Analysis of Material Appearance using Real-World Objects and Rendered Images

January 2016

Midori TANAKA

Graduate School of Advanced Integration Science

CHIBA UNIVERSITY

（千葉大学審査学位論文）

Analysis of Material Appearance using Real-World Objects and Rendered Images

January 2016

Midori TANAKA

Graduate School of Advanced Integration Science

CHIBA UNIVERSITY

ABSTRACT

In recent studies of pertaining to fields such as human perception, human recognition and image engineering, the topic of shitsukan (the appearance of materials) has attracted much attention. Most of the conventional studies have used object images as stimuli for shitsukan analysis. However, in most cases, the displayed images usually give a different perception compared with real-world objects, and the relationship between the appearance of materials for real-world objects and rendered images has not been discussed. As a fundamental investigation, the author conducted color naming experiments for two-dimensional (2D) and three-dimensional (3D) rendered samples, and found that there were definitely differences for color terms between 2D and 3D samples.

The purpose of this dissertation is to analyze the concept of shitsukan using real-world objects and rendered images. To achieve this purpose, the author conducted two different experiments to investigate the perceptual qualities and appearance harmony using real materials and rendered images.

In the first experiment, the author investigated the perceptual qualities of a material appearance using real materials and degraded image versions of the same materials. The author constructed a real material dataset and four image datasets by varying the chromaticity (color vs. gray) and resolution (high vs. low) of the materials' images. To investigate the fundamental properties of the materials' static surface appearance, the author used stimuli that lacked shape and saturated color information. The author then investigated the relationship between these perceptual qualities and the various types of image representation by performing psychophysical experiments. The results showed that for some materials, the method employed to represent them affected their perceptual

qualities. These cases were classified into the following three types based on whether their (1) perceptual qualities decreased by reproducing the materials as images, (2) perceptual qualities decreased by creating gray images, or (3) perceptual qualities such as “Hardness” and “Coldness” tended to increase when the materials were reproduced as low-quality images. By performing methods such as principal component analysis (PCA) and k-means clustering, the author found that material categories are more likely to be confused when materials are represented as images, especially gray images. An analysis between physical properties such as the bidirectional reflectance distribution function (BRDF) and psychophysical evaluations are also discussed.

In the second experiment that was carried out to investigate the harmony of the material appearance, the author investigated the appearance harmony of various materials by conducting psychophysical experiments aimed at collecting quantitative data. The author conducted three sub-experiments using 435 round-robin pairs of 30 samples made from 10 actual materials. In the first sub-experiment, in addition to surface appearance, subjects were allowed to tilt the pair of samples to obtain a comprehensive determination of the harmony based on the reflectance properties of the actual surface. In the second sub-experiment, the samples were placed such that their surfaces and the viewing direction were perpendicular to the subject. In the third sub-experiment, static sample images were displayed on a monitor. The results indicated that the sample pairs with similar surface properties were viewed as harmonious, although their materials were different. The appearance harmony of the materials differed among static real samples, tilted samples, and the displayed static images. In particular, the appearance harmony of some materials was significantly affected by the reactions of subjects to visual information regarding the samples with/without observations of the monitor, rather than tilting a sample. The PCA results indicated that

the harmony among categories of glossy materials was more likely to change when the materials were displayed as images. An analysis between physical properties such as the anisotropy and psychophysical evaluations is also discussed.

CONTENTS

1. Introduction.....1

1.1 Motivations 2

1.2 Contents and Structure of This Dissertation 6

2. Fundamental Study of Shitsukan Perception under Different Viewing Conditions 9

2.1 Introduction 10

2.2 Experimental Methods 14

2.2.1 Basic Color Terms in Modern Japanese..... 14

2.2.2 Image Rendering of Objects 14

2.2.3 System Construction 15

2.2.4 Procedure 18

2.3 Experimental Results 19

2.3.1 Distribution of the Color Terms 19

2.3.2 Color Term Transition..... 25

2.3.3 Frequency of Color Terms..... 27

2.3.4 Reaction Time..... 28

2.3.5 Modal Color Terms 30

2.4 Discussion 35

2.4.1 Position of Illumination 35

2.4.2 Brightness Level..... 37

2.5 Conclusions..... 40

3. Investigating Perceptual Qualities of Static Surface Appearance..42
3.1 Introduction 43
3.2 Experimental Stimuli 46
3.2.1 Material Dataset 46
3.2.2 Image Dataset 49
3.3 Experimental Methods 50
3.3.1 Experiment 1: Visual Judgments of Perceptual Qualities using Material Dataset 51
3.3.2 Experiments 2-5: Visual Judgments of Perceptual Qualities using Image Datasets 54
3.4 Experimental Results 56
3.4.1 Intra- and Inter- Participant Variances 56
3.4.2 Ratings for Each Material Class 58
3.4.3 Ratings for Each Perceptual Quality 61
3.4.4 Correlations between Perceptual Qualities 64
3.4.5 Correlations between Material Classes 67
3.4.6 Distributions of Material Classes in the Space of Perceptual Qualities 68
3.5 The Effect of Binocular Parallax for Material Perception 86
3.6 Analysis between Physical Properties and Psychophysical Evaluations ... 89
3.7 Conclusions 97

4. Investigating Appearance Harmony of Materials.....99
4.1 Introduction 100
4.2 Experimental Stimuli 102

4.2.1 Material Dataset 102
4.2.2 Image Dataset 106
4.3 Experimental Methods 106
4.4 Results and Discussion 108
4.4.1 Intra- and Inter- Subject Variance 108
4.4.2 Perceptual Harmony Ratings within and across the Categories of Materials 110
4.4.3 Changes in Harmony among Experiments 112
4.4.4 Distributions of Samples in the Appearance Harmony Space 122
4.4.5 Consideration for the Color Effect 129
4.5 Analysis between Physical Properties and Psychophysical Evaluations . 131
4.6 Conclusions 138

5. General Conclusions and Future Works 140
5.1 General Conclusions 141
5.2 Future Works 144

Acknowledgements 146

References 147

Contributions 154

Chapter 1. Introduction

1.1 Motivations

Based on recent developments in imaging technology, we can easily acquire and display real-world scenes as digital images. In the digital imaging world, accurate image reproduction is very important. Color management technologies have enabled us to realize true-color reproduction using different color-imaging devices. High-dynamic-range (HDR) imaging technologies have realized realistic image reproduction under HDR scenes. The development of high resolution display devices has enabled us to realize precise image reproduction. However, by using these current imaging technologies, it is still difficult to accurately represent the material appearance of real-world objects on a display device. Figure 1.1 shows that perceptual qualities such as glossiness represented as graphs are different for a car in the real-world and a rendered image of the same car. In our everyday life, there are many experiences that result in changes to the impression of real-world objects by displaying them on a monitor.

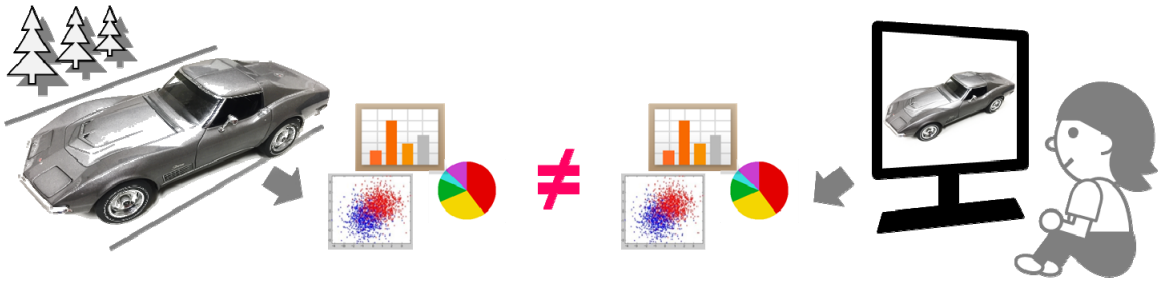


Figure 1.1. Different impressions are obtained for a real-world car and a rendered car.

Recently, a material appearance has been considered to be a part of "shitsukan," and the mechanism of shitsukan perception has been investigated in fields such as brain science, image engineering, and psychology. "Shitsukan" is a Japanese word, but there is

no exact corresponding English word. In this dissertation, the author uses the term "shitsukan" to present material perception, which includes perceptual qualities (glossiness, transparency, etc.). The study of shitsukan is a treasure trove of technological innovation for computer science, but the recognition of shitsukan has had very difficult problems. There are several unknown phenomena, such as the kind of stimulation input, the kind of reaction to shitsukan, and the kind of calculation principle used to estimate them. There are many unknown phenomena, and it has high academic significance to challenge.

The study of shitsukan has attracted much attention in several fields such as vision science, brain science, and image engineering. In Japan in 2010, a shitsukan-related research project, involving brain and information science was started, and the author participated in the project as a study cooperator. The aim of this project was to create a new scientific field about shitsukan in cooperation with engineering, psychophysics, and brain science. It was also a leading project for several other interdisciplinary projects of various scales, which commenced afterwards. In addition, studies regarding shitsukan have been conducted actively all over the world, and a collaborative investigation project called PRISM (Perceptual Representation of Illumination, Shape & Material) was performed in Europe in 2012, and in 2014, an international meeting about material perception was held at the IS&T Electric Imaging Conference in the United States.

Examples of current results of such projects are as follows. In the field of psychology, Wiebel et al. investigated the degree of correspondence between the visual and the haptic representations of different real-world materials for a wide variety of material properties (Wiebel et al., 2013a). They asked subjects to rate 84 different materials falling within seven material categories (plastic, paper, fabric, animal materials (fur and leather), stone, metal, and wood) for 10 qualities (roughness, elasticity, colorfulness,

texture, hardness, three-dimensionality, glossiness, friction, orderliness, and temperature). They found that all material samples were similarly organized within the perceptual space. Further, a subsequent procrustes analysis confirmed that the visual and haptic material spaces are closely linked. Baumgartner et al. also investigated the degree of correspondence between the visual and the haptic representations using real-world objects having different materials (Baumgartner et al., 2013). They asked subjects to both categorize and rate different material properties for 84 different real materials. Based on their results, they found that although the haptic sense appears to be crucial for the perception of real materials, the information that it can gather may not by itself be sufficiently fine-grained and rich to enable perfect material recognition. Furthermore, Wiebel et al. investigated the speed and accuracy of material recognition using images of natural scenes (Wiebel et al., 2013b). Their results showed that material categorization could be as fast as basic-level object categorization, but was less accurate. In the field of brain science, the presence of a region that distinguished glossiness in the brain was found by measuring reactions using functional magnetic resonance imaging (fMRI) when the images of objects with/without glossiness were shown to monkeys (Okazawa et al., 2012).

These studies used either real-world objects or rendered images as the stimuli. As described before, perceptual qualities between both stimuli are usually different. However, with respect to the material appearance, the relationship between real-world objects and rendered images has not been frequently investigated. Furthermore, they performed conventional analysis considering the direction from stimuli to perception. To control the perception of shitsukan in the engineering field, studies in the reverse direction from perception to stimuli have become important. For example, it is important to find image features in order to regulate the perceptual glossiness.

In this dissertation, the author focuses on the analysis of a material appearance using real-world objects and rendered images in order to understand the mechanism behind shitsukan. Before the author deeply investigates the problem, the author investigates the color appearance of objects as fundamental shitsukan characteristics. Color is an important factor in the perception of shitsukan. In this study, the author focuses on color naming. Existing color naming experiments generally utilize two-dimensional (2D) samples such as color patches. However, it should be noted that most objects that are visible in real-world scenes have three-dimensional (3D) curved surfaces instead of 2D flat surfaces. The author investigates the color appearance between 2D and 3D objects. In order to reproduce exact same color between both object, in this experiment, the author used rendered images instead of real-world objects.

This dissertation focuses on two main areas related to static material appearance, and appearance harmony of static and moving objects. The first area focuses on perceptual quality. In a previous work, Fleming et al. reported that perceptual qualities and material classes are closely related based on the use of projected images (Fleming et al., 2013). Their study was conducted using only projected images without any physical information. In this dissertation, the author proposes and conducts new experiments using real-world objects and rendered images for the analysis of relationship between real-world objects and rendered images in shitsukan. The second goal involves investigating the appearance harmony of a pair of materials using real-world objects and rendered images. When we treat multiple materials, it is important to consider the combination of shitsukan-related effects. Thus, the author focuses on the analysis of the effects of different combination on a material appearance. In previous studies, for both the first and second goals, the effect of color has long been considered as color emotion or color harmony. Therefore, the author studies the influence of other factors pertaining

to shitsukan, with the exception of the color. In addition, the author obtains the physical properties such as the BRDF of real-world objects and anisotropy of rendered images, and investigates the relationship between them and psychophysical assessments.

As mentioned above, it is important to analyze the relationship between physical properties and psychological assessments to obtain the material appearance using a few display methods to apply to manufacturing systems that can control human's sensibility based on a constructed shitsukan management technology. By performing this dissertation, the investigation of the shitsukan perception using real-world objects and rendered images is believed to be of benefit to the development of new image-reproduction technology in the field of engineering.

1.2 Contents and Structure of This Dissertation

This dissertation is organized as follows.

Chapter 2 presents a fundamental study of shitsukan perception for different viewing conditions. The author conducted color naming experiments that were performed using 2D and 3D rendered samples. First, an introduction is described in Section 2.1. In Section 2.2, the experimental method, which consists of four sub sections, is as follows: basic color terms in modern Japanese, image rendering of objects, system construction, and experimental procedure, are introduced. Next, Section 2.3 shows the experimental results by analyzing the distribution of the color terms, color term transition, frequency of color terms, reaction time and modal color terms. A discussion about the position of the illumination and brightness level is given in Section 2.4, and the author concludes the study about color naming in Section 2.5.

Chapter 3 introduces the investigation about the perceptual qualities of static appearance using real materials and displayed images. In Section 3.1, the author introduces a previous study of shitsukan using a psychophysical approach, and describes the difference from a previous study that is cited in this dissertation. Section 3.2 presents the experimental stimuli that were developed using a material dataset and image dataset. The experimental methods related to the visual judgments of perceptual qualities using both datasets are presented in Section 3.3. Section 3.4 presents the experimental results that were obtained through analysis based on intra- and inter-participant variances, ratings for each material class, ratings for each perceptual quality, correlations among perceptual qualities and material classes, and distributions of material classes in the space of perceptual qualities. In Section 3.5, an additional analysis between physical properties and psychophysical evaluations is also discussed. The chapter ends with a summary in Section 3.6.

Chapter 4 presents an investigation of the appearance harmony of materials using several display methods. Section 4.1 introduces the overview and describes the motivation of the study into appearance harmony. In Section 4.2, experimental stimuli including the material dataset and image dataset are introduced. Section 4.3 presents the experimental methods using three different display methods. Further, the author presents the results and discussion by showing results such as intra- and inter- subject variances, perceptual harmony ratings within and across the categories of materials, changes in the harmony among experiments, distributions of samples in the appearance harmony space and consideration for the color effect in Section

4.4. An additional analysis between physical properties and psychophysical evaluations is also discussed in Section 4.5. Finally, conclusions are shown in Section 4.6.

Chapter 5 provides general conclusions of the method used for the shitsukan analysis, and discusses future works.

Chapter 2. Fundamental Study of Shitsukan Perception under Different Viewing Conditions

2.1 Introduction

Conventional shitsukan studies have used either real-world objects or rendered images as the stimuli. As described in the previous section, perceptual qualities between both stimuli are usually different. However, with respect to the material appearance, the relationship between real-world objects and rendered images has not been frequently investigated. As a fundamental study, in this section, the author focuses on “color” as an important factor in the perception of shitsukan, and investigates color naming under different viewing conditions.

Although color naming is less specific than color identification through numerical representation, it is more intuitive and easily understood. Human beings respond to an enormous number of real-world color stimuli by using color names. Therefore, investigations of color vocabulary are important for not only color science (Regier, Kay & Khetarpal, 2007, 2009) but also applications in various fields such as linguistics (Roberson, Davies & Davidoff, 2000; Kay & Regier, 2007), anthropology (Kay, 1999; Webster & Kay, 2007), and ethnology (Bornstein, 1973). In recent image processing studies, color naming was sometimes used as an effective tool for image analysis and display. In fact, it is intuitively easier to understand color specification by using language rather than numerical representation based on colorimetric systems such as tristimulus values and RGB values. For instance, effective usage of color names has been reported in image segmentation (Mojsilovic, 2002) and color gamut mapping (Motomura, 2002) studies.

