MASTER

Preferred and acceptable color gamut for reproducing natural image content

de Volder, R.J.

Award date:
2009

Link to publication

Disclaimer
This document contains a student thesis (bachelor's or master's), as authored by a student at Eindhoven University of Technology. Student theses are made available in the TU/e repository upon obtaining the required degree. The grade received is not published on the document as presented in the repository. The required complexity or quality of research of student theses may vary by program, and the required minimum study period may vary in duration.

General rights
Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

• Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
• You may not further distribute the material or use it for any profit-making activity or commercial gain
Preferred and acceptable color gamut for reproducing natural image content

R.J. de Volder
0591167
Preferred and acceptable color gamut for reproducing natural image content

R.J. de Volder
0591167

Master Thesis
November 2009
Preface
In front of you lies my Master thesis in Human Technology Interaction. For my Master thesis I have been conducting research in the area of color science at the Philips High Tech campus in Eindhoven in the Visual Experiences department. I have been working on measuring and analyzing the preferred and acceptable amount of colors for natural image content. The knowledge gained from this work is aimed to support the development of an algorithm for the proper use of the wide range of colors of so-called wide-gamut displays.

During my research study, I learned how to conduct and report a good research project, and I have gained much experience in the field of image processing, color perception, and lighting engineering.

My stay at the Visual Experiences department was most pleasant and educational with much personal contact and an excellent atmosphere. I would especially like to thank my mentors Wijnand IJsselsteijn, Ingrid Heynderickx and Dragan Sekulovski for their excellent input and support during my research. I also would like to thank Michael Murdoch and Cansu Atalay for their extensive contributions to this master thesis. Finally I would like to thank all my colleagues at Philip Research, my parents, family, and friends for their extensive (personal) support.

Robert de Volder
Eindhoven, November 26th 2009
Summary

Nowadays, it is common good for people to convey information through the use of digital displays. It is important that these displays convey their information in a manner that is pleasant, comfortable, correct, and clear. Developments in display technology recently resulted in displays which have an increased range of colors. This study will focus on the appropriate use of this increased range of colors, the so-called wide color gamut.

Most of the previous studies have been dedicated to mapping standard color gamuts into wide color gamuts, but those studies merely use the actual user preferences for developing their mapping algorithms. Through a series of psychophysical experiments this study first explores the natural color gamut boundaries. The natural color gamut boundaries can be defined as the average preferred colorfulness of a wide range of colors.

The first experiment was designed to find the preferred and maximum accepted color gamut boundary. During the experiment, participants were presented one single-hue image (stimulus) at a time on a wide color gamut display. They could adjust the amount of chroma of each stimulus once to their own preference, and once to their level of maximum acceptance. The stimulus could be changed from nearly achromatic to very colorful. The results from the first experiment show that the standard color gamut is not large enough to accommodate the chroma preference of the participants. This shows there is a perceptual need for wide-gamut displays. Furthermore, the results show a ceiling effect for some of the measured hues. For these hues there is thus a need for even wider color gamut displays. Finally, analysis of the results shows that preference and maximum acceptability of the user’s color gamut is dependent on at least 3 factors; namely hue, image content and participant.

An intermediate experiment was conducted to verify whether the tuning methodology as used in the first experiment could have led to some bias in the result due to by chromatic adaptation. This follow-up experiment used a method which was a combination of a paired comparison and a staircase methodology. A subset of 10 participants from the first experiment was used to verify the results with the different methodology. No significant differences were found between the results of both experiments.

The user-preference information gathered by the first experiment was used to create a gamut extension algorithm. The natural color gamut boundaries were re-measured by means of a third experiment which used this gamut extension algorithm. This third experiment used the same tuning methodology as the first experiment. However, only chroma preference was analyzed. Both single-hue stimuli as well as stimuli with a wide range of hues were included. Using an additional painting task, participants were also requested to indicate where the gamut extension algorithm added too much or too little chroma. Results from this last experiment reflect the importance of individual difference. The effect for hue, however, has been accounted for to a great extent. Image content seems to have a large, yet unexplained, effect on the results. Attempts to relate physical image properties to the results failed. The effect of content is most probably not directly related to the size of high-chroma regions, nor to a contrast as defined in this study.

Using the results from this set of experiments, it was shown that there is a natural boundary for the gamut expansion that is wider than the standard gamut, but smaller than the one of an example wide gamut system. Using this knowledge, all existing gamut expansion algorithms can be improved.

Future studies should investigate user sensitivity to over- and under-saturated chroma areas of different sizes. This would aid gamut extension algorithm research to be able to cope better with the yet unexplained influence of image content. A comparison experiment, between state of the art gamut expansion algorithms with and without the inclusion of the proposed boundary, can show the improvement.
## Contents

Preface ......................................................................................................................... iv  
Summary ....................................................................................................................... vi

1 Introduction into gamut extension research ................................................................ 1  
  1.1 Introduction ........................................................................................................ 1  
  1.2 Issues in extending color gamuts ....................................................................... 2  
  1.3 Goal ..................................................................................................................... 3  
  1.4 Thesis overview .................................................................................................. 3  

2 Literature review ....................................................................................................... 5  
  2.1 General color introduction .................................................................................. 5  
  2.2 Color spaces ....................................................................................................... 6  
   2.2.1 RGB and sRGB color space ...................................................................... 6  
   2.2.2 CIE XYZ and CIE xyY color space ........................................................... 7  
   2.2.3 CIE Lab color space ................................................................................. 8  
   2.2.4 CIE LCh color space ............................................................................... 8  
  2.3 Image capture and reproduction ........................................................................ 9  
  2.4 Standard and wide-gamut displays .................................................................... 10  
  2.5 Prior art in gamut mapping and extension algorithms ......................................... 11  
  2.6 Hypotheses ........................................................................................................ 14  

3 Experiment on the preferred and maximum acceptable color gamut for single-hue images .... 17  
  3.1 Introduction ........................................................................................................ 17  
  3.2 Method .............................................................................................................. 17  
   3.2.1 Design ........................................................................................................ 17  
   3.2.2 Participants ................................................................................................. 17  
   3.2.3 Equipment and setup ............................................................................... 17  
   3.2.4 Stimuli ....................................................................................................... 17  
   3.2.5 Measurement ............................................................................................ 20  
   3.2.6 Procedure ................................................................................................. 20  
  3.3 Results ............................................................................................................... 21  
   3.3.1 Preferred color gamut .............................................................................. 21  
   3.3.2 Preferred color gamut versus physical image properties ....................... 26  
   3.3.3 Maximum acceptable color gamut .............................................................. 29  
  3.4 Discussion ......................................................................................................... 34
4 Verifying tuning methodology ........................................................................................................37
  4.1 Introduction ..................................................................................................................................37
  4.2 Method .........................................................................................................................................37
    4.2.1 Design .....................................................................................................................................37
    4.2.2 Participants .............................................................................................................................37
    4.2.3 Equipment and setup ..............................................................................................................37
    4.2.4 Stimuli .....................................................................................................................................37
    4.2.5 Measurement ..........................................................................................................................38
    4.2.6 Procedure ...............................................................................................................................38
  4.3 Results ..........................................................................................................................................38
  4.4 Discussion ......................................................................................................................................38

5 Experiment on the preferred color gamut for single- and multiple-hue images .........................39
  5.1 Introduction ..................................................................................................................................39
  5.2 Method .........................................................................................................................................39
    5.2.1 Design .....................................................................................................................................39
    5.2.2 Participants .............................................................................................................................39
    5.2.3 Equipment and setup ..............................................................................................................39
    5.2.4 Stimuli .....................................................................................................................................40
    5.2.5 Measurement ..........................................................................................................................44
    5.2.6 Procedure ................................................................................................................................44
  5.3 Results ..........................................................................................................................................44
    5.3.1 Preferred color gamut for single- and multiple-hue images ....................................................46
    5.3.2 Preferred color gamut for single- and multiple-hue images versus physical image properties ....52
    5.3.3 Painting experiments ..............................................................................................................55
  5.4 Discussion ......................................................................................................................................57

6 Conclusion and future work ..............................................................................................................59

7 References .........................................................................................................................................61
Appendix A: Color terminology .............................................................................................................. 63
A.1 Intensity ............................................................................................................................................. 63
A.2 Brightness ......................................................................................................................................... 63
A.3 Luminance (Y) ................................................................................................................................. 63
A.4 Gamma ................................................................................................................................................ 63
A.5 Lightness (L*) .................................................................................................................................. 63
A.6 Chromaticity (CIE xy) ..................................................................................................................... 63
A.7 Correlated color temperature ......................................................................................................... 64
A.8 Hue .................................................................................................................................................... 64
A.9 Chroma ............................................................................................................................................... 64
A.10 Saturation ........................................................................................................................................ 64

Appendix B: instructions for the first experiment on preferred color gamut ........................................ 65
Appendix C: instructions for the first experiment on maximum acceptable color gamut ................ 66
Appendix D: instructions for experiment the second experiment ....................................................... 67
Appendix E: instructions for experiment the third experiment ............................................................ 68
1 Introduction into gamut extension research

1.1 Introduction
In the current information era, there is an increasing need to share information between humans. An increasing amount of this information is conveyed through digital displays like computer screens, mobile phone displays or televisions. Displays thus are important tools in the information era. Moreover, they often serve as a means of entertainment. It is therefore important that displays convey their information in a manner that is pleasant, comfortable, correct and clear. Technological developments in display technology lately focus on producing displays which are energy efficient, thin, have a large display area, produce a brighter image and have an increased range of colors they can display. This thesis will focus on the correct use of this increased range of colors, the so-called wide color gamut. A color gamut refers to the entire range of colors which a device like a display or a printer can produce. A wide color gamut is a color gamut which has a substantial larger range of colors than a standard color gamut.

Because the new generation displays have a wide color gamut, they can reproduce an original image more richly and realistically. There are two main tracks underlying this technological development of wide gamut displays. The first one, mainly driven by the use of LED backlighting, is to use more saturated primaries. The second is to enlarge the range of colors by the use of more than three primaries. Such displays are called multi-primary display. Even though most of the developments in this area are driven by a technology push instead of a user need (Laird & Heynderickx, 2008), there are perceptual reasons for the need of a wider color gamut which can produce a richer, better experience.

The data gathered by Pointer (1980) gives one of the perceptual reasons for a wider gamut. Pointer measured a large set of real surface colors. Figure 1.1 shows Pointer’s data in the CIE xyY color space (for an explanation, see subsection 2.2.2) together with a representation of a standard color gamut (white line) and an example wide-gamut display (white dotted line). Most of the real surface colors are covered by the standard color gamut existing of the color gamut standardized by the European Broadcast Union (EBU), or the similar HDTV Rec. 709 standard (“Parameter values for the HDTV standard for production and international programme exchange”, 2002) which are commonly used for image representation. There is, however, a large part of the real surface colors, especially in cyan, red, magenta and yellow, where the standard color gamut is smaller than the gamut of real world object reflections. A wide-gamut display covers a much larger set of the real surface colors. The wide color gamut displays thus have a clear added value concerning the display of real surface colors.
Figure 1.1: an xy plot of the CIE xyY color space (colored area) showing the boundary within which real surface colors can occur (black line), the boundary of the EBU gamut (white line) and boundary of the larger wide-color gamut (white dotted line). The plot shows the maximum (x,y) over all luminance levels. Adapted from Pointer (2009).

A second reason, showing the need for an expanded color gamut, is that for well known objects with a specific color, people expect a color which is typically more saturated than the object in the real world, a phenomenon called memory color (Bruner, Postman & Rodrigues, 1951 and Bartleson, 1960). Given the above arguments, it is thus not unreasonable to expect a need for a wider gamut.

Despite the fact that these newer displays have a wider color gamut, the color gamut of most media used is often smaller. Most media formats, like for example HDTV, use the color gamut as defined in the EBU standard or a similar one. Although some standards are being developed, like the xvYCC (Kim, 2008), which have a sufficiently wide gamut for these displays, they are currently far from being adopted into common end-user applications. As a consequence, in practical application the images sent to a display thus contain less color-information than could be displayed with wide-gamut displays. As a result, a method is needed to map this smaller color gamut of the input signal to its destination size of the wider-gamut display.

1.2 Issues in extending color gamuts

In order to map colors from a smaller color gamut to a wide color gamut, mathematical algorithms, so-called gamut extension algorithms (GEA), are used. The simplest way to extend a signal to a wide color gamut is to use the GEA which maps the primaries of the original gamut linearly to the wide color gamut. This GEA, called same drive signal, or SDS, does not require any image processing as it uses the original driving signal on the wide gamut display. This method, however, can result in unpleasant and unnatural image reproduction. Another, more complex, gamut extension method is thus needed to use the additional part of the wide color gamut. There are various studies which introduce, develop and evaluate a wide diversity of GEAs, for example Kuang, Muijs & Heynderickx (2005), Kim, Shin & Um (2005), Hyun Wook et al. (2006) and Kang et al. (2003). It is, however, difficult to extract conclusive guidelines from these works for several reasons. Firstly various studies use different images to develop and test their algorithms, while some of those same studies actually show that GEAs should take image content into account. Besides different content, most studies also use different hardware, which consequently results in having a different size of the wide gamut.
The wide color gamut of figure 1.1 (dotted line) can thus have another triangular shape (but still remain outside the EBU color gamut). The gamut extension algorithms used in these studies are therefore not generally applicable. There is thus no general reference to which gamut extension algorithms should be extended.

1.3 Goal
Previous gamut extension studies (for example Kuang, Muijs & Heynderickx, 2005, Kim, Shin & Um, 2005, Hyun Wook et al., 2006, and Kang et al., 2003) used the entire color gamut available in the displays and images they used for their studies. None of the proposed gamut extension algorithms were very successful and some of them even performed worse than the original gamut images. Using the entire wide color gamut can thus result in unpleasant and unnatural image reproduction. Laird & Heynderickx (2008) hypothesized that independent from the way the gamut is extended, there is a preferred gamut for the reproduction of natural color images. Such a boundary could serve as a basis for all gamut extension algorithms. Gamut extension algorithms are then easier to develop and will be more universally applicable. The goal of this study is therefore to observe and characterize the preferred and maximum acceptable gamut extension boundaries for different image content. This enables us to design displays which are built on human perceptual needs instead of on available technological applications.

1.4 Thesis overview
For readers who are not very familiar with the terminology and concepts of color science, chapter 2 provides the basic background knowledge they need to understand this thesis, followed by an overview of gamut extension algorithms and the hypothesis of this study. Chapter 3 describes the first experiment, which explored the preferred and maximum acceptable color gamut by means of a psychophysical tuning experiment on single-hue images. To verify the tuning methodology used in the first experiment, a second experiment, described in chapter 4, was performed. A paired comparison methodology combined with a staircase methodology was used for this experiment. Chapter 5 describes the third experiment, designed to evaluate a simple gamut extension algorithm, based on the results of the first experiment. More complex images containing multiple hues were tested together with single-hue images. Finally, the main conclusions and suggestions for future work are provided in chapter 6.
2 Literature review
This chapter provides an introduction to the research area of color science and previous literature. This introduction is merely an aid in understanding this thesis. More detailed information can be found in the cited literature and appendix A also provides some basic explanations about color terms which are used within this study. This chapter first describes what color actually is and how it is related to the human visual system. After that, color in image capture and reproduction is discussed, followed by the explanation of the difference between normal and wide-gamut displays. Different methods to extend color for wide color gamut displays are then discussed together with prior studies. Finally, the hypotheses used within this study are introduced.

2.1 General color introduction
Color is the perceptual response of the human visual system to electromagnetic radiation with a wavelength from approximately 400 to 700 nm. That part of the electromagnetic spectrum is called light, or visible light. Physically, light can be described by the power of the electromagnetic radiation at different wavelengths, or the spectral power distribution. The human visual system can however only distinguish a subset of all possible spectra. Due to the limited number of relatively wide electromagnetic sensors in the basic processing part of the human visual system, light sources with physically different spectra can be perceived as having an identical color. Spectra are called metameric and the effect is called metamerism. Metamerism is one of the basic enablers of capturing and reproducing colors and images. To reproduce an image, only a metameric stimulus needs to be created. There is thus no need to reproduce a stimulus with the exact same spectral power distribution.

Under most normal viewing conditions, three types of color-sensitive cells, called cones, are active in the human eye. The wavelengths to which these cells are most sensitive are spread over the visible spectrum. Based on their peak sensitivity, the color sensitive cells are called the short, middle and long cones. Their maximum sensitivity roughly corresponds to colors that are perceived as blue, green and red. By combining the responses of these three color receptors, a wide range of colors becomes perceivable.

Despite of the fact that basic color vision is based on three color receptors, humans do not communicate color using proportions of the short, middle and long wavelength content. They instead perceive and communicate color in three basic dimensions, namely brightness, colorfulness and hue. Brightness indicates the amount of light the human visual system perceives, which is often described as, for example, being bright or dim. Colorfulness can be described as the purity of a color. An often-washed t-shirt probably has a lower colorfulness than when it was new. Finally hue indicates which actual ‘color’ humans perceive. This can be, for example, red, purple, orange, or blue.

