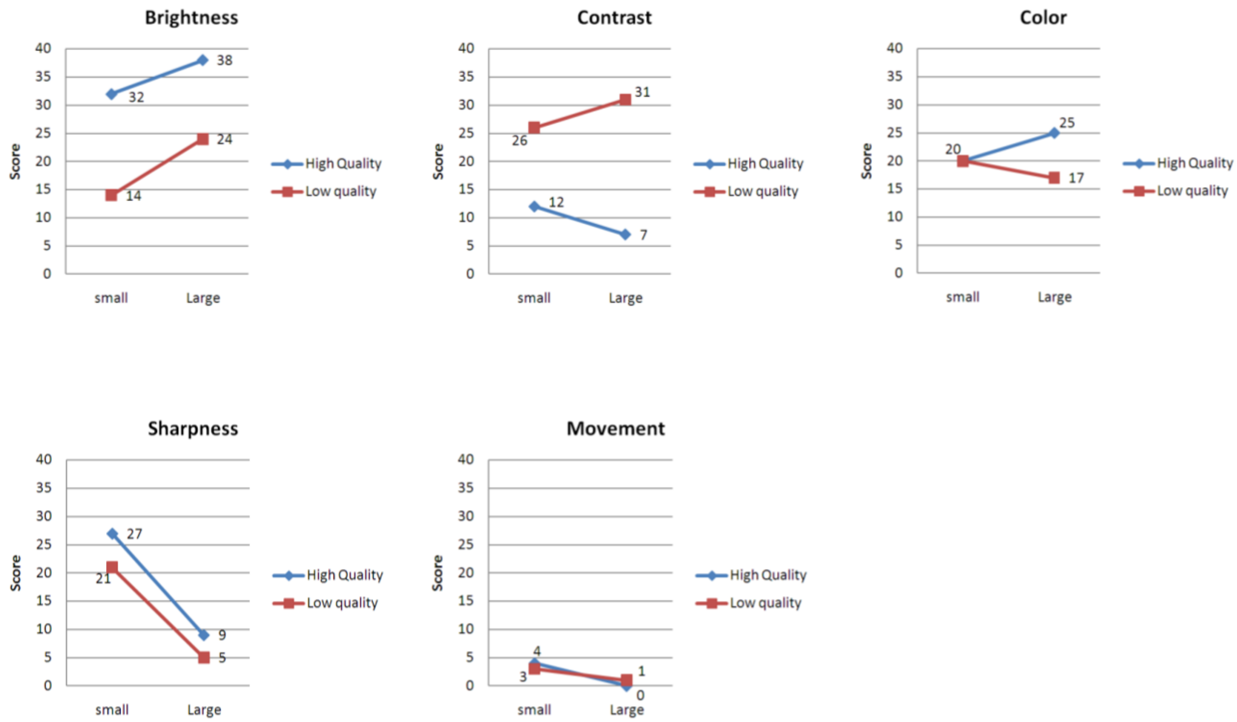


Figure 9 Interaction graphs of the field study (the score on the y-axis represents the amount of times the image quality attribute was mentioned).

Using SPSS, a paired sample t-test was performed on each pair of data for small and large viewing distance, i.e. for both the high and low quality questions for each attribute. Only for the attribute sharpness, a significant difference between a small and large viewing distance was found: for a high quality level of the TV, sharpness was mentioned significantly more often in case of a small distance (*M= 0.47, SE = 0.066*) than in case of a large distance (*M = 0.16, SE = 0.048, t(57)=3.611, p=0.01, r=0.43*); for a low quality level of the TV, sharpness was mentioned significantly more often in case of a small distance (*M = 0.36, SE = 0.064*) than in case of a large distance (*M = 0.09, SE=0.037, t(57)=3.584, p=0.01, r=0.43*). The other differences were not significant.

The measured contrast ratio, luminance and u' v' color points for red, green and blue as measured for the TVs that were selected to have the best or worst perceived image quality, are presented in Table 2. This table also shows the number of times a specific attribute was mentioned to be responsible for the quality of the best or worst TV. Here, the scores for the small and large viewing distance are summed, since it is only important to evaluate the extent to which the attributes mentioned by the participants were indeed related to the actual physical image characteristics of the TVs. The latter is included for brightness, color and contrast only, since we were not able to characterize the other attributes (sharpness and motion) with a simple, portable measurement device in the shop.

Table 2 Measurements and participant scores of the TVs

TV	Luminance	Contrast	Red		Blue		Green		Brightness		Contrast		Color	
			u'	v'	u'	v'	u'	v'	high	low	high	low	high	low
1	86.48	22.82	0.447	0.525	0.173	0.171	0.107	0.572	11	0	3	0	6	0
2	58.23	19.05	0.455	0.526	0.169	0.192	0.102	0.574	2	4	2	3	1	8
3	236.70	170.29	0.456	0.524	0.167	0.162	0.116	0.5661	17	0	1	0	13	1
4	63.21	22.68	0.447	0.525	0.170	0.181	0.110	0.572	0	2	0	4	0	0
5	299.20	99.11	0.439	0.526	0.172	0.149	0.117	0.560	0	1	0	3	1	0
6	442.90	170.81	0.448	0.526	0.165	0.168	0.121	0.565	2	1	3	0	3	0
7	74.30	2.70	0.455	0.527	0.168	0.185	0.109	0.572	1	10	2	25	1	7
8	451.90	376.58	0.428	0.529	0.176	0.176	0.111	0.565	13	1	2	0	7	0

In Figure 10, the left graph for brightness shows the results for the high quality ratings. Here it is for example presented that TV 3 in table 2, is mentioned 17 times with respect to brightness and that it had a luminance value of 236,7. Contrast on the other hand is used more as a determinant for low quality, so here the results for the low quality rating are shown in the right graph of Figure 10. For brightness and contrast, no significant correlations were found between the measurements and the participant counts (meaning how often participants noted a specific factor for specific TVs). The correlation graphs in Figure 10 however all show a small slope in the expected direction. For brightness for example it is expected that with higher luminance, a TV is mentioned more often for the high quality question.

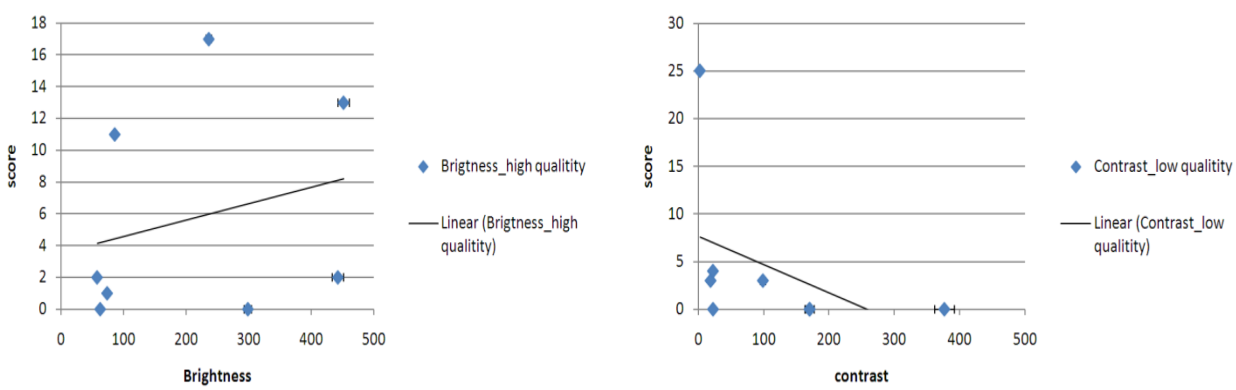


Figure 10 Relation between the measurements of brightness [cd/m2] and contrast [-] with the number of times the attribute was mentioned

Color was a more difficult aspect to compare to the amount of times a color was mentioned as a determinant for image quality, since from the measurements it cannot be unambiguously be determined which TV has the best color gamut. Furthermore, the measured u' and v' values for red, green and blue are all relatively close to each other as can be seen in the u'v'-color space graphs in Figure 11. An image of the full u'v'-color space can be found in Appendix A. Therefore the relation between the measurements of the color characteristics and the amount of times color was mentioned was not further analyzed.

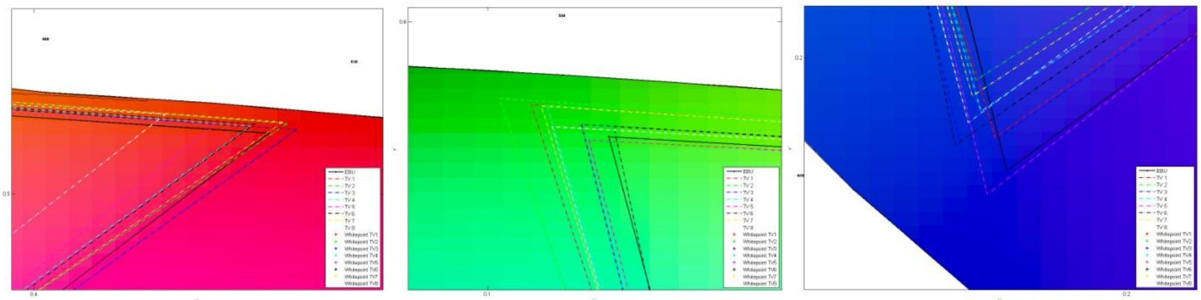


Figure 11 u' v' color space. Plotted are the measured u'v' values for each TV for red, green and blue.

2.3 Discussion

From the results it is clear that brightness is used more often as a determinant for high image quality perception than for image quality perception. Most of the time people mentioned that an image quality was good because it had a ‘nice brightness’ or that it was ‘bright’. People also reported that ‘the brightness is too high’ or ‘it is too bright for my eyes’, but these statements were made much less often than positive remarks about brightness. Contrast on the other hand is used more often as a determinant for low quality image perception. People claimed that image quality was low when the image looked ‘grayish’ or that ‘the difference between black and white is low’. Color, sharpness and movement were mentioned about equally often for the best and worst quality TV.

Even more important for the sake of this study is the effect of viewing distance on each of the attributes. From the results it can be seen that **brightness** is reported more often as a determinant for quality at a large distance than at a small distance. This indicates that people are more sensitive for brightness differences at a large distance than at a small viewing distance. This however remains an indication since this field study was too limited to draw hard conclusions. For a low quality TV people mention **contrast** more often at a large viewing distance than at a small distance. For a high quality TV on the other hand, the reverse is true. These statements are based on the trends found; the insignificance will be further evaluated in controlled experiments discussed in chapters 3 and 4. For both a high and low image quality TV, **sharpness** is mentioned significantly more often in the case of a small viewing distance than in case of a large viewing distance. Color and movement are not reported more often for either a small or a large viewing distance. It should be noted that there might be a systematic order effect in this study. People always began at the large distance and then moved to the small distance which could influence people in mentioning other factors as being determining for image quality.

Looking at the brightness, contrast and color measurements of the TVs, there seems to be a minor trend in the relation between the measurement values and the amount of times a factor was mention for specific TVs. This trend indicates that participants of this study were possibly able to correctly identify brightness and contrast as the determinant for their perception of high or low image quality. However, based on this study, this cannot be unambiguously determined since no significant results were found and the trend was not very reliable.

In our study, the color aspect was not found to be very dependent on viewing distance and hence, only the attributes brightness, contrast and sharpness are considered to be of highest importance. Therefore, for the sake of this study, brightness, contrast and sharpness are considered as the most important image quality attributes and these attributes will be examined further with the use of controlled psychophysical experiments.

3 Experiment 1: just noticeable difference

In this section, the methodology and the results of the first controlled experiment are presented and discussed. This psychophysical experiment was designed to answer the first research question: 'Is the JND of the separate image quality attributes dependent on viewing distance (and ambient illumination)?'.

3.1 Methodology

In order to reject or confirm the hypotheses related to this research question, a setup was designed that allowed to measure the JND of white level changes, black level changes and changes in sharpness of displayed natural images. As discussed in the introduction, white level and black level are physical image characteristics that influence the contrast and brightness of the displayed image.

3.1.1 Participants

The participants of this study consisted of a sample of 24 Philips workers and interns (12 male, 12 female) from different disciplines such as physiotherapy, engineering, marketing, etc. The mean age of the participants was 28 (S.D. 8.1). The youngest participant was 21 years of age and the oldest 54.
The participants were tested on their interest in new technologies with the use of a confidential Philips tool that assesses people's interest and assigns them to one of the following categories: front runner, trend setter, selectives, authentics, status seekers, classics, happy accepters, basics. This tool includes questions about their interests. In total there was one front runner, nine trend setters, six selectives, five classics and one happy accepter. This distribution might be not completely typical since most of the people who joined the experiment had an engineering background. Furthermore the participants were tested on their visual acuity with a Landolt C-chart and color blindness with the Ishihana color blindness test. All participants had a visual acuity of 1.5 or more according to the Landolt C test and had normal color vision.

3.1.2 Setup

The setup allowed for a just noticeable difference experiment at different viewing distances and in different ambient illumination conditions. It also evaluated the effect of content (images) on the JND. Pictures of the setup used can be found in Figure 12. The setup consisted of a TV on which the images were presented, one pc with a keyboard and software, a database of images used during the experiment and a chair at 1, 3 and 7 meters. Furthermore a uniform fluorescent ambient illumination with screen armatures was present. The ambient illumination level could be changed by the experiment leader with the use of a remote control.

Figure 12 The setup for the JND experiment consisted of a TV which could be viewed at different environmental conditions.

3.1.2.1 Equipment

The TV used was a Sony Qualia XDM-4000Q, 40" LCD TV with wide color gamut, LED backlight and a resolution of 1360 x 786 pixels. All processing on the TV itself was bypassed, and hence the incoming signal was never changed by the TV itself. The homogeneity in luminance of the TV was measured with an imaging colorimeter (i.e. the PM-1423E-1 from Radiant imaging) on a full screen white image. The measurement was done in a dark room, with the camera aligned exactly in the middle of the screen. The result of this measurement can be seen in Figure 13. The high frequency variation is due to aliasing between the camera pixels and the pixels on the Sony Qualia. It therefore is a measurement artifact and not actually present on the screen. The green line corresponds to a line measurement at the bottom of the screen from left to right. The red line represents the measurement result at the top of the screen and the blue line the measurement in the middle of the screen. It can be seen that the luminance profile is very uniform from left to right with a maximum difference of about 10 cd/m^2 at the very edges of the screen, which is acceptable.

Also the average luminance of the screen was measured for a totally saturated white image and a black image. The average luminance of the screen for the white image was 334.6 cd/m^2 and for the black image, 0.4 cd/m^2.

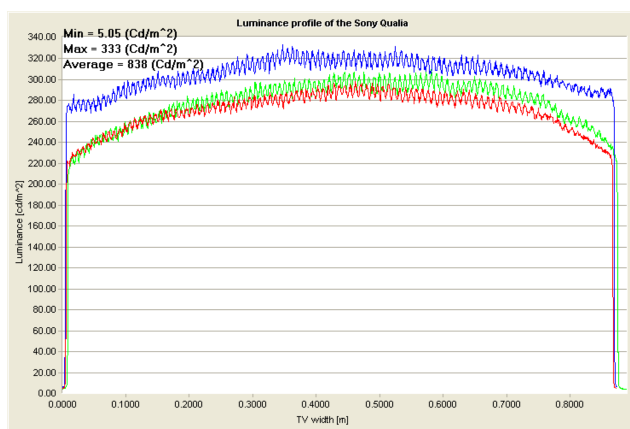


Figure 13 The luminance profile of the Sony Qualia measured in a horizontal direction on the TV with ambient illumination turned off.

3.1.2.2 The software

The software on the streaming PC was written in java and displays two mirrored images next to each other as can be seen in Figure 12. One of the images is the so called reference, which is the unprocessed version of the image. The other image is a processed image which is loaded from a database of 60 processed images on white level, back level and sharpness (see the next paragraph for an explanation about the processing and the images).

The software first presents the reference and the most processed image next to each other. Then it is programmed to go through a two-down-one-up staircase (Young Park et al., 2009), which converges to a 71% correct discrimination, by evaluating the input of the left and right arrow buttons of a keyboard which was connected to the streaming pc. When the button corresponding to the position of the reference is pressed twice (two up), a less processed image is shown. When however, the button corresponding to the processed image is pressed once (one down), a more processed image is shown next to the reference. This process is presented in Figure 14. On the trajectories going downwards, the button corresponding to the reference is pressed twice, and so a less processed image was presented next to the reference. On the trajectories goring upwards however, the button corresponding to the processed image is pressed once, and a more processed image is shown. During this staircase, the position of the reference is changed randomly. Whenever in the program a reversal from showing a less to a more processed image takes place, like at the numbers 5, 6 and 8 in Figure 14, the interval in the processing parameter decreases by 2 units. The interval in the processing parameter counts 12 before the first reversal. It than is changed to 10 at the first reversal, to 8 at the second reversal and so on. During the final stage of the staircase, the processing parameter is changed with one unit at a time. The starting value was 60.