One of the most important contributions to color naming is the study of basic color terms, conducted in 1969 by Berlin and Kay (Berlin & Kay, 1969). Their study was based on psycholinguistics and ethnology, and these authors suggested that most world

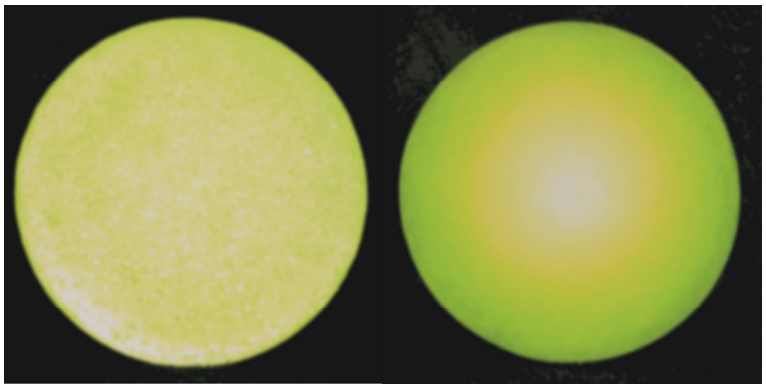
languages have a common system for fundamental color specification. That is, there are cross-linguistic universal tendencies in color naming. These authors claimed that 11 basic color terms (white, black, red, green, yellow, blue, brown, purple, pink, orange, and gray) are found in all human languages. This finding was widely supported by many studies (Kay & MacDaniel, 1978; Boynton & Olson, 1987; Regier, Kay & Cook, 2005). In Japan, a study on color terms in modern Japanese found that a set of 15 color terms, including gold, silver, turquoise, and yellow-green in addition to Berlin and Kay's 11 basic color terms could form a stable set of important color terms used in Japanese daily life (Tominaga, Ono & Horiuchi, 2010).

The study of color naming has gained momentum in recent years. Traditionally, most investigations on color naming were based on the use of color samples or reference patches. More recently, approaches such as experiments on color naming via the World Wide Web (Moroney, 2003; Weijer, Schmid & Verbeek, 2007) and learning of color names using a large set of image data from Google and eBay were introduced. These approaches enable the collection of a huge volume of color name data from the general public in a relatively short time. However, these approaches have several limitations. The collected color names may be influenced by the display devices used, viewing environments, and demographic characteristics of participants (gender, age, culture, etc.).

Existing color naming experiments generally utilize two-dimensional (2D) samples such as color patches. However, it should be noted that most objects we see in real-world scenes have three-dimensional (3D) curved surfaces rather than 2D flat surfaces. In such cases, the 3D object surfaces often include shading, therefore, the appearance is affected by various illumination effects. To illustrate such differences, the author prepared real objects, a 2D disk and a 3D sphere, which were made of the same material as shown in

Fig. 2.1. The material used was foamed polystyrene, and both surfaces were painted with the same paint of CIELAB color space values (L*, a*, b*) = (71, -48, 47). The objects were illuminated from the front by the light source, a daylight lamp. As shown in Fig. 2.1, the appearances of the two objects differed remarkably. Therefore, the author assumed that the specified color names would be different for the 2D and 3D objects, such as “yellow” or “yellow-green” for the 2D object and “green” or yellow-green” for the 3D object, despite the fact that the surfaces were painted with the same color and were observed under the same viewing conditions.

This paper describes a color naming experiment using both 2D and 3D rendered color samples on a display device. The relationship between the 15 basic Japanese color terms (Berlin & Kay, 1969) and object surfaces and the illumination effect was analyzed. The author rendered the color images of a flat disk for a 2D sample and a sphere for a 3D sample on a calibrated display device. The author set the condition of the surfaces for both 2D and 3D samples to be smooth and matte and ensured that both objects have the same diameter, color, and viewing and illumination conditions. The images of sample objects were produced using the Lambertian model, in which the viewing direction and the illumination direction were set to the front of the objects and the light source was D65, as shown in Fig. 2.2. By controlling all these factors, the author verified that the difference in the appearance between the two samples is influenced by the surface shape and subsequent shading effect. The author also explored the shading effect for color naming by changing the illumination angle relative to the surface normal from 0° to 45°. Furthermore, it was hypothesized that the reaction time might depend on the surface shape. Therefore, the author also investigated the reaction time for 2D and 3D objects.



(a) 2D flat disk. (b) 3D sphere.

Figure 2.1. Real 2D and 3D objects painted with the same color paint of (L*,a*,b*)=(71, -48, 47) observed under the same viewing conditions.

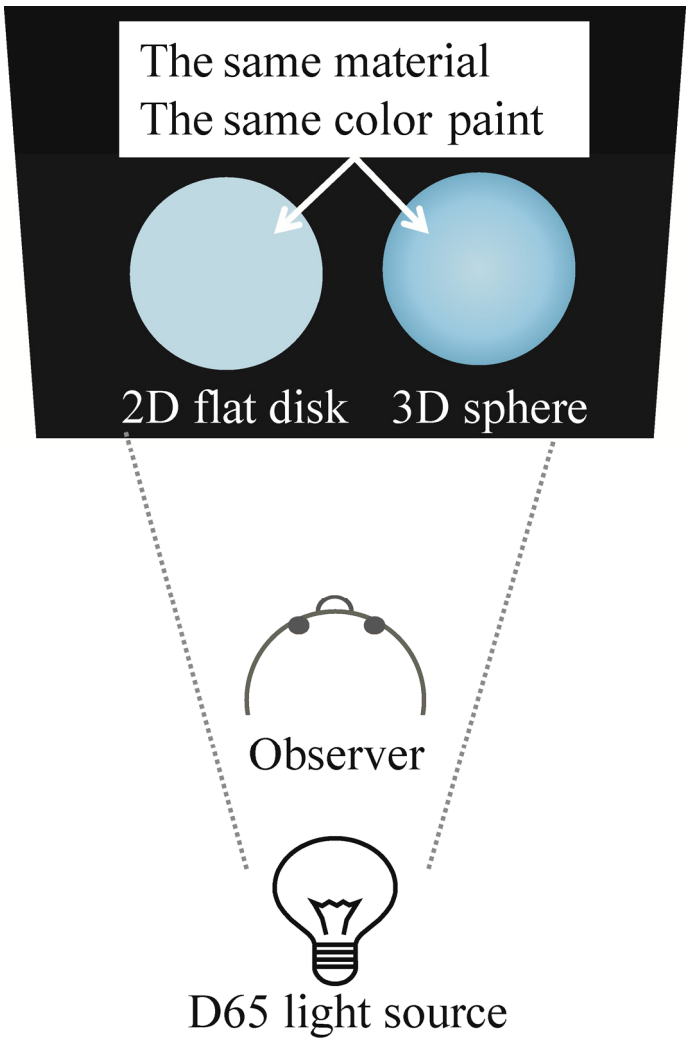


Figure 2.2. Observation condition of 2D and 3D objects set in this paper.

2.2 Experimental Methods

This subsection describes our experimental setup to investigate color terms and reaction times for 2D and 3D color samples.

2.2.1 Basic Color Terms in Modern Japanese

A set of 15 basic color terms that appear in modern Japanese was used in the color naming experiments in this study. These color terms were identified from a previous study reported in 2010 (Tominaga, Ono & Horiuchi, 2010). In the vocabulary test, the participants were asked to write more than 20 color terms that were commonly encountered in daily life on a paper sheet distributed to them. The color terms were recorded, and their frequency was statistically evaluated. The important color terms identified were Shiro (white), Kuro (black), Aka (red), Midori (green), Ki (yellow), Ao (blue), Cha (brown), Murasaki (purple), Pinku (pink), Orengi (orange), Hai (gray), Mizu (turquoise), Ki-midori (yellow-green), Kin (gold), Gin (silver). It was noted that the first 11 color terms were identical to the 11 basic color terms suggested by Berlin and Kay.

2.2.2 Image Rendering of Objects

The Lambertian reflection model, which is often used as a light reflection model with only the diffuse component discounting the gloss effect, was used to render the 2D and 3D color samples. Lambert's law states that the amount of reflected light is proportional to the cosine of the angle of incidence. The spectral radiance distribution $Y(\theta, \lambda)$ from a sample's surface is then described as a function of the wavelength λ and the angle of incidence θ as follows:

$$Y(\theta, \lambda) = \alpha \cos\theta S(\lambda) E(\lambda), \tag{2.1}$$

where α, $S(\lambda)$ and $E(\lambda)$ represent the reflection coefficient, the surface-spectral reflectance and the illuminant spectral-power distribution, respectively. The author set the illuminant $E(\lambda)$ to be D65.

2.2.3 System Construction

The display device used in this study was an EIZO ColorEdge CG221. Fig. 2.3 shows the measured color gamut of the display in the CIELAB color space. The tristimulus values of white origin on the display are (*Xw*,*Yw*,*Zw*) = (94.77, 100.0, 109.1), measured by using a spectroradiometer (Photo Research Inc., PR655), of brightness 156 cd/m^2, and the tristimulus values corresponded to (L*,a*,b*) = (100,0,0) in the CIELAB color space.

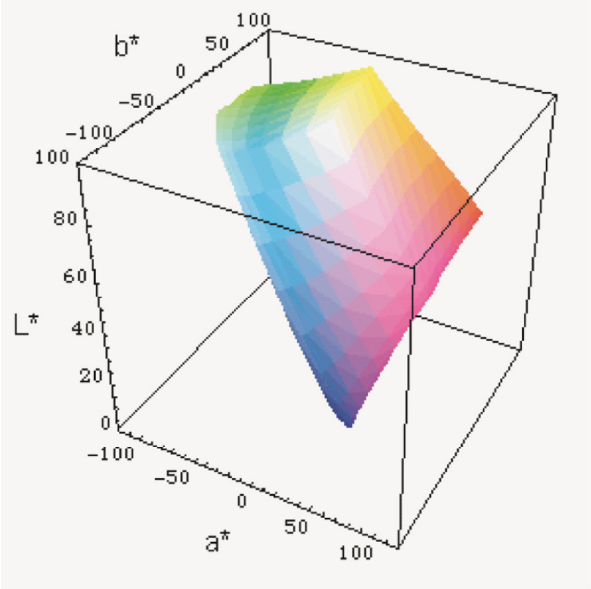


Figure 2.3. Display color gamut.

The display had a wide color gamut that reproduced 98% of the colors displayed using the Adobe RGB model. The author could produce 218 test colors at grid points within the display gamut, where the grid points were sampled in steps of 20 along each

of L*, a*, and b* axes, so that L* = 10, 30, 50, 70, 90; a* = … , -20, 0, 20, … ; and b* = … , -20, 0, 20, … Fig. 2.4 shows all 218 sampled color points in the CIELAB color space.

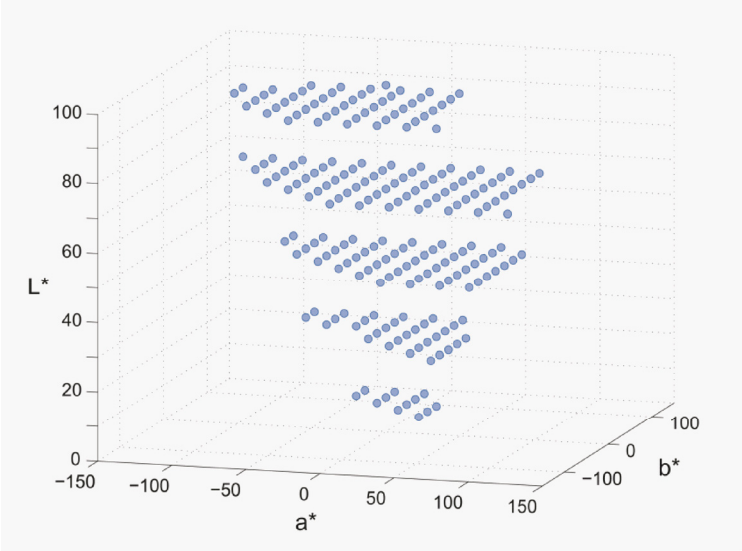
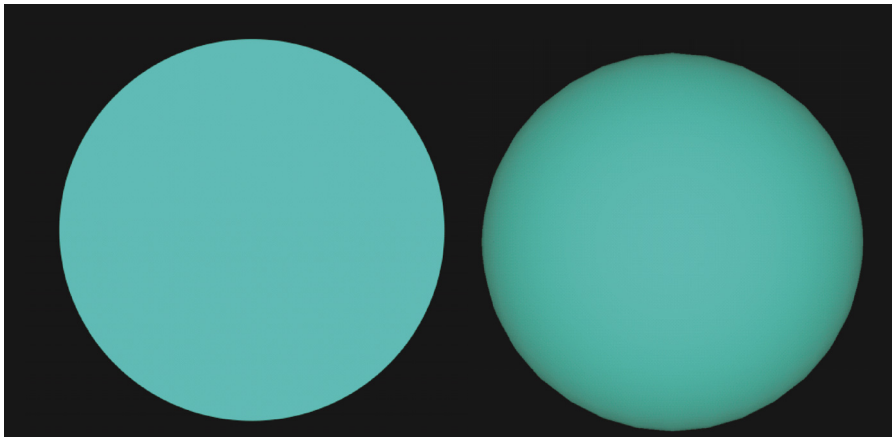


Figure 2.4. 218 sample color points in the CIELAB color space.

The author examined the rendered color images on 2D disks and 3D spheres, for different illumination directions. First, the author produced two sets of rendered color images at the sample points. Set-1 was a set of 2D disks, and Set-2 was a set of 3D spheres with shading effect caused by the diffuse reflection component. For both sets, the light was set on the viewing axis. Since all color sample images were rendered using the Lambertian model, the brightness changed according to the angle of incidence. The brightness decreased with increasing distance from the center of the surface of the 3D sphere, and the shading effect was introduced.

Figure 2.5 shows an example of a pair of rendered images. The author noted the same surface-spectral reflectance $S(\lambda)$ for the 2D and 3D objects, as described by Eq.(1). This means that in Fig. 2.5(b) the color values at the center region of the sphere image were the same as those of the disk image Fig. 2.5(a). Since the observation conditions

were identical for both objects, the differences in the appearance between the two sample images would be due to the shading effect introduced by the surface shape.



(a) Set-1 (disk). (b) Set-2 (sphere).

Figure 2.5. Example of a pair of rendered color samples.

The author designed to a color term collection system based on the above principle. Figure 2.6 shows a screen shot of the system. The distance between the display device and the subject was about 700 mm. A color sample of a diameter 145 mm (viewing angle was about 12°) was displayed in the center of the screen, and the 15 basic color terms were displayed in Japanese, on the right-hand side. In our preliminary experiment, stable responses were obtained by displaying 15 color terms as radio button graphics on the right. As pointed out, a gray/white reference may have affected the results. Therefore, by giving the subjects a large stimulus at a viewing angle of 12° and directing them to pay attention only to the stimuli, the author hypothesized that the influence of reference might be reduced to insignificance.

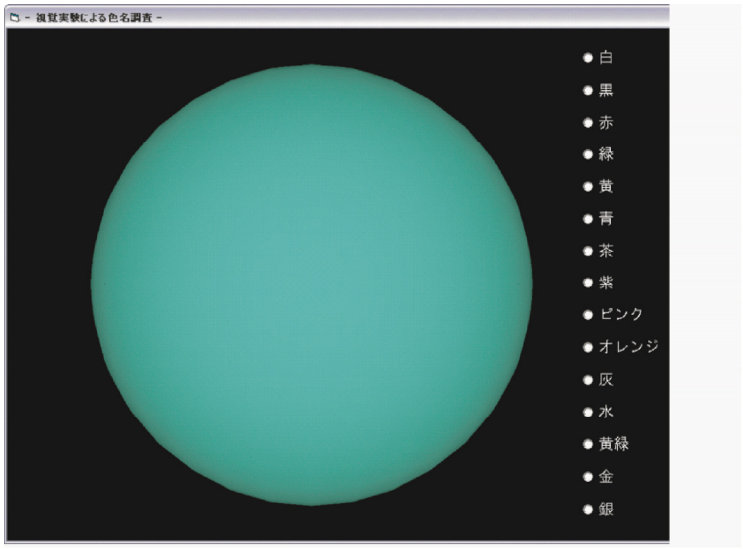


Figure 2.6. Screen-shot of a color term collection system.

2.2.4 Procedure

Color naming experiments were performed using the 2D (Set-1) and 3D (Set-2) color sample images. Each color sample was randomly selected from the set of 218 samples and displayed against the black background. The experiment was conducted with 10 participants in a darkroom. To eliminate the effects of gender and age, all participants were native Japanese men in their early twenties.

A subject was guided into the experimental room and seated on the chair in front of a monitor used to display the color samples. No training tasks were provided. Each experiment started after the subjects were given 2 min. to adapt to the darkness. Since this study focused on the color difference on the display, only cone adaptation was performed. During the adaptation, the following instruction was provided: "Select the most appropriate color term for the displayed test color samples from among the 15 basic color terms on the right of the screen. Click the next button to move to the next test color sample. There are 218 test color samples for one session." The subject selected the most appropriate color term for the displayed test color samples. The system

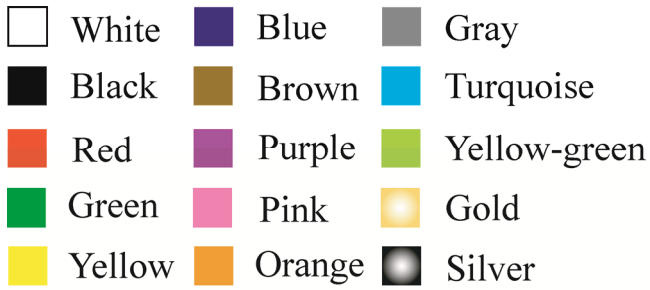
automatically recorded the color term and the reaction time, showing how long the subject took to determine the color term. All subjects had normal color vision, which was confirmed by using Ishihara plates.