Light can reach the human eye in two different ways. Firstly, light from a light source (like the sun) can reach (illuminate) our eye directly. Secondly, light can reach our eye after having been in contact with an object. Contact with an object often alters light by three different means, namely absorption, transmission and reflection (Ryer, 1997). Absorption of light indicates that a certain part of the spectrum of light is transformed into heat in an object, while transmission refers to the part of the spectrum of light that passes through an object. When a red semitransparent sheet is held in front of the sun all colors but red are absorbed, while red itself is transmitted. Reflection of light can be relative to the angle the light hits the surface, like a mirror, which is called specular reflection. It can also be reflected diffusely, which means that the light is reflected in all directions. Finally most of the reflection occurs as a combination of specular and diffuse reflection and is called spread reflection. Spread reflection beholds limited diffuse reflection on an angle relative to the angle the light hits the surface.
Because light can interact with objects, the term color is also used to describe the influence that the object surface has on the spectral power distribution of the incident light. An object that reflects only light with long wavelengths and absorbs all others is usually called “red”. Due to this duality in the way color is defined, the result of mixing different colors also depends on the actual use of the word. When different object colors are mixed, like when mixing paints, the absorption of the different paints is added, which in effect “subtracts” the absorption spectrum of each paint from the spectral power distribution of the incident light. This type of color mixing is called subtractive. Subtractive colors always start with white which is then filtered. For example, when yellow paint on a white surface is mixed with magenta it will appear red. When subsequently cyan is added, the paint will appear black since all colors are absorbed. When lights from different light sources are mixed, the spectral power distributions of the lights are added. This type of color mixing is called additive. If a red lamp is put in a room without any light sources, the room will appear red. When a green lamp is added this will produce a yellow light. If subsequently a blue lamp is added the room will appear to be illuminated white.

2.2 Color spaces
Color is easily communicated through human language. But, to capture it in a form that can be stored and reproduced to match an original, a mathematical definition of the perceptual basis of color is needed. The models of color representation are traditionally called color spaces. Each color space has its own function and origin. However since so many different color spaces exist, only the ones used within this thesis are being described in this section.

2.2.1 RGB and sRGB color space
Based on the fact that human color vision is based on three color sensitive receptors each with their own peak sensitivities, the idea of using three base colors (red, green, and blue) and representing all other colors as their combination, is a natural one. The resulting color space is called RGB, based on the first letters of the base colors, also called primary colors. As every other color is represented as a weighted linear combination of the primaries, these primaries can be interpreted as a coordinate in a three dimensional space. Even though the space of colors is a continuous one, in practical applications, each dimension has a limited set of steps (often 256). In total 16 million possible combinations of red, green and blue can be produced. An example of the RGB color space can be seen in figure 2.1.

Figure 2.1: visual representation of the RGB color space. Adapted from Colontoni (2006).
Although at first glance the RGB color space could appear to be very useful, it also has several drawbacks. Firstly, RGB is a device dependent color space, i.e. the actual color, defined by its RGB coordinates, depends not only on the coordinates themselves, but also on the properties of the capture or the reproduction system. Full red on one display can appear different on another display, while full red captured by one camera does not mean it is captured as full red when another camera is used. Furthermore humans do not perceive color as quantities of red, green and blue. They use quantify color in terms of hue, saturation and lightness.

sRGB (Stokes et al. 1996) accounts for the problem that RGB lacks information on how its primary dimensions should be defined. sRGB is a standardized color space, which is designed to prescribe how RGB should be displayed on a common monitor or television. sRGB is defined with a white point of CIE D65 and has an encoding curve equivalent to a gamma of 2.2. For computer monitors most common images and video material are defined in the sRGB color space. However, common broadcasted television uses a different color space. The color space used for encoding video material depends on the area, with PAL and NTSC being the most used formats in Europe and the USA. The ITU-R BT.709-2 (“Rec. 709”) standard is used for the encoding of so called HD television broadcast.

2.2.2 CIE XYZ and CIE xyY color space
To provide a base color space for the description of color, the CIE XYZ color space has been developed by the CIE consortium. It is a mathematical model based on measurements of the human visual system. It consists of three dimensions, namely X, Y and Z, which encapsulate all colors visible to the human eye. The CIE XYZ color space is mainly used as an intermediate, device-independent color space to transform colors between color spaces. From the CIE XYZ color space the CIE xyY color space can mathematically be derived. This color space makes a distinction between luminance (Y) and chromaticity (x,y). Chromaticity can be described as color without taking brightness into account. Luminance is more complex to describe. This is because the human visual system has different sensitivities to electromagnetic radiation of different wavelengths. Correcting the intensity measure (which is the physical measure of energy over some interval of the electromagnetic spectrum which is radiated by a surface) to this sensitivity function results in the measure called Luminance. The CIE xyY color space can be seen in figure 2.2. Often a top view of the CIE xyY color space is displayed as can be seen in figure 1.1.

Figure 2.2: CIE xyY color space with the transformed RGB values of figure 2.2 drawn into it. Displayed colors are only indicative and do not represent the actual color. Adapted from Colantoni (2006).
2.2.3 CIE Lab color space

Even though CIE XYZ is very useful as a base definition of absolute color, it has poor performance in predicting differences between colors. Two color pairs with the same difference in their CIE XYZ coordinates within the pair can have a very different perceived difference. Several attempts to quantize and capture the perception of color differences have been made over the years. The CIE Lab color space is a derivative of the CIE XYZ color space and the coordinates of a color in it can be calculated using its CIE XYZ coordinates. Additionally, the perception of differences between color pairs depends on their surroundings. This dependency is captured in the CIE Lab model by normalization to the color of a perfectly white object under the same light condition. In this context, the color of the perfectly white object is called the white point. One of the color spaces that have been derived with this goal in mind, is CIE L*a*b* or CIE Lab in short. The L* dimension represents the perceptual measure lightness, which represents the brightness of an object judged relatively to the brightness of a white object under the same illuminant. The a* dimension represents a perceptual uniform color transition from red to green over gray, while the b* dimension represents a perceptual uniform color transition from blue to yellow over gray. Since CIE Lab entails both lightness and uniform color transitions it is often used for color (re)production.

The CIE Lab color space does not predict the entire color range correctly. Colors with a low colorfulness and a blue-like hue are often perceived to have a purple hue.

![CIE Lab color space with the transformed RGB values of figure 2.2 drawn into it. Displayed colors are only indicative and do not represent the actual color. Adapted from Colantonii (2006).](image)

2.2.4 CIE LCh color space

Although CIE Lab provides a good model for calculating differences of colors, it does not entail a set of intuitive dimensions. As mentioned before, users often describe colors in 3 qualities, namely brightness, colorfulness and hue. The CIE LCh (also named CIE L*C*h*) color space, which is based on the CIE Lab color space, is an approximation of these more intuitive measures. The CIE LCh L* is identical to the L* of the CIE Lab color space. However, CIE LCh encompasses the measure h* (hue) and C* (chroma) instead of a* and b*. Hue as defined in the CIE LCh color space is related to color, but it differs since hue consists of one of the colors red, yellow, green and blue or any combination of two of these colors (Wyszecki & Stiles, 1982). Chroma can be defined as “the colorfulness of an area relative to the brightness of a reference white” (Hunt, 1978). Physically, it relates to the amount of light of one specific wavelength. Hence if more light particles of a single wavelength are emitted, chroma increases. In more simple words this means that chroma indicates the amount of chromatic content. Figure 2.4 shows the CIE LCh color space. Note that in this figure the hue axis is linear, so that the difference in chroma for each hue is better visible. Often hue is displayed on a radial axis, since it is expressed in an angle from 0 to 360 degrees.
2.3 Image capture and reproduction

This section describes the process an image has to go through from being recorded until being correctly reproduced on, for example, a television screen. A wide range of color spaces can be used for this reproduction process, but only the color spaces as described in section 2.2 are used to explain this process.

MacDonald (1993) describes a five-stage transform for the accurate reproduction of images (figure 2.5). An RGB image produced by, for example, a camera should first be adapted so that it represents the real (device-independent) colors it recorded. In the first step (1) the RGB values from the recorded image will thus be transformed into CIE XYZ values using the characteristics of the camera. The so-called forward device model takes care of this transformation. Although the real recorded colors are obtained after this first step, the viewing conditions at which the original image was taken should also be accounted for. The forward appearance model (2) accounts for the viewing conditions of the camera (by encompassing its white point) and produces CIE Lab values, which can directly be translated into CIE LCh values. At step three (3) the color gamuts of the recording and reproduction device (i.e. a television) are compared and the gamut of the recorded image is mapped on the gamut of the reproduction device (either being reduced or extended). This thesis focuses on the latter case which is often referred to as gamut extension.

Figure 2.4: CIE LCh color space with the transformed RGB values of figure 2.2 drawn into it. Displayed colors are only indicative and do not represent the actual color. Adapted from Colantoni (2006).

Figure 2.5: the five-stage color reproduction transform. In between each step the color space used within this study is given. Adapted from MacDonald (1993).
After having mapped the CIE LCh values to the gamut of the reproduction device, the image has to be transformed back into RGB value for the television. The inverse appearance model (4) takes care of transforming the CIE LCh values through CIE Lab back into CIE XYZ values using the white point of the television. Finally the inverse device model (5) translates the CIE XYZ values back into RGB by using the characteristics of the television.

2.4 Standard and wide-gamut displays
Since this thesis presents a study on wide-gamut displays, this section will briefly describe the difference between standard and wide-gamut displays.

The first generations of televisions and monitors used so-called Cathode Ray Tubes (CRT). Within these CRT displays electrons collide on a special phosphor layer (Robinson, Chen & Sharp, 2005). Figure 2.6 shows a simplified overview of a Cathode Ray Tube. At the back of the tube an electron gun is mounted which fires a beam of electrons towards the screen. Several techniques ensure that the electrons arrive at the correct position on the screen. On this screen a single layer of tiny phosphor dots has been placed. Each dot can either emit red, green or blue light depending on the type of phosphor. A wide range of colors can then be reproduced by sending the right amount of electrons to each red, green and blue dot.

![Figure 2.6: simplified overview of a Cathode Ray Tube (CRT). An electron gun fires an electron beam to a screen covered with phosphor dots. A dot can only emit red, green or blue light. By adjusting the strength of the electron beam to each dot a wide range of colors can be produced. Adapted from Robinson, Chen & Sharp (2005).](image)

A more recent technique used for displays is the Liquid Crystal Display (LCD), which is shown in figure 2.7. This much thinner displaying technique uses a light source with a broad spectrum (thus producing a bright ‘white’). The light from this source is first equally spread over the surface of the display. Secondly the light passes a complex color filter layer, which controls the amount of colored light to each small part of the display, its pixels, separately. Every single pixel is divided into 3 so-called subpixels, each of which has a red, green or blue hue. Since the intensity for each hue can be controlled, a wide range of colors can be produced.
Figure 2.7: simplified overview of a Liquid Crystal Display (LCD). Wide spectrum light gets guided through a light guide, passes a color filter layer which splits each pixel in a red, green and blue sections of which the light intensity can be controlled independently. Adapted from Robinson, Chen & Sharp (2005).

Most standard displays use primaries as defined in the EBU or sRGB standards. In order to create a wide-gamut display it is important to use primaries for the red, green and blue subpixels which have a narrow spectrum combined with a high intensity. In the ideal case the primaries should consist of a single wavelength of light with a very high intensity. Most lasers possess these characteristics, but lasers are not (cost-effectively) scalable to a normal display size. The primaries produced by phosphor (as used for CRT displays) often have a high intensity, but the spectrum, of mostly green and blue phosphors, is relatively broad. Because most LCD displays use a backlight with a wide light spectrum, they can produce high intensity colors when saturation is low. If a highly saturated color has to be produced, a large part of the light spectrum needs to be absorbed by the color filter layer. Therefore, highly saturated colors have a relatively low intensity when a wide spectrum light source is used.

With the introduction of high power LEDs (Schubert, 2006) practical and affordable high intensity light emitting devices with a narrow spectrum became available. When the light source of a LCD screen is replaced by an array of red, green and blue high power LEDs, the spectrum of the primaries is narrower than the spectrum of conventional LCDs. Because of this narrower spectrum, the LEDs create a wider gamut for the display. A slightly different method to create wide-gamut display is by introducing more primaries into an LCD such as a five channel multi-primary display, as shown by Kim, Shin, and Um (2005).

2.5 Prior art in gamut mapping and extension algorithms

Research in traditional gamut mapping mainly concerned mapping color gamuts to smaller color gamuts. Smaller gamuts are often found in printing processes and certain types of displays (for example mobile telephone displays). Methods were developed to use this smaller color gamut optimally to maintain a good overall color appearance. In order to achieve good overall color appearance, most gamut mapping algorithms operate in perceptual color spaces. Examples of such spaces are CIE Lab (see subsection 2.2.3), CIE LCh (see subsection 2.2.4), CIE LUV (Morovic & Luo, 2001) and IPT (Ebner & Fairchild, 1998). Since gamut mapping algorithms should minimize hue shifts (MacDonald, 1993), gamut mapping is mostly done in the lightness and chroma dimensions (or similar ones). In this lightness-chroma plane, gamut mapping can be done mainly in two ways; using gamut clipping and using gamut scaling (shown in figure 2.8). In gamut clipping, the colors which cannot be produced by the smaller gamut, are decreased to the maximum chroma of the smaller gamut. The value of the chroma of the original gamut relative to the maximum chroma of the original gamut can also be used to scale the chroma in the smaller gamut, which is called gamut scaling. Gamut scaling is characterized by the direction and by the type of extension (i.e. linear or non-linear), which will be discussed in the light of gamut extension.
More information about several classic gamut mapping algorithms can be found in the works of Morovic & Luo (2001) and Morovic (2002). Zolliker & Simon (2006) reviewed the state of the art in 2006.

Gamut extension algorithms, which extend small and standard color gamuts to wider color gamuts, are largely based on knowledge gained in gamut mapping. Because gamut clipping is only possible for mapping into smaller color gamut, gamut extension algorithms are based on scaling to the wide color gamut. As mentioned before the direction of this scaling can differ. Figure 2.8 shows four commonly used chroma extension directions. The most basic direction is to extend only in chroma (indicated as chroma in figure 2.8), and thus leave lightness at its original level. Extension based on vector direction, also called saturation direction, keeps saturation constant by changing lightness and chroma along radial lines through the origin. Another extension direction, centroid, is based on extension on a line from the centroid of the lightness axis (at L*=50) through the original color. Finally, an adaptive extension direction extends chroma depending on the lightness level of the original color. It boosts more chroma for dark colors and more lightness for unsaturated bright colors (Kuang, Muijs & Heynderickx, 2005).

![Figure 2.8: lightness-chroma plot with a gamut scaling and gamut clipping example and four examples of chroma extension methods. The solid blue line indicates a fictive standard color gamut while the dotted-striped line indicates a fictive wide color gamut. The green striped line indicates a fictive small color gamut. Black dots with a gray border indicate mapped colors while solid black dots represent the original color. Adapted from Kim, Shin & Um (2005).](image)

Each of these extension directions, can be combined with different extension types. Kuang, Muijs & Heynderickx (2005) describe several extension types which are also shown in figure 2.9. Extension types describe the way original colors are extended along the direction of the extension method. Linear extension, for example, extends the smaller gamut linearly to the wide color gamut. Another extension type is piece-wise linear extension, which preserves low chroma colors while it linearly extends higher chroma colors. This method of preserving low chroma colors is used because most natural objects have a low to medium chroma which should not be extended too far (for example skin colors). High chroma colors are often associated with artificial colors and can therefore be boosted more. High chroma boost (HCB) functions similarly to piece-wise extension, preserving low chroma color while higher chroma colors are boosted relatively more. HCB, however, extends non-linearly using a second-order function. In contrast to HCB, low chroma boost (LCB) extends low chroma colors more than higher chroma colors and it is intended to reduce very high chroma levels. LCB also extends using a second-order function. Finally sigmoid extension, which uses a sigmoid function, is an extension type which preserves low chroma colors while it boosts high chroma similar to LCB. Sigmoid extension thus creates the largest contrast in chroma in the wide color gamut.
Several articles report evaluations of existing and developed gamut extension algorithms (GEAs). Kim, Shin & Um (2005) describe their own developed GEAs for five channel multi-primary displays for HDTV applications. Their algorithms are based on chroma extension, vector extension and adaptive extension and are developed with hardware limitations in mind. They used wide color gamut images which were first mapped back into the Rec 709 color space. They developed their algorithm not only for the uniform CIE Lab color space but also for the YWV color space, which is computationally less demanding, but at the same time perceptually less uniform. Furthermore their algorithms are designed to be used by one specific display. They compared their algorithms to the Rec 709 standard and the original wide gamut images. They concluded that the CIE Lab color space is more robust for optimizing gamut mapping. No preferred gamut extension algorithm was found since their performance was image dependent. Furthermore the performance of the extension algorithm based on adaptive extension was largely dependent on the actual angle used to map the original colors. Hyun Wook et al. (2006) developed a method to map the Rec. 709 color gamut into their multi-primary wide color gamut display. Their GEA mainly focused on how to map the 3 primaries of Rec. 709 into their 5 primaries display, especially on how to create smooth transitions between colors when using a multi-primary display. They used linear chroma extension to extend the Rec. 709 color space, but no conclusive results about this extension method were discussed.