The program constantly logs the counts at which reversals take place together with the number of the processed image at which this reversal take place. After 8 reversals, it calculates and logs the average of the processed image numbers at the final 6 reversals and then the program quits. Thus in case of the example given in Figure 14, the average is (7+18+10+13+9+10)/6= 11.17.

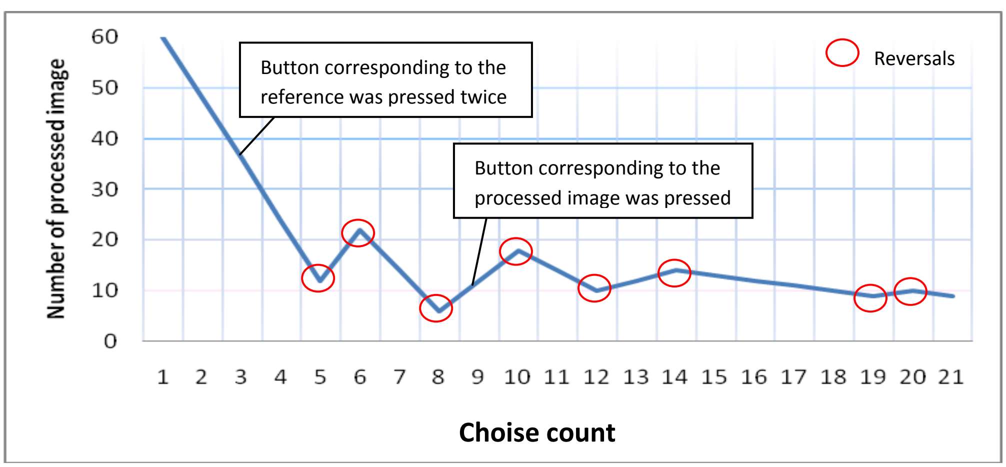


Figure 14 example of a staircase

3.1.2.3 The images and their processing

In the introduction it was mentioned that brightness and contrast can be changed in several different was, for example by changing the white and black level of a TV. Since it is practically difficult to change the white and black level of a TV in a fully controlled way during a perception experiment, it was decided to change the white and black level of the input signal instead, i.e. the image material by means of simple signal processing.

The images were selected on their ability to be adjusted in white level, black level and sharpness. For white level, it was important that the image contained a lot of bright grey levels. Figure 15 shows the grey level histogram for the image “Polar bear, i.e. one of the images selected to evaluate the JND in white level.

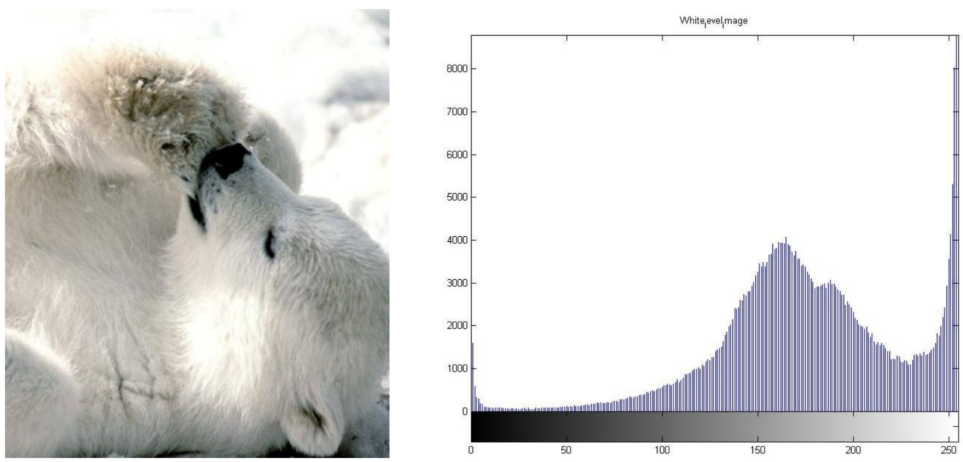


Figure 15 An original image together with its grey level histogram

The white level of this image can be changed by compressing all grey levels to a smaller range, i.e. reducing the highest grey value, while keeping the zero saturation point (black level) constant. In Figure 16 the resulting image is shown, together with its histogram. The peaks in the compressed histogram are a result of the quantization error, during grey level compression. In this image, which appears darker, the 255 white level is shifted to a ‘white level’ of R=G=B=195 and the rest of the grey levels is linearly compressed accordingly. This white level compression is done by means of a MatLab script which can be found in Appendix B. The equation used to generate the images is:

$$(RGB)_{new} = (RGB)_{old} \cdot \left(\frac{(255-i)}{255}\right)$$

This formula adjusts the white level by a value of ‘i’ and compresses the other grey levels accordingly. With the use of the MatLab script, 61 images were produced ranging in white level reduction ‘i’ from 1 to 60. These images were used in the experiment.

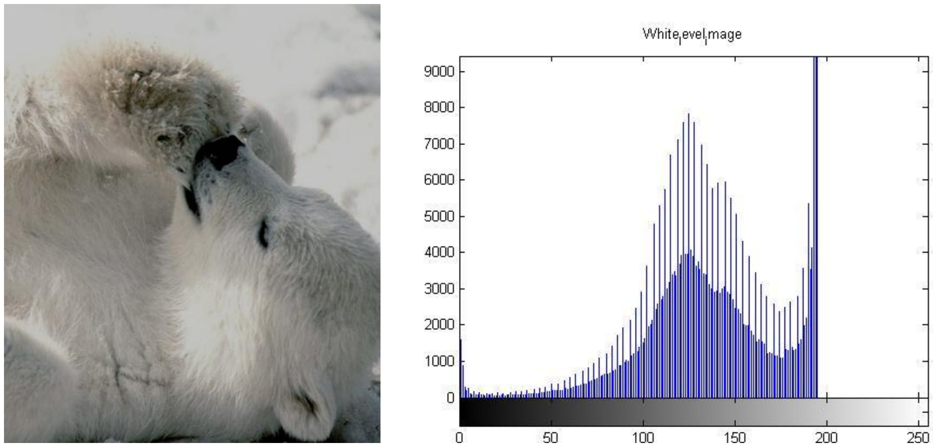
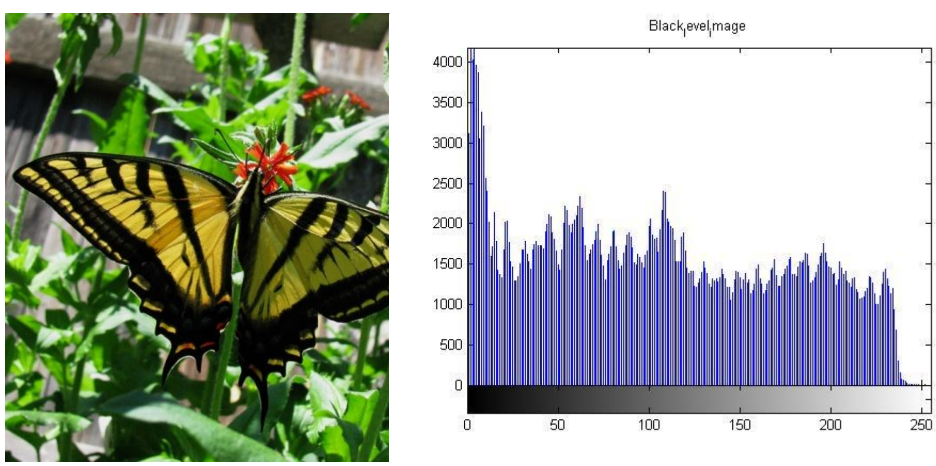


Figure 16 A processed image of which the white level is decreased by 60 grey levels together with its grey level histogram

For the black level, images that contain a lot of dark grey levels were selected. The same kind of processing was applied to generate images with various black levels. Figure 17 shows what happens to the histogram of an image when its black level is increased to an R=G=B value of 60, while the rest of the grey levels are linearly compressed accordingly. The black level compression is also implemented in a MatLab script which can be found in Appendix B. Here the equation used to generate the images is:

$$(RGB)_{new} = (BL)_{new} + (RGB)_{old} \cdot (\frac{(255-i)}{255})$$

In which $(BL)_{new}$ is the grey value of the new black level and 'i' represents the shift from the old to the new black level. Again, 61 images were produced ranging in black level increase from 1 to 60, and were used in the experiment.



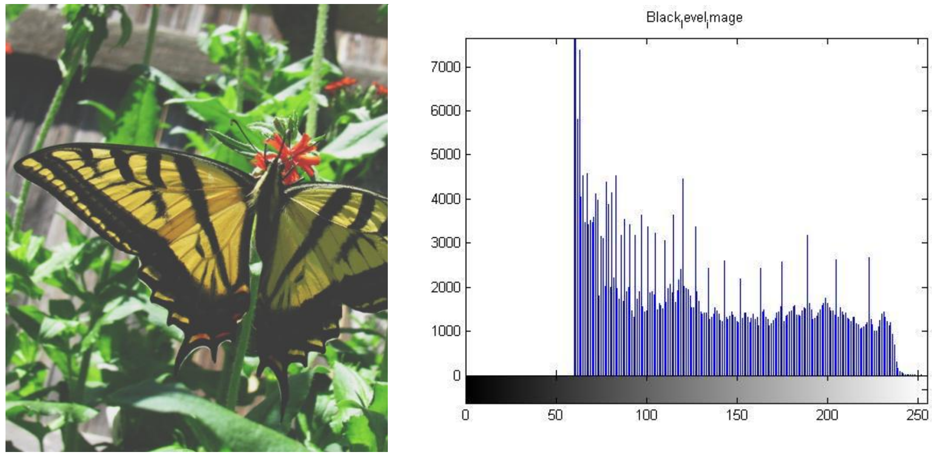


Figure 17 (top) an original image together with its grey level histogram, (bottom) a processed image of which the black level is increased by 60 grey levels together with its grey level histogram

Grey levels can be converted to luminance (in cd/m^2) for the sake of comparison with earlier research in which luminance values were used to express the JNDs. This can be done by using a lookuptable (LUT) which contains the luminance value for each grey level. This LUT is unique for every TV and was measured for the Sony Qualia. The LUT contains the translation of RGB values to normalized XYZ values measured at the screen so. To convert these XYZ values to luminance they need to be multiplied with the primary matrix, so that the resulting Y value represents the luminance in cd/m^2. The MatLab code used to convert grey levels to luminance values can be found in Appendix B. Figure 18 shows the luminance values against grey levels for the Sony Qualia. It can be seen that a change in white level in terms of RGB value elicits a relatively large luminance change as compared to the same change in RGB values for the black level.

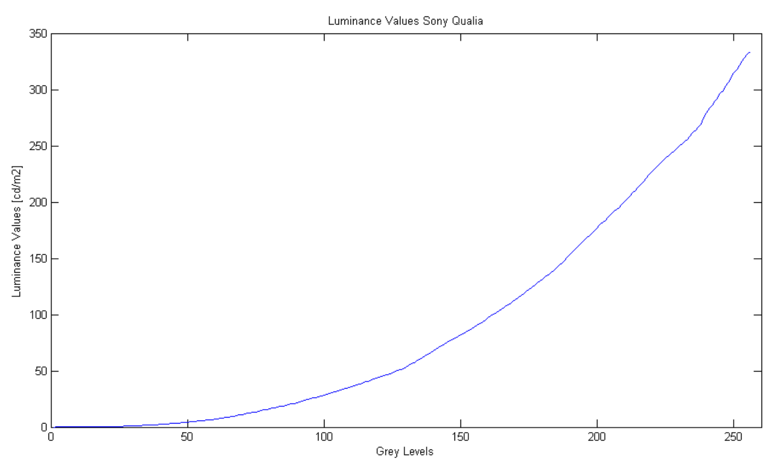


Figure 18 Luminance values of the Sony Qualia plotted against grey levels from 0 (white) to 255 (black)

For white level JNDs an extra action needs to be taken. Once the white level JND is found in terms of grey levels, the JND of luminance change has to be calculated by subtracting the calculated luminance from the maximum luminance which is 334.03 cd/m^2.

For evaluating differences in sharpness the image was chosen to be so that the original image was always the sharpest image and the rest of the images were blurred. The blurring was again done using a MatLab script (Appendix B: MatLab scripts) which contained a filter that convolved the image with a rotationally symmetric Gaussian low pass function. This filter reduced detail in the image by reducing the high frequency components. The images that were used in this experiment hardly contained any noise, so the noise reducing characteristic of the Gaussian filter in this case was reduced to a minimum. The Gaussian function which was used to blur the images is defined by

$$G(x, y) = \frac{1}{2\pi\sigma^2} e^{-\frac{x^2+y^2}{2\sigma^2}}$$

in which the x and y are the horizontal and vertical dimension of the filer (in pixel units). Sigma is the standard deviation of the distribution also expressed in pixel units. For this experiment, the Gaussian filter was used to produce 61 images. For all images the size of the distribution was kept constant at 25 by 25 pixels, but the sigma ranged from 0.05 to 3.05 with intervals of 0.05. In case of a small sigma, the image is blurred only minimally since the Gaussian filter limits almost no (high) frequencies. As the sigma gets larger the image is blurred more since an increasing number of frequencies are filtered by the function. In Figure 19 an original image, the image blurred with a Gaussian filter with sigma 1.50 and the image blurred with a sigma of 3.00 are shown. It can be seen that the details in the image become less visible reduced with increasing sigma.

The sigma value can be converted to visual angle, again for sake of comparison with other research. For this, the sigma value is multiplied by the visual angle of 1 pixel. The visual angle is calculated using the pixel width of the Sony Qualia, which is 0.0595 cm. This gives a visual angle for 1 pixel of 3.4081° at 1 meter viewing distance, of 1.1363° at a viewing distance of 3 meters and of 0.48701° at a viewing distance of 7 meters. So when, for example, the JND of sharpness at 1 meter viewing distance in terms of sigma value is 0.4, the visual angle JND is 0.4*3.4081° = 1.363°.

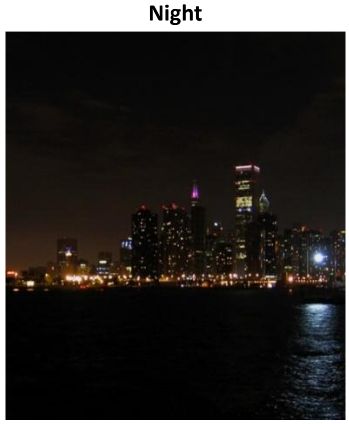
For every image quality factor tested in the JND experiment, 2 images were used. These images are presented in

presented in

Table 3. For all these images, 61 versions with reduced white level, increased black level and reduced sharpness, were produced using the MatLab scripts.

Figure 19 (left) an original image, (middle) a blurred image with a Gaussian function with sigma 1.5, and (right) a blurred image with a guassian function with sigma 3.0

Table 3 Images used in the JND experiment

White level	Black level	Sharpness
Snow	**Night**	**Terrace**
		
Polar bear	**Butterfly**	**Parrot**
		

3.1.2.4 Ambient illumination

The room was illuminated by fluorescent lamps in screen armatures. The room in which the experiment was done is used for perception experiments regularly and is especially designed to realize a uniform illumination. Since the TV was placed exactly in the middle of the room and parallel to the wall also the illumination pattern on the TV should be uniform. This is checked by comparing the measurements of the luminance distribution on the TV with the ambient lighting turned on (at 400 lux measured perpendicular to the middle of the screen), as given in Figure 20 to the luminance distribution with the ambient lighting off which was already given in Figure 13. Apart from the luminance values being on average 10 cd/m^2 higher, no differences were found indicating that the illuminance of the ambient on the screen was very uniform. Note that the high frequency variations are again a result from the interference between the camera resolution and the TV resolution.

Two ambient illumination settings were used in the experiment, namely 30 and 400 lux measured perpendicular at the middle of the screen. These two conditions were used to determine the influence of ambient illumination on the JND of black level, white level and sharpness.