2.3 Experimental Results

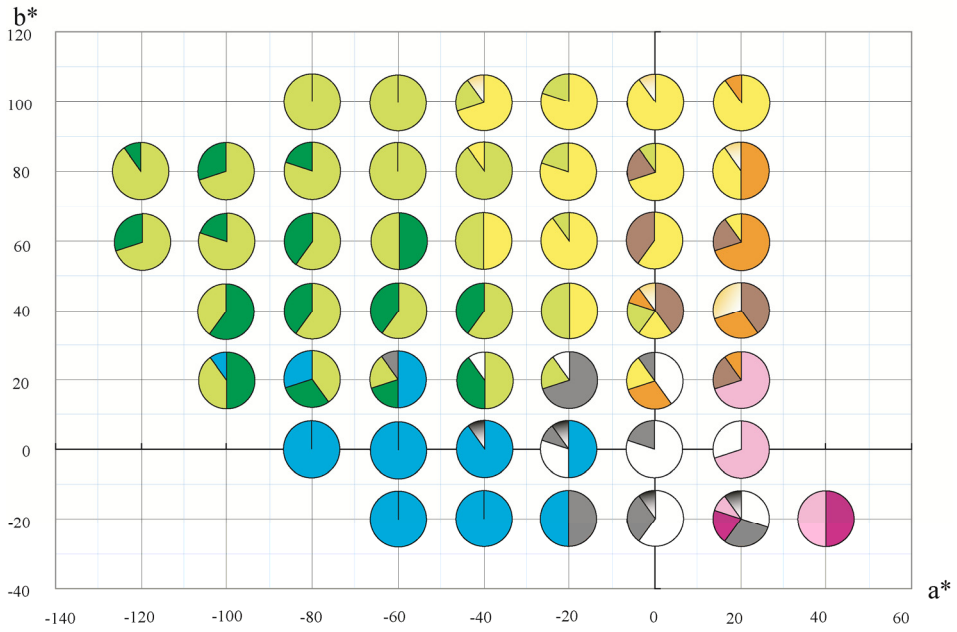
A typical session lasted about 30 min. The author analyzed differences in the reported color terms for 2D and 3D samples.

2.3.1 Distribution of the Color Terms

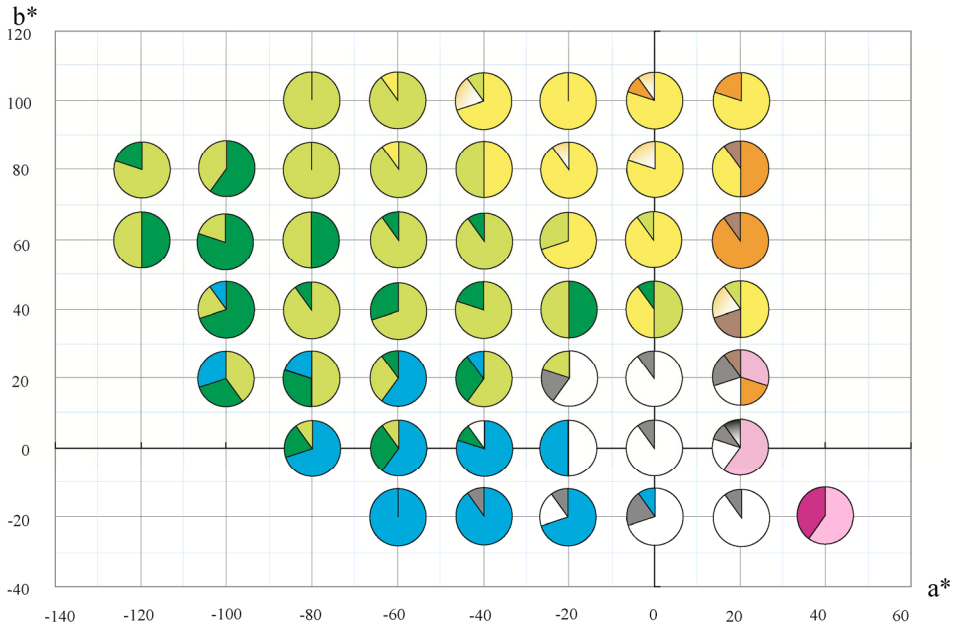
Figure 2.7 shows the distribution of the color terms for all 218 samples in the CIELAB color space. Each pie chart shows the ratio of the specified color terms for each sample placed at the grid points in the (a*, b*) plane. It can be seen that the color gamut is roughly segmented into 15 color term regions. The author compared the response to the 2D disk samples with a different color categorization that was reported in Ref. 20. Although the positions of color terms were dependent on L* in Fig. 2.7, in the case of L*=30, about 74% (23/31) of the samples were assigned to the same category. There seemed to be little difference in the color terms specified for the 2D and 3D samples. However, a transition of color terms between the 2D and 3D samples was observed at the boundary regions. Here, the boundary region is defined as the region surrounding the point where the maximum reported color term changes.



(a) Indices for modal color terms.

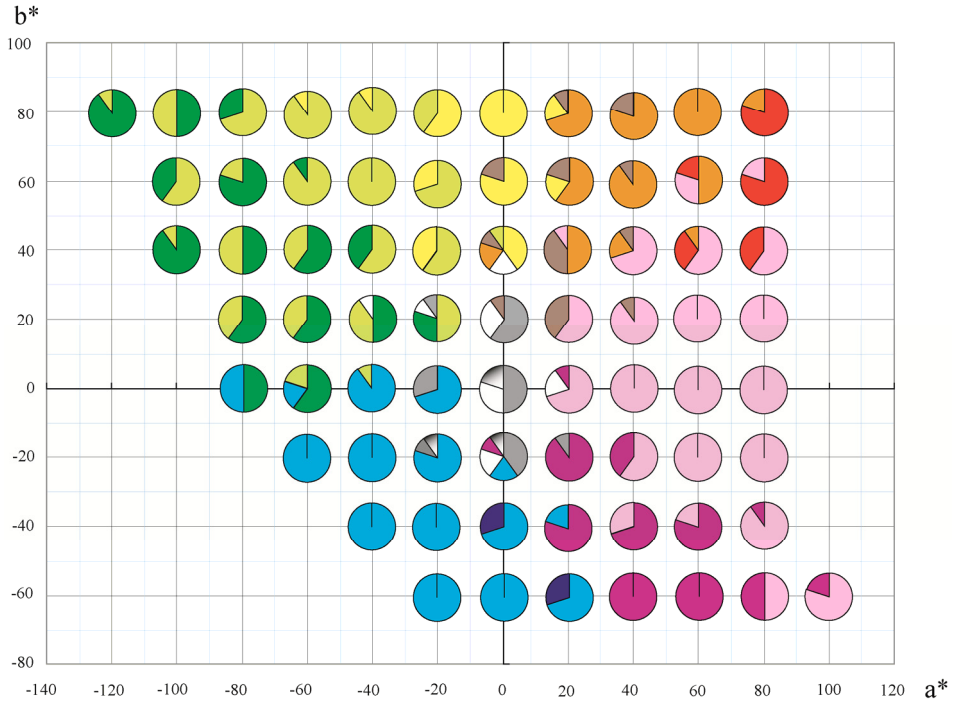


2D samples (Set-1)

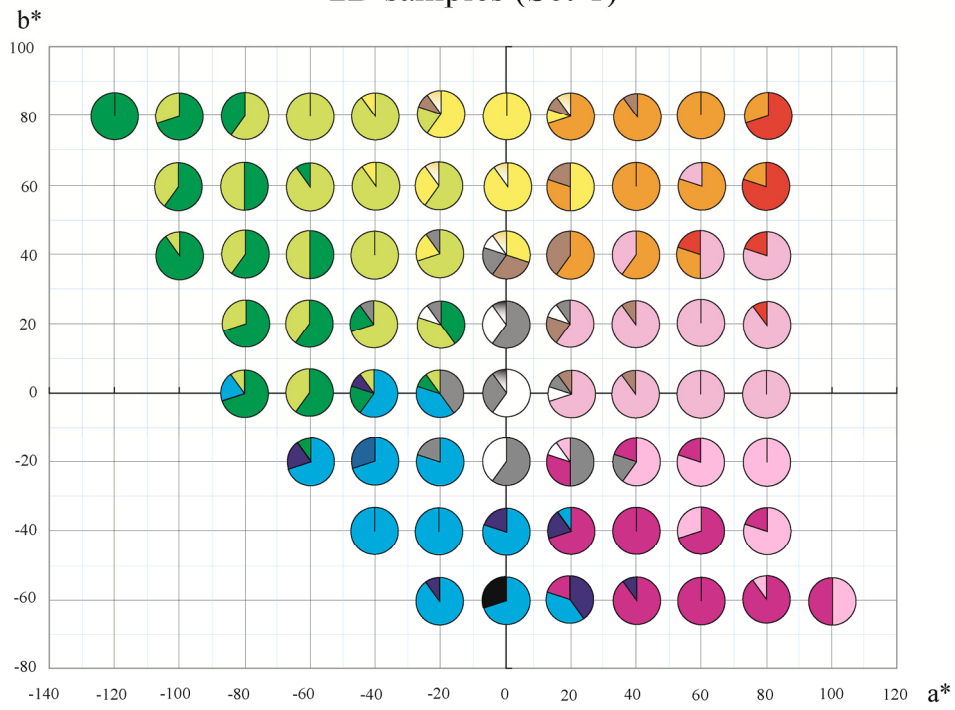


3D samples (Set-2)

(b) L*=90.

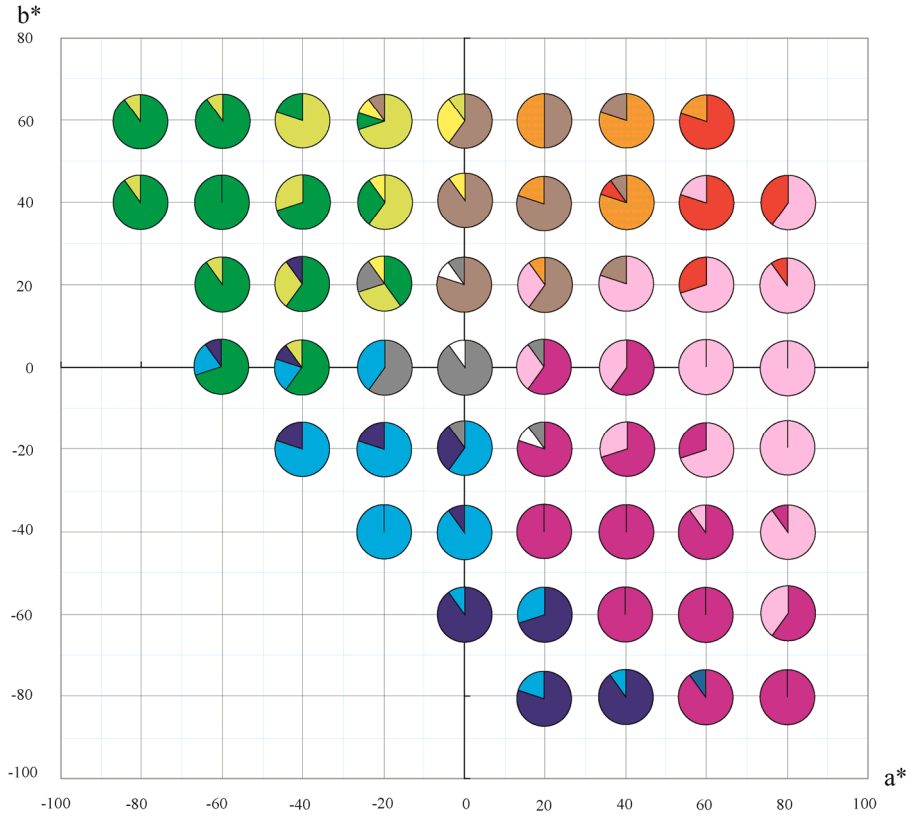


2D samples (Set-1)

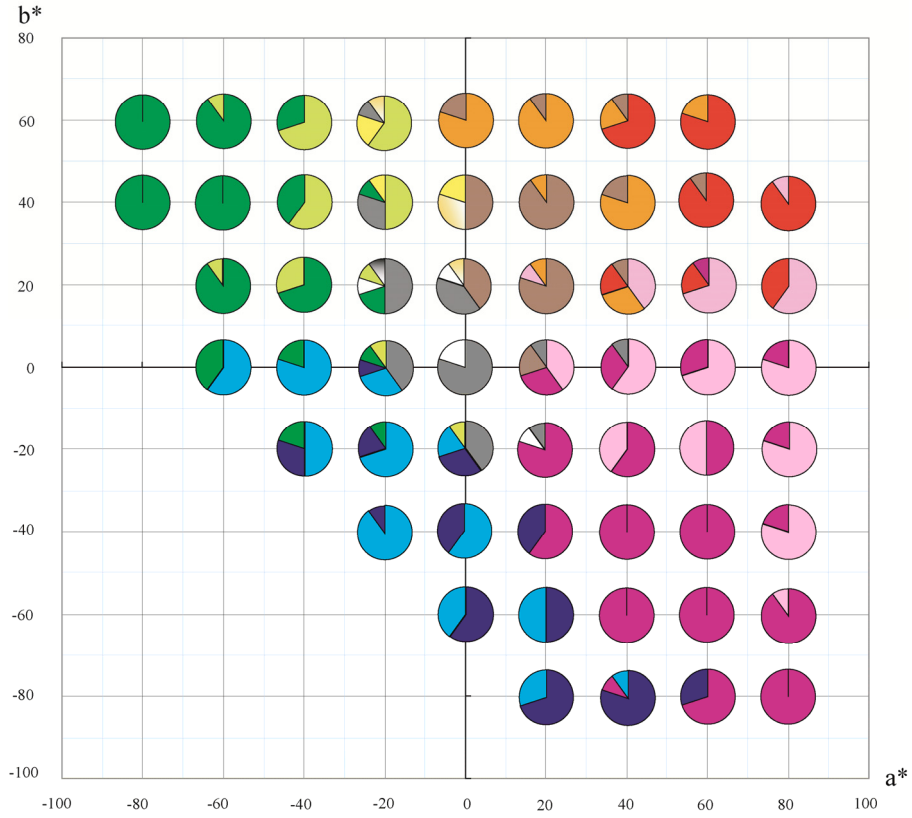


3D samples (Set-2)

(c) L*=70.

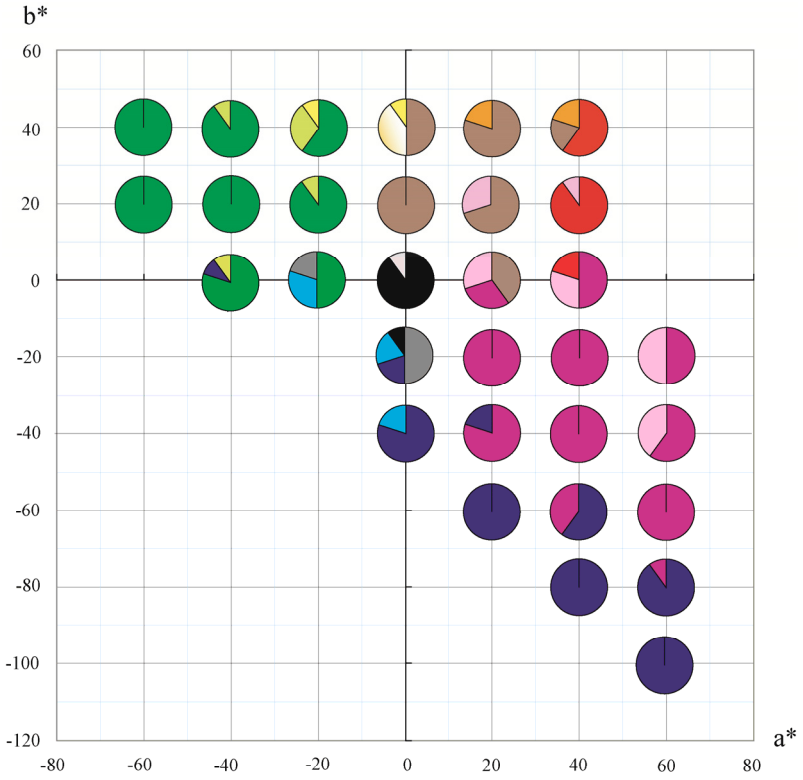


2D samples (Set-1)

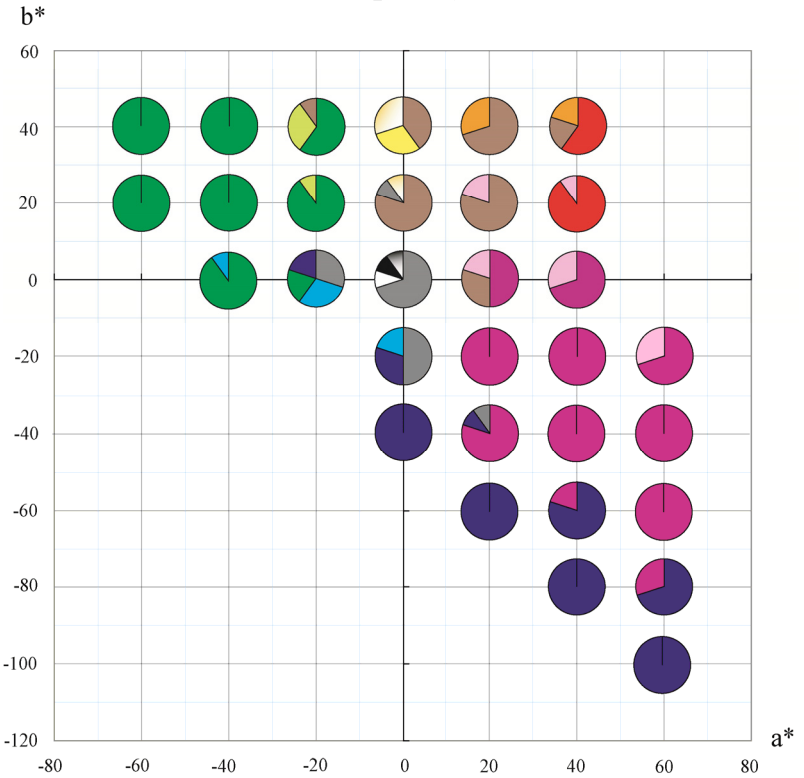


3D samples (Set-2)

(d) L*=50.