Kuang, Muijs & Heynderickx (2005) developed two new GEAs in the CIE Lab color space and evaluated them on end-users together with 3 existing algorithms. Their developed GEAs consisted of a chroma extension and a lightness-adaptive extension algorithm which both used HCB extension. The other three GEAs consisted of a colorimetric mapping algorithm, same drive signal (SDS), and Wide Gamut Color Mapping algorithm (WGCM). The colorimetric mapping algorithm tries to reproduce the original image as accurate as possible and served as a baseline for the other GEAs. For SDS, as discussed in section 1.2, no image processing was performed. This way the primaries of the original color gamut are directly mapped to the wide color gamut. WGCM uses techniques from both the colorimetric and SDS algorithms. It preserves color for the lower chroma colors, but extends towards the primaries, like SDS, when chroma increases. Their results indicate that participants did not have a pronounced preference for any of the five algorithms. The adaptive GEA, however, provided the most promising results. Furthermore, it was found that performance of the GEA was dependent on the images used.
Kang et al. (2003) also developed and evaluated several GEAs in the CIE Lab color space, which were all based on different variations of chroma extension. They used a paired comparison methodology to first test how their GEAs performed. After receiving this feedback they adjusted their GEAs and added two more. The results from the first experiment were used to predict the outcomes of the GEAs, which were tested using in a second experiment. The results of the second experimental show that participants preferred more colorful images than they initially indicated during the first experimental part. Although they used only four images in their experiment, a clear image dependency was found. They also concluded that memory colors (Bruner, Postman & Rodrigues, 1951) had a large impact and should be accounted for in further experiments.

None of the above mentioned studies provides conclusive guidelines to develop a universal GEA. Some of the developed GEAs do not even outperform content with a standard color gamut. Other GEAs, however, are slightly preferred over content with a standard color gamut, but since they are developed for specific hardware, they are still largely incompatible. In order to unify the development and implementation of GEAs it is thus important to have a color gamut boundary which described the extent to which GEAs should use a color space. Although a very large color gamut boundary could be used to develop a universal GEA, it is more logical to use a user-dependent color gamut boundary since that will incorporate the actual needs of the users of the technology than the mere technology itself. Laird & Heynderickx (2008) conducted a first study to find such a user-preferred gamut boundary. They changed both the lightness and chroma of two single-hue natural images to find that the typical preferred gamut for natural content is outside the EBU gamut, but closer to the EBU standard than to the wide-gamut boundary of their wide-gamut display. They conclude that the more saturated green of the wide-gamut displays is most useful, and that for particular types of content, the entire wide gamut can be used (like neon lights and fireworks). Sakurai et al. (2008) mapped wide color gamut content to a wide gamut display with a smaller color gamut and indeed found that for some images the entire wide gamut can be used, while for the more natural images participants were more sensitive. They thus found an image dependency for the amount of gamut extension, like in most studies mentioned above.

This study continues the work of Laird & Heynderickx (2008) to find the preferred color gamut boundary. Although they found interesting results, their results were limited by the use of only two, monochrome, images. This study includes not only more content, but also multiple-hue images.

2.6 Hypotheses

Previous studies did not find a generally preferred GEA. This can indicate that there is no need for a wider gamut. The studies of Laird & Heynderickx (2008) and Sakurai et al. (2008), however, indicate that at least for some image content the wider gamut is preferred. Based on this we formulated the following hypothesis:

1. The boundary for the extension that people prefer (“preferred” gamut) is larger than the standard TV gamut

Based on observations and on the results found in previous studies, which indicate that colors are easily too saturated, the following hypothesis was formulated:

2. The boundary past which people stop accepting the extension (“acceptable” gamut) is smaller than the wide-gamut display used in this study

The study of Laird & Heynderickx (2008) found that image content affected the chroma preference, even though they conducted their experiment on a small set of images. The studies of Kang et al. (2003) and Kuang, Muijs & Heynderickx (2005) also found that their GEA evaluations showed a dependency on image content. Based on this the following hypothesis was created:

3. The preferred gamut boundary is content dependent
The study of Laird & Heynderickx (2008) furthermore found that hue affected the chroma preference, thus different hues were assessed differently. Also Kang et al. (2003) indicated that content dependent memory color can largely affect the evaluations of images. Based on these findings the following hypothesis was formulated:

4. The preferred gamut boundary has an irregular shape and therefore is hue dependent

Finally, the study of Laird & Heynderickx (2008) indicated that also the participants in their experiment influenced the results on chroma preference. Based on this knowledge the following hypothesis was created:

5. The preferred gamut boundary is person dependent
3 Experiment on the preferred and maximum acceptable color gamut for single-hue images

3.1 Introduction
This experiment consists of two related experimental parts. In the first experimental part (E1a) the gamut boundary for the preferred level of chroma is explored, while in the second experimental part (E1b) the maximum level of acceptable chroma is examined. To assess chroma preference and maximum chroma acceptance, users were exposed to stimuli with a single hue, with varying content depicting real world scenes.

3.2 Method

3.2.1 Design
This psychophysical experiment uses a within-subject design. Since a wide variety of content was assessed, which can be rather time consuming and tiring, a fast methodology, namely tuning, was used to measure both the preferred chroma level and the maximum acceptable chroma level.

3.2.2 Participants
A total of 42 persons (21 male and 21 female) from the Philips High Tech Campus Eindhoven, of which 24 persons had a Dutch nationality, participated in both experimental parts. The average age was 31 (SD=11), varying between 21 and 61. All participants were tested negative for color deficiency according to the Ishihara (1999) test.

3.2.3 Equipment and setup
The experiment ran in a dim room with one homogenously lit wall (22 lux measured on the wall, 3.5 lux measured on the display surface). Close to this wall a 40" Sony Qualia LCD wide-gamut display with a LED backlight, a D90 white point and a maximum luminance of 341 cd/m² was placed. The CIE (x,y)-chromaticities of the primaries of the display were (0.699, 0.291) for red, (0.188, 0.700) for green and (0.146, 0.063) for blue. The participants sat behind a table at approximately 3 meters distance from the display surface. Stimuli were displayed one at a time on a gray background, corresponding to 50% lightness of the maximum white of the display. The stimulus order of appearance was randomized, as well as the starting point of the chroma level. Three stimuli were repeated for reliability testing. Participants could change the chroma of the stimulus in front of them by means of pressing the arrow keys on a keyboard. By pressing the enter button the next stimulus would appear. Participants could tune each stimulus to their level of color without any time constraints.

3.2.4 Stimuli
A total of 12 images (Figure 3.1) with different content were used to create the stimuli for both parts of the experiment. Images were selected on the criteria of having a large high-chroma region and having small hue differences for the pixels with a chroma not close to 0.
Figure 3.1: all the images used in both experimental parts. The four top images were presented in 5 single-hue versions. As shown: window, rose, room, paint, water drop, fish, grass, grasshopper, leaf, berry, field, and flower.

Four of the images, namely window, rose, room and paint, were modified into 7 single-hue versions for each image. The hues used to create these images consist of the primaries of the display, which are red (R), blue (B) and green (G). Furthermore the secondaries of the display were selected, being yellow (Y), cyan (C) and magenta (M). Finally an additional red (R_{EBU}) was selected which corresponds to the red primary of the standardized EBU gamut. This R_{EBU} was included since pilot testing showed a clear difference in preference between R and R_{EBU}. An overview of the selected hues and their corresponding CIE LCh hue angle are presented in table 3.2. Two of the remaining images were shown a second time with a different hue than the original. The image water drop was also shown with the C hue and the image raspberry with the R hue. The remaining 6 images were shown in their original hue only. This experiment thus used in total 38 images with a different hue-content combination.

<table>
<thead>
<tr>
<th>Hue name</th>
<th>hue label</th>
<th>CIE LCh hue angle (CIE 90 white point)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>R</td>
<td>39</td>
</tr>
<tr>
<td>Green</td>
<td>G</td>
<td>149</td>
</tr>
<tr>
<td>Blue</td>
<td>B</td>
<td>308</td>
</tr>
<tr>
<td>Yellow</td>
<td>Y</td>
<td>104</td>
</tr>
<tr>
<td>Cyan</td>
<td>C</td>
<td>191</td>
</tr>
<tr>
<td>Magenta</td>
<td>M</td>
<td>335</td>
</tr>
<tr>
<td>EBU red</td>
<td>R_{EBU}</td>
<td>44</td>
</tr>
</tbody>
</table>

Table 3.2: the 7 hues used to create the hue rotated stimuli for this experiment. The hue label is used throughout this report.
In order to change the hue of the original images, the CIE LCh h-value was rotated at a fixed angle for all pixels in an image. For each image of which the hue was shifted, the lightness levels were corrected accordingly. Since the maximum chroma for each hue is located at a different lightness level, chroma values were mapped to account for the difference in gamut shape. Given the old lightness \( L \), chroma \( C \) and hue \( h \) coordinates of every pixel, and the target hue \( h_t \), the new L’C’h’ coordinates of the pixel are given by equations 1 and 2. The hue of every pixel was set to the target hue. The chroma value was scaled according to the ratio of the chroma levels of the maximum chroma points (cusps) of the lightness – chroma plane for the old hue \( (C_{cusp}) \), and the target hue \( (C'_{cusp}) \). The lightness is adapted in two steps. First, the total transformation \( (L') \) was computed. To compute the total transformation, the lower and upper parts of the lightness axis were divided in two parts by the point with a lightness level of the cusp \( (L_{cusp} \text{ and } L'_{cusp} \text{ in the original and the target hue plane}) \). Next, the corresponding parts were linearly scaled to match the target hue plane lightness distribution. This transformation changes the lightness of achromatic, or low chroma pixels, which introduces tone scale changes to the parts of the image that should not be adapted. To correct this, the final lightness level \( L' \) was computed as a convex sum of the original lightness \( L \) and the total transformation lightness \( L' \), depending on the relative chroma of the pixel \( C/ C_{cusp} \). As a result, pixels on the lightness axis (chroma or zero), kept their original lightness, while pixels on the cusp had the totally transformed lightness.

\[
\begin{align*}
    h' &= h_t, \\
    C' &= C \cdot \frac{C'_{cusp}}{C_{cusp}}, \\
    L' &= L' + \frac{C}{C_{cusp}} + L \left( 1 - \frac{C}{C_{cusp}} \right),
\end{align*}
\]  

where \( L' \) is given by

\[
L' = \left\{ \begin{array}{ll}
L - L_{cusp} \times \frac{100 - L'_{cusp}}{100 - L_{cusp}} + L'_{cusp}, & \text{if } L > L_{cusp} \\
L \times \frac{L'_{cusp}}{L_{cusp}}, & \text{if } L \leq L_{cusp}
\end{array} \right.
\]  

Figure 3.3: chroma-lightness plots of the image rose in R hue (left) and Y hue (right), red dots represent the pixels of the image while the blue line represents the displays boundary at the maximum chroma value. As can be seen the gamut is shaped different for both hues and the images are adapted to the shape of both boundaries.
In order to create stimuli from these images with different hues and different levels of chroma, these original images need to be transformed from their RGB value into a CIE LCh value. Hence a model of the display device was needed. Because the colorimetry in the experiment is more important than the simplicity of the model, nonstandard models were used. The CIE XYZ values at all 256 digital values for the primaries were measured using a colorimeter. Using this data, a forward and a reverse model was developed. The RGB values output were verified by measuring them on the display. From the model the CIE Lab and CIE LCh (Wyszecki & Stiles, 1982) color gamuts for the display were derived using the CIE D90 white point and the forward model of the wide-gamut display. Once the CIE LCh value was modified, RGB values were calculated using the inverse model. For the inverse model the CIE LCh values were reverted -through the CIE Lab space- to XYZ values for each primary which were then transformed into the RGB values for the display. RGB values were obtained by first doing a local search and if necessary followed by interpolation to obtain the required value. If a RGB value outside the 0 to 256 range was obtained, the CIE LCh chroma value was decreased until a valid RGB value could be produced. Figure 3.4 shows the pixel transformation process as used for the gamut extension algorithm.

Figure 3.4: the trajectory for each pixel as it is being transformed from its original RGB value to a RGB value suitable for the experiment.

From each image, stimuli with different chroma levels were produced using a chroma multiplier. A multiplier of 0 resulted in a grey-scaled stimulus, while a stimulus with a multiplier of 1 was at equal chroma level of the original image. Chroma differences were spread over 80 stimuli per image, in steps of 0.05 of the original chroma. For a multiplier larger than 1, every next 20 steps the chroma was extended 100% from the original value.

3.2.5 Measurement
For this experiment two independent variables were used; namely image content and hue. There were 12 different types of image content and 7 different hues. In the first part of the experiment the dependent variable was preferred chroma, while in the second part of the experiment maximum chroma acceptance was the dependent variable.

3.2.6 Procedure
Participants were first tested for color deficiency after which they were presented with a written instruction (see appendix B for the first experimental part and appendix C for the second experimental part). Additional oral explanations were provided by the experimenter. In the first experimental part participants were instructed to “adjust the color of the image to your preferred level”. In the second experiment, which was conducted at least one week later, participants were instructed to “select the highest level of color to which the image –in your opinion- still looks acceptable.” At the start of each experimental session, two trials were presented. After each session, which on average lasted about 8.1 minutes (SD=3.4), participants were given the opportunity to provide oral feedback to the experimenter.
3.3 Results

The results show the preferred and maximum acceptable chroma level per image per participant. In total each participant provided 41 data points per experimental part. The 3 reliability measurements were excluded from the analysis, except for the reliability test, to keep the number of data points per image equal for the analysis. The results of both experimental parts will be first presented separately (subsection 3.3.1 and 3.3.3) after which they will be discussed (paragraph 3.4). Since the original images used to create the stimuli have large high chroma areas, they almost reach their overall maximum chroma value of the display at a chroma multiplier of 1 and thus do not change much if their chroma multiplier is increased. Therefore results higher than a chroma multiplier of 1 have been clipped to a chroma multiplier of 1. Furthermore, results have been rescaled to be relative to the wide-gamut display. Therefore the chroma multiplier ranging from 0 to 1 has been linearly transformed to the range of the display which goes from 0% to 100% of the maximum chroma of the display at each representative hue. It should be noted that this scaling factor is different per hue.

To test if the results are reliable, a paired sample t-test was conducted to compare the results of the first measurement to the results from the second -repeated- measurement. For the first experimental part the results show no significant differences for the image paint with M hue ($t(41)=0.07$, $p = 0.946$), the image rose with B hue ($t(41)=0.70$, $p = 0.486$) and the image window with R hue ($t(41)=0.04$, $p = 0.970$). Also the results of the second experimental part show no significant differences for the image paint with M hue ($t(41)=0.25$, $p = 0.804$), the image rose with B hue ($t(41)=-0.63$, $p = 0.530$) and the image window with R hue ($t(41)=-0.54$, $p = 0.596$). These results of both experimental parts suggest that the repeated stimuli were assessed similarly in both cases and that the measurement method was reliable.

3.3.1 Preferred color gamut

The results from the first part of the experiment are shown in figure 3.5. Results are shown per content and per participant. The red lines indicate the median result while the blue bars indicate 25% of the results above the median and 25% of the results below the median. A first glance at the results shows a low preference for the window images and a large difference between participants.

![Figure 3.5: Box plots of the results of the first experimental part in which chroma preference was assessed. The left box plot shows the results per stimulus while the right one shows the results per participant. The percentage of chroma is relative to the hue dependent maximum amount of chroma which can be produced by the wide-gamut display.](image_url)
The median results per hue for all content are as shown in figure 3.6 together with the EBU gamut boundary (single red horizontal lines). This figure shows that over 50% of the results were out of the EBU gamut for five of the seven hues. The B (blue) hue is one of the exceptions, as can be seen in figure 1. The R (red) hue also is, on average, preferred with a lower chroma, which could be caused by its rather magenta appearance. Figure 3.6 also shows that, when preference is assessed, most participants do not prefer stimuli which contain the maximum chroma of the wide-gamut display. As can be seen, at least 75% of the participants did not pick a stimulus which was at the boundary of the wide gamut of the display for the R, G, M and R_{EBU} hues. So at least for the images used in this experiment most of the hues produced by the wide-gamut display are sufficient. Using the same drive signal for mapping image content into the wide color gamut of the display is not preferred.

![Box plot of the results per hue over all content of the first experimental part in which chroma preference was assessed. The single red horizontal lines indicate the EBU gamut boundary. The percentage of chroma is relative to the hue dependent maximum amount of chroma which can be produced by the wide-gamut display.](image)

In order to see in more detail how images were assessed within the EBU gamut and the wide-gamut display their actual distribution of pixels in the lightness-chroma space was investigated. The results as presented above, are only an indicator of the distance of the maximum chroma in a stimulus to the gamut boundary of the wide-gamut display. These new results thus take many more pixels within a stimulus into account. The median preferred stimulus per image was selected and the lightness-chroma coordinates of all pixels of this stimulus were plotted in a lightness-chroma plot. For an example of a lightness-chroma plot see figure 3.3. A lightness-chroma plot was created for these stimuli. From each lightness-chroma plot the average of 5% of the highest chroma pixels were calculated around a 2% offset above and below the lightness level at the maximum level of chroma (of wide-gamut display gamut). The same average was also calculated halfway in between both the 0 and 100 lightness boundaries and the lightness level at the maximum level of chroma. This resulted in 3 chroma averages at 3 lightness levels for each stimulus. Combining these averages over all images grouped per hue provides figure 3.7. Both the EBU gamut boundary (green line) as well as the wide-gamut display gamut boundary (blue line) was drawn into the figure. Please notice that some data points are missing since there are no pixels in that lightness region. Furthermore some pixels may appear to be out of the gamut in the figure, which can be explained by the fact that the pixels in that given lightness region have a slightly different hue angle and thus are accompanied by the wrong gamut boundaries at that lightness.
Figure 3.7: lightness-chroma plots of the first part of the experiment (E1a) per hue including all stimuli (see legend). The EBU gamut (green/inner line) and the wide-gamut display gamut (blue/outer line) are indicated as well. Data points represent the average of 5% of the highest chroma pixels, of the median preferred stimulus, at the given lightness levels, using a 2% lightness offset.