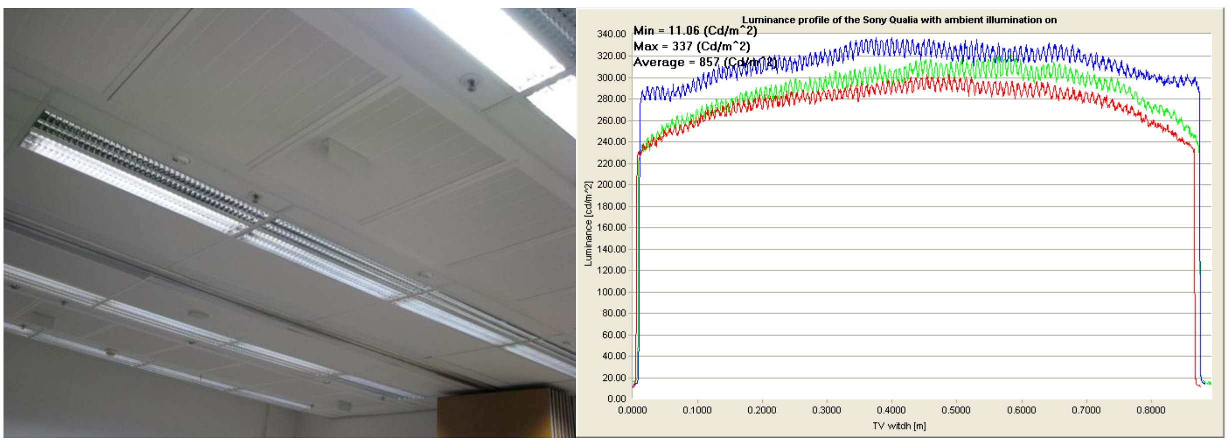


Figure 20 Ambient illumination in the experiment room on the left, and the measurements of the luminance profile of the Sony Qualia with ambient illumination at 400 lux on the right

3.1.3 Design

To be able to evaluate the influence of the independent variables viewing distance (N=3), ambient illumination (N=2) and content (N=2) on the dependent variable JND for three different image quality attributes, a list of (3x2x2x3)=36 conditions was made for each of the 24 participants before the experiment. In this list the ambient illumination, image order and attributes under test (white-, black level, sharpness) were balanced. The order of distance was varied randomly between participants and lighting conditions. A table presenting all the conditions for each participant can be found in Appendix C.

A program was written that guided the participants through the steps of ambient illumination, viewing distance, image quality factor and image. It selected the appropriate staircases and images belonging to a given condition.

3.1.4 Procedure

A participant was invited to join an experiment about visual perception of natural images with the information that it would take about one hour of their time. In case they liked to join, they were invited in the experiment room at a specific time and date.

Once inside the room, they were asked for their age and background. This was noted by the experimental leader together with their gender and the color of their clothes (to be able to evaluate possible reflection on the screen influencing the JND). Then they were tested on their visual acuity by means of a Landolt C test in the Netherlands. This was done with an ambient illumination of 400 lux. Furthermore they were tested on color blindness using the Ishihara test. Finally, before starting the experiment they were asked to answer a couple of questions to evaluate their interest in technology. These questions are generated by Philips and confidential.

Once they finished the questions, the actual JND experiment began. Participants were presented with the following question on the screen.**' indicate the image which contains the whitest white (for the white level images), the blackest black (for the black level images) or which image is sharpest.'** Before every staircase, the corresponding characteristics of the condition (ambient illumination, viewing distance, image quality attribute under test) were shown on the screen. The experimenter adjusted the lighting to the appropriate level and the participant had to move to the correct distance and perform 2 staircases on this distance for all three image quality attributes.

Next, participants had to press a button to continue with the first staircase. Participants could select the image of their choice using the left and right arrow key. The reference was always the whitest, blackest or sharpest image. When people indicated the reference to be the whitest, blackest or sharpest image twice in a row, a less processed image was shown next to the reference in a subsequent evaluation step. This means, that participants had to be correct twice before a less processed image was shown allowing for a more reliable JND. When the participants selected the processed image as being the reference, a more processed image was shown next to the reference in the subsequent evaluation step. This process was repeated until eight reversals from more to less processed or visa versa. The resulting JND, being the average over the last six reversals was logged before continuing to the next staircase. Hence, for every participant 36 JND data points were logged.

Once the participant performed the first 18 staircases (the first lighting condition), they were asked to take a 5 minute break before continuing to the next lighting condition. When they were finished with all 36 staircases, they received a small incentive (candy) to thank them for their participation.

3.2 Results

With the experimental setup discussed above, the JND at different viewing distances, ambient illumination conditions and for different content was found for all 24 participants. This was done for the three different image quality attributes; white level and black level (for brightness and contrast) and sharpness. In case of the white and black level JND, this difference is expressed in terms of luminance in cd/m^2. In case of sharpness, the JND is expressed in terms of sigma, being the width of the Gaussian blur filter. A univariate ANOVA (full factorial for the independent variables image, viewing distance and ambient illumination) was used to investigate the effect sizes and significance levels of the independent variables. Graphs were made to clarify these effects. Furthermore, for each image quality attribute the JND and its 95% confidence interval are given.

3.2.1 White level JND

For white level, small and non-significant effects of viewing distance ($F(2,48)=0.985$, $p=0.381$, $\eta^2=0.039$), ambient illumination ($F(1,24)=1.912$, $p=0.180$, $\eta^2=0.074$) and content ($F(1,24)=1.235$, $p=0.277$, $\eta^2=0.049$) were found. This can also be seen in the 95% confidence interval graphs of the white level JND related to these independent factors as shown in Figure 21. In this graph all error bars, for both images, overlap. This again shows that the differences in JND as a function of viewing distance, luminance and content are indeed insignificant.

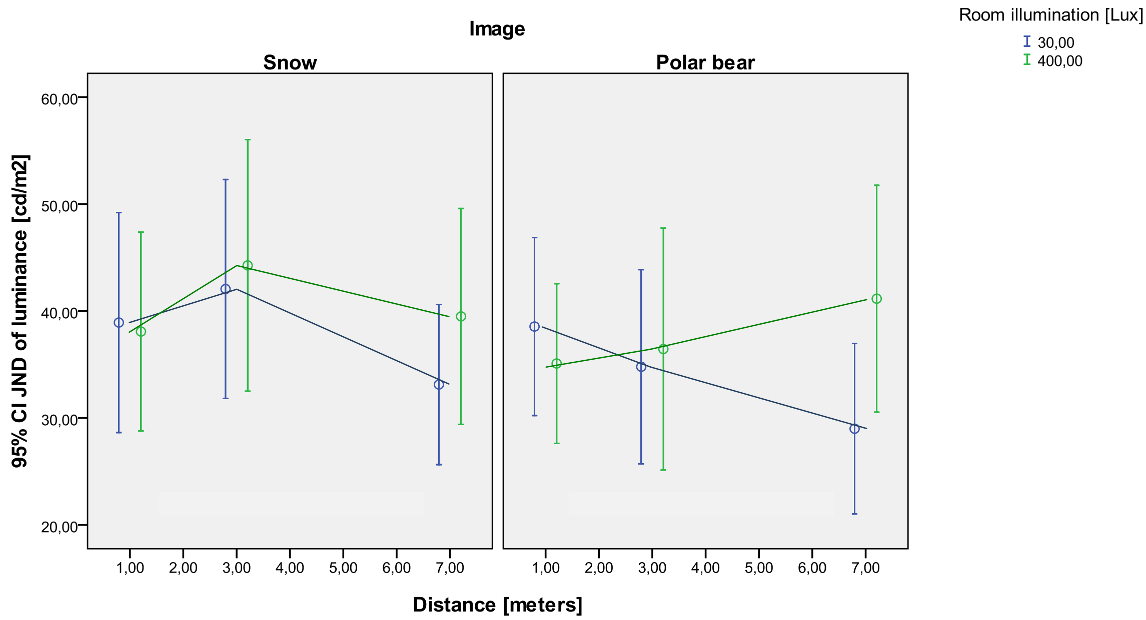


Figure 21 Mean white level JNDs in terms of luminance [cd/m^2] plotted against the viewing distance. The different lines represent the ambient illumination settings of 30 and 400 lux.

Since no significant effects of the environmental variables were found, only the overall mean JND for white level is determined. This JND in terms of luminance is **37.6 ± 2.6 cd/m^2 (CI 95%)**.

3.2.2 Black level JND

For black level, all three independent factors had a significant and large effect on JND. The viewing distance had a large and significant effect on the black level JND ($F(2,48)=9.899$, $p<0.001$, $\eta^2=0.292$). Furthermore, a Tukey post hoc test for viewing distance showed that the difference in JND for the distances 1 and 3 meters ($p<0.05$) and 1 and 7 meters ($p<0.001$) were significant, but the difference in JND for the distances 3 and 7 meters was not significant ($p=0.230$). The JND as a function of distance is presented in the graphs of Figure 22. The confidence interval for the JND at the different distances is shown in Table 4.

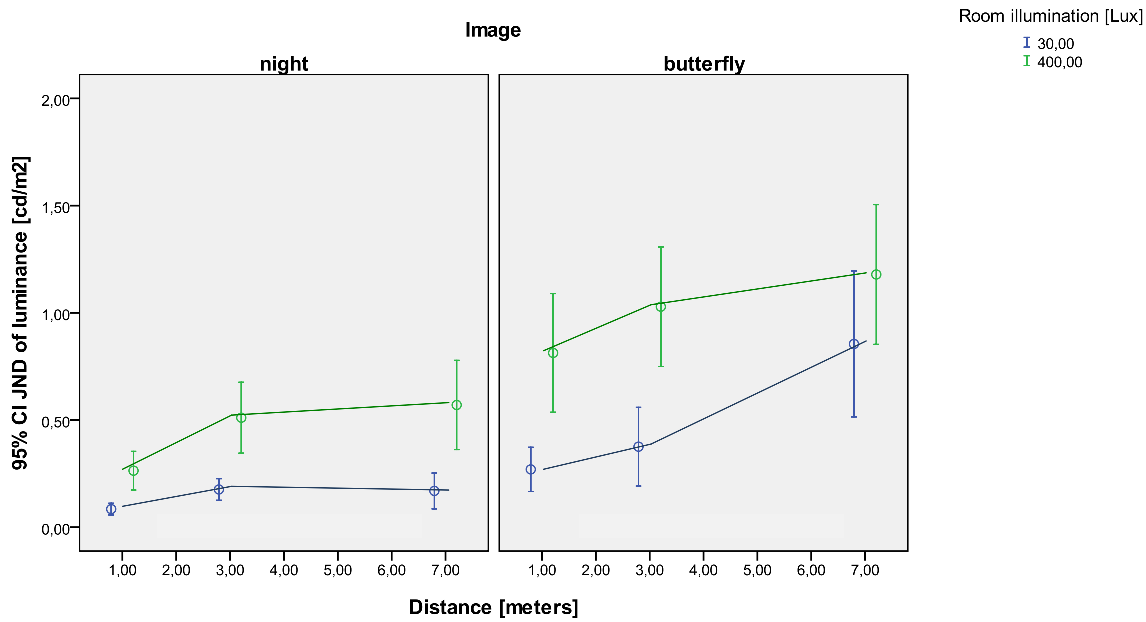


Figure 22 Mean black level JNDs in terms of luminance [cd/m^2] plotted against the viewing distance. The different lines represent the ambient illumination settings of 30 and 400 lux.

Also the ambient illumination had a large and significant effect on the JND ($F(1,24)=61.739$, $p<0.001$, $\eta^2=0.720$). This can also be seen in the graphs of Figure 22. In these graphs, the blue line corresponds to the JND measured at 30 lux and the green line to the JND measured at 400 lux. Hence, the JND for black level becomes larger for a brighter room illumination. The confidence interval for the JND at the ambient illuminations is shown in Table 4.

For the black level, also the content which is presented during the experiment had an effect on the JND ($F(1,24)=51.632$, $p<0.001$, $\eta^2=0.683$). This can again be seen in the graphs of Figure 22 by comparing the graph of the image “night” to the graph of the image “butterfly”. The mean JND for the image “butterfly” is larger than for the image “night”. The confidence interval for the JND of the different images is shown in Table 4.

Table 4 JNDs for black level

Variable	Value	Mean JND [cd/m^2]	Lower bound [95% CI]	Upper bound [95% CI]
Viewing	1 meters	**0.36**	0.27	0.45
	3 meters	**0.52**	0.41	0.63
	7 meters	**0.69**	0.55	0.84
Ambient	30 lux	**0.32**	0.24	0.40
	400 lux	**0.73**	0.62	0.83
Image	Night	**0.30**	0.24	0.35
	Butterfly	**0.75**	0.64	0.87

The overall mean JND for black level in terms of luminance is **0.52 ± 0.07 cd/m^2 (CI 95%)**. This is obviously significantly lower than the JND for white level of 37.6 ± 2.6 cd/m^2 (CI 95%).

3.2.3 Sharpness JND

For sharpness, the ambient illumination did not have a significant effect on the JND ($F(1,24)=0.013$, $p=0.912$, $\eta^2=0.001$), but the other two factors did. Distance had a very large, linear effect on the JND ($F(2,48)=78.549$, $p=0.000$, $\eta^2=0.766$) and a post hoc test indicated that the JND significantly differed for all distances ($p<0.005$). The JNDs are presented in the graphs of Figure 23 (in terms of the number of the images) and the confidence intervals for the JNDs are presented in Table 5. Here, the mean JND for sharpness is expressed in terms of the sigma used in the Gaussian blur filter. Hence the JND for the graphs is different from the JNDs in the table.

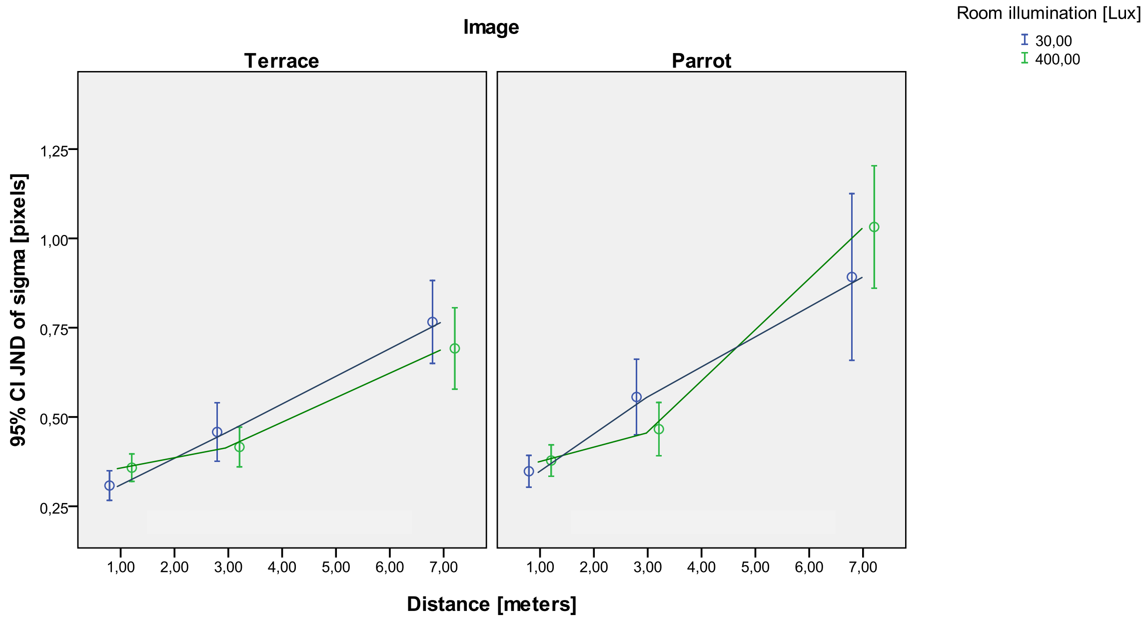


Figure 23 Mean sharpness JNDs in terms of sigma [pixels]. The different lines represent the ambient illumination settings of 30 and 400 lux.

The content also had a significant effect on the sharpness JND ($F(1,24)=13.341$, $p=0.001$, $\eta^2=0.357$). This is also shown in the graphs of Figure 23. The JND for the image “Parrot” is larger than for the image “Terrace” especially at a viewing distance of 7 meters. Furthermore the JND for the image “Parrot” seems to increase more rapidly with distance (the slope is steeper) than for the image “Terrace”.

Table 5 JNDs for sharpness

Variable	Value	Mean JND in terms of sigma	Mean JND in visual angle	Lower bound [95% CI]	Upper bound [95% CI]
Viewing distance	1 meters	0.35	**1.186**	0.33	0.37
	3 meters	0.47	**0.539**	0.43	0.51
	7 meters	0.85	**0.485**	0.76	0.93
Image	Terrace	0.50	**0.876**	0.46	0.54
	Parrot	0.61	**1.073**	0.55	0.68

The overall JND for sharpness in terms of sigma [pixels] is 0.56 ± 0.04 (CI 95%) pixels.

3.3 Discussion

The difference between the overall white level and black level JND in terms of luminance as found in chapter 3.2.1. and 3.2.2. respectively is supported by earlier literature in the sense that for example Fechner's law (Stevens, 1957) states that the eye is more sensitive to luminance differences in dark areas than in bright areas.