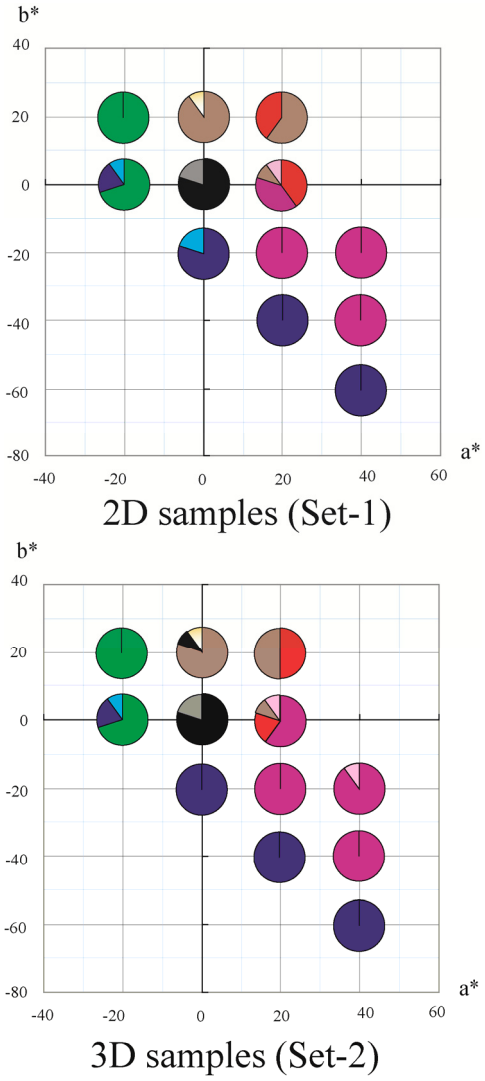


2D samples (Set-1)



3D samples (Set-2)

(e) L*=30.



(f) L*=10.

Figure 2.7. Distribution of the reported color terms.

Concerning the intra-observer variability, the author verified the repeatability for 11 subjects by using 135 color patches in the preliminary experiment. The average and minimum rates of repeatability were 90% and 82%, respectively. Owing to the experiment ethics guidelines of our university, an individual experiment session could not take longer than 30 min. Therefore, in our main experiment, stimuli were not used to check the repeatability. Regarding inter-observer variability, the maximal number of each reported color term was averaged across 218 samples. The numbers of 2D and 3D

samples were 7.62 and 7.53, respectively. This result shows that the variation in the reported color term among 10 subjects is almost identical between 2D and 3D samples.

2.3.2 Color Term Transition

First, the author investigated the color term transition of achromatic colors (a* = b* = 0). Figure 2.8 shows examples of the color term transition for achromatic color samples from 2D to 3D samples. The color term "black" was mostly chosen for the darkest 2D sample, while "gray" was chosen most often for the 3D sample painted with the same color. For achromatic colors, the author found that the subjects were likely to choose a brighter color name for the 3D samples than for the 2D samples. In the case of L* = 10, there was no difference in color name because both samples were too dark.

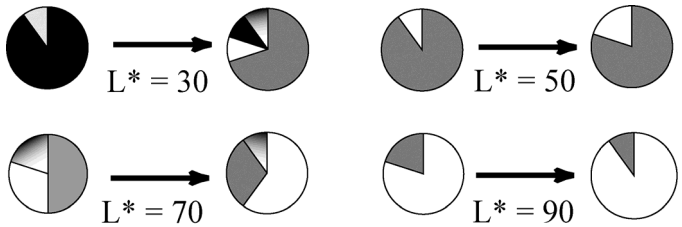


Figure 2.8. Examples of color term transition for achromatic color samples (a* = 0, b* = 0). Black gradation represents "Silver" color.

The author subsequently investigated the color term transition of chromatic colors having low saturation. In the low saturation condition, participants were found to choose achromatic color terms more often for 3D samples than for 2D samples, as shown in Fig. 2.9. The author also investigated the color term transition of other chromatic colors. Sometimes, 2D samples in the boundary regions of color names were named as color terms with low brightness. For instance, Fig. 2.10 shows the pie charts for two examples of high chrominance colors. In Fig. 2.10(a), the color term "pink" was chosen for the 2D

sample of (L*, a*, b*) = (70, 40, -40), while "purple" was chosen for the 3D samples of the same color. Note that the pie chart for the 3D sample is identical to the pie chart for the 2D sample of L* = {10, 30, 50}. In Fig. 2.10(b), the color term "turquoise" was chosen for the 2D samples of (L*, a*, b*) = (90, -60, 0), while "turquoise" and then "green" were chosen for the 3D samples. The pie chart of the 3D sample is close to the pie chart for the 2D sample of (L*, a*, b*) = (70, -60, 0). In these examples, the perception of "pink" generally tends to turn into "purple," and the perception of "turquoise" tended to turn into "green" and "yellow-green" when the brightness decreased. The author speculates that the subjects determined color names on the basis of the dark shading area of the 3D samples.

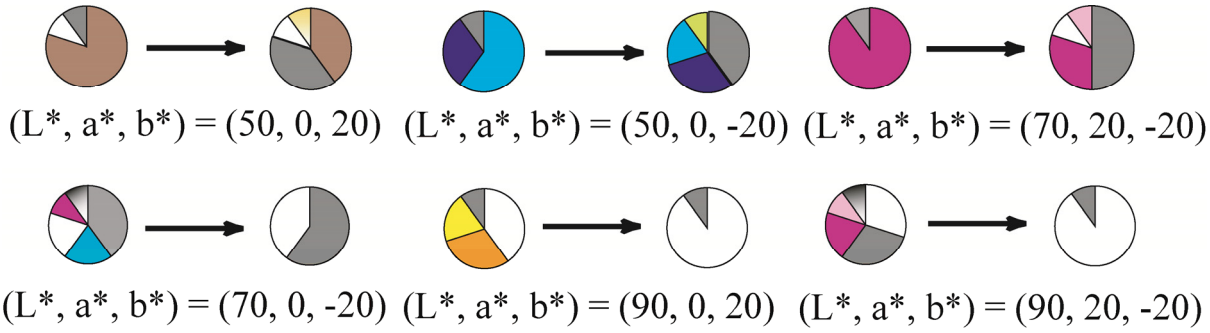
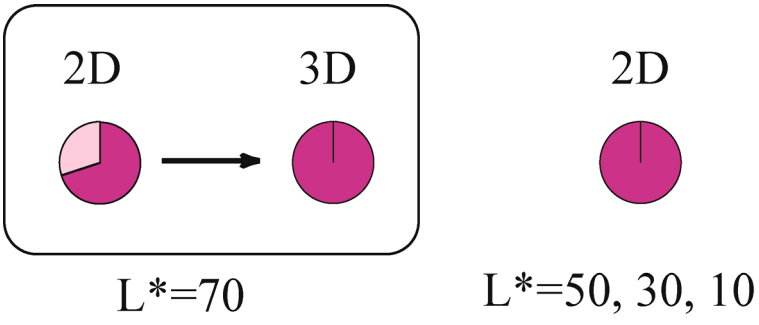
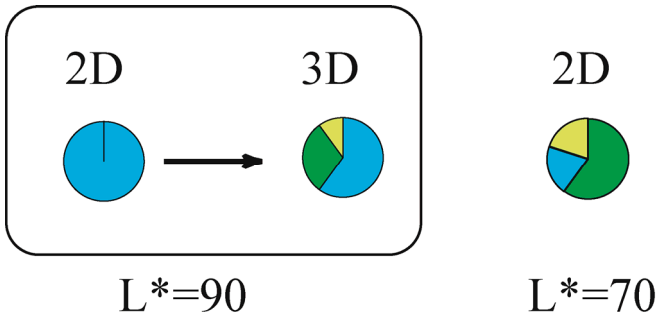


Figure 2.9. Examples of color term transition from a chromatic color term to an achromatic color term.



(a) (a*, b*) = (40, -40).



(b) (a*, b*) = (-60, 0).

Figure 2.10. Examples of color term transition for chrominance samples.

2.3.3 Frequency of Color Terms

Figure 2.11 shows the frequency of the color terms reported for each color sample. The black and white bars denote 2D and 3D samples, respectively. A difference in the number of color terms between the 2D and 3D samples was noted. In Fig. 2.11, the author divided the color terms into two groups. In the group to the left of the vertical dotted line, the color terms were used more often for the 2D samples rather than for the 3D ones. In contrast, in the group presented to the right of the vertical dotted line, the color terms were used more often for the 3D samples rather than for the 2D ones. In particular, the frequency of the terms “brown” and “pink” decreased from 2D to 3D, and the frequency of the terms “white” and “orange” increased. Detailed analysis of the color transitions suggested that “pink” changed to “red,” “brown,” and “purple,” while “brown” changed to “orange” and “yellow.”

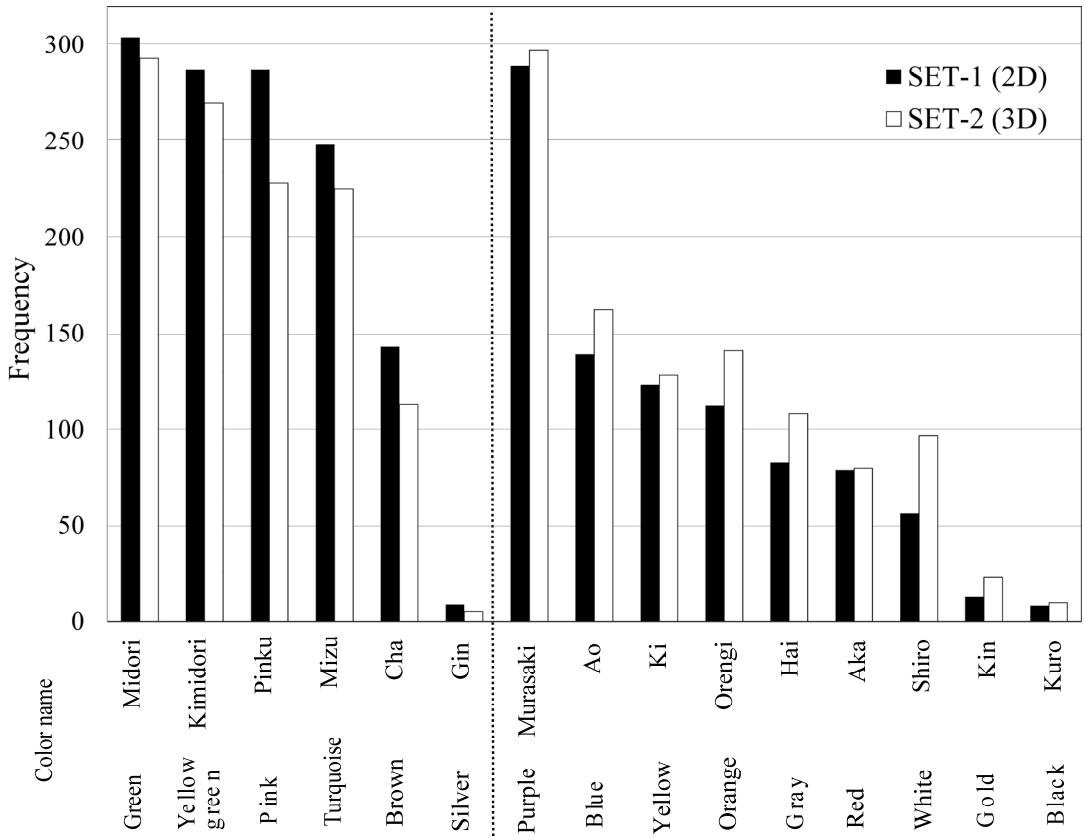


Figure 2.11. Frequency of the color terms.

The author further analyzed the reaction time, i.e., the time elapsed between the display of a color sample and selection of a proper color term for the sample. First, the Smirnov-Grubbs rejection test was used to exclude the outlier data for each color term. Then, the one-tailed t-test was performed to test for any significant differences between the 2D and 3D samples.

2.3.4 Reaction Time

Figure 2.12 shows the average reaction time for selecting each color term. The black and white bars denote the 2D and 3D samples, respectively. As shown in Fig. 2.12, the reaction time for selecting the color term “pink” was the shortest in both samples. The reaction time for selecting “purple” and “orange” were also short among all samples.

Conversely, the reaction times for selecting “gray,” “gold,” and “yellow” were long for all samples. Many of the short reaction times for color terms in the 2D samples were significantly shorter than the reaction times for the 3D samples. For “pink,” “purple,” “turquoise,” “yellow-green,” “brown,” and “green,” the reaction times for the 2D samples were shorter, and there was a statistically significant difference of 5% between the reaction times for 2D and 3D samples. Conversely, for “orange” and “silver,” the reaction times for the 3D samples were shorter, and there was a statistically significant difference of 5% between the reaction times for the 2D and 3D samples.

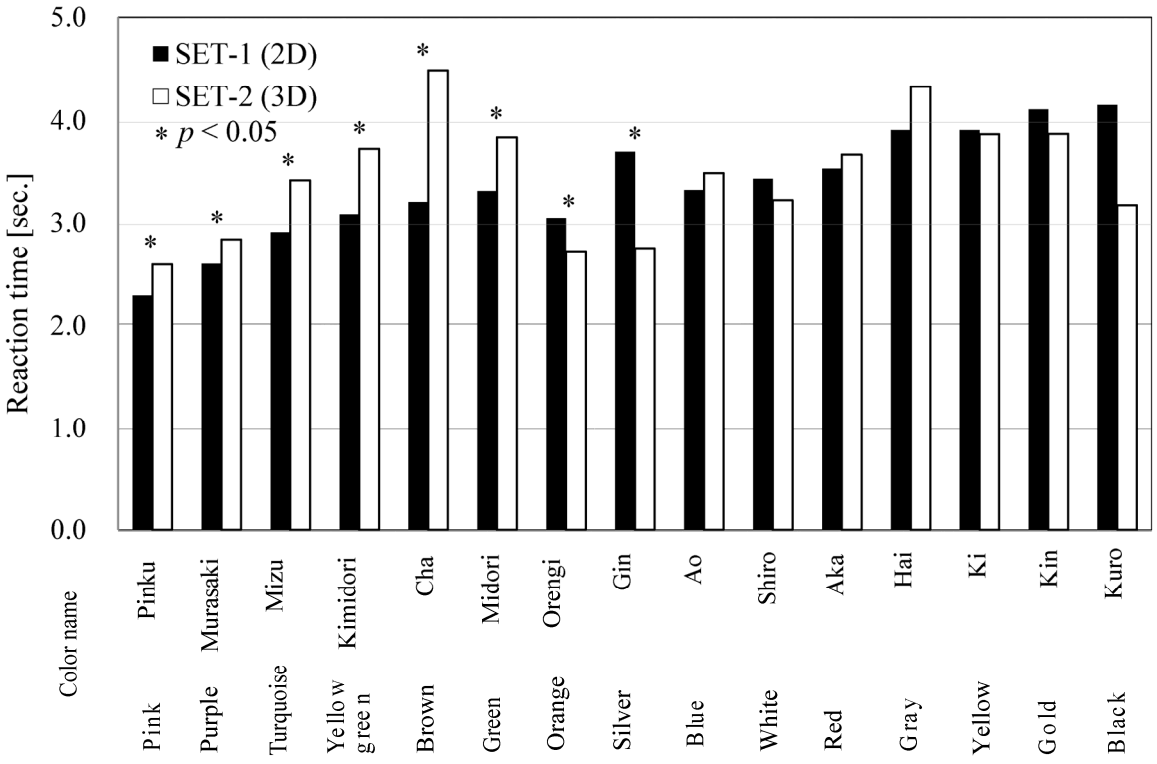


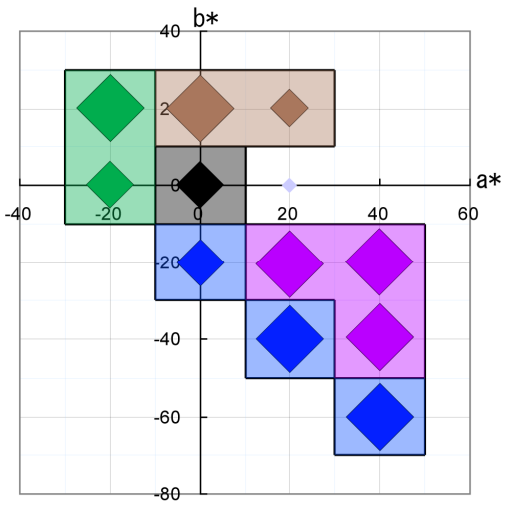
Figure 2.12. Average reaction time to identify each color term.