As can be seen in figure 3.7 the median preferred stimulus for the image window is always within the EBU gamut for all lightness levels, which confirms earlier findings that indicate that the window image is rated below the EBU gamut boundary. For the image paint, however, the average chroma level of 5% of the highest chroma pixels is for most hues within the EBU gamut boundary at high lightness levels while at lower lightness levels it is mostly out of the EBU gamut. This means that although results indicate that the median result is out of the EBU gamut, there still are a vast amount of pixels within those stimuli which are within the EBU gamut. Furthermore figure 3.7 shows that for a lot of images the median preferred chroma is indeed far out of the EBU gamut boundary like indicated by the box plots. These results confirm the need for a wider color gamut for the selected images.

In order to quantify the influence of hue, content and participant on the results, a general linear model based on the analysis of variance (ANOVA) was used. In order to perform an ANOVA to the dataset, the data should comply with four criteria (Field, 2005); (1) the data should be normally distributed, (2) the variances throughout the data should be homogeneous, (3) interval data should be used and (4) data from different participants should be independent. As can be seen in the histogram of figure 3.8, the data is not normally distributed. However the normality in an ANOVA is required on the error, or the residuals of the data (obtained by omitting the 3-way interaction of hue, image and participant). As can be seen in figure 3.8, a normal distribution is found for the residual. We can thus conclude that criterion 1 is satisfied.
Because the results consist of a large set of 1596 samples the Levene’s test is significant (p<0.001) for both experimental parts, as can be expected. The variance ratio of hue, however, is 1.6. The violation of the homogeneity of variance will therefore have a negligible effect on the analysis (Field, A., 2005). Criterion 3 is satisfied since the chroma steps can be considered perceptually homogeneous because they have equal steps in the homogeneous CIE LCh color space. The images at which chroma was extended for more than 100% perceptually changed less, because most of the pixels were clipped at 100% chroma. This can be ignored since the results of these images would be treated as being extended to 100%. Criteria 4 can be satisfied since the subjects participated separately.

Because the 4 criteria were met, ANOVA was applied to the hue rotated images. For these images all hues were used for each type of content. Preferred chroma was the dependent variable, while image content and hue were the independent variables. Participant was a random factor and both the main effects and two-way interactions were taken into account. The results from the ANOVA are displayed in table 3.9. Image content had a significant effect (F=38.1, df=3, p<0.001, $\eta^2_p =0.482$). Significant effects of hue (F=14.9, df=6, p<0.001, $\eta^2_p=0.267$) and participant (F=4.97, df=41, P<0.001, $\eta^2_p =0.608$) were also found. Besides all main effects being significant, also all two-way interactions were significant, including the interaction between hue and content (F=2.47, df=18, p<0.001, $\eta^2_p=0.057$), between hue and participant(F=1.27, df=246, p=0.010, $\eta^2_p=0.297$) and between content and participant(F=4.35, df=123, p<0.001, $\eta^2_p=0.420$). The effect of participant is the strongest followed by content, indicating that hue has a smaller but still significant effect on the results. Although the interaction between hue and content is significant, the effect size is so small that it is interaction effect is negligible.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>P</th>
<th>$\eta^2_p$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>38.1</td>
<td>3</td>
<td>&lt; 0.001</td>
<td>0.482</td>
</tr>
<tr>
<td>Hue</td>
<td>14.9</td>
<td>6</td>
<td>&lt; 0.001</td>
<td>0.267</td>
</tr>
<tr>
<td>Participant</td>
<td>4.97</td>
<td>41</td>
<td>&lt; 0.001</td>
<td>0.608</td>
</tr>
<tr>
<td>Content*participant</td>
<td>4.35</td>
<td>246</td>
<td>0.010</td>
<td>0.420</td>
</tr>
<tr>
<td>Hue* participant</td>
<td>1.27</td>
<td>123</td>
<td>&lt; 0.001</td>
<td>0.297</td>
</tr>
<tr>
<td>Content*hue</td>
<td>2.47</td>
<td>18</td>
<td>&lt; 0.001</td>
<td>0.057</td>
</tr>
</tbody>
</table>

Table 3.9: results from the ANOVA of the preferred gamut boundary (E1a). A * indicates an interaction.

The Tukey-HSD post-hoc test on content shows that the images rose and room are grouped (figure 3.10). The window and paint images are in a group of their own and are significantly different from the other content. This is however not surprising for the image window since its colors are most probably expected to be pastel. More surprising is that the rose and the room content are not significantly different. Although naturalness was expected to be different for both images it was not fully assessed this way.
Participants indicated that the *rose* could also be perceived as a piece of art, especially with the blue and cyan hue, explaining the non-significant difference. The image *paint* was rated, on average, the highest. This is probably due to its content, since paint can appear very colorful.

<table>
<thead>
<tr>
<th>window</th>
<th>rose</th>
<th>room</th>
<th>paint</th>
</tr>
</thead>
<tbody>
<tr>
<td>48.7</td>
<td>70.0</td>
<td>73.4</td>
<td>78.8</td>
</tr>
</tbody>
</table>

*Figure 3.10:* grouping results of the Tukey-HSD test for content from the ANOVA of the preferred chroma (E1a). The numbers below the image names indicate the average preferred chroma per image.

When the Tukey-HSD post-hoc test is ran over hue, three groups become visible. The hues G and M group together, also M, C and R form a separate group as well as B, Y and R_{EBU} (see figure 3.11). Most interesting to see is that the R and R_{EBU} are in different groups, while their actual CIE LCh hue angles are relatively similar. Because the effect of participant was found significant in the ANOVA test, a Tukey-HSD post-hoc test was also applied to participant in order to detect any potential (group of) outliers. An additional ANOVA was applied to the dataset, having hue, content and participant as a fixed factor. No outliers have been identified.

<table>
<thead>
<tr>
<th>G</th>
<th>M</th>
<th>C</th>
<th>R</th>
<th>B</th>
<th>R_{EBU}</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>59.8</td>
<td>62.4</td>
<td>65.5</td>
<td>65.5</td>
<td>73.0</td>
<td>73.6</td>
<td>74.4</td>
</tr>
</tbody>
</table>

*Figure 3.11:* grouping results of the Tukey-HSD test for hue from the ANOVA of the preferred chroma (E1a). The numbers below the hue labels indicate the average result per hue.

In order to describe the structure or inter-person variation on hue, the results of the hue-rotated images were analyzed using principle component analysis (PCA). The chroma preference of all hue variations per image per participant were used in the PCA. The seven per hue results for all image-participant pairs were taken as data samples in a seven dimensional space with the hues being the axes. The PCA finds new dimensions, also called components, in this seven dimensional space which have the largest variance of the projection of all data points on them. Components were selected for having an eigen-value higher than 1, resulting in only one component. This component explained 65% of the variance. The normalized factor loadings of this component, which are shown in table 3.12, describe the relation between hues and the new axis.

<table>
<thead>
<tr>
<th>normalized factor loadings</th>
</tr>
</thead>
<tbody>
<tr>
<td>PCA component 1</td>
</tr>
<tr>
<td>hue</td>
</tr>
<tr>
<td>-----</td>
</tr>
<tr>
<td>R</td>
</tr>
<tr>
<td>G</td>
</tr>
<tr>
<td>B</td>
</tr>
<tr>
<td>Y</td>
</tr>
<tr>
<td>C</td>
</tr>
<tr>
<td>M</td>
</tr>
<tr>
<td>R_{EBU}</td>
</tr>
</tbody>
</table>

*Table 3.12:* normalized factor loadings per hue for E1a resulting from a principle component analysis.
3.3.2 Preferred color gamut versus physical image properties

Results from the first experimental part, in which chroma preference is assessed, show a significant, but yet unexplained, effect for content. Since it is impossible to map, model and compare all possible content, an exploratory analysis on the physical properties of the images was conducted to identify the source of this effect on content.

Firstly a hierarchical cluster analysis grouped content based on preferred chroma. The created groups indicated that stimuli were firstly grouped by content and secondly by hue as can be seen in table 3.13. An unexplained difference was found when comparing the images field and flower. They both have the same Y (yellow) hue, but are grouped in considerably different groups. Both contrast and the size of the high chromatic region can be a possible cause for this difference. Derived from these two findings, a contrast filter and a high-chroma region filter were constructed and used to compare several measures to explain the results. First the contrast filter will be discussed followed by the high-chroma region filter.

<table>
<thead>
<tr>
<th>Group 1</th>
<th>Group 2</th>
<th>Group 3</th>
<th>Group 4</th>
<th>Group 5</th>
<th>Group 6</th>
<th>Group7</th>
</tr>
</thead>
<tbody>
<tr>
<td>Berry - R</td>
<td>Fish - C</td>
<td>Water drop - B</td>
<td>Room - G</td>
<td>Room - Y</td>
<td>Rose - Y</td>
<td>Window - R</td>
</tr>
<tr>
<td>Berry - R_{EBU}</td>
<td>Paint - Y</td>
<td>Water drop - C</td>
<td>Room - G</td>
<td>Room - Y</td>
<td>Rose - C</td>
<td>Window - G</td>
</tr>
<tr>
<td>Field - Y</td>
<td>Grasshopper - G</td>
<td>Room - G</td>
<td>Room - M</td>
<td>Room - G</td>
<td>Room - C</td>
<td>Window - B</td>
</tr>
<tr>
<td>Grass - G</td>
<td>Paint - R</td>
<td>Room - Y</td>
<td>Rose - G</td>
<td>Room - C</td>
<td>Room - R_{EBU}</td>
<td>Window - Y</td>
</tr>
<tr>
<td>Paint - G</td>
<td>Paint - B</td>
<td>Rose - B</td>
<td>Rose - M</td>
<td>Room - R_{EBU}</td>
<td>Flower - Y</td>
<td>Window - C</td>
</tr>
<tr>
<td>Paint - C</td>
<td>Paint - M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Window - M</td>
</tr>
<tr>
<td>Paint - R_{EBU}</td>
<td>Rose - R</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Window - R_{EBU}</td>
</tr>
<tr>
<td>Rose - R_{EBU}</td>
<td>Rose - M</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Leaf - M</td>
</tr>
</tbody>
</table>

Table 3.13: seven groups produced by hierarchical cluster analysis on the results per stimulus of first experimental part. Per group the names of the images are given with their respective color label.

To assess contrast an exact measure is needed with which we can evaluate the stimuli. Contrast can be defined in several ways and it can be applied to several properties of an image. Peli (1990) describes a predictor of local band-limited contrast, which can be applied to complex images. This predictor is defined as a band pass over low pass filter in a certain frequency range. According to Bex and Makous (2002) humans are most sensitive to contrast in complex images for spatial frequencies between 1 and 2 degrees visual acuity. The band-pass of the local band limited contrast filter is thus set to be in between 1 and 2 degrees visual acuity. Contrast is determined on the CIE Lab lightness level. A low-pass filter based on a Butterworth filter (Butterworth, 1930), which cuts off unwanted spatial frequencies to null, was used to create the band-pass filter. Only stimuli with a chroma multiplier of 1 were processed; an example of such processing is shown in figure 3.14.
Figure 3.14: the images flower (left) and field (right) with a chroma multiplier of 1 (top) and their corresponding band-pass limited local contrast image (bottom). Please note that the stimuli are intended to be displayed on the experiment display in its original size and that the band-pass image has been normalized for visualization purposes.

After obtaining the local band-limited contrast, which still entails a value for each pixel, several measures were applied to create one single measure out of the values of all pixels. These measures were the maximum, the absolute maximum, the minimum, the mean, the median, the standard deviation, and the mean of 10% of the highest value pixels. The value of each measure per image was correlated to the average preferred level of chroma per image, as indicated in experimental part. The results of these correlations are shown in table 3.15. This table shows a correlation of 0.58 for absolute maximum and a correlation of -0.59 for minimum. Although these correlations are not very high it could potentially indicate an effect over the results.

<table>
<thead>
<tr>
<th>Measure</th>
<th>E1a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>0.39</td>
</tr>
<tr>
<td>Absolute maximum</td>
<td>0.58</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.59</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.40</td>
</tr>
<tr>
<td>Median</td>
<td>0.13</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.49</td>
</tr>
<tr>
<td>Mean of 10% of the maximum data</td>
<td>0.39</td>
</tr>
</tbody>
</table>

Table 3.15: the correlation (r) of average preferred chroma against each measure of the local band-limited contrast over all images.

When viewing the actual correlation plot of minimum value of the local band-pass limited contrast filter versus preferred chroma (figure 3.16) we see that there indeed is a possibility that a meaningful correlation exists. However, the results from the third experiment will show that, when a wider variety of images is used, this correlation has little value.
As mentioned before, differences in chroma preference over content could be caused by the actual size of the high-chroma regions in the images. To test if this actually is the case, a high-chroma region filter was developed. Stimuli with a chroma multiplier of 1 were selected and blurred using a Gaussian blur with a width of 10 pixels. For the high-chroma filter only pixels containing a minimum 88% of the highest chroma value in the image were selected. Both the shape of the Gaussian blur as well as the high chroma threshold was found using an empirical search.

This empirical search was conducted by combining different Gaussian blurs, sizes varying from 0 to 50 pixels, with different chroma threshold from 0 to 100% of the maximum chroma. Both the selected chroma threshold as well as the size of the Gaussian blur had a tolerance of approximately 6% and 7 pixels respectively at which the highest correlating value did not deviate much. Results of the filter applied to the images field and flower can be seen in figure 3.17.

![Figure 3.16: the average preferred chroma (in % relative to the maximum amount of chroma the wide-gamut display can produce) against the minimum of the local band-limited contrast per image. The blue dots indicate the images while the red line indicates the correlation (r=0.59).](image)

![Figure 3.17: the images flower (left) and field (right) with a chroma multiplier of 1 (top) and their corresponding high-chroma region filtered image (bottom).](image)

After applying the high chroma filter, again several measures were calculated over the chroma of the remaining pixels. These measures are the maximum, the absolute maximum, the minimum, the mean, the median, the standard deviation, the mean of 10% of the highest value pixels and the size of the high chroma area in pixels. Table 3.18 shows the correlations for each measure.
<table>
<thead>
<tr>
<th>Measure</th>
<th>E1a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>-0.25</td>
</tr>
<tr>
<td>Absolute maximum</td>
<td>-0.25</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.09</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.20</td>
</tr>
<tr>
<td>Median</td>
<td>-0.19</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.11</td>
</tr>
<tr>
<td>Mean of 10% of the maximum data</td>
<td>-0.20</td>
</tr>
<tr>
<td>Size of high chroma area</td>
<td>-0.79</td>
</tr>
</tbody>
</table>

Table 3.18: the correlation factors (r) of average preferred chroma against each measurements of high-chroma region over all images.

As can be seen from table 3.18, only the size of the high chroma area correlates high for E1 (r=-0.79). Also the plot of the high chroma area against the result (figure 3.19) suggests that the correlation is genuine. This correlation indicates that when the size of the high chroma area decreases the chroma preference increases.

![Figure 3.19: the average preferred chroma (in % relative to the maximum amount of chroma the wide-gamut display can produce) against each measure of the high-chroma region per image. The blue dots indicate one image while the red line indicates the correlation (r=-0.79).](image)

These findings thus suggest that both the local band-pass limited contrast filter as well as the high-chroma region filter could possibly model the effect of content on the results. However, only a small amount of content has been used for this experiment. Results from the third experiment (see chapter 5), in which a much wider range of content is used, show that both filters do not explain the effect of image content on the results.

3.3.3 Maximum acceptable color gamut

The results of the second part of the experiment are shown in figure 3.20. The results are shown per content and per participant. The red lines indicate the median result, while the blue bars indicate 25% of the results above the median and 25% of the results below the median. At first sight the results show that for some of the images the color gamut for the wide-gamut display is not even large enough since the results are clipped. Maximum acceptable chroma is thus, on average, also higher than chroma preference. Furthermore the image window is again rated lower on average, like in the case of chroma preference. The differences per participant are smaller for maximum acceptable chroma than for chroma preference, but this is probably caused by the large amount of clipped data. This again indicates that the gamut for maximum acceptability is larger than the gamut of the wide-gamut display.
Figure 3.20: box plots of the results of the second experimental part. The left box plot shows the maximum acceptable chroma per image while the right one shows the same results per participant. The percentage of chroma is relative to the hue dependent maximum amount of chroma which can be produced by the wide-gamut display.

Where for preference at least 50% of the results of 5 of the 7 hues (G, Y, C, M, R_EBU) were above the EBU boundary, in the case of maximum acceptance even 75% of the results lie above the EBU boundary for those hues as can be seen in figure 3.21. Even for the two remaining hues, R and B, more than 50% of the stimuli were selected out of the EBU gamut. Furthermore, the results seem to be clipped very strongly for Y (yellow), which indicates that the wide color gamut is too small for that hue. Also the results for the B, C and M hues show that 50% of the participants selected a chroma level at the border of the display, indicating that the wide color gamut of the display is too small at least for half of the people. For the R and G hues, however, results indicate that the used color gamut is largely sufficient in size since most results lie within the gamut of the wide-gamut display.

Figure 3.21: box plot of the maximum acceptable chroma per hue. The single red lines indicate the EBU gamut boundary. The percentage of chroma is relative to the hue dependent maximum amount of chroma which can be produced by the wide-gamut display.
Figure 3.22 shows the lightness-chroma plot per median accepted stimulus sorted by hue. For a detailed explanation of this graph please consult subsection 3.3.1. From this figure it can be clearly seen that most data points are out of the EBU gamut at all given lightness levels, showing that not only the maximum chroma pixels, but also the 5% highest chroma pixels are out of the EBU gamut. Furthermore, a lot of data points are at the maximum gamut boundary of the display, indicating that a large majority of pixels are located at the maximum chroma boundary. For the green hue, the data points for the window and room images are relatively close, while the box plot shows that their median results are 76% and 100% respectively. Apparently the top 5% of pixels for the image room with G hue are not at the maximum gamut boundary as could be expected. This is probably due to the large gamut extension for the green hue, since the EBU gamut is almost twice as small as the gamut of the wide-gamut display.