In a study by Heynderickx et al. (2007) also natural images were used to find the JNDs of black level and white level with a staircase experiment. They used a single viewing distance of 1.2 meters and a fixed ambient illumination of 20 lux. Hence, their trends and values would be very well comparable to our values obtained at a viewing distance of 1 meter and an ambient illumination of 30 lux. First of all, they found a similar trend as in our experiment, namely that the JND for black level was much smaller than the JND for white level when expressed in luminance. Secondly the values were very similar. The average white level JND found in our study for the condition mentioned above was 38.7 ± 3.2 cd/m^2 as compared to a value ranging from 30.8 – 41.9 cd/m^2 as found by Heynderickx. The black level JND in this condition found in our study was 0.177 ± 0.03 cd/m^2 as compared to a value ranging from 0.19 – 1.52 cd/m^2 as found by Heynderickx. Also the sharpness JND of 0.33 ± 0.01 pixels found in our study corresponds to the sharpness JNDs found by Heynderickx ranging from 0.50 – 57 pixels. In another study by Heynderickx et. al. (2004), the sharpness sigma JND of viewing natural images was found to range from a value of 0.3 pixels to 0.8 pixels, which again corresponds to the values found in our experiment.

Our values can also be compared with the DICOM grayscale standard display function (Samei, 2007). This curve specifies a standard display function taking the non-linear nature of human perception into account (for greyscale images), as measured by a side-by-side comparison task for homogeneous images. In this function, the JND of luminance is plotted against luminance and therefore can be used as comparison with our values. For the white level images used in our experiment, the luminance output of the TV was on average 320 cd/m^2. According to the DICOM curve, the JND for white level would be about 3.80 cd/m^2 and for black level, with an average light output of about 1 cd/m^2, the JND that follows from the DICOM curve is 0.05 cd/m^2. These values are obviously much lower than the average JNDs found in our study (38.7 cd/m^2 for white level and 0.177 cd/m^2 for black level), but since natural content was used in our experiment, which also contained intermediate grey levels, this was to be expected. Intermediate grey levels are effected by white- and black-level changes to a lesser extend.

Below, the effects of the environmental aspects, distance, content and ambient illumination, on the JNDs of the different attributes is discussed.

3.3.1 White level

From this experiment it was found that none of the effects of viewing distance, ambient illumination and content on the white level JND were significant. This means that it would not have any additional value to adjust the size of the white level changes with different environmental conditions since changes in white level don't become more or less visible at different viewing distances or with different room illuminations; at least for images comparable to the ones used in this experiment.

No effect of distance on the visibility of white level differences was found. This disagrees with the fact that at a smaller distance (i.e. broader visual field), more cones and rods would be active allowing for more sensitivity in changes in brightness and contrast (Blake, 2006).

The absence of an effect of ambient illumination on the JND of white level can possibly be explained by the fact that for these white (and bright) images the higher ambient illumination hardly affected the appearance in brightness and contrast of the image since the actual percentage of luminance increase was very low. This means that for the visibility threshold of white level changes, the ambient illumination should not be taken into account in case of images with a high overall brightness like the images used in this experiment.

The absence of an effect for content can be explained by the fact that the images used in our experiment looked very much alike. They both contained mostly bright grey levels and only minor intermediate and dark grey levels. Therefore they were both influenced by the grey level changes by a similar extend. With other content, the content independency would maybe disappear as indicated in the results for black level were we did use two images that were less similar in their grey level histogram.

3.3.2 Black level

For black level, the just noticeable difference increases with larger viewing distances. This means that changes in contrast and brightness as a result of a change in black level would be less visible at a large than at a small viewing distance. Therefore, bigger steps in black level have to be made at a large viewing distance to make the difference visible. Interestingly, the extensive document for the 'assessment of display performance for medical imaging systems' (Samei, 2002), in which they strive to make a standard for imaging systems to make the images presented appear the same, does not discuss the influence of viewing distance. From our experiment however, it appears that the viewing distance would have an effect on the way brightness and contrast are perceived. Therefore this environmental characteristic should be taken into account when making images appear alike in different environments.

Also content had an effect on the JND of the black level. For image "butterfly", the mean JND was a lot higher than for the image "night". This was expected since the image "night" contains mostly dark grey levels resulting in a lot of change in brightness and contrast when the black level changes. The image "butterfly" on the other hand, contains grey levels more evenly spread over the histogram, meaning that changes of the black level have relatively little impact on the brightness and contrast appearance of the image. A content dependent effect was also found in the study by Heynderickx et al. (2007). They found indications that the JND became substantially larger for images with a lot of intermediate grey levels, which is in agreement to the findings of our experiment.

This effect of content means that when the black level has to be changed perceptually (for example different black levels could be preferred at different distances), the content should be taken into account. Possibly an algorithm can be designed that analyzes the grey level histogram of an image and adjusts the black level changes accordingly. It should be noted that the images used to evaluate the JND in black level were more different in their grey level histogram than the images used to evaluate the JND in white level. This could possibly explain the insignificant effect of content when measuring white level. Furthermore we would like to remark that as a consequence of the content dependency, the values found for the JND in black level are only valid within this experiment, i.e. for the images used in this test. To be able to get a good average of the JND in black level, the experiment needs to be repeated with many more images.

At higher room illuminations, black level changes were less noticeable. This means that the black level would have to be changed more when room illumination is high to provide a similar difference in perception. This also is connected to the amount of ambient illumination falling on the TV. Adding 10 cd/m^2 to 350 cd/m^2 (for the white level test) is negligible, while adding 10 cd/m^2 contributes much to, for example, 1 cd/m^2 (for the black level test).

3.3.3 Sharpness

A very large effect of distance on the JND of sharpness was present. As the distance increased, the sharpness changes were far less visible. The relation of JND with viewing distance seemed to be linear, which corresponds to the basic visual acuity theory that states that it is harder to discriminate between two points when distance increases (Blake, 2006). This means that steps in sharpness have to be larger at a larger distance to be perceivable.

Furthermore, content has an effect on the JND of sharpness. The sharpness differences in the image of the parrot were less visible than the terrace image. This might be due to the clearly defined edges of the chairs on the terrace. During the experiment participants indicated that they used these edges mostly in their determination of the sharpest image. The experimenter did not ask for this specifically, but this was what people mentioned during the test.

Room illumination did not have any effect on the JND of sharpness. This might be a result of the fact that the images used for this test were all quite bright images when being displayed on the Sony Qualia. This means that only a small percentage of increase in the actual luminance of the screen was present. Moreover, no effect was expected to be found since the eyes cones allowing for sharp vision are already active at 10^{-2} cd/m^2 (Blake, 2006), way below the luminance value of the screen presenting the images.

3.3.4 Discussion of hypotheses

Is the level of JND for a specific image quality attribute dependent on viewing distance?	The JND of brightness is larger at a large distance than at a small distance, since the area of the eye perceiving the change is larger at small viewing distances. More cones and rods in the human eye respond to the signal (Blake, 2006).
	The JND of contrast is larger at a large distance than at a small distance, since the area of the eye perceiving the change is larger at small viewing distances. More cones and rods in the human eye respond to the signal (Blake, 2006).
	The JND of sharpness is larger at a large distance than at a small distance, since the visual angle is smaller at large viewing distances, making it harder to discriminate between two points (Blake, 2006).

As the JND of the white level, did not depend on viewing distance, the hypotheses of brightness and contrast will be rejected or accepted based on the results for the JND of the black level. Here, we will use the fact that (1) contrast gets smaller when the black level is adjusted to higher grey levels (since the dark grey levels become closer to the lighter grey levels) and that (2) the brightness increases for higher black levels (since the averaged luminance increases). Since the black level JND became larger with increasing distance, also the JND of brightness and of contrast becomes larger with increasing viewing distance which confirms the hypotheses. For sharpness the hypothesis that the JND of sharpness changes is smaller on small viewing distances can be accepted based on the results.

Is the level of JND for a specific image quality attribute dependent on ambient illumination?	The JND of brightness is larger with a larger room illumination (Samei, 2002)
	The JND of contrast is larger with a larger room illumination (Samei, 2002)
	The JND of sharpness is not dependent on room illumination (Graham, 1965)

For ambient illumination, effects were found for black level with respect to the ambient illumination but they were not found for the white level. The JND of black level is larger at higher room illumination, which means that the JND of brightness and the JND of contrast are also higher at higher room illumination, so the hypotheses can be accepted.
The hypothesis on the JND of sharpness can also be confirmed based on the results since no significant effect of ambient illumination on the JND of sharpness was found for the images used in this experiment.

Is the level of JND for a specific image quality attribute dependent on content?	The JND of brightness is dependent on content (Heynderickx, 2007)
	The JND of contrast is dependent on content (Heynderickx, 2007)
	The JND of sharpness is dependent on content (Radun, 2007)

Since the JND of black level is dependent on content, while the JND of white level is not, also the JND of brightness and the JND of contrast needs to be dependent on content. The JND of sharpness also depends on content so that also this hypothesis is confirmed.

It should be noted that the conclusions about the JND of brightness and contrast can vary if in future studies it is found that, despite the results found in our experiment, there is a significant effect of viewing distance, ambient illumination or content on the JND of white level. When this is the case, both aspects (i.e. the effect of viewing distance on both black level and white level) can possibly interact and compensate each other.

4 Experiment 2: Attribute preferences

From the previous experiment it is now known to what extent the physical image characteristics have to be changed to elicit a perceptual difference. What is not known is what the optimal settings for contrast, brightness and sharpness are for different settings of the environmental variables ambient illumination, viewing distance and content. This will be answered in this chapter. The methodology and the results of the second controlled experiment are presented and discussed here. This psychophysical scoring experiment is designed to answer the second research question: ‘Are the settings of the image quality attributes belonging to an optimal perceived image quality dependent on viewing distance?’.

It should be noted that for this research question, the optimal settings for image quality were determined via actual algorithms in commercially available TVs. This implies that the various physical image characteristics were less independently controlled than in chapter 3, discussing the JND experiment. This choice was made to make our results more useful for implementation in the current TV sets which vary the dynamic contrast of the display.

4.1 Methodology

In order to reject or confirm the hypotheses of this research question, a setup was designed that allowed participants to score the image quality of displayed natural images with various contrast and sharpness levels. Contrast and sharpness were chosen as image quality factors to be evaluated, because these two attributes relate to the black level, white level and sharpness which were also evaluated in the previous JND experiment. As mentioned above, contrast and sharpness were each changed via a specific parameter in a processing chain of a TV, as will be explained in more detail in chapter 4.1.2.3.

4.1.1 Participants

The participants of this study consisted of a sample of 24 Philips workers and interns (12 male, 12 female) from different disciplines. The mean age of the participants was 25.99 (S.D. 3.707). The youngest participant had an age of 21 and the oldest was 35. 10 Participants also joined the first experiment.

The participants were tested on their interest in new technologies with use of a confidential Philips tool that assesses people’s interest and assigns them to one of the following categories: front runner, trend setter, selectives, authentics, status seekers, classics, happy accepters, basics. This tool includes questions about their interests. In total there was one trend setter, ten selectives, three authentic, two status seekers, 4 happy accepters and 4 basics.

Furthermore the participants were tested on their visual acuity using a Landolt C-chart and color blindness using the Ishihana test. All participants had a visual acuity of 1.5 or more and had normal color vision.

4.1.2 Setup

Pictures of the setup can be found in Figure 24. This setup consisted of two Philips 98903H LEDLUX TVs. The TVs were controlled by means of a PC with software that could manipulate parameters of image processing algorithms used in the TV set to improve picture quality. This PC is called the TV signaling PC and was connected to the TVs using a USB Multimedia connection. A database of images to be used during the experiment was stored on an external streaming media device. This device was connected to the TVs via an HDMI splitter and HDMI cables. Another PC was connected to the signaling PC via LAN. This PC had a keyboard, mouse and monitor and was used to operate the TV signaling PC remotely. Furthermore a chair at 1, 3 and 7 meters and a uniform fluorescent ambient illumination with standard screen armatures were available.

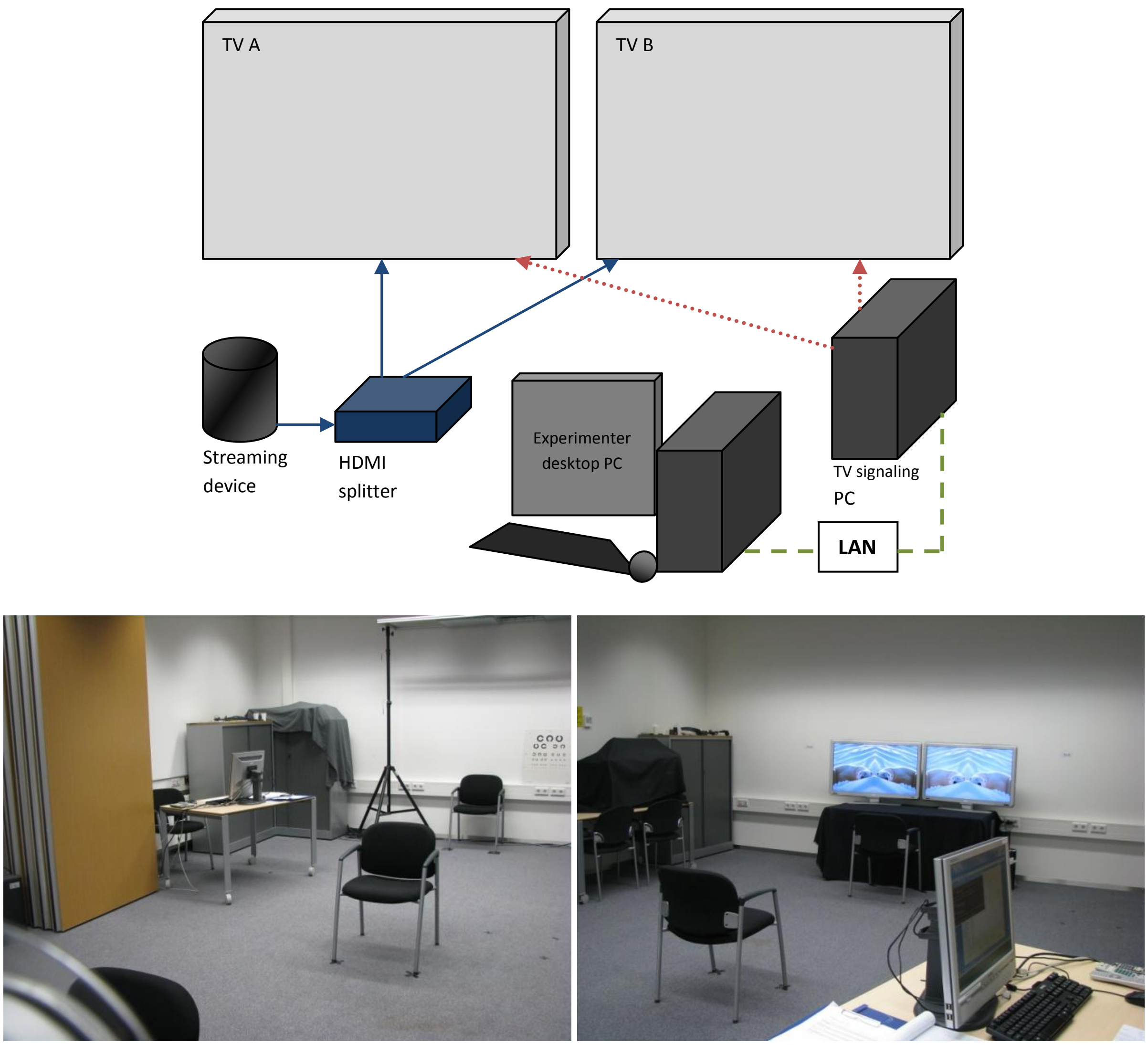


Figure 24 the setup of the scoring experiment. Top: a schematic representation, bottom: pictures of the setup

4.1.2.1 Equipment

The two TVs were identical 42" Philips LEDLUX TVs with a resolution of 1920 x 1080 pixels. They were tested on their equality by the Philips Consumer Lifestyle department by means of a small perception experiment in which people were asked about the differences between the two TVs for a limited set of image material. Based on this experiment, no differences between the two TVs were reported. All processing in the TVs was turned off, hence the incoming signal was never processed by the TV itself without controlled intervention of the experimenter. Since the TVs were reported to be identical, homogeneity was tested for one TV only in the same way as was done for the Sony Qualia which was used in the JND experiment. In Figure 25, the result of this homogeneity test on a totally saturated white screen can be found. The green line represents a measurement at the bottom of the screen. The red line gives the measurement at the top of the screen, while the blue line is a measurement in the middle of the screen. It can be seen that the luminance profile is very uniform with a maximum difference of about 20 cd/m^2 at the very end of the middle of the screen, which is about 3.5% of the maximal luminance, and therefore acceptable (a change of 10% is considered to be acceptable). Figure 25 also shows that apart from the peaks the screen is quite uniform in luminance. The peaks are the result of the segmented backlight of the new Philips LEDLUX TVs.