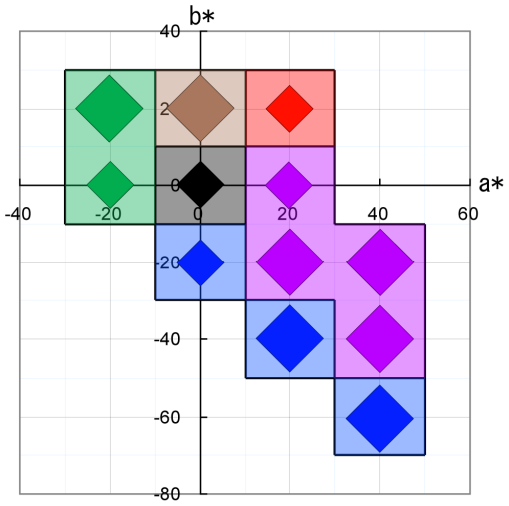
However, the average reaction time for the 3D sample sets was shorter than that for the 2D color samples. Human beings usually encounter 3D colored objects more often

than 2D colored objects in daily life. This may have effectively helped color naming for the 3D objects.

2.3.5 Modal Color Terms

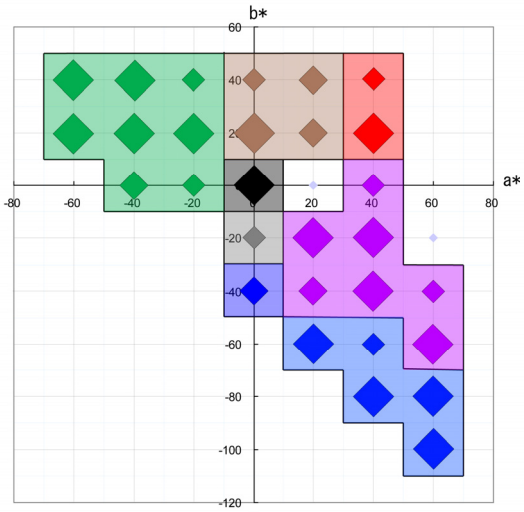
The author investigated the modal color term for each sample. Figure 2.13 shows the modal color terms. Each square in the figure is colored by the modal color term. In the case L*=10 (Fig. 2.13(a)), brown changed to red from the 2D to the 3D samples. In the case L*=30 (Fig. 2.13(b)), each color stabilize between 2D and 3D samples but black of 2D samples disappeared in the 3D samples. In the case L*=50 (Fig. 2.13(c)), pink changed to purple from 2D to 3D samples. In the case L*=70 (Fig. 2.13(d)), from the 2D to the 3D samples, orange changed to red and brown, pink changed to red and purple, and turquoise changed to gray. Interestingly, a new brown area appeared in the 3D result. In the case L*=90 (Fig. 2.13(e)), yellow-green changed to green from 2D to 3D samples. As shown in these figures, the author confirmed a definite change in the color area between 2D and 3D samples.

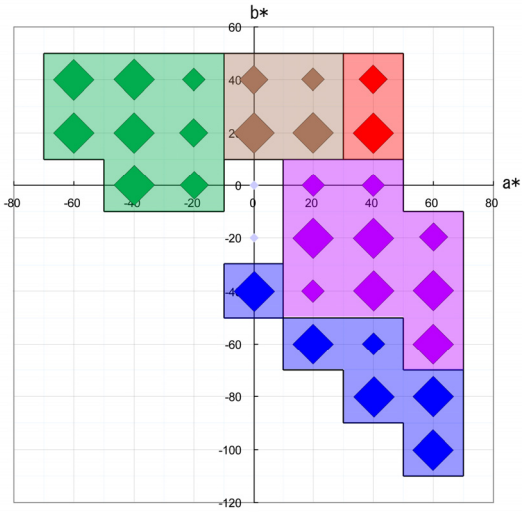




2D samples (Set-1) 3D samples (Set-2)

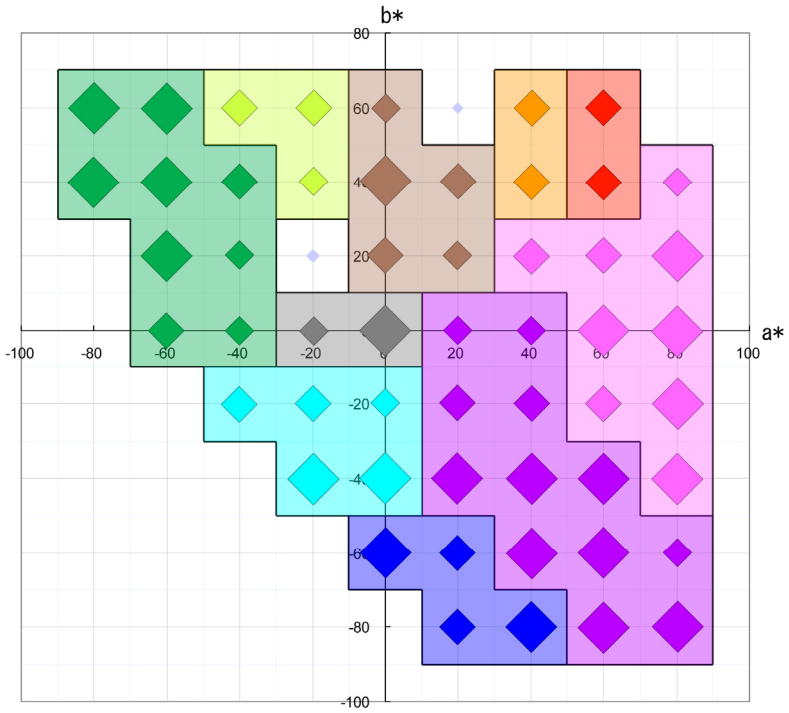
(a) L*=10.



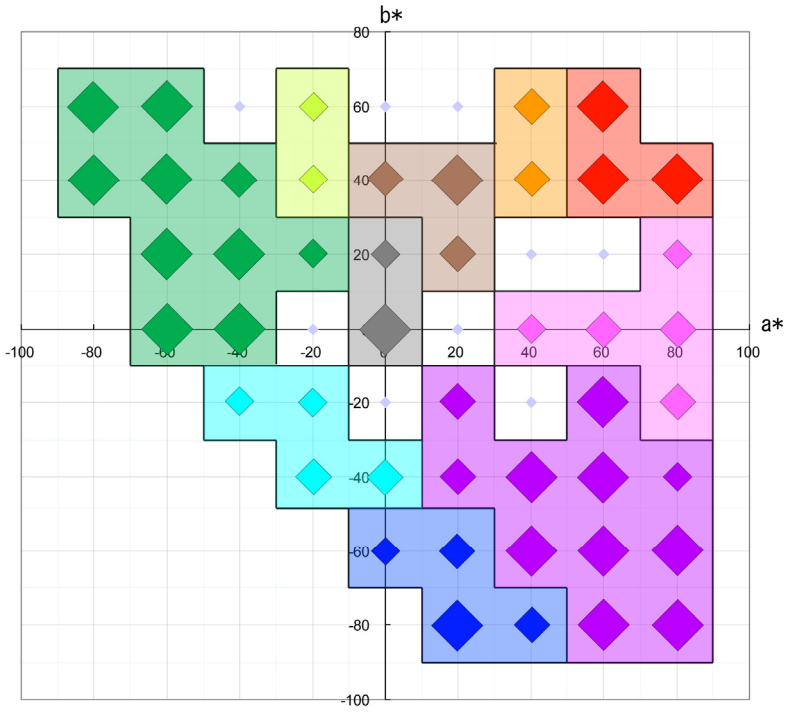


2D samples (Set-1) 3D samples (Set-2)

(b) L*=30.

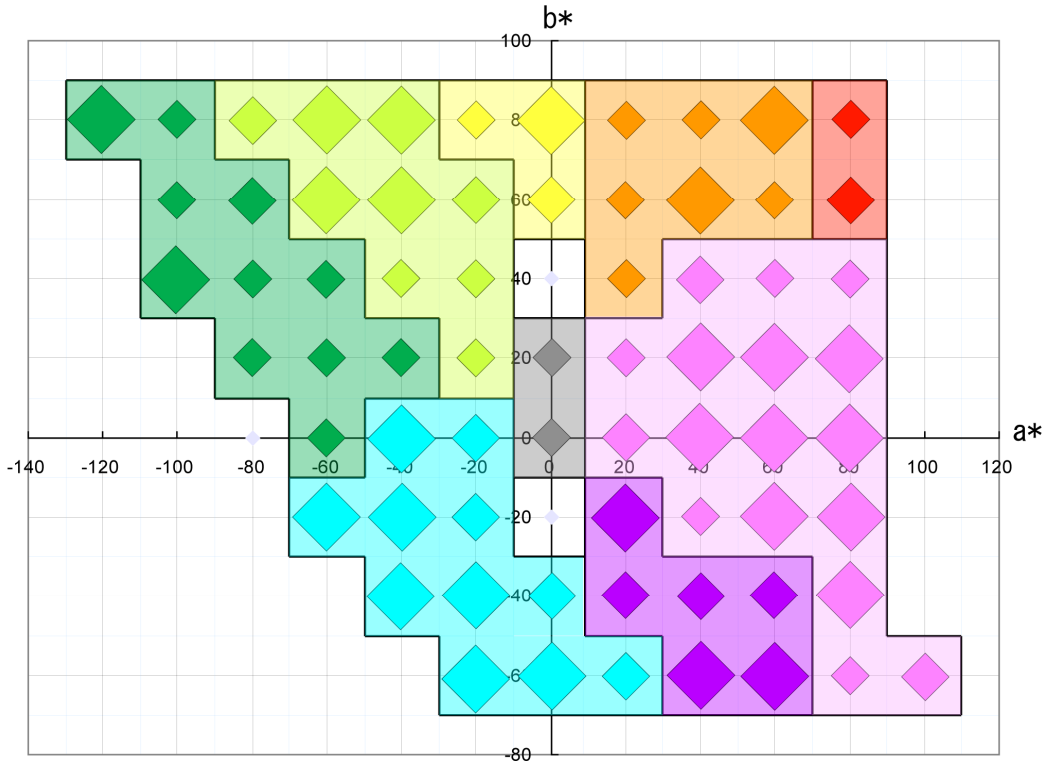


2D samples (Set-1)

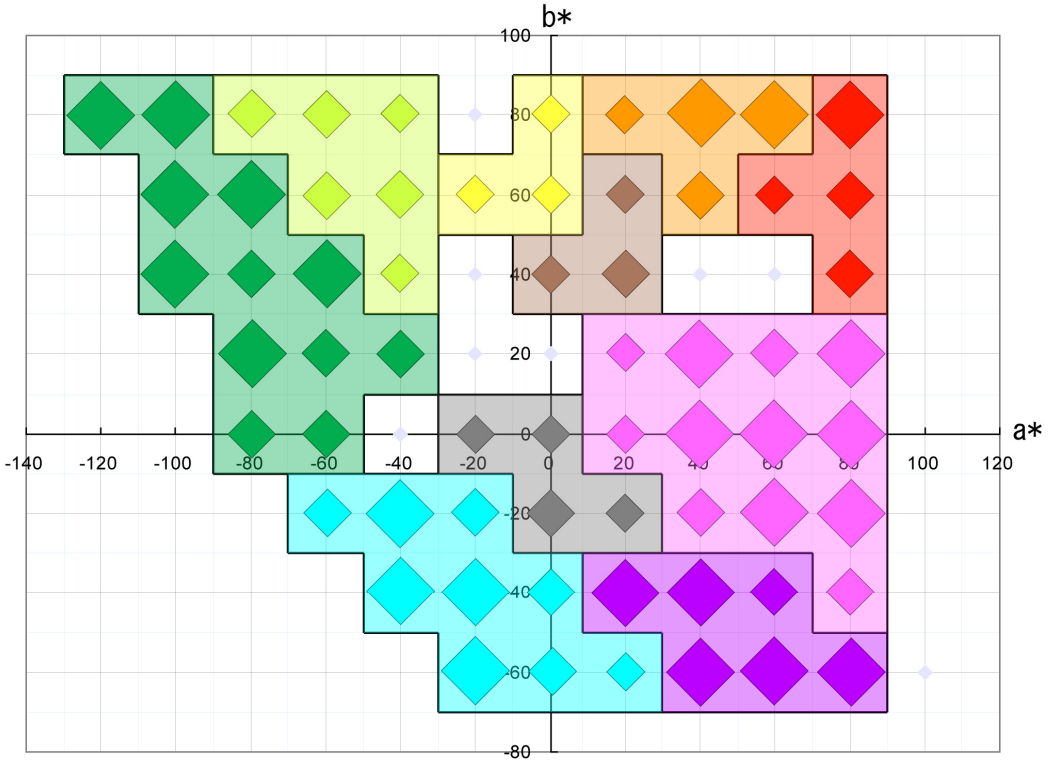


3D samples (Set-2)

(c) L*=50.

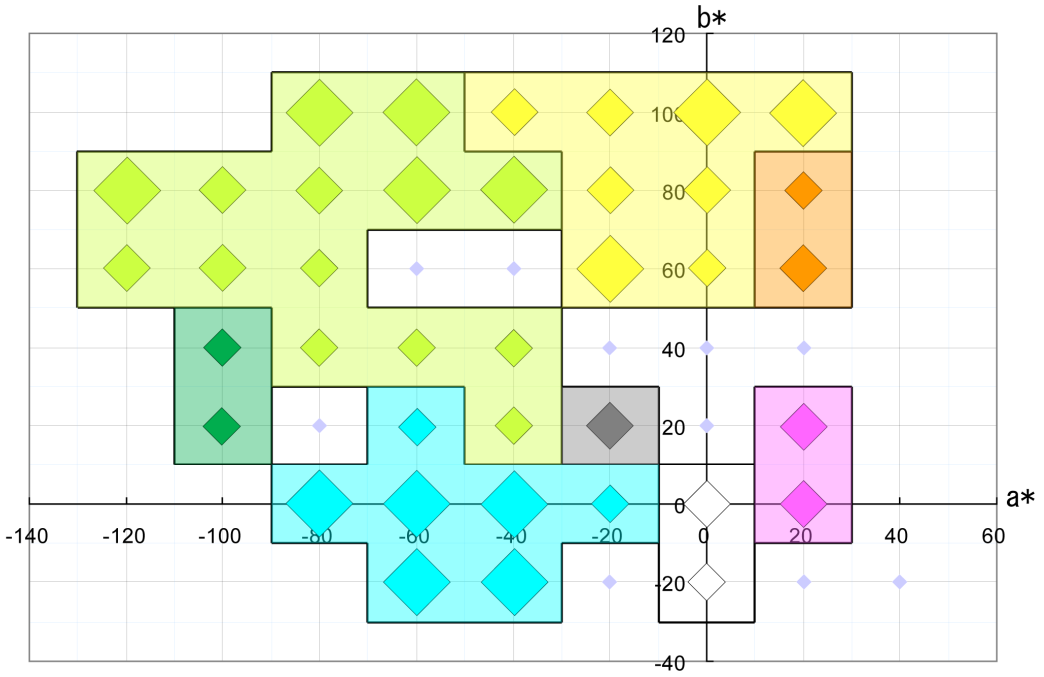


2D samples (Set-1)

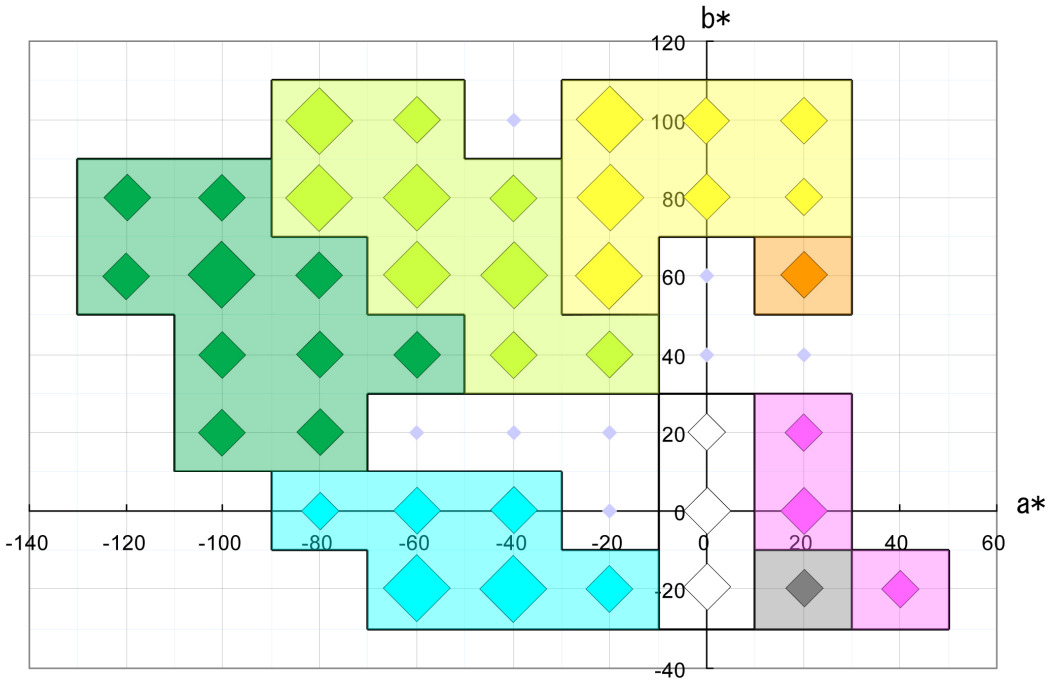


3D samples (Set-2)

(d) L*=70.