In order to perform parametric analysis to the dataset (like ANOVA), the data should comply with four criteria (Field, 2005) as mentioned before. Again the data of the results is not normally distributed (see figure 3.23). However the normality in an ANOVA is required on the error, or the residuals of the data obtained by omitting the 3-way interaction of hue, image and participant. The residuals are normally distributed from which we can conclude that criterion 1 is satisfied.
Figure 3.23: histogram of the maximum accepted chroma (left) and a histogram of the residuals of the maximum accepted chroma (right). A normal curve is plotted on top of the histogram of the residuals (red line).

Since we have a large set of 1596 samples the Levene’s test is significant, as can be expected. Since the variance ratio of hue is 2 we can safely assume that the violation of the homogeneity of variance will have a negligible effect on the analysis (Field, 2005). Criteria 3 and 4 are already satisfied as mentioned before.

Since the 4 criteria are met, ANOVA was applied to the four images that were used with the 7 different hues. Preferred chroma was the dependent variable, while image content and hue were the independent variables. Participant was set as a random factor and both the main effects and two-way interactions were taken into account.

The ANOVA indicated a significant effect for content (F=34.5, df=3, p<0.001, \( \eta^2_p =0.457 \)), hue (F=20.0, df=6, p<0.001, \( \eta^2_p =0.328 \)) and participant (F=5.91, df=41, P<0.001, \( \eta^2_p =0.652 \)). The effect size of image content was bigger than hue, suggesting that image content was more important for gamut extension than the actual hue of the image. The results from the ANOVA are shown in table 3.24. Significant effects were also found for the two-way interaction between content and participant (F=1.21, df=246, p=0.025, \( \eta^2_p =0.289 \)) and for the two-way interaction between hue and participant (F=5.90,df=123, p<0.001, \( \eta^2_p =0.496 \)). The effect of the two-way interaction between content and hue (F=1.23,df=18,p=0.228, \( \eta^2_p =0.029 \)) was not significant. This implies that, when maximum chroma acceptance level is assessed, participants’ selected level of chroma does not depend on a specific color in a specific image.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>( \eta^2_p )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>34.5</td>
<td>3</td>
<td>&lt; 0.001</td>
<td>0.457</td>
</tr>
<tr>
<td>Hue</td>
<td>20.0</td>
<td>6</td>
<td>&lt; 0.001</td>
<td>0.328</td>
</tr>
<tr>
<td>Participant</td>
<td>5.91</td>
<td>41</td>
<td>&lt; 0.001</td>
<td>0.652</td>
</tr>
<tr>
<td>Content*participant</td>
<td>1.21</td>
<td>246</td>
<td>0.025</td>
<td>0.289</td>
</tr>
<tr>
<td>Hue* participant</td>
<td>5.90</td>
<td>123</td>
<td>&lt; 0.001</td>
<td>0.496</td>
</tr>
<tr>
<td>Content*hue</td>
<td>1.23</td>
<td>18</td>
<td>0.228</td>
<td>0.029</td>
</tr>
</tbody>
</table>

Table 3.24: results from the ANOVA of the maximum acceptable chroma (E1b). A * indicates an interaction.

The Tukey-HSD post-hoc analysis on content shows that only images room and paint are grouped (see figure 3.25). This thus indicates that the images window and rose are assessed significantly different from the other images. When maximum acceptance is assessed, the image room thus groups with the image paint instead of with the image rose, as was the case where preference was assessed. Apparently the image room can be stretched relatively more than the image rose.
Because participants had a significant effect on the results, a Tukey-HSD post-hoc test was applied to participant in order to detect any potential (group of) outliers. No outliers have been identified. When the Tukey-HSD post-hoc test ran over hue, three groups become visible (see figure 3.26). The hues R and G group together, also M, R$_{EBU}$, C and B form a separate group as well as Y. Most interesting to see is that again the R and R$_{EBU}$ are in different groups, while their actual CIE LCh hue angles are relatively similar.

![Figure 3.26: grouping results of the Tukey-HSD test for hue from the ANOVA of the maximum acceptable chroma (E1b). The numbers below the hue labels indicate the average maximum accepted chroma per hue.](image)

Similar to the results of the preference data, the acceptance results were also analyzed using PCA. Components were selected for having an eigen-value higher than 1, resulting in only the first component being extracted. This first component explained 71% of the variance. The normalized factor loadings of this component, which are shown in table 3.27, describe the relation between hues.

<table>
<thead>
<tr>
<th>hue</th>
<th>E1b</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.84</td>
</tr>
<tr>
<td>G</td>
<td>1.00</td>
</tr>
<tr>
<td>B</td>
<td>0.83</td>
</tr>
<tr>
<td>Y</td>
<td>0.62</td>
</tr>
<tr>
<td>C</td>
<td>0.85</td>
</tr>
<tr>
<td>M</td>
<td>0.97</td>
</tr>
<tr>
<td>R$_{EBU}$</td>
<td>0.77</td>
</tr>
</tbody>
</table>

Table 3.27: normalized factor loadings per hue for E1b resulting from a principle component analysis.
3.4 Discussion

The results show that the EBU color gamut is clearly too small when assessing user preference or acceptance. Especially when maximum acceptance is assessed, more than 75% of the results per hue are outside the EBU gamut for five of the seven hues. For the remaining two hues (R and B), over 50% of the results are out of the EBU gamut. Also when chroma preference is assessed over 50% of the results per hue are out of the EBU gamut for five of the seven hues. Figure 3.28 shows these results compared to the EBU gamut for both experimental parts. The use of wide-gamut displays is thus preferred. Hypothesis 1, which states that the natural boundary for the extension that people prefer (“preferred” gamut) is larger than the standard TV gamut, can thus not be rejected. Furthermore, the color gamut of the wide-gamut display seems to fall short for several (B, Y, C, and M) hues when maximum acceptance is assessed. For those hues at least 50% of the results per hue are at the maximum amount of chroma the display can produce (figure 3.28). The maximum acceptable color gamut boundary is thus probably wider for some hues than measured during this experiment. For the R and G hues at least 75% of the results per hue are within the wide color gamut. This means that hypothesis 2 (which states that the boundaries at which people don’t accept the extension (“acceptable” gamut) is smaller than the available wide gamuts) cannot be rejected for the R and G hues.

![Figure 3.28](image)

Figure 3.28: results for both parts of the experiment per hue for all images. E1a refers to the chroma preference experimental part and E1b refers to the maximum acceptable chroma experimental part. The single red lines indicate the EBU gamut boundary. The percentage of chroma is relative to the hue dependent maximum amount of chroma which can be produced by the wide-gamut display.

The ANOVA results in both experimental parts showed a significant effect of hue. Furthermore, as seen by the significant difference between the two red hues in the Tukey-HSD analysis, a small change in hue can have a large impact on the results between those hues. These results thus do not enable us to reject hypothesis 4, which states that the preferred gamut boundaries have an irregular shape. Furthermore these results indicate that user-preference can change dramatically when unexplored hues are assessed. The ANOVA also indicate that image content is even more important than hue. It is thus more important for gamut extension algorithms to account for the effect for content than for the effect of hue. This can furthermore be confirmed by the fact that participants assessed the image rose as being artificial, which could probably have influenced the results. Possibly the color memory for the image window made participants rate this image much lower for all hues. Using these findings hypothesis 3 (the preferred gamut boundaries are content dependent) cannot be rejected.
The preferred and maximum acceptable chroma fluctuated much between participants as could be seen in both box plots (see figure 3.5 and 3.20) as well as in the outcome of the ANOVA. Hypothesis 5 (the preferred gamut boundaries are person dependent) thus cannot be rejected. To further investigate the differences between hues a PCA was applied to the results of the hue-rotated images for both parts of the experiment. The PCA showed that approximately 65% for preferred chroma and 70% for maximum acceptable chroma of the variance of the results can be explained for both experimental parts by a largely homogenous preference over hue.

We can thus conclude that content, hue and participant have a significant effect on the preference as well as the maximum acceptable chroma. According to the results of the ANOVA, participants also seem to have a preference for combinations of hue and content in both parts of the experiment. A gamut extension algorithm should thus take the effect of hue, content and participant into account when expanding the gamut. To be able to extend the gamut according to the yet unexplained effect of content, two filters which measured contrast and high-chroma regions were created. The output of these filters was quantified using several measures. For the contrast filter the minimum and absolute maximum measures showed possible candidates which could explain the effect of content. For the high-chroma region filter the size of the filtered area seemed to influence the results.

Finally the results are largely in line with the findings of Laird & Heynderickx (2008), while this experiment uses a much large set of image, which were also more complex. Laird & Heynderickx also concluded that the typical preferred gamut for natural content is outside the EBU gamut, though not to the full extent of the wider-gamut boundary. This experiment thus shows that the preferred gamut boundary is largely out of the EBU gamut. Furthermore, Laird & Heynderickx showed that both image content, hue and the participant influenced the preferred chroma. Each of these findings was confirmed by this experiment. They assessed preference for the same image window as used in this study. The results for the image window are on average lower than other content which is probably due to the expectancy of memory color.
4 Verifying tuning methodology

4.1 Introduction
To test if the tuning methodology used in the first experiment biases the outcomes, a supplementary experiment (E2) was conducted. Outcomes from the first experiment could be biased by so-called chromatic adaptation. According to Fairchil (2005), chromatic adaptation refers to “the human visual system’s capability to adjust to widely varying colors of illumination in order to approximately preserve the appearance of object colors” (p. 146). Thus when being exposed to a stimulus the human visual system adapts to the stimulus according to the expected appearance. Since during a tuning experiment participants view one stimulus for an extended amount of time, it is possible that they adapt to this stimulus and adjust it accordingly. To test whether that is the case, the experiment was repeated with a different methodology, in which the various images were shown interweaved. Thus, a paired comparison experiment was performed for comparison.

4.2 Method

4.2.1 Design
A paired comparison methodology in combination with a staircase methodology was performed to measure chroma preference using a within subject design. This methodology was chosen to minimize the continuous exposure time of each image in order to reduce chromatic adaptation.

4.2.2 Participants
A subset of 10 participants that also performed the first experiment using the tuning methodology was selected. They were selected because they had, on average, the smallest difference from the mean preference per image.

4.2.3 Equipment and setup
The equipment is the same as was used the first experiment (see section 3.2.3). After participants selected one of the two stimuli, their preference was acquired by presenting the same content hue combination of the stimuli multiple times using a staircase method. After each choice a stimulus with different hue and/or content was presented at random in order to change content as quickly as possible and thus to reduce chromatic adaptation. Both stimuli, which were shown simultaneously on the screen, had the same hue and image content; however, one of the stimuli had a lower chroma multiplier than the other. The position of the low and the high chroma stimuli, which was left or right, was randomized. Participants could choose between the two stimuli in front of them by means of the left or right arrow buttons on a keyboard, after which the next stimulus appeared. Participants could choose without any time constraints. One trial staircase was presented to test whether participants understood their task.

A pilot experiment, which was conducted prior to this experiment, showed that when the first stimuli had a low chroma multiplier, the results were biased. This pilot experiment used a staircase methodology which accidentally found a local maximum at approximately 0 chroma. Some participants thus seem to have a preference for low chromatic images. Because the overall chroma preference for nearly gray-scaled images was lower than the high chroma images for the tuning methodology of the first experiment, the chroma multiplier was set to 0.75 to avoid the local maximum at 0 chroma.

4.2.4 Stimuli
A subset of stimuli was selected from the first experiment. It includes 20 different image-hue combinations, each having 81 levels of chroma. The selected content consisted of the images paint (in R, G, B, Y, C, M and R_EBU), rose (in R, M and R_EBU), room (in G, B, Y and C), berry (in R and R_EBU), water drop (in B), fish (in C), grasshopper (in G) and the image field (in Y).
4.2.5 Measurement
For this experiment two independent variables were used; namely image content and hue. There were 8 different types of image content and 7 different hues. In the first part of the experiment the dependent variable was preferred chroma, while in the second part of the experiment maximum chroma acceptance was the dependent variable.

4.2.6 Procedure
Participants were asked to read the instructions (added here in appendix D). Additional oral explanations were provided by the experimenter. Their task was to “select the image which you prefer the most”. After being instructed, one trial paired comparison was presented to see whether participants understood their task. After the experiment, which on average lasted about 13.3 minutes (SD=5.1), participants were given the opportunity to provide oral feedback to the experimenter.

4.3 Results
The preferred chroma obtained with this new methodology was compared to the preferred chroma obtained with the tuning methodology using a paired sample t-test. On average, no significant difference between the means of both experiments (t(199)=0.319, p=0.750) was found. The average preferred chroma, M=82.6 (SD=20.8) when using a tuning method and M=82.1 (SD=20.1) when using the interweaved staircase method, indeed are remarkably stable between both experiments (see figure 4.1).

![Box plot of the preferred chroma for 10 participants who participated in both the first and the second experiment. These results show a resemblance between the methodologies used in both experiments.](image)

4.4 Discussion
Because the results of this interweaved staircase experiment are not significantly different from the tuning experiment, it is probable that the preferred chroma obtained with the tuning methodology was not biased by chromatic adaptation. The hypothesis that the results obtained with the tuning methodology were not influenced by chromatic adaptation thus cannot be rejected. This experiment also shows that the interweaved staircase methodology is more time consuming (13.3 minutes on average for 20 images) than the tuning methodology, which only takes 8.1 minutes on average for 38 images. Furthermore, a pilot study prior to this experiment showed that a local maximum was found at approximately 0 chroma. There thus seems to be a preference for gray-scaled images besides their colorful versions. This preference was not found during the tuning experiment, indicating that the preference for gray-scaled images is lower than their colorful versions.
5 Experiment on the preferred color gamut for single- and multiple-hue images

5.1 Introduction
From the first tuning experiments, information was gained about chroma preference (E1a) and maximum chroma acceptance (E1b). These results showed that participants preferred and accepted more chroma than defined in the EBU standard for most hues. As concluded before, there is thus a need for wide-gamut displays, and correspondingly, a gamut extension algorithm. Because participants provided chroma preference information during the first experiment, this information can be used as a basis for a gamut extension algorithm. For this third experiment such a gamut extension algorithm was developed as will be explained further on. The primary goal of this experiment, from here on referred to as E3, is to find out how this gamut extension algorithm performs and to what extent it can be applied to more complex images. The experiment consists of three parts. The first part, E3a, is a tuning experiment similar to E1a. The second and third part (E3b and E3c) consist of an image painting task as described in section 5.2.3.

5.2 Method

5.2.1 Design
Two different experimental designs were used within this experiment, each having its own purpose. First, a tuning methodology was used to assess chroma preference. It has already been demonstrated (see chapter 4) that the tuning methodology used yields highly comparable results to the interweaved staircase methodology. Because the tuning methodology is faster, this methodology was chosen again. The second methodology used in this experiment provides a qualitative measure of the performance of the gamut extension algorithm. Participants were requested to indicate which part of their tuned stimulus had an inadequate amount of chroma by means of painting the particular area in the stimulus.

5.2.2 Participants
In total, 43 persons (27 male and 16 female) from the Philips High Tech Campus Eindhoven participated in this experiment. The average age was 30 (SD=11), varying between 20 and 64. All participants were tested negative for color deficiency according to the Ishihara (1999) test. Twelve participants that already performed the first experiment also participated in this experiment.

5.2.3 Equipment and setup
The equipment is the same as was used in the previous experiments (see section 3.2). This experiment was divided in three parts which were assessed consecutively in one session. In the first part (E3a), participants tuned the chroma of each image to their preferred level. Again, the images were shown one at a time. Participants could change the chroma of the image in front of them by means of pressing the arrow keys on a keyboard. By pressing the enter button the next image appeared. In the second and third part of the experiment (E3b and E3c), a fixed subset of 11 images was shown in random order at the preferred level of chroma for each participant (as obtained in part E3a). Also here, the images were presented one at a time on a gray background which corresponded to 50% lightness of the device. The participants were able to indicate which part of the given image had too much or too little chroma by means of a digital painting tool. The painting was thus performed on the display using the mouse cursor. Painted areas were indicated using a semi-transparent layer on top of the image. Figure 5.8 illustrates the result of using the paint-tool on an image. For both painting parts participants were instructed that painting was not compulsory. In case all areas in the image had the right chroma level according to their opinion, they could leave the image as it was. There was no time constraint on any of the three parts of this experiment.
Figure 5.8: illustration of the result of using the paint-tool on an image. The light-red areas indicate the regions in the image that were painted. The mouse cursor was represented by the circle.

5.2.4 Stimuli
The images for the first part of the experiment (E3a) consisted of 51 multiple-hue images (partly shown in figures 5.1 and 5.3) and a set of 30 single-hue images. The single-hue images were created using 3 original images (see figure 5.2), each rotated over 10 hues (see table 5.4). Additionally a fixed subset of 5 images was presented twice to measure reliability.

Figure 5.1: five of the 51 multiple-hue images. From left to right: car with graffiti, parasol, heart with marbles, temple, and beach toy.

Figure 5.2: the 3 original images which were rotated over 10 hues to create the 30 hue rotated images. From left to right: multiple bottles, bottle, and paint.