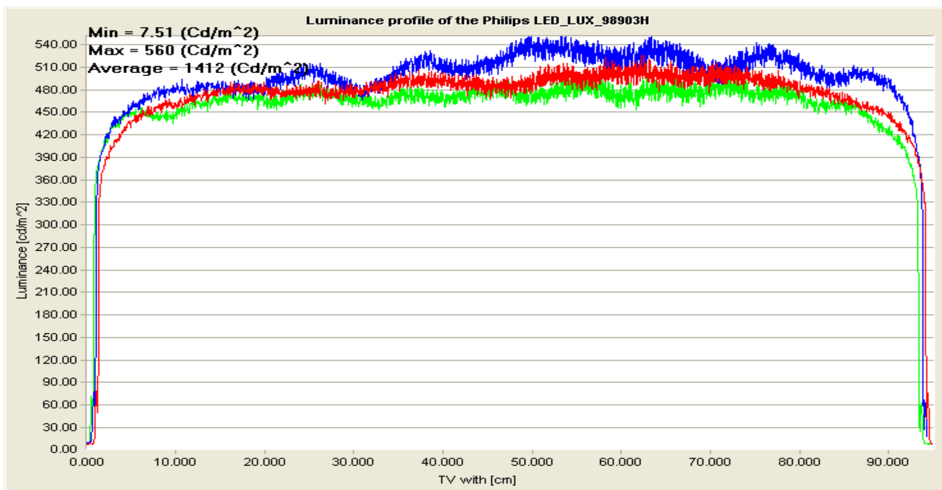


Figure 25 Luminance profile of the Philips LED_LUX as measured with ambient illumination turned off

The images used in the experiment were loaded on a Dvico TivX 5130 HD Media player, which provided easy access to the digital images used for this experiment. This media player was connected to the TVs HDMI connection via an HDMI splitter.

4.1.2.2 The software

The software on the TV signaling PC consisted of three cooperating parts. One part was a program that was able to send signals to the internal algorithm of the TV to adjust the contrast settings and sharpness algorithm settings of the TV. More specifically and as discussed in the introduction, sharpness is enhanced by a peaking algorithm, boosting the high frequencies in an image. As a consequence the sharpness algorithm is hereafter also referred to as the peaking algorithm. This program is called the TV setting program. Another part was a script which allowed for easy adjustment in the TV setting program during the experiment. This is called the adjustment script. To make the experiment even more resistant to possible errors made by the experimenter, a third script was written that automatically gave the correct input for the adjustment script, just by pressing the ‘space’ button once for each stimulus.

4.1.2.2.1 TV setting program

In the TV setting program, that was developed for the algorithm chip in the LEDLUX TVs, values could be inserted which were loaded to the algorithms in the TV. For each algorithm, the program had different tabs were the values could be changed. The only algorithms that were adjusted during this experiment were the contrast algorithm and the peaking algorithm.

4.1.2.2.2 Adjustment script

This script asks for the type of experiment that should be done, i.e. the sharpness experiment (then key 1 should be pressed) or the contrast experiment (key 2 should be pressed). When either of the two experiments was selected, the choice was communicated to the TV setting program so that it could go to the corresponding tab. Then, the script asked which TV should be the reference, TV A (key l) or TV B (key p), meaning that this TV should remain constant at the middle value for the sharpness or contrast setting during the experiment. Finally, by pressing the keys q, w, e, r or t different contrast and sharpness settings were loaded to the TV setting program. By pressing n, the program resets all settings of the TV setting program and restarts.

4.1.2.2.3 Automating script

This script executes the keys of the adjustment program and generates a predefined series of numbers and letters. When executing the script it asks for the number corresponding to the rule of the series in a text-file that it has to execute. For every participant a series of numbers and letters were predefined and included in this text-file, according to the balanced list of conditions which is discussed in chapter 4.1.3.

4.1.2.3 The images and the TV processing settings

For this scoring experiment, the images were selected on a couple of aspects. To be able to indicate which aspects of the images were important, first the sharpness and contrast processing that is done in the TV needs to be is explained.

Sharpness

For sharpness, 5 different settings were used in the experiment for both the adjusted TV and the reference TV. These settings were chosen in cooperation with Philips image quality perception experts so that they represented realistic values, varied enough to make differences perceptually, and that intervals between the settings were approximately perceptually equal. The peaking algorithm was used in profiles 4 and 5 to increase the sharpness of the edges with respect to the original image. Gaussian blurring was used for profiles 1 and 2 to blur the image. Therefore, for images that were used in the sharpness experiment, it was important that they included interesting features that were positively or negatively influenced by either peaking or blurring. In the adjustment script, pressing key ‘q’ makes the reference TV go to sharpness setting three (corresponding to the middle setting) and the adjusted to setting 1 (the most blurry profile), pressing ‘w’ means adjusted TV setting 2, reference: setting 3, pressing ‘e’ means adjusted TV setting 4, reference: setting 3, pressing ‘r’ means adjusted TV setting 5, reference: setting 3, and when ‘t’ is pressed both the reference as the other TV are adjusted to setting 2 in order to evaluate a zero measurement which will be further discussed in chapter 4.1.3.

The three images used for the sharpness experiment because of features sensitive to peaking or blurring are shown in Figure 26. These features are artifacts mosquito noise mentioned in the introduction.

Talkshow **Marathon** **Meeting**

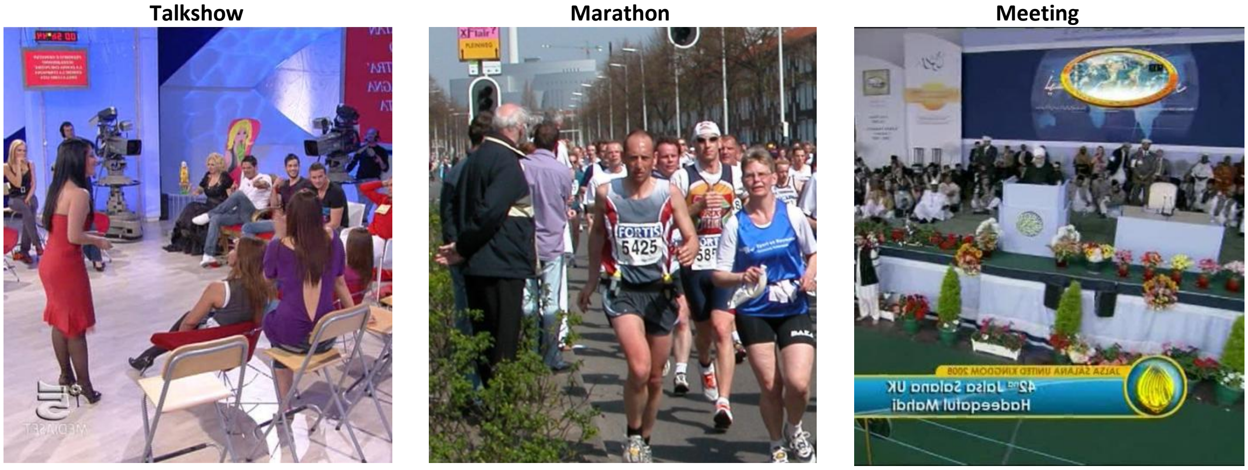


Figure 26 Images used for the sharpness experiment

Since with this type of processing the level of sharpness cannot be characterized with a single, simple value, the effect of the sharpening manipulations is illustrated via extracting parts of the processed images. Additionally, a short explanation of what becomes more or less visible with the various sharpness manipulations is given. The opposite is true for increased peaking. In row 1 of Figure 27 it can be seen that the amount of noisy artifacts increases when peaking increases (i.e. for profiles 4 and 5). On the other hand, in the more blurry images (i.e. profiles 1 and 2), there is loss of detail, for example, in the hair of the man, or in the background. Row 2 of Figure 27 shows that details of the plant become better visible as peaking increases. This is considered a positive effect of the peaking algorithm. Row 3 of Figure 27 illustrates the effect of ‘electronic ageing’, as can be induced by a peaking algorithm, on the leg of the man. The leg appears wrinkled as peaking increases. Mosquito noise can be seen as ringing around the circle in row 4 of Figure 27 as peaking increases which might relate negatively to perceived image quality. On the other hand, much of the details in the flowers are lost when in the image is more blurry. In row 5 of Figure 27, the ringing artifact is clearly visible the images of profile 4 and 5, especially around the head of the woman. With the blurry images (i.e. profile 1 and 2), this artifact is less clearly visible, but here detail is lost in, for example, the faces.

These examples generally show that as the image becomes blurrier details are lost which is hypothesized to be negatively related to image quality perception. On the other hand, there is the accompanying loss of visibility of compression artifacts, which is positively related to image quality.

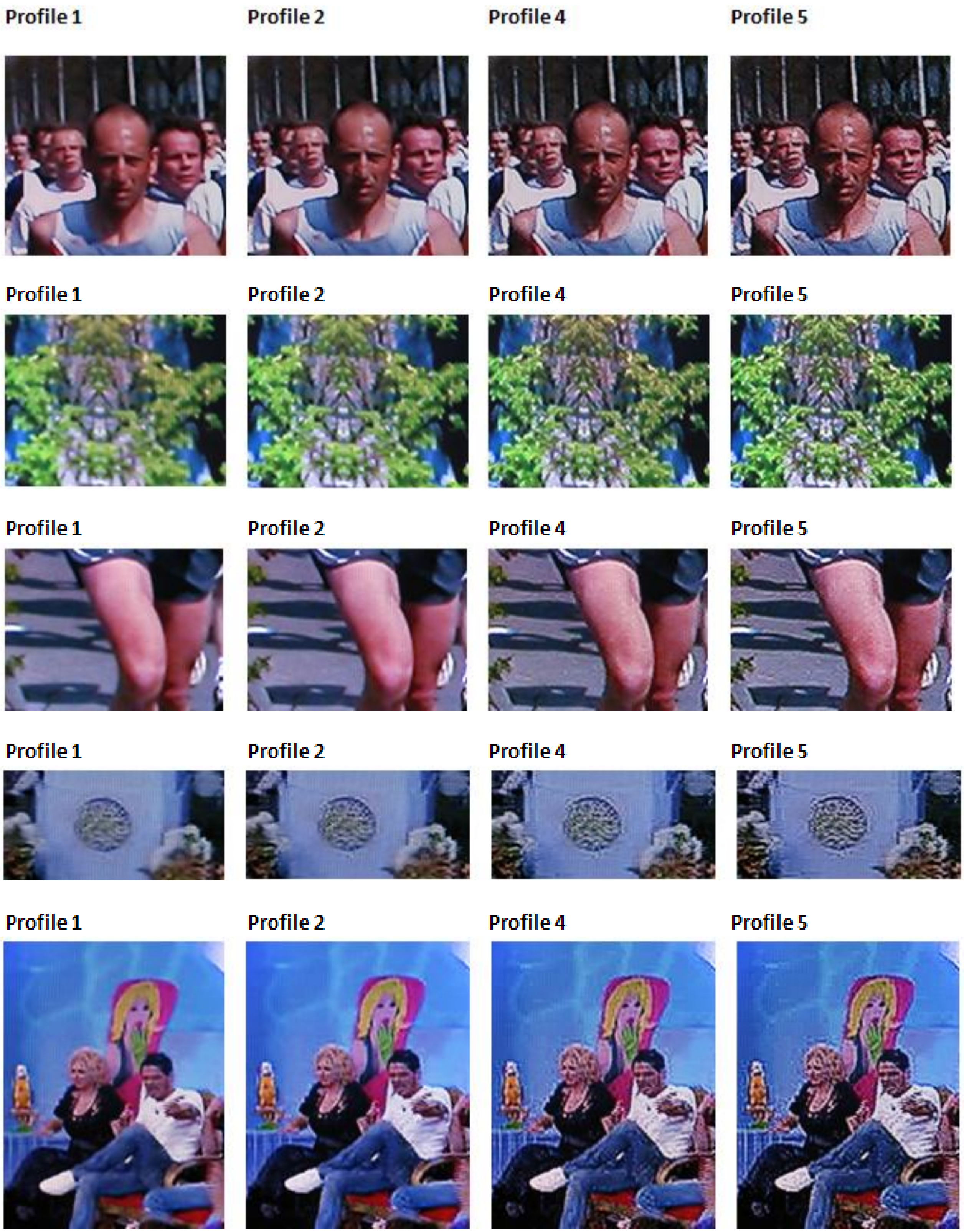


Figure 27 examples of clearly visible effect of peaking and blurring on images used in the experiment

Contrast

For contrast, 5 different profiles were used. The reference TV was always set to profile 3. When pressing key 'q' the other TV adjusted to profile 1, w=adjusted TV profile 2, e=adjusted TV profile 4, r=Adjusted TV profile 5, t=Adjusted TV reference profile 3. These profiles were also chosen in cooperation with Philips image quality perception experts, so that they represented realistic values, and varied enough to make differences visible. To illustrate the effect of 5 different profiles, TV measurements were performed with an imaging colorimeter (i.e. the PM-1423E-1 from Radiant) on a grey level gradient (left = totally saturated white, right = black). The effect of the five contrast settings on the luminance profile of the gradient is shown in Figure 28. In this figure, line C1 represents the luminance profile corresponding to the lowest contrast between black and white, while line C5 corresponds to the highest contrast profile. Line C3 represents the reference, and is also the profile currently used in the Philips LEDLUX TV with the Vivid setting in contrast setting menu of the TV. From these measurements it is clear that the 5 contrast profiles indeed vary in contrast with the largest variations occurring near the lowest and highest grey levels. Therefore, theoretically, for the contrast experiment, images should be selected with a lot of bright and dark grey levels.

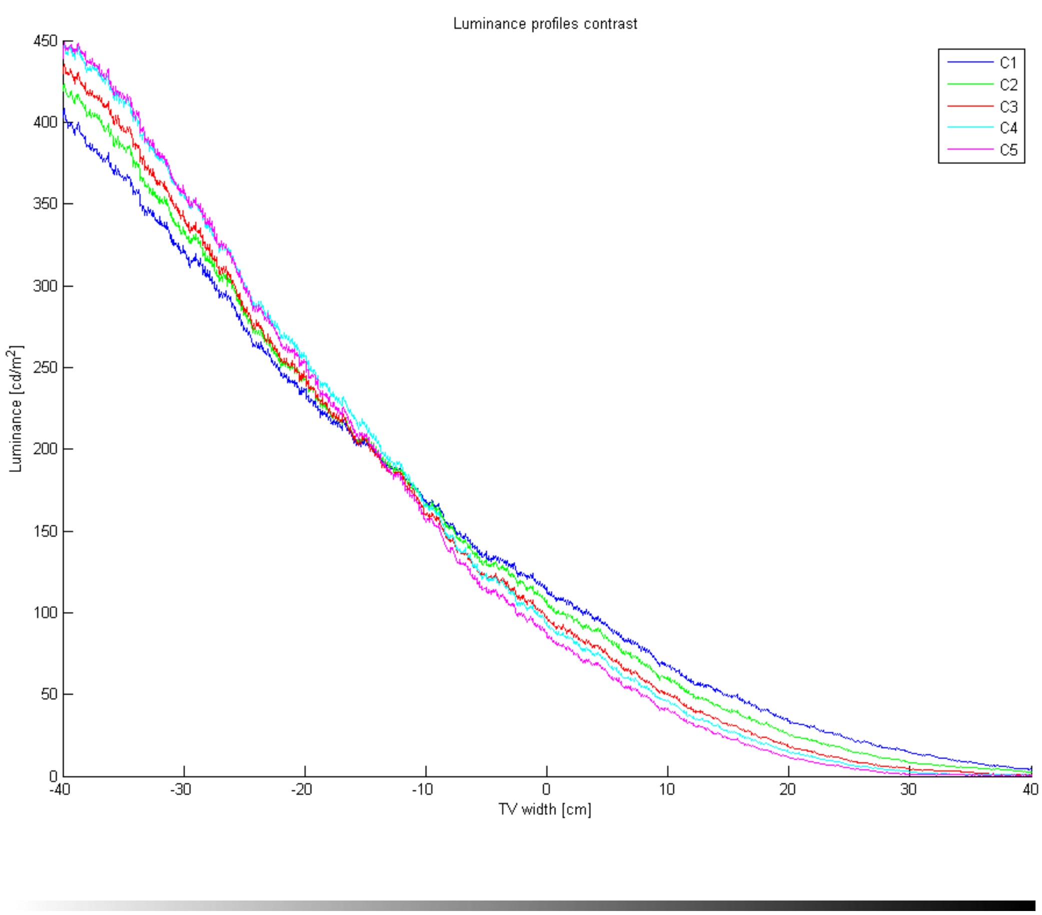


Figure 28 Luminance measurement on a gradient presented on the Philips LEDLUX for all five contrast profiles

Based on this need, two images were selected for the contrast experiment. They are shown in Figure 29 together with their histogram. The image “polar bear” contains a lot of bright grey levels, while the image “grass” contains mostly medium grey levels.