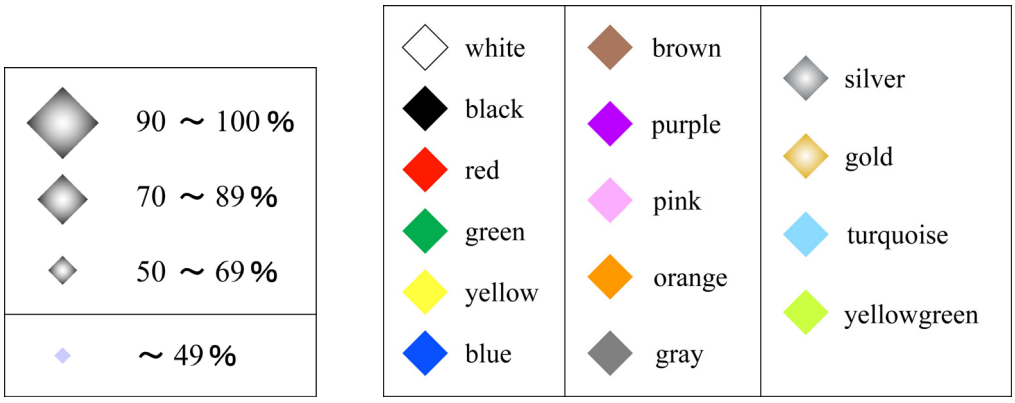


2D samples (Set-1)



3D samples (Set-2)

(e) L*=90.



(f) Indices for modal color terms.

Figure 2.13. Modal color terms.

2.4 Discussion

As described in previous section, shading effects seem to influence the color term transition between 2D and 3D rendered samples. In this section, the author discusses the shading effects by changing illumination conditions.

As described previously, shading effects cause differences of color categorization between the 2D and 3D rendered samples. In this section, the author discusses the shading effects under different the illumination conditions.

2.4.1 Position of Illumination

In our experiment, 3D samples were rendered by illuminating the object surface from an angle of 0° relative to the surface normal. The author changed the illumination angle to 45° relative to the surface normal and then rendered an additional 3D sphere sample, Set-2', composed of the diffuse reflection component with 45° incidence. Figure 2.14 shows an example of Set-2' in which the 3D sphere has a high contrast by strong shading effect, compared with the sample Set-2 shown in Fig. 2.5(b). The same psychophysical experiment was then conducted by using Set-2'.

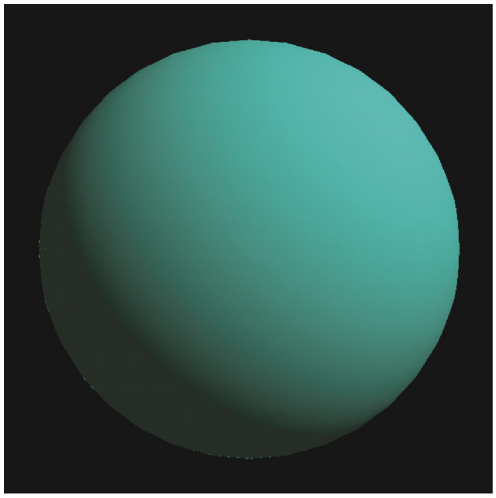


Figure 2.14. Additional 3D sample (Set-3).

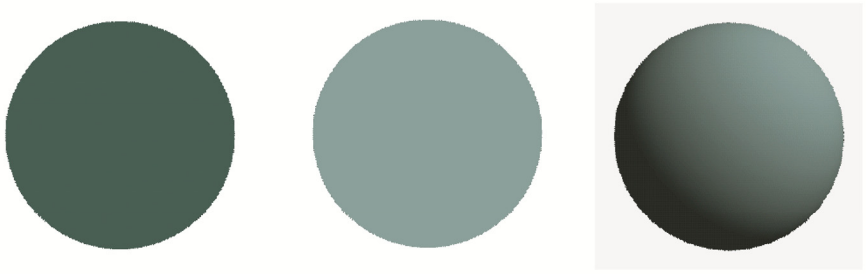
Table 2.1 summarizes the experimental results, where the frequency of color terms is listed. Interestingly, the frequencies for Set-2 were between those of Set-1 and Set-2'. In the first group from "green" to "black," the number of answers for Set-2 was greater than that for the 2D samples of Set-1, but less than that for the 3D samples of Set-2'. In the second group from the rows for "pink" to "orange," the number of answers for Set-2' was generally less than that for Set-1 and Set-2. Most of the color term occurrences for Set-2 existed between those for Set-1 and Set-2'. It may be possible that the color term shift toward darker terms occurred due to the shading effect of the 3D samples. Thus, it may be inferred that color categorization changes with different illumination positions directed on 3D surfaces (Set-2 and Set-3).

Table 2.1. Frequency of color terms.

Color Name	In English	2D(Set-1)	3D(Set-2)	3D(Set-2')
Midori	Green	304	293	390
Murasaki	Purple	289	297	304
Cha	Brown	143	113	170
Ao	Blue	139	162	143
Hai	Gray	83	108	117
Aka	Red	79	80	117
Kin	Gold	13	23	34
Kuro	Black	8	10	25
Pinku	Pink	287	228	209
Ki-midori	Yellow green	287	270	181
Mizu	Aqua	248	225	215
Ki	Yellow	123	128	111
Orengi	Orange	112	141	96
Shiro	White	56	97	62
Gin	Silver	9	5	6

2.4.2 Brightness Level

The author also investigated the relationship between color terms and the average brightness level of reflected light. In our experiment, the color of the 2D sample was the same as the brightest pixel in the 3D sample. Therefore, the average brightness level of the 3D samples was lower than that of the 2D samples when both samples had the same chromatic component. Figure 2.15(a) shows a 2D sample with the same average brightness level as a 3D sample. The normal 2D disk patch in Set-1 and the 3D rendered sample in Set-2' are shown in Figs. 2.15(b) and 15(c), respectively. All samples in Fig. 2.15 have the same chroma but different brightness levels. Nevertheless, each sample appears different in this figure when observed in surface mode.



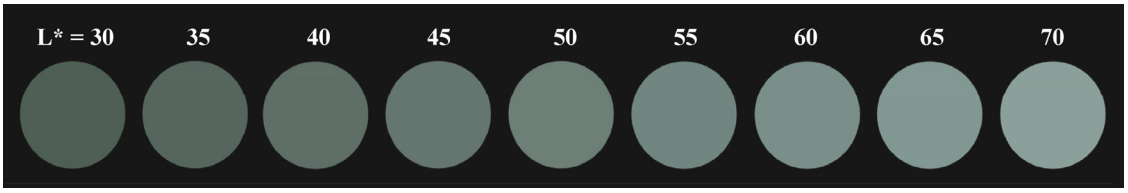
(a) (b) (c)

(a) 2D painted with the average brightness of 3D,
(b) Normal 2D (Set-1), (c) 3D (Set-2').

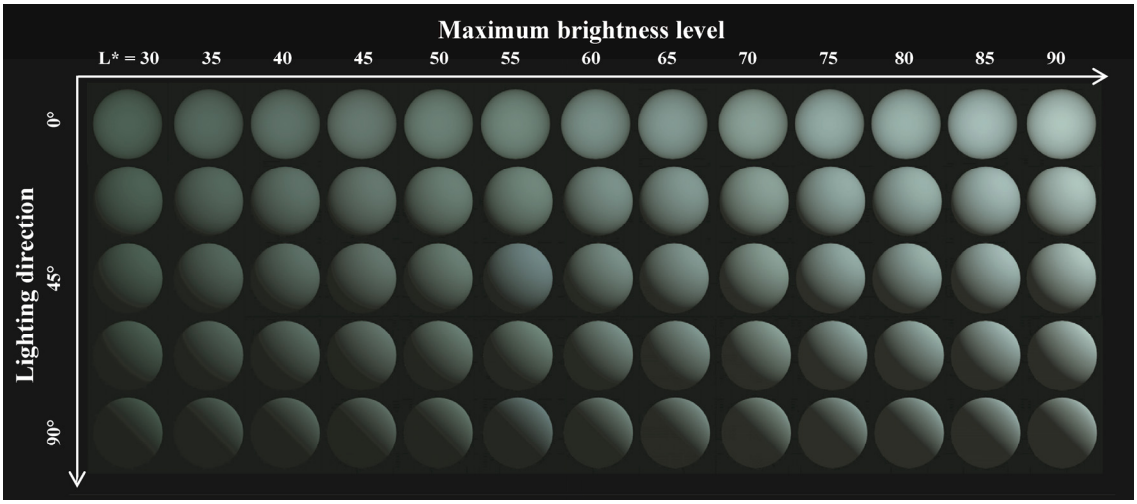
Figure 2.15. Samples for brightness comparison.

The author prepared many 2D samples with the same chroma but different brightness levels, as shown in Fig. 2.16(a). The author also prepared many 3D samples with the same chroma, (a*, b*) = (-20, 0), but illuminated under different lighting directions, as shown in Fig. 2.16(b). These 3D samples were rendered by using the Lambertian reflection model. The author calculated the average L* value for each surface and leveled each sample with the average brightness (L* average) in numerical form.

The same visual experiment was performed to collect color terms for the additional 2D and 3D samples presented in Fig. 2.16. The corresponding results in Fig. 2.17 show that the color of each sample corresponds to the modal color term, and each number in Fig. 2.17 denotes the L* value of each sample. Here, the author termed the most frequently reported color term for each color sample as the "modal color term." The 2D sample with L* = 30, as shown in Fig. 2.17(c), was recognized as "green," but the 3D samples with the same average brightness level did not always elicit this response. This indicates that samples with the same brightness level may be ascribed different color terms depending on whether they are perceived as 2D or 3D samples. Indeed, this interesting hypothesis merits further investigation.



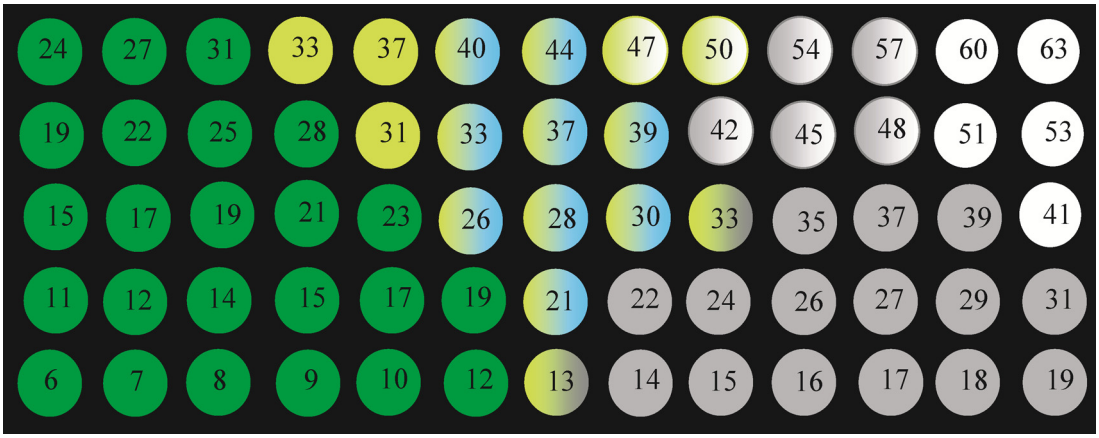
(a) 2D samples with different brightness level.



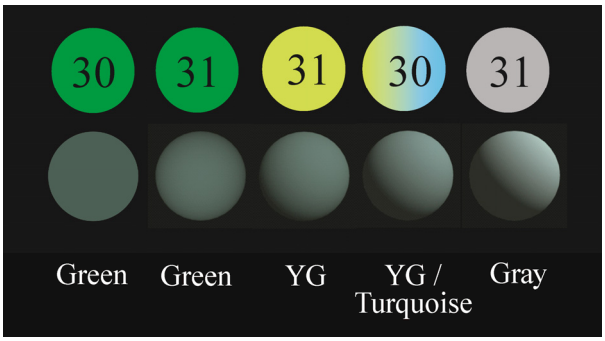
(b) 3D samples with different brightness levels and different lighting directions.

Figure 2.16. Additional rendered samples having the same chroma.

(a) 2D samples.



(b) 3D samples.



(c) Examples of samples with the same brightness level: (1st row) 2D disk patch: (2nd to 5th row) 3D samples illuminated from 0°-45° to the surface normal.

Figure 2.17. Color terms for samples in Fig. 2.16.

2.5 Conclusions

As a fundamental study of shitsukan perception for different viewing conditions, the author has focused on "color", and has described a color naming experiment using 2D and 3D rendered color samples on a calibrated display device. The 2D and 3D samples had matte surfaces with the same object size and color, which were obtained under the same viewing angle and illumination. The scene images of sample objects were produced using the Lambertian reflection model, in which the viewing and illumination directions were towards the front of the objects and the light source was D65. The difference between the appearances of the two types of samples was then based primarily on the shading effect. The author subsequently analyzed the relationship between color terms and object surfaces.

Our analyses yielded the following important findings:

(1) Brighter color terms tend to be chosen more often for 3D samples than 2D samples when observing achromatic colors;

(2) Achromatic color terms are chosen for 3D samples having low saturation;

(3) For chromatic colors, darker color terms are generally chosen for 3D samples in comparison to the corresponding 2D samples of the same color.

Further, the author investigated the color term transition by changing the illumination conditions. An experiment was performed by using a 3D sphere with a shading effect by a diffuse reflection component illuminated at 45° relative to the surface normal. Interestingly, changing the illumination position changed the color terms and the average brightness level of the sample. However, the samples with the same brightness level were ascribed different color terms depending on whether they were 2D or 3D. Furthermore, an analysis of the reaction time showed that the reaction time for "pink" was the shortest for both types of samples. The reaction time for six color terms was significantly shorter for the 2D samples than for the 3D samples, and the reaction times for two color terms were significantly longer for the 2D samples than for the 3D ones. These results show that humans perceive color differently for 2D and 3D objects and that they identify the same color using different color terms.

To avoid the influence of gender and age, all participants of our experiment were native Japanese men in their early twenties. The author will perform similar investigation for different gender and age groups. Although 3D stimuli seemed to be perceived as a surface color mode in our experiment, 2D stimuli with a high luminance value L* might be perceived in an aperture color mode. The relationship between the appearance mode and color naming is an interesting issue. Furthermore, the author will perform visual experiments using 3D color samples with large variations in shape, illumination direction, and colored light such as non-Lambertian 3D objects.

From the above investigation, the importance of the shitsukan analysis of a material appearance using real-world objects and rendered images was confirmed in order to understand the mechanism behind shitsukan.

Chapter 3. Investigating Perceptual Qualities of Static Surface Appearance

3.1 Introduction

In everyday life, we can distinguish object categories without difficulty by recognizing different shapes and the functions of the objects based on visual information. For example, a rocking chair and a sofa can be clustered within the category "chair" if they are grouped according to the function of "sitting down"; such grouping enables precise discrimination and can be used in the field of computer vision (Andreopoulos & Tsotsos, 2014; Prasad, 2012). Building on this and the rich scientific information, recent studies have been undertaken on material perception, which contributes to the perception of objects.

In the field of computer vision, most object recognition systems have relied on low-level material invariant features such as color; for example, the Scale Invariant Feature Transform (SIFT: Koenderink & Doorn, 1987; Lowe, 2004) or the Histogram of Oriented Gradients (HoG: Dalal & Triggs, 2005) have tended to ignore material information altogether. The Bidirectional Reflectance Distribution Function (BRDF: Debevec et al., 2002; Marschner et al., 2005), the Bidirectional Texture Function (BTF: Dana et al., 1999) and the bidirectional surface scattering reflectance distribution function (Jensen et al., 2001) seem to be trivial features for representing surface properties, depending on the materials. However, it is nearly impossible to estimate such features from a single image without employing simplifying assumptions (Debevec, 2000; Dror, Adelson & Willsky, 2001). Recently, a few approaches have been proposed in order to directly study the relations between image features and several perceptual attributes and to estimate the attribute values for a given image (Abe, Okatani & Deguchi, 2012; Dror, Adelson & Willsky, 2001; Varma & Zisserman, 2009; Liu et al, 2010;) this research was performed using a large image dataset such as the Flickr

Materials Database (Sharan, Rosenholtz & Adelson, 2009).