For the painting tasks in the experimental parts E3b and E3c, a subset of 11 images (figure 5.3) was manually selected. Images were selected so that sky content and a wide range of hues were included.

Figure 5.3: the images used for the painting task in the experimental parts E3b and E3c. They were shown with a chroma multiplier obtained as being the preferred chroma multiplier in the first pair of this experiment. As shown from left to right: textile tubes, xylophone, cupcakes, sky and flower field, glass unions, yellow flower, back of boats, 7 paint bottles, sailboats, sky and balloons and jelly cupcakes.
The ten hues which were used to create the hue-rotated images are displayed in table 5.4, together with their corresponding CIE LCh hue angle and the hue label as used throughout this report. The first seven hues mentioned in the table are identical to the ones used in the first experiment. As mentioned before, the R, G, and B hues correspond to the primaries of the wide-gamut display and the Y, C, and M hues match its secondaries. \( R_{EBU} \) and \( G_{EBU} \) correspond to the red and green primary as defined in EBU, respectively. Earlier results showed that there was a significant difference in preferred chroma for \( R \) and \( R_{EBU} \), which may be due to color memory for \( R_{EBU} \). Since this effect may also happen near the other primaries, \( G_{EBU} \) was added. No additional blue hue was introduced because EBU blue and the B primary of the display have almost the same hue angle. The BC and RM hues were added in the largest hue gaps, between B and C (117 degrees), and R and M (64 degrees). These additional hues provide information about how the stretching algorithm performs in the previously unexamined hue regions between the B and C hues and the R and M hues.

<table>
<thead>
<tr>
<th>Hue name</th>
<th>hue label</th>
<th>CIE LCh hue angle (white point of CIE D90)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Red</td>
<td>( R )</td>
<td>39</td>
</tr>
<tr>
<td>Green</td>
<td>( G )</td>
<td>149</td>
</tr>
<tr>
<td>Blue</td>
<td>( B )</td>
<td>308</td>
</tr>
<tr>
<td>Yellow</td>
<td>( Y )</td>
<td>104</td>
</tr>
<tr>
<td>Cyan</td>
<td>( C )</td>
<td>191</td>
</tr>
<tr>
<td>Magenta</td>
<td>( M )</td>
<td>335</td>
</tr>
<tr>
<td>EBU red</td>
<td>( R_{EBU} )</td>
<td>44</td>
</tr>
<tr>
<td>Blue-cyan</td>
<td>( BC )</td>
<td>240</td>
</tr>
<tr>
<td>Red-magenta</td>
<td>( RM )</td>
<td>10</td>
</tr>
<tr>
<td>EBU green</td>
<td>( G_{EBU} )</td>
<td>136</td>
</tr>
</tbody>
</table>

Table 5.4: the 10 hues used to create the hue rotated image. The hue label is used throughout this report. Note that the top 7 hues are identical to the ones used in the first experiment (see table 3.2).

Each of the 3 original images shown in figure 5.2 was rotated over the 10 hues, resulting in 30 hue-rotated images. The rotation process was identical to the rotation process as previously described (see section 3.3). In short, all the pixels in the original image were rotated by a fixed amount in hue while lightness levels were corrected for each hue as well.

All 81 images, which include all hue rotated and multiple-hue images, were processed in order to create several stimuli per image with different levels of chroma. Because only single-hue images were used during the first experiment, interpretation of these hues was arbitrary and same drive signal was used. This third experiment uses more complex multiple-hue images, which ideally should be interpreted as being displayed on an average screen. sRGB (Stokes et al., 1996) is a standard which can be used to interpret images as intended to be displayed on an average computer display. The sRGB standard is defined with a white point of CIE D65, a set of standard primaries and has an encoding curve equivalent to a gamma of 2.2. However, using the sRGB primaries introduced artifacts which were caused by the differences in chromaticities with the primaries of the display. Therefore the display primaries were used instead of sRGB like in the first experiment, while the gamma as defined in sRGB was used. For this third experiment, the display device did not change; therefore the device model of the display as described in chapter 3 has been used. A hybrid forward model was created using the encoding curve of sRGB and a color primary matrix computed with the wide-gamut display primaries and white point of D90. This was nearly the same as using same-drive-signal, as was done in the first experiment, but the tone scale is more accurately interpreted with the sRGB encoding curve. Using this model, the sRGB values were converted to CIE 1931 XYZ tristimulus values (Wyszecki & Stiles, 1982).
The white point of CIE D90 was used to create CIE Lab LCh values. After acquiring the LCh values, chroma, C*, was stretched using the chroma multiplier per hue (CMh) as will be described in the following section. Using the LCh values, the L*a*b* and XYZ values were calculated using the D90 white point. Finally, these XYZ values were transformed into RGB values for the wide-gamut display using the inverse model of the display in combination with interpolation and local search. Chroma clipping was detected at the transformation of XYZ to RGB values. If an invalid RGB value was obtained, the LCh chroma value was decreased until a valid RGB value could be produced. Figure 5.5 shows the transformation process visually.

![Diagram](image)

*Figure 5.5: the processing steps for each pixel as it is being transformed from a sRGB value to a RGB value suitable for the experiment.*

As described before, the results described on hue dependent chroma preference are used here in the gamut extension algorithm. To use these data, several problems should be accounted for. Firstly, the data gained from the first experiment are only representative for the seven hues which were measured. No detailed chroma preference information is thus available for the intermediate hues. Secondly, only single-hue images were used, while a real gamut extension algorithm should be able to handle images with many hues at the same time. Finally preferred chroma differed between participants, meaning that the gamut extension algorithm should be customized for different participants. A gamut extension algorithm which tries to account for all these problems is thus needed.

During the experiment participants should increase or decrease just one parameter (the overall chroma multiplier, or OCM). By changing this OCM, the chroma at all hues within the image needs to be scaled. Instead of scaling the chroma of all hues equally, the information on the hue dependent chroma preference as obtained from the first experiment was used. An OCM of 1 was defined as the preferred level of chroma according to the preferred chroma per hue over all hue-rotated images from the first experiment. Each of the seven hues used in of the first experiment had a different chroma preference (CP) which is given in table 5.6. CP is expressed, for each hue, as the percentage of chroma the wide-gamut display can maximum produce. Since the OCM is based on the average preferred level of chroma per hue from the first experiment, it is expected that participants select on average an OCM of 1.
<table>
<thead>
<tr>
<th>Color</th>
<th>CP</th>
<th>PCA₂n</th>
</tr>
</thead>
<tbody>
<tr>
<td>R</td>
<td>0.66</td>
<td>0.68</td>
</tr>
<tr>
<td>G</td>
<td>0.63</td>
<td>0.97</td>
</tr>
<tr>
<td>B</td>
<td>0.73</td>
<td>0.85</td>
</tr>
<tr>
<td>Y</td>
<td>0.78</td>
<td>0.81</td>
</tr>
<tr>
<td>C</td>
<td>0.67</td>
<td>1.00</td>
</tr>
<tr>
<td>M</td>
<td>0.62</td>
<td>0.96</td>
</tr>
</tbody>
</table>

Table 5.6: details of the hues from the preferred chroma of the first experiment and their CP (the average preferred level of chroma from E1a relative to the gamut boundary of the wide-gamut display) and PCA₂ (the normalized PCA factor loadings per hue, taken from table 3.12).

As seen in the first experiment, preferred chroma was person dependent. Apart from differences in the mean preferred chroma, there was also more variance for the C hue than for, for example, the R hue. This indicates that participants were more sensitive to changes in the R hue than in the C hue. To account for this difference in sensitivity, the normalized PCA factor loadings (PCA₂) as shown in table 5.6) were used as a measure for scaling the chroma multiplier at each hue. Using the PCA₂ as a slope together with both the OCM and CP we can define the chroma multiplier per hue (CMh) as given in equation 3.

\[
CMh = CP \times ((OCM - 1) \times PCAn + 1)
\]

(3)

After being computed separately for each of the hues of the first experiment, the CMh was approximated for the intermediate hues using cubic interpolation as shown in figure 5.7.

Figure 5.7: values of CMh for all hues. The x-axis for the left plot and the circular unit for the right plot represent LCh hue angle. The y-axis for the left plot and the circular dotted lines for the right plot indicate CMh. The red line corresponds to CMh for an OCM of 1 and the adjacent lines indicate CMh in steps of 0.4 OCM.

After conducting exploratory research, CMh for each hue was calculated by changing OCM over 4 ranges; starting from 0 to 2 in steps of 0.1, continuing from 2.2 to 6 in steps of 0.2, followed by from 6.4 to 12 in steps of 0.4, and finally OCM was varied from 12.8 to 20 in steps of 0.8. These four ranges added up to 66 levels of OCM, and thus created 66 chroma variations per image for participants to choose from.
5.2.5 Measurement
For the first part of the experiment two independent variables were used; namely image content and hue. There were 54 different types of image content and 10 different hues. In this part of the experiment the dependent variable was preferred chroma. In the second and third part of the experiment the independent variable was image content, which had 11 different kinds of content. The dependent variable for the second experimental part was the image region which is too colorful, while the dependent variable for the third experimental part was areas which are not colorful enough.

5.2.6 Procedure
Participants were first tested for color deficiency after which they were presented with a written instruction (as given in appendix E). Additional oral explanations were provided by the experimenter. For the first part of the experiment (E3a), participants were instructed to “adjust the image to your preferred level of colorfulness”. After having finished the tuning, participants were asked to “paint the area which you perceive as being too colorful”. Subsequently participants were asked to “paint the area which you perceive as being not colorful enough” (i.e. experiment E3c). Trials of all three parts of the experiment were run prior to the experiment. After the three experimental parts, which on average lasted about 14.6 minutes (SD=6.6), participants were given the opportunity to provide oral feedback to the experimenter.

5.3 Results
In total 86 data points on preferred chroma level per image were gathered per participant. The reliability measurements, however, were removed from the analysis, leaving 81 data points per participant for analysis. Furthermore, 11 painted images, which indicating areas with too high chroma and 11 images, indicating areas with a too low chroma, were produced per participant.

The data from the tuning experiment (E3a) show which OCM (overall chroma value) was selected. The OCM, however, is not a representative measure for the actual amount of chroma stretching in an image. As described in chapter 5.2.4, chroma values outside the gamut of the wide-gamut display were clipped to their nearest realizable chroma while preserving hue and lightness. Because pixels were clipped in chroma -for an OCM of at least 1.25- the number of pixels that was actually stretched in chroma decreased as the OCM increased. In practice this means that images gradually changes less when more chroma was requested during the tuning experiment, while their OCM kept increasing. On top of that each image had a unique set of pixels which mostly clipped at different OCM values depending on their LCh values before processing. Thus each image clipped at a different OCM, making it difficult to use OCM values for the analysis.

In order to have a more representative measure, a new multiplier was defined. The corrected chroma multiplier (CCM) was defined as the averaged amount of OCM over all pixels. If a pixel was clipped, its individual OCM was set to the OCM at the point of clipping. Thus, if an image had many low chroma pixels, these pixels were clipped at a higher OCM, increasing the average OCM of the image. However, when using this technique on a white or black background, very low chroma pixels were clipped at a very high OCM or even not clipped at all (if chroma=0). These pixels affected the CCM significantly, while they had no or very little effect on the amount of visible change. In order to counteract these very low chroma pixels, pixels having a chroma lower than 12.5, before applying equation (3) were omitted in determining the CCM. This low chroma boundary was experimentally determined by examining all images, in which low chroma pixels were removed, for different low chroma boundaries. Figure 5.9 shows the OCM to CCM relationship for all 81 images. This figure shows that some of the images clipped (their slopes became 0) almost instantly around an OCM of 1.5, while other images clipped at a much higher OCM. Interesting to notice is the upper red line which represents the image beach toy (see figure 5.1). The high clipping point for this image was caused by its large medium-chroma background, which can be stretched to a large extent.
Figure 5.9: the relationship between OCM (overall chroma multiplier) and CCM (corrected chroma multiplier) for pixels with a chroma above 12.5. Each line represents one image while the dotted line represents an imaginary image of which no pixels needed to be clipped.

As mentioned before, five stimuli were repeated to check for reliability. The box plot of the repeated stimuli (see figure 5.10) shows that their difference with the original preferred CCM is relatively small. The mean difference in CCM for the image sky & balloons is 0.03, for the image green bottles with triangles 0.06, for the image sailboats 0.07, for the image red flower with butterfly 0.005, and for the image multiple bottles 0.14. To test if the results are reliable, a paired sample t-test was conducted to compare the results of the first measurement to the results from the second repeated measurement. The results show no significant difference for the images sky & balloons ($t(21)=-0.75$, $p = 0.464$), green bottles with triangles ($t(21)=0.86$, $p = 0.401$), sailboats ($t(21)=1.69$, $p = 0.106$), red flower with butterfly ($t(21)=0.10$, $p = 0.925$) and multiple bottles with BC hue ($t(21)=-1.64$, $p = 0.115$). These results suggest that the repeated stimuli are assessed similarly and that our measurement method thus again seems reliable.

Figure 5.10: box plot of the results from the repeated stimuli. The vertical axis represents the result (CCM).
5.3.1 Preferred color gamut for single- and multiple-hue images

The box plot of preferred chroma level (expressed in CCM) per image is shown in figure 5.11. It illustrates that both the spread and the median in preferred chroma largely differ between images. Most remarkable is the fact that the majority of images have a median CCM larger than 1, which indicates that participants preferred a higher chroma on average than estimated from the first experiment (E1a). The median results for the hue rotated versions of the image *multiple bottles* are visibly lower than for most of the other images. The median results for the hue rotated versions of the image *paint* are most stable among images, which is not surprising since the chroma extension algorithm was partly based on the preferred chroma for these images as obtained from the first experiment (E1a). The various versions of the image *multiple bottles* have a rather low preferred chroma, while the various versions of the image *bottle* have a rather high preferred chroma. The content of the image thus seems to be a more important attribute than its hue.

![Box plot of preferred chroma per image](image.png)

*Figure 5.11: box plot of the preferred chroma (in terms of CCM) per image.*

Figure 5.12 shows the box plot of the preferred chroma (in CCM) per hue. The median preferred chroma is relatively constant over all hues. Apparently, the low preferred chroma for the images *multiple bottles* and *bottle* compensate each other similarly for all hues. The multiple-hue images (indicated as “multiple” in figure 5.12) were on average a slightly higher median than most individual hues. As can be seen in figure 5.11 this is mostly due to the low preferred chroma for the images *multiple bottles* which reduces the averaged preferred chroma for each individual hue. The Y (yellow) and BC (blue-cyan) hues are stretched more in chroma than the other hues, which was already found for the image *paint* in the first experiment. The gamut extension algorithm thus should stretch more in Y and BC.
Figure 5.12: box plot of the preferred chroma (in terms of CCM) per hue over all hue rotated images. The column “multiple” indicates the box for all multiple hue images.

The preferred chroma for the hue-rotated images relative to the EBU gamut boundary and the wide-gamut display gamut boundary is shown in figure 5.13. This figure shows that the participants again preferred the chroma for some images out of the EBU gamut boundary. When comparing the results for the image paint (triangle shape) to the results for the preferred chroma obtained in the first experiment for the same image (see figure 3.7) they are very similar. The results for the image paint also seem to clip at both the Y and C hue, suggesting that more chroma than the display can provide is preferred. Furthermore the results for the image bottle (square shape) are biased to some extent by the vast amount of low chroma pixels at a high lightness level in that image (for the original image see figure 5.2). However, for the low lightness levels the image bottle is for all hues close to the wide-gamut display gamut boundary. Users indeed indicated that the content of the image bottle could be extended relatively far. For the image multiple bottles (circle) the preferred chroma clearly within the EBU gamut for almost all hues. Participants often indicated that this particular content was most artistic or surreal, and therefore was rated so low on preferred chroma. Another explanation can be that the preferred chroma for these images is so low since the size of the high chroma area is unusually large, which can create a surrealistic appearance.
Figure 5.13: lightness-chroma plots of the first part of the third experiment (E3a) per hue including the hue rotated images. The EBU gamut (green/inner line) and the wide-gamut display gamut (blue/outer line) are shown for reference. The circles represent data for the image multiple bottles, the squares for the image bottle and the triangles for the image paint. Data points represent the average of 5% of the highest chroma pixels at the given lightness level, using a 2% lightness offset, for all pixels of the median preferred stimulus. A more detailed explanation can be read in subsection 3.3.1.

In order to test if hue, content or participants significantly influence the preferred chroma, an ANOVA was performed. Since 51 of the 81 images used for tuning the preferred chroma contained multiple hues, it was not possible to compare these images with the remaining 30 images in one analysis. Therefore, the analysis is divided into an ANOVA over the multiple-hue images (ANOVA$_m$) and an ANOVA over the hue-rotated images (ANOVA$_r$) separately. A repeated measures ANOVA was not used because it does not provide insight on the effect of participant.

As described earlier (see chapter 3), in order to perform an ANOVA four criteria have to be met (Field, A., 2005). The criteria are (1) the data should be normally distributed, (2) the variances throughout the data should be homogeneous, (3) interval data should be used and (4) data from different participants should be independent. To satisfy criterion (1) the residuals for ANOVA$_m$ and ANOVA$_r$ were computed by omitting the interaction between hue, content and participant for ANOVA$_m$ and omitting the interaction between content and participant for ANOVA$_r$. As can be seen in figure 5.14, the residuals follow a normal distribution for both the ANOVA$_m$ and ANOVA$_r$. 
Figure 5.14: histograms of the residue of the ANOVA\textsubscript{m} (left) and ANOVA\textsubscript{r} (right). As can be seen both distributions are normal. The red line represents the fit of a normal distribution through the histogram.