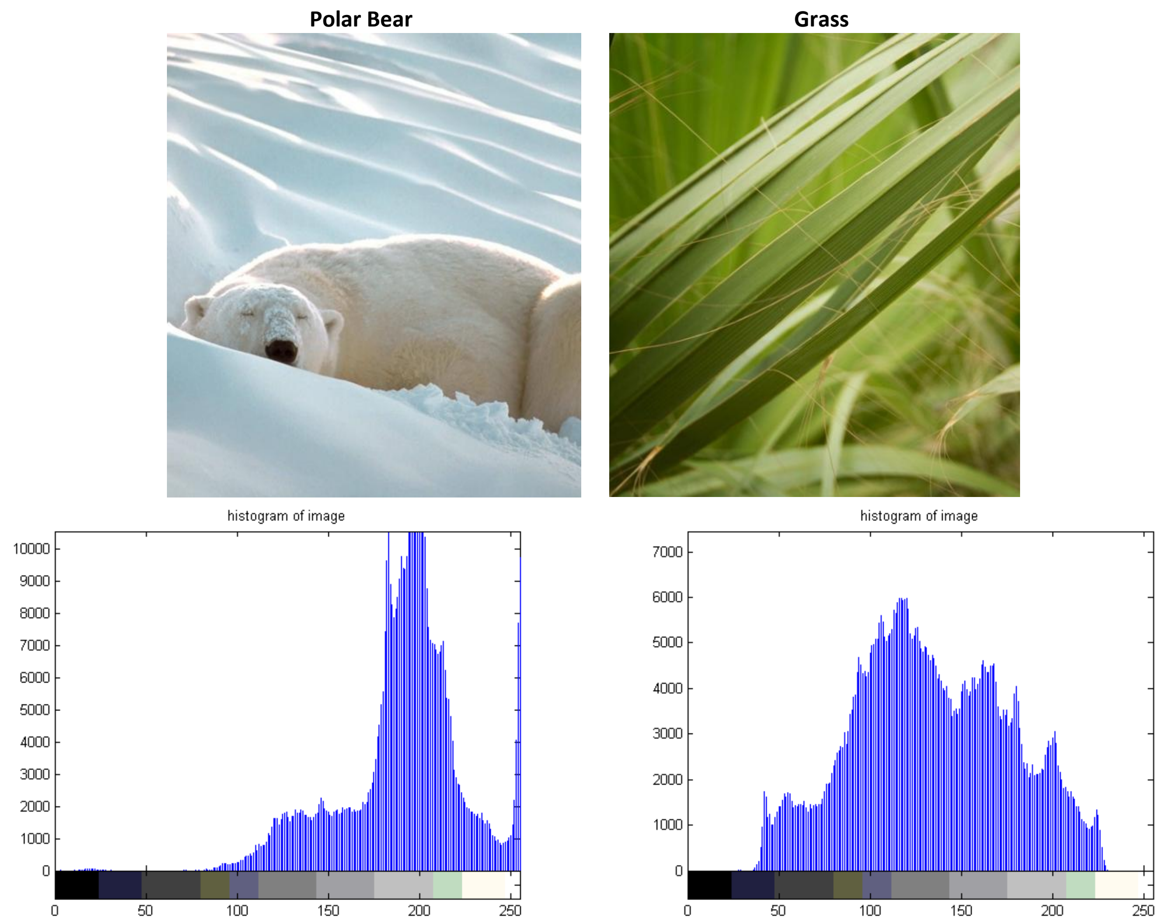


Figure 29 Images used for the contrast scoring experiment together with their grey level histograms. It can be seen that the image “grass” contains mostly intermediate grey levels, while the image “Polar bear” consists mostly of bright grey levels.

4.1.2.4 Ambient illumination

The ambient illumination for the scoring experiment was exactly the same as for the JND experiment. Hence, the low illumination setting corresponds to 30 lux measured perpendicular to the screens, while the high illumination setting corresponds to 400 lux.

4.1.3 Design

The scoring experiment evaluated the influence of the independent variables viewing distance, ambient illumination, sharpness and contrast settings and image content on the image quality perception. Therefore, people were asked to score the image quality of the adjusted TV with respect to the image quality of the reference TV on a scale from -5 to 5 in all conditions. A negative grade implied that the adjusted TV had a poorer image quality than the reference TV, while a positive grade implied that the adjusted TV had a better image quality than the reference TV. When people did not see any difference, or did see a difference, but not in image quality, they were asked to give a score of 0. For each participant the reference TV position was kept at the same position during the whole experiment. For 12 participants the reference TV was on the left (TV A) and for the other 12 participant, the reference TV was positioned on the right (TV B).

Just as with the JND experiment, a list of conditions per participant was made up front. In this list the independent variables viewing distance, room illumination and the position of the reference TV were balanced. The order of the setting for sharpness and contrast and the related images were randomized among participants. The list of conditions per participant can be found in Appendix C.

The lighting and the distance were controlled manually by the experimenter and the participant respectively. The reference position, image quality attributes (i.e. sharpness or contrast) and profile were all arranged by the software. A list of conditions was loaded into the automating program and all the experimenter had to do was press the space to go through all conditions. The experimenter also had to manually select the correct picture by pressing the next button on the remote control of the TivX streamer. Between every scoring, a black image was shown on the screen for 1 second. Every participant had to do a total of 150 scorings: 5 profiles for 5 pictures (3 for sharpness and 2 for contrast) at 3 distances and 2 lighting conditions (5*5*3*2). This took them about 45-60 minutes to complete.

4.1.4 Procedure

The invitations and the beginning of the procedure were similar as for the JND experiment (see chapter 3.1.4).

Then, participants were trained to get an indication of the scoring task and on how to use the -5 to +5 scoring scale for the range on image quality variation used in the experiment. They were asked to have a seat on the three meter chair and to look at the TVs. They were then presented with 5 pictures with two extreme settings on the non-reference (i.e. adjusted) TV and the same picture with the reference setting on the reference TV. During this training they were asked if they saw any differences in image quality between the two TVs, and if so whether they could give a score for the non-reference TV compared to the reference TV. Once people saw all five images with their extreme settings the real experiment began.

They were told that during the experiment they got to see the same pictures on both TVs but with changing image quality. They were then asked:

‘Please rate the image quality of the image shown on the non-reference TV (TV A or TV B) relative to the one shown on the reference TV. It’s about your opinion and your perception.’

Next they were asked to take a seat at the distance corresponding to their first distance condition and the lighting was set to the appropriate level. Then, the participant received the scoring list on paper and a pen. The experimenter took his place behind the ‘experimenter desktop pc’ to control the automating program and pressed the space button to get to the first image quality attribute, pressed the space button again to select the correct reference TV and again to select the correct sharpness or contrast profile while a black picture was displayed on the TVs. The experimenter pressed the next button on the TivX remote to display the first image. Then the participant was asked to note his score (on a paper scoring form) for the relative image quality of the non-reference TV. When this was done, the experimenter pressed the next button of the TivX to display a black image and he pressed the space button to get to the next profile. This was continued until all 150 cases were completed. When finished, the participant received a small incentive (candy) and was thanked for his or her participation.

4.2 Results

This experiment evaluated whether the optimal settings, i.e. the ones with the highest score, for the specific image quality attributes depend on viewing distance and ambient illumination. It is also interesting to know if this optimum depends on the content shown on the TV. Here the results following from the scoring experiment are presented for each of the two attributes separately.

4.2.1 Sharpness

For sharpness a zero measurement was performed, i.e. the same stimulus was shown on both TVs to test user's accuracy in scoring. Profile 2 was loaded on both TVs and thus, the participants were expected to give a score of 0. The mean score for this stimulus was 0.07 ± 0.08 (95% CI), and so, not statistically significantly different from 0. The scoring histogram of this stimulus is presented in Figure 30. Furthermore, when we checked whether the outcome of the zero measurement was dependent on whether the reference was placed right or left. The scores were 0.04 ± 0.10 (95% CI) and -0.06 ± 0.13 (95% CI) respectively, and so, in both levels not statistically significantly different from zero. This indicates that people do not have a specific preference for the right or the left TV.

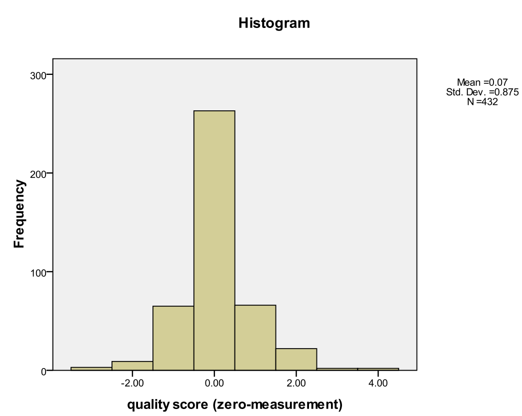


Figure 30 frequency of image quality scores for the zero measurement of the sharpness experiment

A univariate full factorial ANOVA was done on the data to evaluate the size and significance of the effects. The dependent variable was the stimulus score and the independent variables used in the analysis were viewing distance, sharpness profiles, content and ambient illumination. The effects of ambient illumination was found to be $F(1,1690)=0.369$, $p=0.544$, $\eta^2<0.001$ and for content

$F(2,1690)=0.211$, $p=0.810$, $\eta^2<0.001$ on the sharpness scores. Hence these effects were not significant. Distance on the other hand had a significant effect on the sharpness score, but with a very small effect size of $F(2,1690)=4.994$, $p=0.007$, $\eta^2=0.006$. The different sharpness profiles had a large and significant effect on the score of $F(3,1690)=474.233$, $p<0.001$, $\eta^2=0.457$ as can be seen in the left graph of Figure 31. Here it is shown that overall the more blurry stimuli (i.e. profiles 1 and 2) received a higher quality score than the peaked stimuli (i.e. profiles 3 and 4). It should be noted that profile 3 is not included in this sharpness analysis, because it only served as the reference. Also, the effect of the interaction between profile and distance on the sharpness score was large and significant ($F(6,1690)=51.972$, $p<0.001$, $\eta^2=0.156$). This is shown in the right graph of Figure 31, i.e. with increasing viewing distance, the differences between the sharpness scores became smaller.

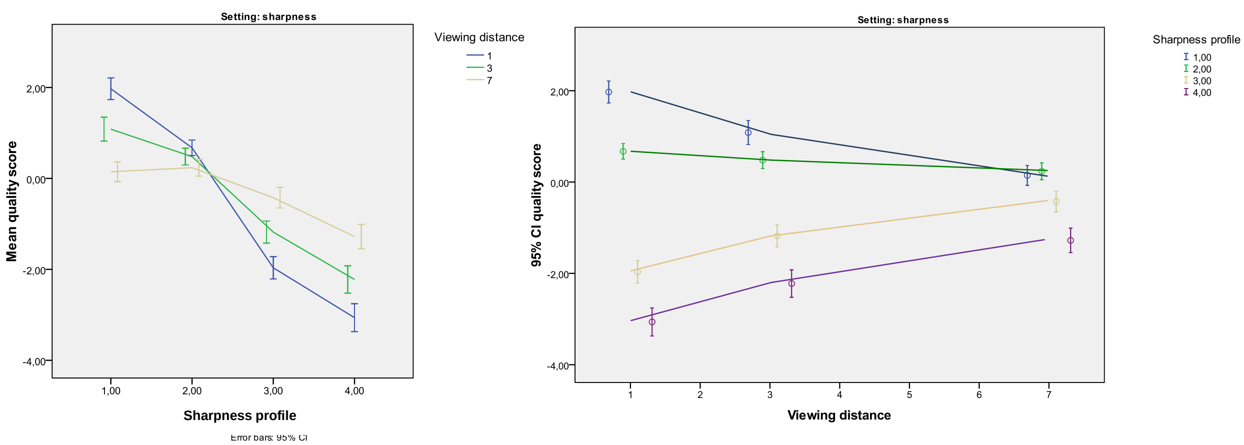


Figure 31 (left) Mean scores for the sharpness profiles with the different lines representing the distances (right) mean scores for the distances with the different lines representing the sharpness profiles

Apart from the overall effect we would like to know the optimal profile at a specific distance. Therefore, the data file was split on distance, and again a univariate ANOVA was performed. This analysis it showed that at all distances, the effect of profile was large and significant, but decreasing with larger distance as can again be seen in the right graph of Figure 31. At 1 meter, the effect of sharpness profile on the score was $F(3,546)=379.330$, $p<0.001$, $\eta^2=0.649$, at 3 meters the effect was $F(3,546)=146,831$, $p<0.001$, $\eta2=0.447$ and at 7 meters the effect was $F(3,546)=38.787$, $p<0.001$, $\eta2=0.176$. To evaluate the significant effect of profile at each distance further, Tukey post-hoc tests were performed. At 1 and 3 meters the scores were significantly different between all profiles ($p<0.005$). At 7 meters, only the scores for profiles 1 and 2, i.e. blurred profiles, were not significantly different.

The scores belonging to a given profile at each of the viewing distances together with their 95% confidence interval are shown in Table 6. From this it can be clearly seen that the scores for the more blurry profiles (i.e. profiles 1 and 2) are higher that the scores for the peaked profiles (i.e. profiles 3 and 4). People obviously prefer the more blurry profiles, which will be discussed in chapter 4.3.1.

Table 6 image quality scores for the different sharpness profiles at different viewing distances

Distance [m]	Sharpness profile	Mean score	Lower bound	Upper bound
1	1	1.97	1.73	2.21
	2	0.67	0.50	0.85
	3	-1.96	-2.21	-1.72
	4	-3.06	-3.37	-2.76
3	1	1.08	0.82	1.35
	2	0.48	0.29	0.66
	3	-1.18	-1.42	-0.94
	4	-2.22	-2.52	-1.92
7	1	0.15	-0.07	0.36
	2	0.24	0.05	0.42
	3	-0.42	-0.65	-0.20
	4	-1.28	-1.55	-1.01

4.2.2 Contrast

Here also a zero measurement was performed to evaluate the user's accuracy. Profile 3 was loaded on both TVs and thus, the participants were expected to give a score of 0. The mean score for this stimulus was -0.04 ± 0.16 (95% CI), and so, not statistically significantly different from 0. In this case however, comparing the zero-scores between cases when the reference was placed left or right, a difference is found (respectively -0.3194 ± 0.1615 and 0.2361 ± 0.1470), indicating a small preference for the left TV. Balancing over the 'reference TV'-condition however averages out this effect.

For contrast, a univariate ANOVA was done with as the dependent variable the stimulus score and the independent variables distance, content, ambient illumination, profile and the interactions between profile and distance, and image and profile. The effect of ambient illumination was found to be insignificant $F(1,1396)=2.936$, $p=0.122$, $\eta 2=0.002$. Also the effect of distance on the contrast score was not significant with an effect of $F(2,1396)=0.188$, $p=0.829$, $\eta 2<0.001$. The interaction between profile and distance that was found to be insignificant for contrast score ($F(8,1396)=1.484$, $p=0.158$, $\eta 2=0.008$). The content effect, on the other hand, was small but significant (i.e. $F(1,1396)=6.782$, $p<0.009$, $\eta 2=0.005$), and the interaction between image and profile was large and significant ($F(4,1396)=77.118$, $p<0.001$, $\eta 2=0.181$), indicating that the profiles could possibly have an effect on the separate images. Therefore, the data-file is split on image to analyze the data again in a fully factorial ANOVA for the independent variables profile, image and ambient illumination.

For image "polar bear", the effect of profile was large and significant ($F(4,681)=262,791$, $p<0.001$, $\eta 2=0.607$). For the image "grass" the effect of profile was also significant but with a much smaller effect size ($F(4,681)=8.434$, $p<0.001$, $\eta 2=0.047$). Considering the above, it would be interesting to visualize the profile effects on the scores of the separate images for all distances taken together (i.e. since the distance is not significant). This interaction between profile and image is shown in Figure 32 and the values of the scores with their 95% CI can be found in Table 7. Here it can also be seen that for the image “polar bear”, the score of profile 1 and 2 do not differ significantly, but that the differences between all other scores are significant. This is also confirmed by a post hoc test of profile for image “polar bear”. For image “grass”, the same post hoc test showed that only profile 5 received a significantly lower score than profiles 1, 2 and 3, and that the rest of the profiles did not receive a significantly lower or higher score. Furthermore the graph in Figure 32 and the values in Table 7 show that, at least for the image “polar bear”, profiles 1 and 2 (i.e. the lower contrast profiles) receive a higher score than profiles 4 and 5 (i.e. the higher contrast profiles).