Several studies have investigated the mechanism of how material sensations are processed in the human brain using various approaches such as fMRI (Hiramatsu, Goda & Komatsu, 2011) and psychophysical studies (Motoyoshi et al., 2007). Most of these studies have focused on the visual estimation of the specific properties of materials (Anderson, 2011; Thompson, 2011; Zaidi, 2011), such as glossiness (Fleming, Dror, & Adelson, 2003; Motoyoshi & Matoba, 2012; Nishida & Shinya, 1998), translucency (Fleming & Bülthoff, 2005, Fleming, Jäkel, & Maloney, 2011; Motoyoshi, 2010), or roughness (Padilla, et al., 2008; Pont & Koenderink, 2005; Pont & Koenderink, 2008). Taken together, these findings support the general idea that the human visual system can estimate the properties of materials from relatively low-level vision features.

There is experimental evidence to support the hypothesis that human observers excel at recognizing and categorizing materials. For example, Sharan, Rosenholtz, and Adelson (2009) have shown that participants can identify a wide range of materials from photographs, even after a very brief exposure. Recently, Fleming, Wiebel, and Gegenfurtner (2013) showed participants photographs of materials from different categories and asked them to rate various subjective qualities, such as hardness, glossiness, or prettiness. Although the participants were not explicitly informed that the samples belonged to different classes, the subjective ratings of the samples were systematically clustered into categories, suggesting that the participants could theoretically classify materials by making visual judgments concerning their properties. This study explains, for example, that we can judge qualities such as the hardness or softness of an object from visual information alone, rather than through touching.

As has been shown by many previous studies, both surface property (i.e., color, texture, surface reflectance, etc.) and shape are influential in distinguishing materials by

providing relevant visual information. Furthermore, as has been shown previously (Doerschner et al., 2011; Muryy et al., 2013; Wendt, Faul & Mausfeld, 2008), judgments of specular reflectance are affected by both binocular disparity and motion information. Many studies have analyzed this by presenting the two stimuli simultaneously.

In the current study, which was inspired by the study of Fleming et al. (2013), the author investigated the relationship between material categories and *perceptual qualities*. Fleming et al. used test stimuli with both surface and shape information. In daily life, humans successfully discriminate materials using both types of information. However, it is scientifically relevant to separately investigate the influence of each type of information on perceptual qualities. For the current study, the author eliminated shape information and investigated only perceptual qualities obtained from the static surface properties. Moreover, the author also eliminated saturated color exemplars to avoid the influence of color deviation. The test stimuli consisted of 34 exemplars obtained from 10 different materials. The participants rated nine subjective properties for each material.

The author further investigated the relationship between perceptual qualities and image representations. The majority of conventional approaches used photographs or synthesized images. However, the influence of the representation, such as its resolution and color reproduction, has not been fully considered. To this end, the author conducted four additional experiments, in which the observation conditions (i.e., stimuli's sizes and luminance intensity) were preserved, but the chrominance components and resolution changed. By comparing the responses obtained in the five experiments, the author analyzed the perceptual qualities based on the static surface appearance of materials under different viewing conditions.

3.2 Experimental Stimuli

This subsection describes stimuli used in our experiment to investigate perceptional qualities of static surface appearance.

3.2.1 Material Dataset

To investigate material perception not influenced by shape, the author produced a dataset of 34 exemplars (size 50 × 50 mm). The individual images were selected from 10 material categories —stone, metal, glass, plastic, leather, fabric, paper, wood, ceramic, and rubber— to cover a wide range of appearances for each material. Each category contained two or more exemplars. All materials and their specifications are shown in Fig. 3.1 and Table 3.1, respectively. As described by Albertazzi & Hurlbert (2013), color has a strong influence on perceptual qualities. Since it is difficult to collect uniform material exemplars of various hues, the author collected exemplars with only low saturation. In Fig. 3.2, the symbol "×" represents the location of each exemplar on the CIE xy chromaticity diagram. Since the various exemplars were collected according to the differences in their surface properties, the number of exemplars per material category was uneven.

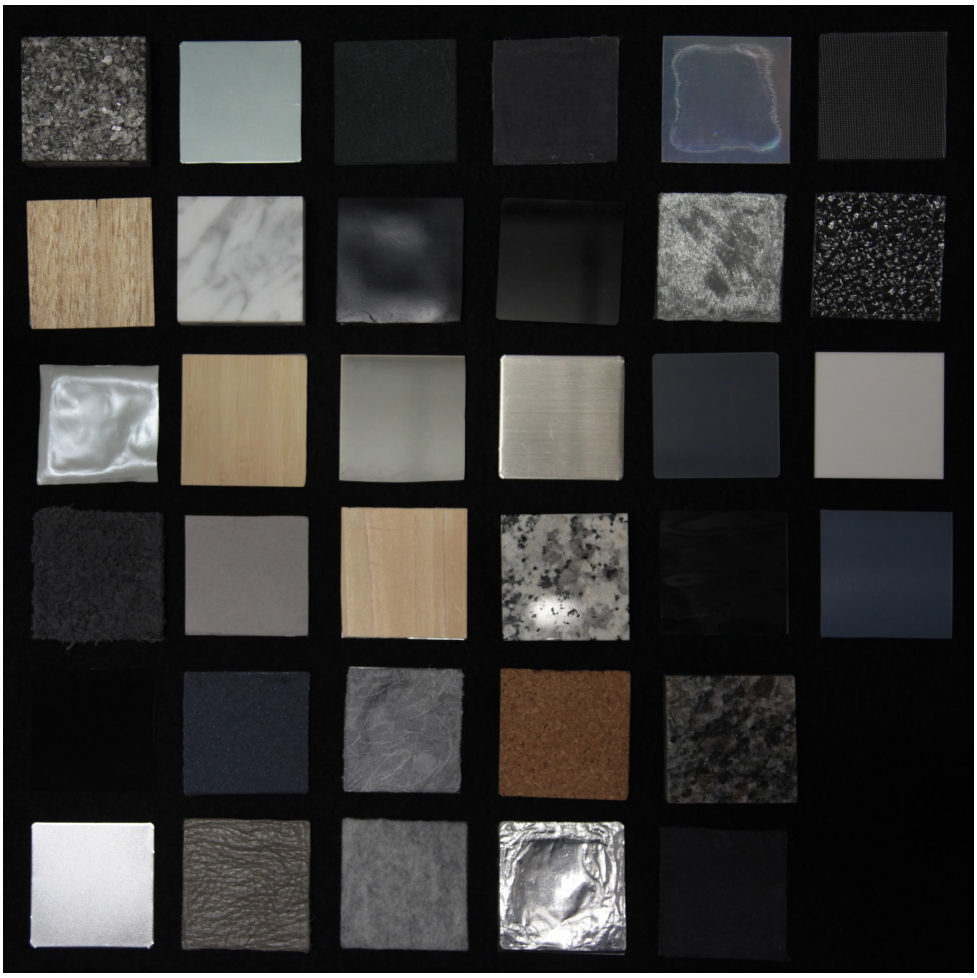


Figure 3.1. Material dataset of the 34 exemplars.

Table 3.1. Specifications of the material dataset.

Category	Remarks
Stone	Rustenburg
	Bianco brouille
	Caledonia
	White pearl
Metal	Almite gray
	Chrome
	SUS HL
Glass	Pearl gray
	Water drop pattern
	Checker pattern
Plastic	Opal
	Black
Leather	Saddle leather matte
	Pig suede
	Calfskin
Fabric	Cotton
	Satin
	Boa
	Crepe
	Felt
	Wool
Paper	H-2
	P-14
	Drawing paper
	Washi (handmade)
	Silver-coated paper
Wood	Paulownia
	Bamboo
	Japanese cypress
	Cork
Ceramic	Glazed tile
	Alumina
Rubber	Styrene
	Silicon

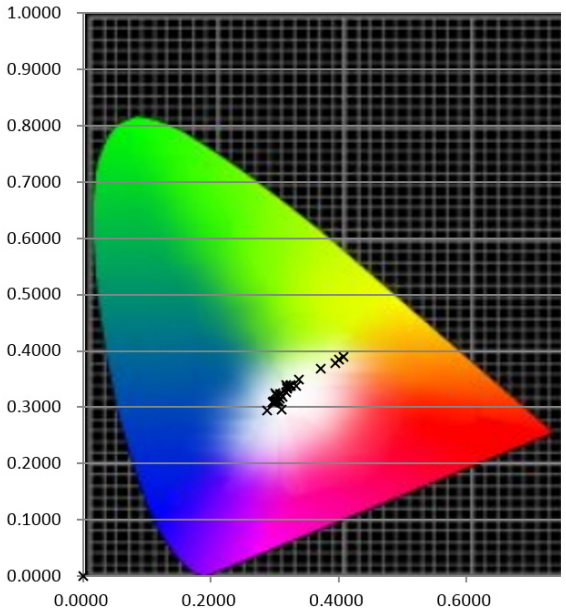


Figure 3.2. Chrominance of materials.

3.2.2 Image Dataset

The author hypothesized that when materials are reproduced on a monitor, the following factors strongly influence the perceptual qualities: (1) intensity, (2) color reproduction, and (3) resolution. In order to realize an accurate reproduction of real-world display materials, the author constructed an imaging system. The camera system was composed of an RGB camera and a standard lens. The camera, which was able to obtain linear camera output, was a Canon EOS 5D Mark II with a *sRAW2* image size of 2784 × 1856 pixels and a quantization level of 14 bits. The author then prepared a color image dataset (A) by capturing the materials set up in a viewing booth.

For the output monitor, the author used an Apple 15.4" MacBook Pro with Retina display. The widescreen, LED-backlit IPS screen has a glossy finish and a native resolution of 2880 × 1800 pixels with 220 pixels per inch.

The author completed the following procedure for reproducing the actual display

scene. Let $\left[\mathrm{R_C} \quad G_\mathrm{C} \quad B_\mathrm{C}\right]^\mathrm{T}$ be a color signal vector of a pixel captured by the RGB system, where T indicates the transpose operator of the vector. The 3-dimensional vector can be converted into CIE-XYZ tristimulus values $\left[\mathrm{X_C} \quad \mathrm{Y_C} \quad Z_C\right]^\mathrm{T}$ by multiplying a 3 × 3 matrix $\mathbf{M}_1$ that was constructed by approximating the CIE1931(2 deg) color matching function using the accumulative camera sensitivity function. The tristimulus values were then converted into linear RGB values $\left[\mathrm{R_L} \quad \mathrm{G_L} \quad \mathrm{B_L}\right]^\mathrm{T}$ by multiplying a 3 × 3 matrix $\mathbf{M}_2$ that was constructed in the display calibration process. Finally, the RGB vector $\left[\mathrm{R_M} \quad \mathrm{G_M} \quad \mathrm{B_M}\right]^\mathrm{T}$ for transmitting the display was obtained by adopting the gamma operator φ. The above procedure can be summarized by the following equation:

$$\left[\mathrm{R_M} \quad \mathrm{G_M} \quad \mathrm{B_M}\right]^\mathrm{T} = \varphi\left(\mathbf{M}_2\mathbf{M}_1\left[\mathrm{R_C} \quad \mathrm{G_C} \quad B_\mathrm{C}\right]^\mathrm{T}\right) \tag{3.1}$$

By using this calibration process, the author verified that the intensity and chromaticity between real materials and the reproduced image on the display were almost equivalent.

In order to investigate the influence of reproduction factors on material perception, the author prepared three additional sets: a gray image dataset (B), a low-resolution color image dataset (C), and a low-resolution gray image dataset (D). The gray image dataset was constructed by replacing $X_C = Y_C$ and $Z_C = Y_C$ before the display calibration process. By applying 4:1 horizontal and vertical down-sampling to the color and gray image datasets, low-resolution color and gray image datasets were created, respectively.

3.3 Experimental Methods

The author conducted five experiments in order to investigate perceptual qualities from

the static surface appearance of materials under different viewing conditions. All experiments were conducted according to the principles laid down in the Helsinki Declaration. Written informed consent was obtained from all participants.

3.3.1 Experiment 1: Visual Judgments of Perceptual Qualities using Material Dataset

In Experiment 1, the participants were asked to make judgments based on perceptual qualities by viewing actual materials. The procedure of the experiment followed that of Fleming et al. The experiment was organized into nine *blocks* of 34 *trials*, and this was defined as a *session*. Ten participants, nine males and one female, participated in this experiment.

For each trial block, a different perceptual quality was assessed, and in each trial within a block, participants rated the quality for a single material that was manually placed in a viewing booth (Macbeth Judge II) under the standard illuminant D65 by the participants themselves. The reason that the participants set up the materials is that our experiment required 612 repetitions (explained below) for evaluation, and it was not realistic that the experimenters would place all the stimuli. This also meant that the participants directly judged each material category in our experiment, unlike in Fleming et al.'s study, wherein the participants had to estimate material categories based solely on visual information. It is important to note that participants could not acquire tactile information, such as temperature or roughness, from the materials. The viewing booth was set in a dark room, and the inside wall was covered with black felt. The participants were instructed to set up each material while wearing gloves. Therefore, the author assumes that tactile exploration did not directly affect the participants' quality assessment.

Figure 3.3 shows the experimental environment in Experiment 1. The viewing distance was 300 mm, and the luminance of a white reference point presented in the same location as the material was 32.4 cd/m^2. Each material was placed perpendicularly to the participant's gaze.

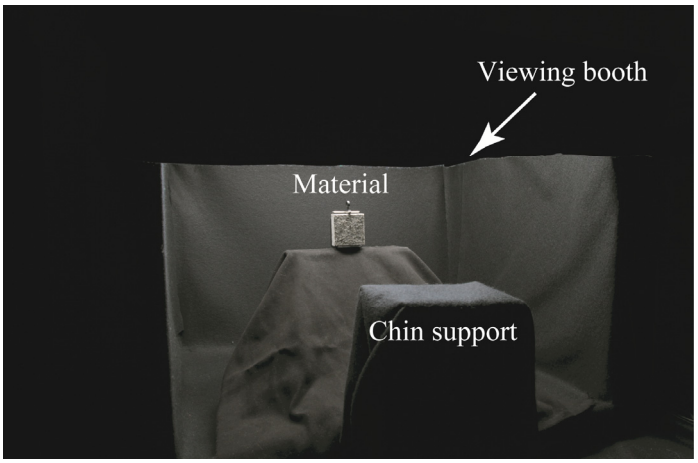


Figure 3.3. Experimental environment using material dataset.

An alternative approach would have been to provide ratings for several perceptual qualities in each trial. The order of the 34 exemplars was scrambled, but they were shown in the same order in each block. In each trial, the participant's task was to assess the material's perceptual quality within the current block and enter a rating from one to six into a spreadsheet in order to record their responses. Having assigned a value for a given perceptual quality to all 34 exemplars, the participants took a short break, and then started the next block (i.e., the next perceptual quality assessment).

Before each block, the perceptual quality to be judged in the forthcoming block was defined, and the polarity of the six-point scale (i.e., what low and high values corresponded to) was explained. The participants were encouraged to ask questions to

clarify their understanding of the material property to be rated and the rating scale. It is important to note that the participants were not informed that the materials were grouped into distinct classes; they were simply instructed to respond to the 34 exemplars of various materials. After finishing the first session's assessment, the second session, which was also organized into nine blocks of 34 trials, was performed to verify the intra-participant rating variances. In this second session, however, the author showed the 34 exemplars in the opposite order in each block. The author used the same definitions of Fleming et al.'s study. The nine qualities were assessed using the following definitions (Fleming et al., 2013):

- Glossiness: How glossy or shiny does the material appear to you? Low values indicate a matte, dull appearance; high values indicate a shiny, reflective appearance.
- Transparency: To what extent does the material appear to transmit light? Low values indicate an opaque appearance; high values indicate that the material allows a lot of light to pass through it.
- Colorfulness: How colorful does the material appear to you? Low values indicate a grayish, monochrome appearance; high values indicate a colorful appearance, which could consist of either a strong single color or several colors.
- Roughness: If you were to reach out and touch the material, how rough would it feel? Low values indicate that the surface would feel smooth; high values indicate that it would feel rough.
- Hardness: If you were to reach out and touch the material, how hard or soft would it feel? How much force would be required to change the shape of the material? Low values indicate that the surface would feel soft; high values indicate that it would feel hard.

- Coldness: To what extent would you expect the surface to feel cold to the touch? Low values indicate that the material would typically feel warm or body temperature; high values indicate that the material would feel cold to the touch.
- Fragility: How fragile or easy to break is the material? High values indicate that a small amount of force would be required to break, tear, or crumble the material; low values indicate that the material is highly resistant and could not be easily broken.
- Naturalness: How natural does the material appear to be? To what extent is the material in its most natural, common state? Low values indicate that the material appears unnatural; high values indicate that it appears natural.
- Prettiness: How pretty or visually attractive is the material to you? Low values indicate the material is ugly or unattractive; high values indicate that it is attractive or beautiful to the eye.

Within each block, the participant manually progressed through the real materials in the absence of any time constraints. There was no communication whatsoever between the participants. They were explicitly instructed not to confer with each other during the experiment. The participants were asked to not adjust their ratings to the materials that were viewed a second time in order to correct their error. From Experiments 2 to 5 in the next section, the same rules applied. Each participant performed 612 (34 × 2 × 9) trials in Experiments 1.

3.3.2 Experiments 2-5: Visual Judgments of Perceptual Qualities using Image Datasets

For Experiments 2 to 5, the participants' assessed the same perceptual qualities by viewing reproduced images of the display materials. The same participants as in Experiment 1 participated in Experiments 2-5.