For both ANOVA\textsubscript{m} and ANOVA\textsubscript{r}, the variance was not homogeneous between participants as found with Levene’s test as can be expected for a large sample set. The variance between participants was even larger than a factor 2 for both the hue rotated images and the multiple hue images. Thus, criterion (2) was not fulfilled. To check the effect of the failure of criteria (2), the ANOVA was repeated after normalizing the results over participants, consequently yielding a variance ratio of 1. The ANOVA results with and without normalizing the data are discussed below.

Criterion (3) was satisfied since CCM, as described in section 5.3, was representative for a homogenous amount of chroma change in the images. Finally criterion (4) was also satisfied since participants participated independently.

ANOVA\textsubscript{r} was run on the raw unnormalized data of the 30 hue rotated images, which were a combination of 10 hues and 3 types of content. For the analysis, CCM was used as dependent factor, hue and content as fixed factors, and participant as a random factor. All main effects and two way interactions were taken into account. From the ANOVA\textsubscript{r} results (see table 5.15) we can see that there is a significant effect of content (\(F=72.4, \text{df}=2, p<0.001, \eta^2_p =0.633\)), hue (\(F=3.29, \text{df}=9, p<0.001, \eta^2_p =0.073\)) and participant (\(F=4.62, \text{df}=42, p<0.001, \eta^2_p =0.697\)). The two-way interactions between hue and content (\(F=2.30, \text{df}=18, p=0.002, \eta^2_p =0.052\)) and between content and participant (\(F=6.61, \text{df}=84, p<0.001, \eta^2_p =0.424\)) are significant as well. The two-way interaction between hue and participant is not significant (\(F=1.05, \text{df}=378, p=0.299, \eta^2_p =0.344\)).

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>P</th>
<th>\eta^2_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>72.4</td>
<td>2</td>
<td>&lt;0.001</td>
<td>0.633</td>
</tr>
<tr>
<td>Hue</td>
<td>3.29</td>
<td>9</td>
<td>&lt;0.001</td>
<td>0.073</td>
</tr>
<tr>
<td>Participant</td>
<td>4.62</td>
<td>42</td>
<td>&lt;0.001</td>
<td>0.697</td>
</tr>
<tr>
<td>Content*participant</td>
<td>6.61</td>
<td>84</td>
<td>&lt;0.001</td>
<td>0.424</td>
</tr>
<tr>
<td>Hue*participant</td>
<td>1.05</td>
<td>378</td>
<td>0.299</td>
<td>0.344</td>
</tr>
<tr>
<td>Content*hue</td>
<td>2.30</td>
<td>18</td>
<td>0.002</td>
<td>0.052</td>
</tr>
</tbody>
</table>

Table 5.15: results of the ANOVA\textsubscript{r} on the preferred chroma over the hue rotated images (E3a). A * indicates an interaction.

The findings of the ANOVA performed so far on the raw unnormalized data are validated by an ANOVA on the normalized data of the hue rotated images, which accounts for the lack of homogeneity in variance between participants. This second ANOVA included the same dependent and independent variables and interactions as the first one. The results (table 5.16) show that the effect of hue is again significant with a small effect size (\(F=3.80, \text{df}=9, p<0.001, \eta^2_p =0.026\)).
Also the two way interaction between hue and content (F=1.64, df=18, p<0.043, η²_p =0.023) is still significant with a small effect size. The effect of content is significant with a substantially larger effect size (F=339.4, df=2, p<0.001, η²_p =0.350). Findings of the first ANOVA on the raw data are thus confirmed.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>P</th>
<th>η²_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>339.4</td>
<td>2</td>
<td>&lt; 0.001</td>
<td>0.350</td>
</tr>
<tr>
<td>Hue</td>
<td>3.8</td>
<td>9</td>
<td>&lt; 0.001</td>
<td>0.026</td>
</tr>
<tr>
<td>Content*hue</td>
<td>1.64</td>
<td>18</td>
<td>0.043</td>
<td>0.023</td>
</tr>
</tbody>
</table>

*Table 5.16: results of the ANOVA on the normalized data (over participant) of the preferred chroma for the hue-rotated images (E3a). A * indicates an interaction.*

Although hue was corrected for, it still has a significant effect on preferred chroma. The effect size however, is small, and thus differences in preferred chroma between hues are rather small. The Tukey-HSD post-hoc analysis on hue shows which hues are different from the others. Two significantly different groups are found. The first group contains all hues except Y, the second group consists of 5 hues (B, Y, M, BC and G_EBU) as can be seen in figure 5.17. Although the Y hue is thus not significantly different from almost half of the hues, removing it will remove the significant effect of hue.

The preferred chroma for the multiple-hue images (M=1.31) is on average higher than for most single hue images, except for images with the Y hue.

<table>
<thead>
<tr>
<th>G</th>
<th>R</th>
<th>RM</th>
<th>C</th>
<th>R_EBU</th>
<th>BC</th>
<th>G_EBU</th>
<th>M</th>
<th>B</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.14</td>
<td>1.14</td>
<td>1.18</td>
<td>1.18</td>
<td>1.19</td>
<td>1.20</td>
<td>1.21</td>
<td>1.22</td>
<td>1.23</td>
<td>1.32</td>
</tr>
</tbody>
</table>

*Figure 5.17: grouping results of the Tukey-HSD post-hoc test on hue for ANOVA*. The numbers below the hue labels indicate the average CCM per hue.

The homogeneity of variance cannot be achieved. Although countermeasures are taken to correct for this lack of homogeneity, results could be slightly biased. Therefore significant results with small effect sizes have little value. As a result, the significant effect of hue and of the two way interaction between content and hue will not be considered of great value because of their low effect size (η²_p=0.073 and η²_p=0.052 respectively).

The effect of participant, however, has a large effect size (η²_p=0.697) on the results. Also, the effect of content and the two way interaction between content and participant are found to be significant. The effect sizes for this effect and interaction effect are relatively high (η²_p=0.633 and η²_p=0.424 respectively). Results over content were analyzed using a Tukey-HSD post-hoc test and the 3 types of content used for the hue-rotated images were found to be significantly different from each other (see figure 5.18).

<table>
<thead>
<tr>
<th>multiple bottles</th>
<th>paint bottle</th>
</tr>
</thead>
<tbody>
<tr>
<td>0.83</td>
<td>1.27</td>
</tr>
</tbody>
</table>

*Figure 5.18: grouping results of the Tukey-HSD post-hoc test on content for the ANOVA*. The numbers below the image labels indicate the average CCM per image.

Since the effect of participant was also found to be significant in ANOVA, participant was analyzed using a Tukey-HSD post-hoc test to find any potential outliers. No outliers were found in the group of participants.
In ANOVA, the 51 multiple hue images were used, CCM was included as the dependent factor, content as a fixed factor and participant as a random factor. The results of ANOVA (see table 5.19) show a significant effect on both content (F=17.8, df=50, p<0.001, η²_p =0.297) and participant (F=45.7, df=42, p<0.001, η²_p =0.478). Since the effect on content and participant was found to be significant for the ANOVA analysis, participant was analyzed using a Tukey-HSD post-hoc test. The results of this analysis show that for both content and participant no outliers or group or outliers were found.

<table>
<thead>
<tr>
<th>Effect</th>
<th>F</th>
<th>df</th>
<th>p</th>
<th>η²_p</th>
</tr>
</thead>
<tbody>
<tr>
<td>Content</td>
<td>17.8</td>
<td>50</td>
<td>&lt; 0.001</td>
<td>0.297</td>
</tr>
<tr>
<td>Participant</td>
<td>45.7</td>
<td>42</td>
<td>&lt; 0.001</td>
<td>0.478</td>
</tr>
</tbody>
</table>

Table 5.19: results from the ANOVA of the preferred chroma over the multiple hue stimuli (E3a).

Although more content was used for this analysis, results confirm the findings for ANOVA, because for both content and participant a significant effect is found. Results are thus largely affected by participant and content.

In order to test whether gender, nationality or age had an effect on the results, the preferred chroma per participant was analyzed using a Monte Carlo permutation test (Rubinstein, 1981). Because the differences in preferred chroma between the participants were so large, as can be seen in the box plot of figure 5.21, a one way ANOVA was not applied since it does not account for the differences between participants.

![Box plot of results per participant for E3a](image)

Figure 5.21: box plot of the results per participant for E3a. Notice the large difference between the chroma preferences of the participants. The vertical axis represents the chroma preference (CCM).

The Monte Carlo permutation test was applied first for gender. The difference between the mean results of each group was calculated. Men had a slightly higher preferred chroma (M=1.29, SD=0.50, N=27) score than women (M=1.23, SD=0.41, N=16). In order to test whether this difference lies within the confidence interval of 5%, in a two-tailed distribution, simulations were run to estimate the distribution of mean differences. The distribution of the mean differences of the responses was created by permuting the result over observers, while each observer’s characteristics were unchanged. Group sizes were kept constant. This way, 1000 groups with permuted results were created, yielding one distribution of mean differences between the groups. To estimate the confidence interval of the significance boundaries, 1000 of such distributions were created.
Figure 5.22 shows the average distribution (solid line), the 2.5% confidence intervals (outer vertical lines), the difference between means found for the original groups (middle red line), and one standard deviation distance (dotted lines). As can be seen in the figure, the mean difference for gender of the original groups does not lie outside of the 95% confidence interval. We can thus conclude that no significant difference is found between genders.

A Monte Carlo permutation test as described above was also conducted for nationality and age. Groups were divided in Dutch and non-Dutch, and younger than 24 years old and older than 25 years old respectively. Dutch participants had a higher preferred chroma on average (M=1.31, SD=0.50, N=25) than non-Dutch participants (M=1.20, SD=0.42, N=18). Also participants which were younger than 25 years preferred on average a higher chroma level (M=1.34, SD=0.49, N=20) than participants which were older than 24 years old (M=1.21, SD=0.44, N=23). The mean difference for both nationality and age of the original groups does not lie outside the 95% confidence interval. We can thus conclude that no significant difference is found for nationality and age.

5.3.2 Preferred color gamut for single- and multiple-hue images versus physical image properties
As found in the first experiment, the results on chroma preference again show a significant effect of content. In chapter 3, we investigated various options to relate the preferred chroma level to physical properties of the images. Both a local band-pass limited contrast filter as well as a high-chroma region filter were defined (see subsection 3.3.2). Measures from both filters showed potential correlations with the preferred chroma level per image. In this section, these measures are checked again for the larger database of images used during this third experiment. First, the results for the local band-pass limited contrast filter will be presented. After that, the results for the high-chroma region filter are described.

As mentioned before, the local band-pass limited contrast filter produces one contrast value per pixel (see subsection 3.3.2). Based on these contrast filtered images, several measures were applied to create one single measure out of the set of pixels. These measures were the maximum, the absolute maximum, the minimum, the mean, the median, the standard deviation and the mean of 10% of the highest value pixels. The resulting value per stimulus was correlated to the average preferred level of chroma for that image. The results of these correlations are shown in table 5.23 together with the correlation values obtained earlier in the first experiment (see table 3.15).
<table>
<thead>
<tr>
<th>Measure</th>
<th>E1a</th>
<th>E3a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>0.39</td>
<td>0.20</td>
</tr>
<tr>
<td>Absolute maximum</td>
<td>0.58</td>
<td>0.38</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.59</td>
<td>-0.37</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.40</td>
<td>-0.18</td>
</tr>
<tr>
<td>Median</td>
<td>0.13</td>
<td>0.00</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.49</td>
<td>0.25</td>
</tr>
<tr>
<td>Mean of 10% of the maximum data</td>
<td>0.39</td>
<td>0.21</td>
</tr>
</tbody>
</table>

Table 5.23: the correlation \((r)\) of the average preferred chroma per image against each measure of the local band-limited contrast filtered image as obtained in the first (E1a) and third (E3a) experiments. Notice that the correlation for E1a is based on \(N=38\) images, while the correlation for E3a is based on \(N=81\) images.

Although correlations with the measures of absolute maximum and minimum did seem promising based on the results of the first experiment (see section 3.3.2), these correlations dramatically decreased when based on the results of the current experiment (E3a). When comparing the actual correlation plot of for example the minimum measure (figure 5.24), a high correlation is found based on the results of E1a, but, for E3a, where more content was used, data points are more scattered over the plot. A clear correlation of the preferred chroma with any of the measures for local band-limited contrast is thus not found.

![Figure 5.24: the average preferred chroma (in % relative to the maximum amount of chroma the wide-gamut display can produce) against each measure of the local band-limited contrast filtered image for the first experiment (left) and this third experiment (right). The blue dots indicate each individual image while the red line indicates the linear model fit (with correlations of -0.59 for left and -0.37 for right).](image)

As mentioned before, differences in chroma preference over content could be caused by the actual size of the chroma regions in the images. To test if this hypothesis is true, a high chroma region filter was developed (see subsection 3.3.2).

After applying the high chroma filter to the images, several measures were taken. These measures were the maximum, the absolute maximum, the minimum, the mean, the median, the standard deviation, the mean of 10% of the highest value pixels and the size of the high chroma area in pixels. Table 5.25 shows the correlations for each measure for both the first (E1a) and the third (E3a) experiment.
As can be seen from table 5.25 the correlations between the high-chroma region filter and the average preferred chroma as obtained in this third experiment are low. This thus indicates that when more complex images are assessed, like in this third experiment, the high-chroma region filter is not a valid predictor for the preferred chroma. Figure 5.26 also shows that most of the included complex images have a relatively small high chroma area. The median amount of pixels left after applying the high-chroma region filter actually decreases from 41967 pixels for the image content used in the first experiment (i.e. only single-hue images) to 12617 pixels for the image content used in the third experiment (i.e. a mixture of single-hue and multiple-hue images). From these observations we conclude that neither of the measures from the local band-pass limited contrast filter nor from the high-chroma region filter is a good predictor for the preferred chroma of an image.

<table>
<thead>
<tr>
<th>Measure</th>
<th>E1a</th>
<th>E3a</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maximum</td>
<td>-0.25</td>
<td>-0.16</td>
</tr>
<tr>
<td>Absolute maximum</td>
<td>-0.25</td>
<td>-0.16</td>
</tr>
<tr>
<td>Minimum</td>
<td>-0.09</td>
<td>-0.21</td>
</tr>
<tr>
<td>Mean</td>
<td>-0.20</td>
<td>-0.22</td>
</tr>
<tr>
<td>Median</td>
<td>-0.19</td>
<td>-0.22</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>0.11</td>
<td>0.04</td>
</tr>
<tr>
<td>Mean of 10% of the maximum data</td>
<td>-0.20</td>
<td>-0.19</td>
</tr>
<tr>
<td>Size of high chroma area</td>
<td>-0.79</td>
<td>-0.27</td>
</tr>
</tbody>
</table>

Table 5.25: the correlation factors (r) of average preferred chroma against each measurement of high chroma region over all stimuli for both the first (E1a) and the third (E3a) experiment. Notice that the correlation for E1a is based on N=38 images, while the correlation for E3a is based on N=81 images.

Figure 5.26: the average preferred chroma (in % relative to the maximum amount of chroma the wide-gamut display can produce) against each measure of high chroma region per stimulus for the first experiment (left) and this third experiment (right). The blue dots indicate each individual image while the red line indicates the linear model fit (-0.79 for left and -0.27 for right).
5.3.3 Painting experiments

After the tuning, participants were asked to paint the areas of a subset of 11 stimuli. These stimuli were selected by the participant during the tuning experiment. Participants first painted areas which were perceived as being too colorful, after that participants painted areas which were perceived as being not colorful enough. In total, 48.2% of the stimuli in the first painting task and 41.4% of the stimuli in the second painting task were painted. Moreover, 22.0% of the exact same stimuli were painted by the same participant in both painting tasks, showing that almost half of the painted stimuli were perceived as having too low and too high chroma levels. Figure 5.27 shows the percentage of paintings per stimulus for both experimental parts.

![Figure 5.27: the percentage of participants per stimulus who painted for both painting task. The gray bars indicate the first painting task (E3b) in which areas were painted which were perceived as being too colorful and the white bars indicate the second painting task (E3c) in which areas were painted which were perceived as being not colorful enough.](image)

Figure 5.28 shows the percentage of stimuli painted for in the first painting task compared to the second painting task per participant. This data indicates that the number of paintings during between painting tasks are positively correlated. Thus participant’s preference to paint stimuli seems to be relatively equal for both experiments. Most participants themselves indicated at the end of the experiment that they mostly enjoyed the painting task and that the task was very intuitive.
Figure 5.28: the relationship between the percentage of painted images between the first painting task (E3b) and the second painting task (E3c) per participant. Blue dots represent participants while the red line represents the correlation between the percentage of images painted per participant between two painting tasks (r=0.57).