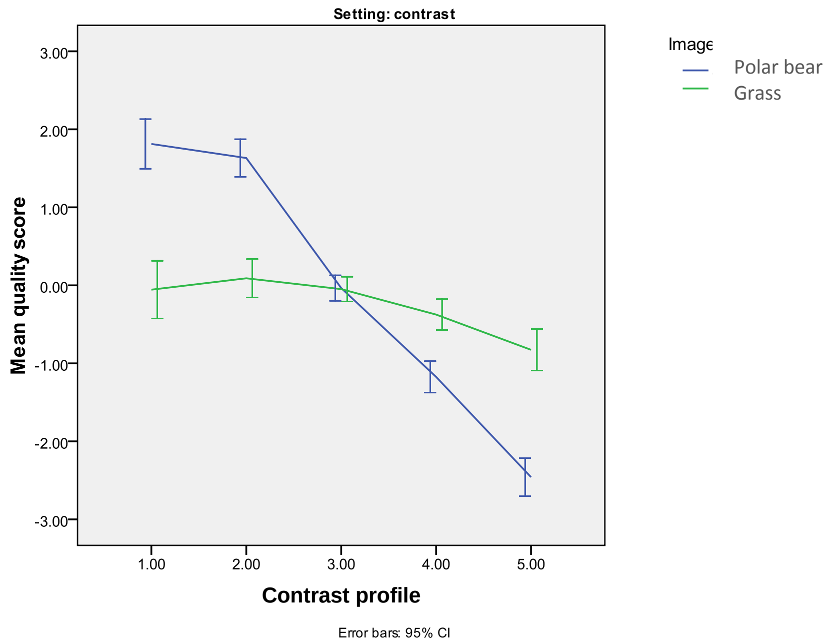


Figure 32 Mean scores for the contrast profiles with the different lines representing the images

Table 7 Image quality scores for the different contrast profiles for the different images

Image	Contrast profile	Mean score	Lower bound	Upper bound
Polar bear	1	1.81	1.49	2.13
	2	1.63	1.39	1.87
	3	-0.03	-0.20	0.13
	4	-1.17	-1.38	-0.98
	5	-2.46	-2.70	-2.21
Grass	1	-0.06	-0.43	0.31
	2	0.09	-0.16	0.34
	3	-0.05	-0.21	0.11
	4	-0.38	-0.57	-0.18
	5	-0.83	-1.09	-0.56

4.3 Discussion.

4.3.1 Sharpness

The profile that received the highest scores at a viewing distance of 1 and 3 meters was the most blurry profile and on 7 meters the most blurry or the slightly blurry profile. From the results on sharpness it can be concluded that, although the researchers did a very thorough job on selecting profiles to assure that a real optimum sharpness profile would be found at all three distances, an optimum was not found. Only for a distance of 7 meters the optimum seems to be around profile 1 or 2, but this cannot be concluded significantly. It appeared that the blurrier image was always preferred over the sharper (or peaked) image. During the tests, participants indicated that they preferred not to see any of the sharpness artifacts and blurring the image helped reducing these artifacts. This observation is in line with Keelan (2002), who stated that the visibility of image quality artifacts in an image is one explanation for variations in perceived image quality. To support this idea, the sharpness test was repeated for 12 participants at 30 lux ambient illumination with two images that did not contain such artifacts. The experiment was limited to one illumination level, since that proved to have no effect on image quality as was found in chapter 4.3.1. In the absence of artifacts, an optimum in the sharpness profiles was expected to be found. If so, the question was whether this optimum changed for the different viewing distances? The images that were used for this quick experiment are shown in Figure 33 together with the graphs showing the scores as a function of profile for each distance and image.

From this short and quick experiment it can be seen that especially for the image “parrot 1” there is a trend towards an optimal profile for each of the viewing distances. At a viewing distance of 1 meter the optimal profile seems to be a slightly blurred one, while at 3 and 7 meters the optimum seems to shift to the reference profile or a slightly peaked one. For the image “scuba diver” however this effect is not that clear. This can again be explained by the visibility of artifacts. When the peaking algorithm was applied to the image “scuba diver”, a lot of artifacts became visible on her cheeks making her skin appear very uneven (i.e. the ageing artifact). Participants indicated that they liked her skin more when it appeared smoother and blurrier, even while this was less realistic than the reference (unprocessed) image.

When looking at the image “parrot 1”, which is, except from movement, representative for current HDTV material, the preferred sharpness profile seems to depend on the viewing distance. It appears that when people are close by, they prefer a slightly blurred profile. This might result from little bits of noise present in the image and disappearing when blurring the image. When the distance increases, the preference seems to shift to the reference profile or a slightly peaked profile. Some participants indicated that they preferred the peaked profile since the image looked sharper. The main goal of a study by Radun et al. (2007), was to find if sharpness was dependent on content and if so, whether differences could be explained by interpretations related to the changes which were perceived. From a free description experiment, they found that for image quality perception, the effect of sharpness variations was dependent on the type of content. In accordance to the present findings, they also found that for some images a low sharpness received a more positive score than for other images. They explained that people use different terms to describe the sharpness of different images. If people want more sharpness they use terms like: irritating, dirty, amateurish, etc. When people want a more blurry image however, they use terms like: artistic, soft, light colors, etc.

Image: scuba diver

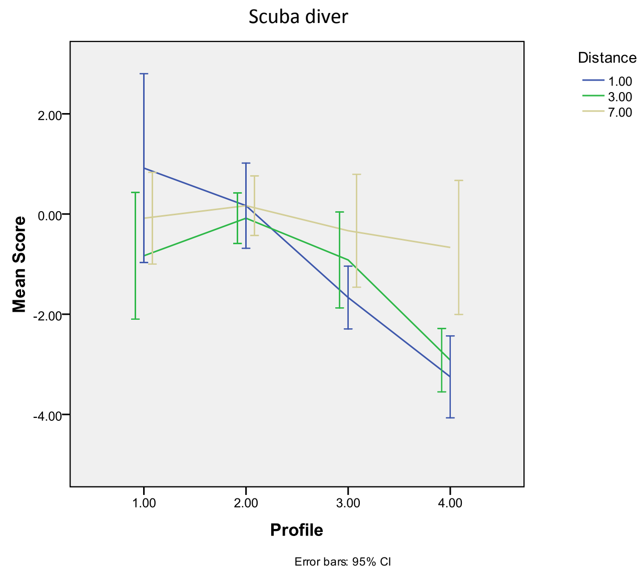


Image: parrot 1

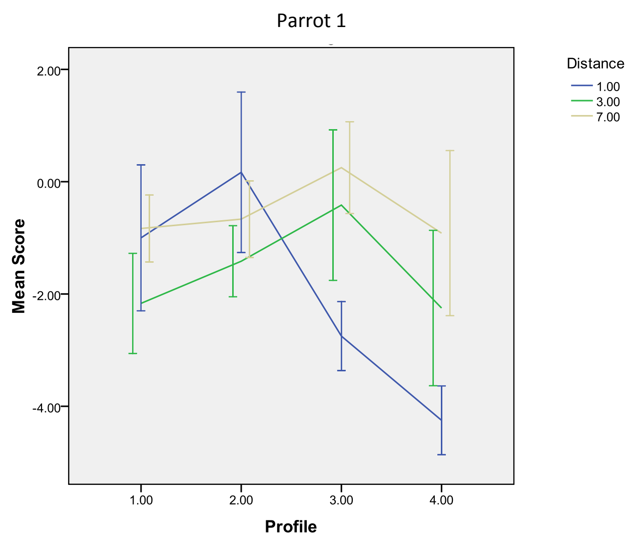


Figure 33 Images and results from the short and quick sharpness experiment

Another finding supported by the visual acuity theory (Blake, 2006) is that we found that the difference in scoring between the blurriest and the most peaked image becomes smaller with distance, becoming almost insignificant at 7 meters for the 42” monitors that were used. This is also consistent with the findings of Neuman (1990) beyond HDTV who found that the preference for different resolutions became equivalent at large viewing distance.

4.3.2 Contrast

The difference found between the images used for the contrast experiment can possibly be explained by the differences in their histograms shown in Figure 29. The histogram of the image “polar bear” contains a lot of pixels with a grey level near to a totally saturated white level, which is the level influenced most by the different contrast profiles. The image “grass” contains mostly grey levels in the middle of the grey level spectrum. Therefore, only small changes occur when changing the contrast profiles, and as a consequence, the scores are expected to be closer together. This implies that content has a very large influence on the perception of contrast of an image. For the image “grass”, people did not score differences between the profiles significantly. For the image “polar bear”, however, people liked the contrast to be lower. It has to be said that the image “polar bear” appeared much brighter when the contrast was low (i.e. with profile 1) and much darker when the contrast was higher (i.e. with profile 5). People also indicated this to be of large importance when judging the image “polar bear”. This indicates that for other content, probably other profiles would be preferred. This should be thoroughly investigated in future studies.

For the images used in the experiment, no effect of distance on the preferred contrast profile was found. At all distances, the two profiles with the lowest contrast were preferred for the image “polar bear”. This indicates that for images identical to those used in this experiment, contrast does not have to be changed related to viewing distance. This agrees with a study by Barbier, L’Hote and Plantier (1996) in which they measure the difference in perceived image quality between two images which differ in contrast and luminance. In this study, Barbier used an image of a symbol, being a white ring on a dark background. A first experiment was done with fixed illumination and at a viewing distance of 0.5 meters. In a second experiment the luminance was fixed, but now the viewing distance changed from 0.5 to 2 meters. In his second experiment, also contrast was changed to find a possible interaction effect between contrast and distance. From the first experiment higher contrast seemed to have a positive effect on image quality and luminance also had a positive, but smaller effect on image quality for a luminance lower than 30 cd/m2. From the second experiment, a large effect of distance on perceived image quality was found, but no significant interaction of contrast and viewing distance on image quality was found. This means that from these experiments they were also not able to conclude that contrast had to be changed for different viewing distances to maintain the optimal perceived image quality.

Furthermore, the effect of ambient illumination on the preferred contrast profile was insignificant at least for the images used in this experiment. This indicates that for similar images, the contrast does not have to be changed for different ambient illumination conditions.

4.3.3 Discussion of the hypotheses

Are the settings belonging to an optimal perceived image quality dependent on viewing distance?	The settings belonging to an optimal perceived image quality at a large viewing distance are different from those at a small viewing distance (small distance: lower brightness wanted , lower contrast required (Montag, 2005), less sharpness required (Damstra, 2009)).

From the results of our experiment it cannot be stated that the optimal setting for sharpness or contrast depends on viewing distance, since the optimum for the images that contain artifacts seems to be even lower than the most blurred profile used in this experiment. From the short additional experiment done on high quality images, however, it can be seen that there is definitely a trend towards distance dependency. It looks like the optimum profile seems to shift towards more peaky profiles as viewing distance increases. With smaller viewing distances people prefer a less peaked profile. These effects are unfortunately not significant, and therefore, no scientific conclusions can be drawn from both experiments.

Are the settings belonging to an optimal perceived image quality dependent on content?	The settings belonging to an optimal perceived image quality are dependent on content (Heynderickx, 2007;Radun, 2007)

This hypothesis cannot be confirmed based on the results from our experiments. No effect of ambient illumination on both the optimal sharpness and contrast profile are found.

Are the settings belonging to an optimal perceived image quality dependent on content?	The settings belonging to an optimal perceived image quality are dependent on content (Heynderickx, 2007;Radun, 2007)

This hypothesis is confirmed for both sharpness and contrast. For sharpness the optimal profile seems to depend highly on the amount of artifacts in the image. For contrast an image with intermediate grey levels seems to be less affected by contrast processing than images with grey levels near the white level. The optimal contrast profile was not found yet, because it seemed to be a lower contrast profile than used in this experiment. Images with black levels are not tested but it is hypothesized that for these images also an optimal profile can be found.

5 Conclusions

In this study it was found that the perceptibility of changes in black level and sharpness depends on several environmental aspects. For the perceptibility of changes in white level, this dependency was not found.

Larger changes in black level are needed to be perceivable at longer viewing distances. Furthermore, the perceptibility of changes in black levels varies with content. From our JND experiment it was found that the visibility of the change is smaller for images that contain a lot of (near)black grey levels than for images that contain intermediate or (near)white grey levels. Finally, from the JND experiment it can be concluded that bigger changes in black level are needed at a larger room illumination.

Sharpness changes have to be substantially bigger to be perceptible at a larger viewing distance. The required sharpness changes also depend on the image content. Ideally some sort of content analysis to measure the variables which elicit this change in perceptibility should be implemented. One of these variables (as we can conclude from our second experiment discussed in chapter 4) is the amount of artifacts present in an image. Other variables could be the amount of edges and the amount of details in an image.

The above is important for manufacturers and researchers since they now know to what extent black level and sharpness have to be adjusted according to these environmental factors to have an effect on perceived image quality. For white level no significant effect of the environmental aspects was observed, meaning that for comparable conditions as used in our experiment, fixed changes in white level needed to make a perceptible change in the related image quality attributes work evenly well in every environmental setting.

When scoring sharpness differences, it can be stated that no optimal profile was found for any of the environmental conditions. People mostly preferred the most blurry image and mentioned that the peaked images appeared to be ‘too sharp’. Especially at 1 and 3 meter viewing distance, the optimal profile seems to be even blurrier than the blurriest profile used in our experiment. Only at 7 meters, a slightly less blurred profile received a slightly higher score, but this difference was not significant. At this distance, the variation in scoring was limited any way, indicating that sharpening or blurring an image does not have a lot of effect at larger viewing distances. Therefore, over-sharpening by peaking, especially for images that include artifacts to the extent of the images used in our experiment, should clearly be avoided. It is clear that people do not prefer a sharper image, when the artifacts become clearly visible. By repeating the experiment with images that did not contain any of the compression and capturing artifacts, it could be concluded that an optimum seemed to shift with varying viewing distance. At a larger distance, people preferred an image which was sharper than the image they preferred at a closer viewing distance. This effect however seemed to be very dependent on the content. More investigations are needed to find for which images sharpness enhancement is beneficial and for which images it is not acceptable.

Also for contrast, the effect of profile strongly depended on content, which can possibly be linked to the grey level histogram of the images used. For one image used in this experiment no effect of the different contrast profiles on the perceived image quality was found, while for the other image a lower contrast resulted in a higher perceived image quality. Furthermore, no effect of distance on contrast preference was found, meaning that contrast does not have to be changed differently for other viewing distances.

Although more investigations are needed on the content dependency of the optimal sharpness and contrast for image quality, it is clear that manufacturers should take the environmental aspects into account in finding the most optimal settings for these image quality attributes. Once the effects are all known, manufacturers can find the optimal settings for their TV sets to be winning over other manufacturers in the very competitive factor of perceived image quality.

6 Future directions

From the second experiment it can be concluded that content has a very large effect on the effects the optimal settings for the image quality attributes. Because of the broad and exploring nature of this experiment (i.e. a lot of environmental aspects were taken into account), the amount of images that could be used was limited to two for each attribute; otherwise the amount of conditions used would be too large. Therefore it is advisable to repeat this experiment in following studies with much more content and try to quantify this content on their differences. The results from this study can be used as a guideline for the setup of these new studies (i.e. which factors can be left out of the equation because of their insignificant influence on the scoring of image quality, and what are important aspects to consider while selecting new images). In quantifying the images on their differences, a possibility, for black level and white level images, would be to realize this by using its grey level histogram to make a measure of the amount of bright and/or dark grey levels in the images. This way, it would be possible to find the optimal settings for all content at a certain distance and ambient illumination.

With the use of the second experiment we found that for some images sharpening the image (with the use of the peaking algorithm) does not work positively on the perceived image quality, and for some it does. In future research it should be investigated if certain aspects of an image can be indicative to the positive or negative effect of the peaking algorithm. From this experiment it at least seems that for images with artifacts, it is not preferred.

In both of the controlled experiments done in this study, still images were used as stimuli. Therefore the conclusions drawn might not yield for moving images. Also, since with moving images some of the artifacts which appear in still images are less obvious you might expect that the results from the short experiment from chapter 4.3.1. in which images were used that did not contain artifacts, might therefore correspond more with the results that will be found with moving images. To be sure, the experiments done in this study would have to be repeated with moving images to be able to draw such a conclusion. From our study we would hypothesized that the sharpness changes have to be bigger before perceivable with the use of moving images.

In future experiments it might be better to do the same type of experiment as our second experiment at different distances on different days. This might strengthen the effects distance on scoring. From our experiment it seemed that distance at which people first began to take the test, significantly determined the scores over all the other distances. When people did not like the over-sharpened image at 1 meter, they mentioned at the other distances things like ‘ah, this is the same setting for this image, so I will give it the same kind of score as before’. Hence, instead of judging the image quality as a novice, they used their earlier findings to score the image at the new distance and again judged is as being bad. This probably made the effect of distance smaller than it would be if the experiment was done on different days for each distance, because then people would not remember the settings that much.

References

Ardito, M. (1994). Studies of the influence of display size and picture brightness on the preferred viewing distance for HDTV programs. *SMPTE journal, 103*(8)*,* 517-522.