For each pixel in each stimulus, the number of times it was painted has been counted. Subsequently, the number of paintings per pixel has been normalized to the number of paintings for each image. These normalized numbers per pixel were translated linearly to gray-scaled images. These images are shown in figure 5.29. From each gray-scaled image a mask was created containing the pixels which were painted by at least 50% of the painters. Following that, the number of actual hues within the original stimulus has been reduced using mean shift clustering (Comaniciu & Meer, 2002). Mean shift clustering, groups similar hues without prior knowledge about the number of hues. The created masks for both painting tasks were used to select the clusters corresponding to pixels which were painted by more than 50% of the painters. Using these masked stimuli, it becomes visible that, from the chosen stimuli, red and green areas were indicated as being too colorful; lime and yellow were painted as being not colorful enough. Natural objects like the skies and flowers in the images, which evoke memory color, are relatively often painted. Because of memory color, most participants have a reference point for these objects which can explain why these parts are painted relatively often. Parts which have low chroma due to high lightness levels are often indicated as being not colorful enough, participants in these cases thus preferred colorfulness over high lightness. These findings, however, are merely a qualitative indicator rather than an exact measure. This is due to the fact that the painted hues are not equal to the hues used in the hue-rotated images. The stimuli on which participants painted depended on their selected level of chroma during the tuning experiment prior to the painting tasks. Thus, participants painted stimuli with differed levels of chroma. And lastly, less than 50% of the stimuli were painted, indicating that more than 50% of the selected stimuli were judged as having satisfactory chroma throughout.
Figure 5.29: for each paint-stimulus 6 images are provided. The original image is given (top left) as well as the color clustered image (bottom left), the cumulative painted areas for the first painting task (top center) and the areas which were painted by at least 50% of the painters for the first painting task (bottom center). Furthermore the cumulative painted areas for the second painting task (top right) and the areas which were painted by at least 50% of the painters for the second painting part (bottom center) are shown. The colors clusters shown in these painted regions are the mean of the corresponding cluster for each region.

5.4 Discussion

The results from tuning experiment indicate that the preferred level chroma lies above a CCM of 1. As a CCM of 1 refers to the preferred level of chroma for E1a, we can thus conclude that chroma preference is higher when multiple-hue images are presented. This conclusion is further strengthened by the fact that the mean CCM level for the multi-hue images was higher than the single hue images, except for yellow. Also the paint stimulus, which has been used in both the first and the third tuning experiments, confirms this higher CCM. This difference in chroma preference between the two experiments can be caused by the image window used in the first experiment which lowered the average outcome of that experiment. We can thus conclude that the first hypothesis (the natural boundary for the extension that people prefer (“preferred” gamut) is larger than the standard TV gamut) cannot be rejected, since the preferred chroma boundary for the third experiment is larger than the preferred chroma boundary of the first experiment on average. The preferred chroma boundary of the first experiment by itself is larger than the EBU color gamut, hence the preferred color gamut found during the third experiment lies out of the EBU gamut.
Since the results for the tuning part of this third experiment also showed that there is a significant effect of content for both the hue rotated stimuli as well as the multiple hue stimuli, the preferred gamut boundaries are thus content dependent like in the first experiment. We thus cannot reject the second hypothesis (the preferred gamut boundaries are content dependent). To investigate this unexplained difference between content, results were analyzed on physical properties. Although the results of the first experiment were promising, the results of this experiment show that no clear causes for the effect of content can be obtained with the approach taken. The two filters which were used for the analysis, a high-chroma region filter and a local band-pass limited contrast filter, thus do not seem to predict the outcome of the experiment when a wider range of content is assessed. The results of the multiple bottles content used within the experiment however suggest that the results are biased if a stimulus contains high-chroma area with almost the same size as the stimulus. However, this is most probably more related to the perceived content of the stimulus rather than its physical properties. Participants indicated that that specific type of content was hard to assess since it appeared to be artificial. In further studies it is thus advisable to be aware of any bias caused by artificial appearance.

The hues from the hue rotated stimuli showed a significant effect for the Y (yellow) hue. Since the other 9 hues did not differ significantly much we can conclude that the shape of the preferred gamut boundaries is irregular since the irregular shape of the color gamut boundary was accounted for by the gamut extension algorithm. However, results are only applicable to the limited set of stimuli used. As could be seen within the results, they differed much between the different content used for this experiment. More results on more single-hue stimuli should provide more insight on the relationship between hues. For now, however, hypothesis 3 (the preferred gamut boundaries have an irregular shape) cannot be rejected.

As can be seen from results, chroma preference differed strongly between participants. This can be both confirmed by visually inspecting the results per participant as well as the significant effect on participants in the ANOVA analyses. This thus means that hypothesis 4 (the preferred gamut boundaries are person dependent) cannot be rejected. One notable difference between the effects of the first and the third experiment was the non significance of the participant-hue interaction effect. This shows that the variability in hue preference between participants can be modeled by a single axis up to the statistical power of the third experiment.

In more than 50% of all painting tasks participants indicated that no chroma corrections were needed, which indicates that the participants either were able to use the gamut extension successfully to control for their own multi hue preference or that the participants simply didn’t notice any irregularities. Furthermore, the painting experiments indicated that mostly the red and green hues were extended too much by the gamut extension algorithm while the yellow and lime hues were rated as being extended to little. The latter hues however are mostly limited by the size of the wide-gamut display’s color gamut. The painting also showed that participants did not have a preference to paint either too low chroma areas or too high chroma areas. Their preference for painting seems to be relatively equal for both cases. Finally, participants indicated that they enjoyed the painting task and perceived it as a very intuitive tool. These findings indicate that further use of this tool, mainly for qualitative measures can prove to be very useful in further studies.
6 Conclusion and future work
This report discussed three experiments. The first experiment explored the preferred and maximum acceptable color gamut by means of a psychophysical tuning experiment on single-hue images. To verify the tuning methodology used in the first experiment, a second experiment was performed, which used a randomized staircase for the same task. The third and last experiment evaluated a simple gamut extension algorithm which was based on the results of the first experiment for additional single and multiple hue images.

The results from the first and last experiments of this study provide evidence that the EBU color gamut is not large enough to satisfy the chroma preference of users measured during psychophysical experiments. Thus there is a perceptual need for wide-gamut displays. Furthermore, results from the first experiment suggest that even for wide-gamut displays the color gamut is not always sufficient; hence there is a need for even wider color gamut displays. However, for some of the hues, in particular the R (red) and G (green) hue, the results indicated that the wide color gamut was sufficient to a large extent. The results of the first experiment confirm earlier findings of Laird & Heynderickx (2008), who also concluded that the typical preferred gamut for natural content is outside the EBU gamut, though not reaching the full extent of the wider-gamut boundary.

Furthermore, analysis on the results of the first experiment suggests that preference and maximum acceptability of the users color gamut is dependent on at least 3 factors: participant, image content and hue. The last experiment confirms these findings on preference of the user color gamut. In both experiments, the effect of participant was the strongest, followed by the effect of content, and the smallest effect was found on hue. Although the effect on hue was the smallest, results showed that a small difference in hue, at least for red hues, can have a large influence on the results. Also Laird & Heynderickx (2008) found the dependency for participants, image content and hue. However, their results indicated that the effect of hue was the strongest, followed by image content and participant. This discrepancy in findings can be caused by the limited amount of images used by Laird & Heynderickx.

The robustness of the tuning methodology used to measure chroma preference and maximum chroma acceptability has been verified by the second experiment (E2) using a combination of paired comparison and staircase methodology. Results show that both experimental methods yield results which are not significantly different when the same participants are assessed. Thus, the tuning methodology has been verified. Furthermore, a pilot study suggests the existence of a bimodal preference for chroma, meaning that chroma preference was found not only at a relatively colorful chroma level but also very close to the zero chroma level. Thus, there seems to be a preference for gray-scaled images besides their colorful versions. This preference was not found during the tuning experiment, indicating that the preference for gray-scaled images is lower than their colorful versions. The results from this experiment also show that the tuning methodology is almost 4 times as fast as the methodology used in the second experiment. For further research it is recommended to use the given tuning methodology, which enables researchers to test a larger number of images.

Using the information gained on user color preference from the first experiment, a gamut extension algorithm has been developed which tries to take the effect of both hue and participant into account. This gamut extension algorithm has been tested by means of the third psychophysical experiment. Results from that experiment indicate that the effect of participant is not sufficiently accounted for; however, the effect of hue has been accounted for to a great extent. Only the Y (yellow) hue seems not to be accounted for sufficiently. This can be explained by the limited color gamut in yellow of the wide-gamut display, which already was proven in the first experiment to be too small. Furthermore, image content seems to have a large, yet unexplained, effect on the results.
Attempts to relate physical image properties to the results failed, showing at least that the effect of content is most probably not directly related to the size of high-chroma regions, nor is it related to contrast as defined in this study. Participants did indicate that when certain images were presented, mostly those with a high chroma area which approximated the size of the image, they were interpreted as being artificial. Examples of these images are the rose (see image 3.1) and the multiple bottles (see image 5.2).

On average, the results of the last experiment indicated that chroma preference is slightly higher when stimuli containing multiple hues were assessed. This finding can be explained by the fact that the first experiment contained one image which was consistently assessed relatively low and thus lowered the average outcome of that experiment. Another explanation could be that the tolerance for high chroma areas becomes higher when images become more complex. During the painting tasks of the second and third part of the last experiment (respectively E3b and E3c), participants could paint parts of the stimuli they selected during the tuning part of that experiment, which were under or oversaturated. In more than 50% of all painting tasks, participants indicated that no corrections were needed. This thus indicated that either the proposed gamut extension algorithm performed sufficiently or that participants indeed are not sensitive enough to notice irregularities. The results of the painting tasks also supported the need for a wider gamut in yellow and it indicated that both red and green hues are sometimes oversaturated. Finally, participants indicated that they enjoyed painting the images and that the painting tool worked intuitively. The painting tool thus potentially could provide a simple and effective way to evaluate gamut extension algorithms in a vast number of images.

The results from the third experiment show that the developed gamut extension algorithm has room for improvement. It is not only possible to improve this gamut extension algorithm using more psychophysical experiments, but also the information gained through this study can be used to limit the use of the wide color gamut in other gamut extension algorithms. Thus previous studies, of which none were able to provide a very successful gamut extension algorithm, can prove to be more successful when colors are stretched to the preferred chroma boundary.

Future studies should also investigate to what extent the users are sensitive to over- and underexposure of chroma of a part of a stimulus. The painting tasks provided a first qualitative measure but do not provide detailed information about user sensitivity. Participants could be relatively sensitive to chroma differences. Using this knowledge about user chroma sensitivity for a part of a stimulus could aid gamut extension algorithms since variance between participants and content can be accounted for (partially) by the insensitiveness. More detailed information should also be gained about less chromatic colors and how these can be stretched. Especially the stretching of colors which appear in faces should be handled with care. More research can include wide color gamut displays which provide yellow hues with a higher chroma level. The results from the first experiment showed that there is a ceiling effect for the yellow hue which suggests that the maximum chroma level produced by the display used in this study is too low.

A future study can also compare original wide gamut content to the image content stretched by the gamut extension algorithm introduced in this study. This can provide more insight in the actual discrepancy between users color gamut preferences, which for example can be biased by memory color, and the most realistic gamut extension. This kind of research can also provide insight on how sensitive the users are to over and under saturation of partial image content.
7 References


Appendix A: Color terminology
This section serves to clarify the meaning of the majority of color terms used within this thesis. More detailed information can be found in Morovic (2002), Wyszeck & Stiles (1982), Hunt (1978) and the rest of the cited literature.

A.1 Intensity
Intensity is the physical measure of energy over some interval of the electromagnetic spectrum which is radiated by a surface. So it expresses how much light (in a specific range of the electromagnetic spectrum) is radiated from a surface. Intensity is expressed in Watts per square meter (W/m²).

A.2 Brightness
According to the Commission Internationale de L’Éclairage (CIE) brightness is described as “the attribute of a visual sensation according to which an area appears to emit more or less light” (Wyszeck & Stiles, 1982). In short this term can be defined as the amount of light the human visual system perceives. It is a perceptual measure.

A.3 Luminance (Y)
The human visual system has different sensitivities to electromagnetic radiation of different wavelengths. Correcting this intensity measure to this sensitivity function results in the measure called Luminance. Even though they are proportional, luminance is non-linearly related to brightness. Luminance is denoted by Y and expressed in candela per square meter (cd/m²). The CIE XYZ color space related luminance is mostly scaled between 0 and 1 or 0 and 100. The reference point for this scale is often indicated by the luminance level (in cd/m²), which refers to the maximum luminance (i.e. 1 or 100).

A.4 Gamma
Many capture and reproduction devices, including the human visual system, have a nonlinear response to a change in light intensity. This nonlinearity is usually represented by Steven’s law (Goldstein, 1999). The exponent of this power law is usually denoted with the Greek letter gamma (γ). Because of this, the value of γ is usually called the gamma of the system. For displays, gamma corrects for the non-linearity between the intensity measured from a display’s surface and its RGB driving values. For conventional displays the intensity of the display is the RGB driving value raised to the power of approximately 2.2 to 2.5.

A.5 Lightness (L*)
Lightness is a perceptual measure which represents the brightness of an object judged relatively to the brightness of a white object under the same illuminant. Lightness is expressed in L* and again ranges from 0 to 1 or from 0 to 100.

A.6 Chromaticity (CIE xy)
Chromaticity can be described as color without taking luminance or lightness into account. It is actually defined by the CIE by means of the CIE x and y chromaticities. When Luminance is added to the CIE chromaticities, the CIE xyY color space (see subsection 2.2.2) is obtained and any color can be represented using these three dimensions.
A.7 Correlated color temperature
When heated to a certain temperature, a hypothetical black body (an object without any reflections) emits electromagnetic radiation of a certain spectral power distribution, and consequently a certain color. The source that this object represents is called a black body radiator, and the color it produces is usually denoted by its temperature in degrees Kelvin. The chromaticity of an achromatic light source can be similar to the chromaticity of a black body radiator, and thus can be also described by the temperature of the black body radiator with a similar chromaticity. This is called the correlated color temperature. A low color temperature (about 2200 K) provides an almost achromatic color with a red chromaticity, while a medium color temperature (about 5000 K) has a yellow chromaticity. A very high color temperature (about 9000 K) is an almost achromatic color with a blue chromaticity.

A.8 Hue
Hue can be defined in two ways. Firstly hue can specify color, like yellow, orange, purple, pink etc. Hue is also used within the CIE LCh color space and its meaning is similar to the more intuitive definition of color. It differs from color since it describes that hue is one of the colors red, yellow, green and blue or any combination of two of these colors (Wyszeck & Stiles, 1982).

A.9 Chroma
Chroma can be defined as “the colorfulness of an area relative to the brightness of a reference white” (Hunt, 1978). Physically, it relates to the width of the color spectrum of a light. If the color spectrum is more narrow the chroma increases. Maximum chroma can thus be achieved by using one single wavelength of a color spectrum.

A.10 Saturation
Although there is no formal description for saturation, saturation can be described as the colorfulness relative to a gray color of the same brightness. Colorfulness refers to the perceived intensity of a color. It depicts to which extent the perceived color content appears to be more or less chromatic. In the CIE LCh color space saturation is often explained as the mathematical equation of chroma divided by lightness.
Appendix B: instructions for the first experiment on preferred color gamut

Instructions for Tuning experiment:

Experiment description
In this experiment, you will see one image at a time on the TV in front of you. For each of the images your task will be to adjust the color of the image to your preferred level.

Controls
To decrease the color press the left arrow key (small steps) or the down arrow key (large steps)
To increase the color press the right arrow key (small steps) or the up arrow key (large steps)
After you find the preferred level of color press enter to continue to the next image.

Trials
Prior to the experiment you will have 3 trials to become accustomed to the experiment and the use of the controls. Please feel free to ask questions during these trails.
Appendix C: instructions for the first experiment on maximum acceptable color gamut

Instructions for Tuning experiment:

Experiment description
In this experiment, you will see one image at a time on the TV in front of you. For each of the images your task will be to select the highest level of color to which the image—in your opinion—still looks acceptable.

Controls (are the same as in the previous experiment)
To decrease the color press the left arrow key (small steps) or the down arrow key (large steps)
To increase the color press the right arrow key (small steps) or the up arrow key (large steps)
After you find the preferred level of color press enter to continue to the next image.

Trials (are the same as in the previous experiment)
Prior to the experiment you will have 2 trials to become accustomed to the experiment and the use of the controls. Please feel free to ask questions during these trails.
Appendix D: instructions for experiment the second experiment

Instructions for paired comparison/staircase experiment:

Experiment description
In this experiment, you will see two images at a time on the TV in front of you. For each of the images your task will be to select the image which you prefer the most.

Controls
To select the left image press the left arrow key
To select the left image press the right arrow key
After each selection the next two images will be shown automatically.

Trials
Prior to the experiment you will have one trial to become accustomed to the experiment and the use of the controls. The same image will be repeated during the trial, during the experiment however a different image will be shown after each selection. Please feel free to ask questions during these trails.
Appendix E: instructions for experiment the third experiment

Welcome!
This experiment consists of 3 parts and will take approximately 30 minutes.

**Experimental part 1**

**Task**
You will see one image at a time on the TV in front of you. For each image your task will be to adjust the image to your preferred level of colorfulness.

**Controls**
UP/DOWN arrow keys: adjust the image using large steps
LEFT/RIGHT arrow keys: adjust the image using small steps
Press ENTER once to confirm your selection and to continue with the next image

**Experimental part 2**

**Task**
Again you will see one image at a time on the TV in front of you. Your task will be to paint the area which you perceive as being too colorful

**Controls**
Hold the LEFT MOUSE BUTTON and DRAG to paint the area or areas
Hold the RIGHT MOUSE BUTTON and DRAG to remove a painted area
Press SPACE to undo all selections on the current image
Press ENTER once to confirm the area(s) and to continue with the next image

**Experimental part 3**

**Task**
This part is similar to part 2, except that your task will be to paint the area which you perceive as being not colorful enough

**Controls**
Identical to part 2

**Trials**
Prior to the experiment some trials will be presented to get accustomed to the controls.

Thank you for your participation!