Bailey, D. G., & Punchihewa, A. (2002). Artifacts in image and video systems: Classification and mitigation. *Proceedings of Image and Vision Computing New Zealand 2002,* pp 197-202.

Barbier, B., L'Hote, F., & Plantier, J. (1996). Multi-scale filtering for image quality on LCD matrix displays. *SID Digest,* 1-5

Blake, R., & Sekuler, R. (2006). Perception Fifth edition. New York, NY: McGraw-Hill

Damstra, A.(2009). Distance Adaptive Picture Quality, Technical note for Philips Consumer Lifestyle. Eindhoven, Netherlands.

Engeldrum, P. G. (2000). Psychometric Scaling: a toolkit for imaging systems development. Winchester: Imcotec Press.

Fish, R. S., & Judd, T. H. (1991). A subjective visual quality comparison of NTSC, VHS, and compressed DS1-compatible video. *Proceedings of the SID, 32(2)*, 157-164.

Ford, A., & Roberts, A. (1998). Colour Space Conversions, text at http://www.inforamp.net/˜poynton /PDFs/coloureq.pdf

Fujio, T. (1985). High-definition television systems. *Proceedings of the IEEE, 73*, 92-100.

Haan, G. (2006). Digital Video Post Processing. The Hague: CIP-data Koninklijke Bibliotheek

Heynderickx, I., Teunissen, C., Qin, S. L., Ge, S., Yin, H. C., Xia, J., et al. (2007). Just noticeable difference of image attributes for natural images. *Proceedings of SID, 38*(1)*,* 326-329.

Heynderickx, I., Xia, J., Qin, L., Liu, L., & Yin, H. C. (2005). The just noticeable difference in chromaticity, black level, white level and contour rendering in natural images. *Proceeding of Int. Disp. Workshops, 12*(2)*,* 1821-1824.

Heynderickx, I., Liu, L., Xia, J., & Yin, H. C. (2004). Visibility threshold in sharpness for people with different regional backgrounds. *Asian Symposium on Information Display No8*, 12(4), 509-515.

Jesty, L. C. (1958). The relation between picture size, viewing distance, and picture quality. *Proceedings of the IEEE, 105B*, 425-439.

Johnson, G. M., & Montag, E. D. (2005). Size matters: the influence of viewing distance on perceived spatial frequency and contrast. *Proc. IS&T/SID's13th Color Imaging Conference, 13,* 399-343.

Keelan, B. W. (2002). Handbook of image quality: Characterization and prediction. Rochester, NY: CRC.

Kirenko, I. O., Muijs, R., & Shao, L. (2006). Coding artifacts reduction using non-reference block grid visibility measure. Proc. *IEEE International Conference on Multimedia and Expo, 2006,* 469-472.

Lund, A. M. (1993). The influence of video image size and resolution on viewing-distance preferences. *SMPTE journal, 102*(5), 406-415.

Meesters, L.M.J., & Martens, J. B. O. S. (1999). Blockiness in JPEG-coded images. *Proceedings of SPIE, 3644,* 245-257.

Neuman, W. R. (1990). Beyond HDTV: Exploring subjective responses to very high definition television. A Research Report for GTE Labs and the TVOT Consortium. MIT, Cambridge, MA.

Radun, J., Virtanen, T., Nyman, G. (2006). Explaining multivariate image quality – Interpretation-based quality approach. *Proceedings of ICIS, 06,* 119-121.

Radun, J., Leisti, T., Nyman, G., Häkkinen, J., Ojanen, H., Olives, J.-L., et al. (2007). Content and quality: Interpretation-based estimation of image quality, *ACM Transactions on Applied Perception (TAP), 4(4),* doi:10.1145/1278760.1278762

Ridder, H., & Willemsen, M. C. (2000). Percentage scaling: a new method for evaluating multiply impaired images. *Proceedings of the SPIE, 3959,* 68-77.

Roufs, J. A. J, & Bouma, H. (1980). Towards linking perception research and image quality. *Proceedings of the SID 21*, 247-260.

Samei, E. (2002). Digital imaging and communications in medicinde (DICOM), *standard-PS 3.14-2003, Part 14: grayscale standard display function*, 21-27.

Shen (1998) Review of postprocessing Techniques for compression artifact removal. *Journal of Visual Communication and Image Representation, 9(1),* 2-14.

Stevens, S. S. (1957). On the psychophysical law. *Psychological Review, 64,* 153-181.

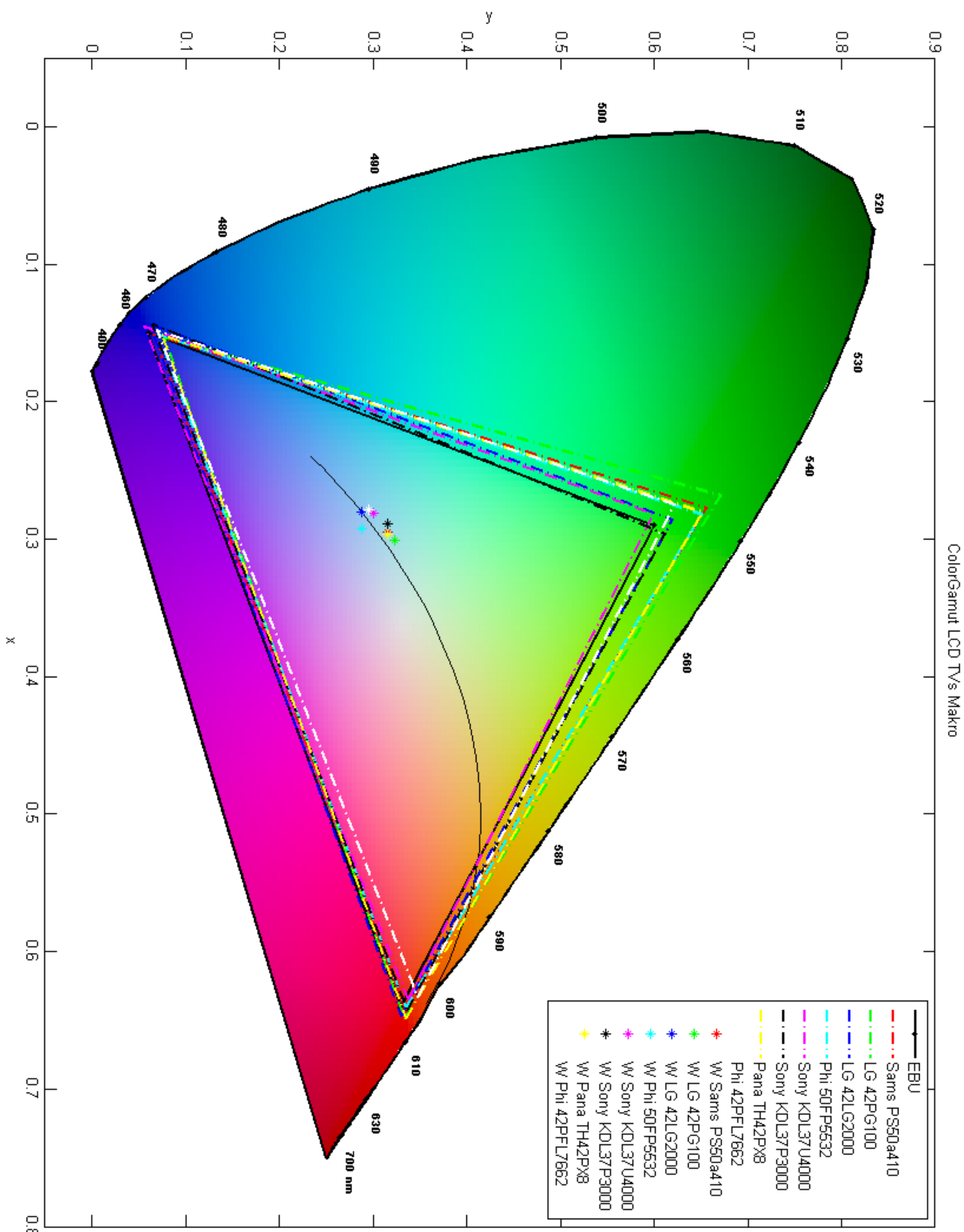
Subedar, M., & Caviedes, J. E. (2007). Review of practical MPRG post-processing methods for the consumer video processing chain. Printed from http://enpub.fulton.asu.edu/resp/vpqm2007/papers/376.pdf

Westerink, J. H. D. M., & Roufs, J. A. J. (1989). Subjective image quality as a function of viewing distance, resolution, and picture size. *SMPTE Journal, 98,* 113-119.

Young Park, J., Triantaphillidou, S., & Jacabson, R. E. (2009). Identification of image attributes that are most affected with changes in displayed image size. *Proceedings of SPIE, 7242,* doi:10.1117/12.805844.

Appendices

Appendix A: the full xy-color space



Appendix B: MatLab scripts

White level adjustment

gamma=2.2; (gamma value of the camera)

I = imread('w1.jpg'); (read original image)

I_cropped = I(1:700,1:600,:); (cut original image to correct size)

figure; imshow(I_cropped); title('Original Image'); (visualize original cropped image)

imwrite(I_cropped,'white_2_0.png'); (save orignal cropped image)

imwrite(I_cropped,'white_2_1.png'); (save again)

I_in = (double(I_cropped)/255).^(gamma); (transform to the linear light domain)

for i=1:1:60

iLin = ((255-i)/255).^(gamma); (adjust the white level of the image and compress the other grey levels accordingly)

I_white = (I_in)*(iLin); (transform back to the original gamma)

I_out=uint8(255*(I_white.^(1/gamma))); (transform back to the original gamma)

imwrite(I_out,sprintf('white_2_%d.png',(i+1))); (save processed image)

end

Black level adjustment

```
gamma=2.2;

I = imread('definitief/b2.jpg');

I_cropped = I(1:700,1:555,:);

imwrite(I_cropped,'definitief/black/black_2_0.png');

imwrite(I_cropped,'definitief/black/black_2_1.png');

I_in = (double(I_cropped)/255).^(gamma);


for i=1: 1: 60

iLin = (i/255).^(gamma);

I_black = I_in*(1-iLin) + iLin;

I_out=uint8(255*(I_black.^(1/gamma)));

imwrite(I_out,sprintf('definitief/black/black_2_%d.png',(i)+1),'PNG');

end
```

Sharpness adjustment

```
gamma=2.2;

I = imread('s5.jpg');

I_cropped = I(1:780,1:680,:);

imwrite(I_cropped(40:740,40:640,1:3),'Sharp_new/blurr_5_0.png')

imwrite(I_cropped(40:740,40:640,1:3),'Sharp_new/blurr_5_1.png')

I_in = double(I_cropped).^(1/gamma);


for i=1:60

Blur = fspecial('gaussian',25,i/10);

I_blurred = imfilter(I_in,Blur,'conv');

I_out = uint8(I_blurred.^(gamma));

I_out_cropped = I_out(40:740,41:641,1:3);

imwrite(I_out_cropped,sprintf('Sharp_new/blurr_4_%d.png',round(i+1)));

end
```

Transformation of grey level to luminance.

```
a = load('DisplayLUTS.txt');

% Primary matrix (XYZs of each channel: Y is 2nd row)
Mat = [  190.5013   58.6013   83.0513;
   79.3156  218.6406   36.0706;
    2.7611   35.1906  449.9555];

% Grey level matrix
rgb = [255 255 255]

for i = 1:3
    % compute the "linear" RGB from the code value
    % note the adding 1 because the index starts at 1, not 0
    rgbLin(:,i) = a( rgb(:,i) + 1 );
end

% XYZ from linear RGB using the matrix
% forgive the transposes: because I like to keep things in columns!
XYZ = (Mat * rgbLin')'
% the 2nd column, Y, is of course luminance, units of candelas per square meter (cd/m2).
% You should see a max value of about 335 cd/m2.
```

Appendix C: tables of conditions

JND experiment

Participant	Lighting (BALANCED)	Distance (RANDOM)	Image factor (BALANCED)	Image (BALANCED)
1	30 lux - 400 lux	371 317	Black - White - Sharp	Image 1 - Image 2
2	400 lux - 30 lux	173 371	Black - White - Sharp	Image 1 - Image 2
3	30 lux - 400 lux	371 173	Black - Sharp - White	Image 1 - Image 2
4	400 lux - 30 lux	317 371	Black - Sharp - White	Image 1 - Image 2
5	30 lux - 400 lux	173 317	White - Black - Sharp	Image 1 - Image 2
6	400 lux - 30 lux	371 317	White - Black - Sharp	Image 1 - Image 2
7	30 lux - 400 lux	713 317	White - Sharp - Black	Image 1 - Image 2
8	400 lux - 30 lux	317 317	White - Sharp - Black	Image 1 - Image 2
9	30 lux - 400 lux	317 317	Sharp - Black - White	Image 1 - Image 2
10	400 lux - 30 lux	137 713	Sharp - Black - White	Image 1 - Image 2
11	30 lux - 400 lux	371 137	Sharp - White - Black	Image 1 - Image 2
12	400 lux - 30 lux	713 173	Sharp - White - Black	Image 1 - Image 2
13	30 lux - 400 lux	317 713	Black - White - Sharp	Image 2 - Image 1
14	400 lux - 30 lux	371 731	Black - White - Sharp	Image 2 - Image 1
15	30 lux - 400 lux	371 731	Black - Sharp - White	Image 2 - Image 1
16	400 lux - 30 lux	731 371	Black - Sharp - White	Image 2 - Image 1
17	30 lux - 400 lux	371 731	White - Black - Sharp	Image 2 - Image 1
18	400 lux - 30 lux	371 731	White - Black - Sharp	Image 2 - Image 1
19	30 lux - 400 lux	317 317	White - Sharp - Black	Image 2 - Image 1
20	400 lux - 30 lux	173 317	White - Sharp - Black	Image 2 - Image 1
21	30 lux - 400 lux	713 137	Sharp - Black - White	Image 2 - Image 1
22	400 lux - 30 lux	173 713	Sharp - Black - White	Image 2 - Image 1
23	30 lux - 400 lux	173 173	Sharp - White - Black	Image 2 - Image 1
24	400 lux - 30 lux	137 731	Sharp - White - Black	Image 2 - Image 1

Scoring experiment

Participant	Reference (BALANCE)	Lighting [lux] (BALANCED)	Distance [m] (BALANCED)	image quality factor (RANDOM)	profile (RANDOM)	sharp-image (RANDOM)	contr-image (RANDOM)
1	L	400 - 30	1, 3, 7	contrast - sharpness		2, 3, 1	4, 5
2	L	400 - 30	1, 7, 3	contrast - sharpness		2, 3, 1	4, 5
3	L	400 - 30	3, 1, 7	sharpness - contrast	this variable	1, 2, 3	5, 4
4	L	400 - 30	3, 7, 1	contrast - sharpness	was randomized	3, 1, 2	5, 4
5	L	400 - 30	7, 1, 3	contrast - sharpness	for every	2, 3, 1	5, 4
6	L	400 - 30	7, 3, 1	contrast - sharpness	picture within	2, 1, 3	4, 5
7	L	30 - 400	1, 3, 7	sharpness - contrast	every image	2, 3, 1	4, 5
8	L	30 - 400	1, 7, 3	contrast - sharpness	quality factor	2, 1, 3	5, 4
9	L	30 - 400	3, 1, 7	sharpness - contrast		1, 3, 2	4, 5
10	L	30 - 400	3, 7, 1	sharpness - contrast	**Examples:**	1, 3, 2	4, 5
11	L	30 - 400	7, 1, 3	contrast - sharpness	qewrt	1, 2, 3	5, 4
12	L	30 - 400	7, 3, 1	sharpness - contrast	twqer	1, 3, 2	4, 5
13	R	400 - 30	1, 3, 7	contrast - sharpness	reqwt	3, 2, 1	4, 5
14	R	400 - 30	1, 7, 3	sharpness - contrast		1, 2, 3	5, 4
15	R	400 - 30	3, 1, 7	contrast - sharpness		3, 1, 2	4, 5
16	R	400 - 30	3, 7, 1	sharpness - contrast		3, 2, 1	4, 5
17	R	400 - 30	7, 1, 3	contrast - sharpness		2, 3, 1	4, 5
18	R	400 - 30	7, 3, 1	sharpness - contrast		2, 3, 1	5, 4
19	R	30 - 400	1, 3, 7	sharpness - contrast		3, 2, 1	4, 5
20	R	30 - 400	1, 7, 3	contrast - sharpness		3, 1, 2	4, 5
21	R	30 - 400	3, 1, 7	contrast - sharpness		1, 3, 2	5, 4
22	R	30 - 400	3, 7, 1	sharpness - contrast		2, 1, 3	5, 4
23	R	30 - 400	7, 1, 3	sharpness - contrast		3, 2, 1	5, 4
24	R	30 - 400	7, 3, 1	contrast - sharpness		2, 3, 1	5, 4