In Experiment 2, the author used the color image dataset (A) described in 3.2. The image size, color, and brightness were adjusted to match the visual environment of Experiment 1. Figure 3.4 shows the experimental environment in Experiment 2. By comparing this setup with the one presented in Fig. 3.3, the author could confirm its accurate reproduction. The image dataset was compiled into a single BMP image, one image per page. The author presented the BMP images to the participants using the Apple Mac application "Preview" in slideshow mode. The participants were permitted to move the slideshow only forward; they could not complete any backward operations. As in Experiment 1, an alternative approach would have been to provide ratings for several perceptual qualities for each trial.

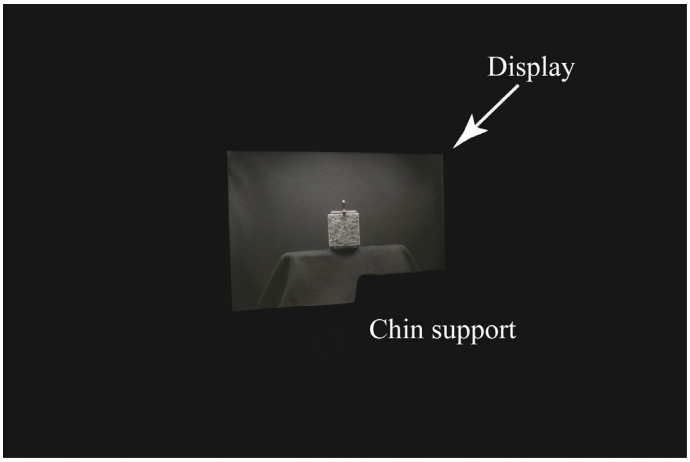


Figure 3.4. Experimental environment using image datasets.

In Experiments 3 to 5, the participants performed the same assessments using the gray image dataset (B), the low-resolution color image dataset (C), and the low-resolution gray image dataset (D). The participants did not assess the aspect of colorfulness of the gray images (Experiments 3 and 5, datasets B and D). Therefore, in

total, each participant performed 2,924 (34 × 2 × (9 × 3 + 8 × 2)) trials throughout Experiments 1 throughout 5.

3.4 Experimental Results

It took on average 74 min, 30 min, 26 min, 25 min, and 23 min, to complete each session in Experiments 1 through 5, respectively. Therefore, it took 356 min on average to complete all five experiments.

3.4.1 Intra- and Inter-Participant Variances

Table 3.2 summarizes intra- and inter-participant rating variances. The intra-participant variance $\sigma^2_{\text{intra}}(i,j)$ is shown in Table 3.2(a) for the *j*-th block in the *i*-th experiment. The intra-participant variance is calculated as the average of rating variances in each block between the two sessions for the ten participants, as follows:

$$\sigma^2_{\text{intra}}(i,j) = \frac{1}{10\times 34\times 2}\sum_{k=1}^{10}\sum_{m=1}^{34}\sum_{l=1}^{2}\left(\overline{a}_{k,m}(i,j) - a_{k,l,m}(i,j)\right)^2, \tag{3.2}$$

where $a_{k,l,m}(i,j)$ means the rated score of the *m*-th trial in the *l*-th session of *k*-th participant, and $\overline{a}_{k,m}(i,j)$ means the average of rated scores between the two sessions. The inter-participant variance is shown in Table 3.2(b) for the *j*-th block in the *i*-th Experiment. The inter-participant variance $\sigma^2_{\text{iner}}(i,j)$ is calculated as the averaged ratings of the two sessions in each block among the ten participants as follows:

$$\sigma^2_{\text{inter}}(i,j) = \frac{1}{10\times 34}\sum_{k=1}^{10}\sum_{m=1}^{34}\left(\overline{b}_{m}(i,j) - b_{k,m}(i,j)\right)^2,$$

$$b_{k,m}(i,j) = \frac{1}{2}\sum_{l=1}^{2} a_{k,l,m}(i,j), \tag{3.3}$$

where $b_{k,m}(i,j)$ is the averaged score of $a_{k,l,m}(i,j)$ between sessions, and $\overline{b}_{m}(i,j)$ is

the average of the scores for the ten participants.

There are two notable aspects of the observed variances. First, the intra-participant variances were lower than the inter-participant variances. This result indicates that the perceptual quality rating was stable within an individual participant. Our second observation is that the rating in Experiment 1 had a higher overall intra-participant variance as shown in Table 3.2(a). This suggests that the rating of perceptual qualities is sensitive to the rich information obtained from real materials. On the other hand, perceptual quality ratings were less sensitive to the poor information obtained from the images alone. The intra-participant variances for visible qualities such as "Transparency", "Colorfulness", and "Roughness" were relatively low. This result is opposite to the intra-participant variances as shown in Table 3.2(b).

This suggests that evaluations of perceptual qualities obtained from the rich information of real materials are similar between individual participants. On the other hand, the evaluations of perceptual qualities obtained from the poor information of images differed between the individual participants.

Table 3.2. Intra- and inter-participant rating variances.

(a) Average of Intra-participant variances for ten participants.

Ex.#	Glo.	Tra.	Col.	Rou.	Har.	Cld.	Fla.	Nat.	Pre.	Total
1	0.35	0.17	0.28	0.31	0.33	0.45	0.63	0.36	0.62	0.39
2	0.39	0.15	0.17	0.28	0.45	0.44	0.43	0.37	0.29	0.33
3	0.34	0.08	-----	0.20	0.34	0.41	0.38	0.30	0.36	0.30
4	0.30	0.06	0.24	0.26	0.39	0.30	0.32	0.34	0.39	0.29
5	0.23	0.13	-----	0.25	0.41	0.34	0.33	0.38	0.27	0.29

(b) Average of Inter-participant variances between ten participants.

Ex.#	Glo.	Tra.	Col.	Rou.	Har.	Cld.	Fla.	Nat.	Pre.	Total
1	0.94	0.31	0.60	0.55	0.71	0.79	1.10	2.36	1.66	1.00
2	1.29	0.68	0.61	0.52	0.79	0.69	1.18	2.18	2.00	1.11
3	1.29	0.83	-----	0.58	0.98	0.75	1.24	2.31	2.02	1.25
4	1.42	0.72	0.77	0.71	0.96	0.95	1.04	2.10	2.01	1.19
5	1.38	0.82	----	0.69	0.93	0.94	1.02	2.15	2.16	1.26

3.4.2 Ratings for Each Material Class

The participants were not informed that the 34 different images consisted of 10 distinct material classes. Fleming et al. concluded that the ratings of different qualities formed a distinctive feature "signature" for each class of materials. The author also investigated the mean ratings of each quality for each material class averaged across all participants. Figure 3.5 shows the mean quality scores for each material class. For each perceptual quality, the average responses for each of the 5 experiments are represented by individual bars in Fig. 3.5. From left to right, the bars represent responses of real objects, color images, gray images, low-quality color images, and low-quality gray images, respectively.

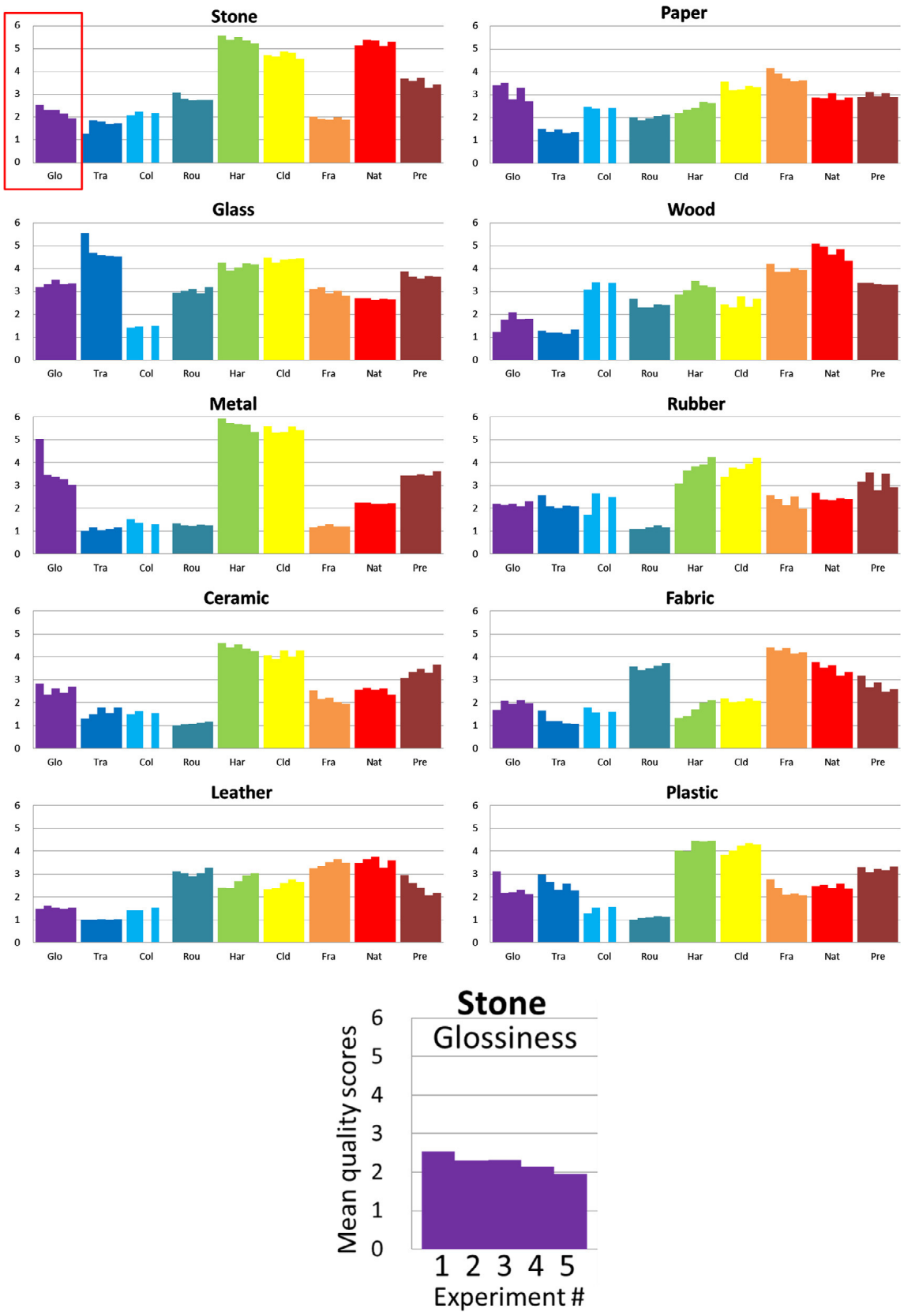


Figure 3.5. Mean quality scores for each material class. The bottom graph shows an enlarged version for one material (stone) and quality (“Glossiness”) to

illustrate the index of represented scores. From left to right, each bar shows scores for real objects (Experiment 1), color images (Experiment 2), gray images (Experiment 3), low-resolution color images (Experiment 4), and low-resolution gray images (Experiment 5).

As expected, the different material classes tended to have distinctive signatures for different qualities. Eight material classes in our experiment were the same as in Fleming et al.'s experiment. The experimental condition of our Experiment 1 was almost identical to Fleming et al.'s study, since the participants became familiar with the materials by viewing the real materials in our experiment and by viewing material images with shape information in Fleming et al.'s experiment. Therefore, the author first compared our results of Experiment 1 with the results of Fleming et al.

The author found that the signature of each material was similar in ours and Fleming et al.'s study. However, the following four responses were significantly different from Fleming et al.'s responses by more than two rating scores. (1) The rating of "Roughness" of stone was lower in our experiment, because the surface condition of our exemplars was flat. (2) The rating of "Glossiness" of leather was lower, because the surface condition of our exemplars was flat and no specular highlights included. (3) The rating of "Hardness" of wood was lower and that of "Fragility" higher, as our exemplar consisted of a processed product and not a tree trunk. (4) The rating of "Colorfulness" of plastic was lower, because the color of our exemplar was de-saturated.

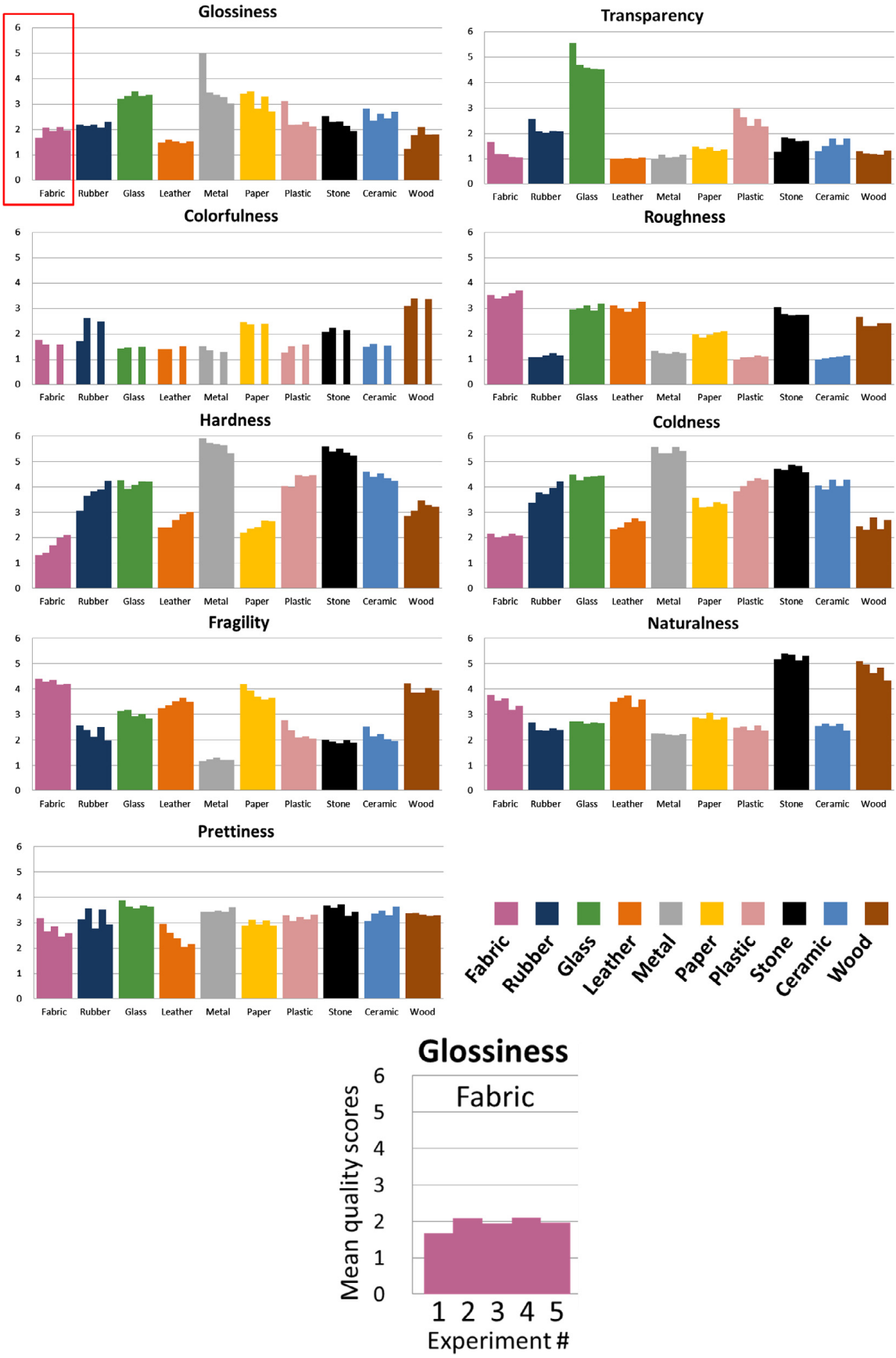
Next, the author compared our results between the five experiments. The signature of each material was similar between the five experiments as shown in Fig. 3.5, but there were some notable differences. The rating of "Glossiness" of metal and plastic materials decreased from Experiment 1 to the remaining experiments, and the rating of "Transparency" of glass materials decreased for the image reproductions. The

“Glossiness” of paper and “Prettiness” of rubber materials decreased when changing from color to gray image reproductions.

Based on these observations, the author suggests that the perceptual qualities of materials differ between real objects and reproduced images.

3.4.3 Ratings for Each Perceptual Quality

The author can plot the same data grouped by perceptual qualities, as shown in Fig. 3.6. Each bar indicates the average rating for a different material class. For each material class, the five responses are represented by individual bars. From left to right represent responses for real objects, color images, gray images, low-quality color images, and low-quality gray images, respectively.


Glossiness
Transparency
Colorfulness
Roughness
Hardness
Coldness
Fragility
Naturalness
Prettiness
Fabric
Rubber
Glass
Leather
Metal
Paper
Plastic
Stone
Ceramic
Wood
0
1
2
3
4
5
6

Glossiness
Fabric
Mean quality scores
0
1
2
3
4
5
6
1 2 3 4 5
Experiment #

Figure 3.6. Mean ratings for each perceptual quality (same data as in Fig. 3.5 but regrouped according to perceptual quality). The bottom graph shows an enlarged version for one material (fabric) and quality (“Glossiness”) to illustrate the index of represented scores. From left to right, each bar shows scores for real objects (Experiment 1), color images (Experiment 2), gray images (Experiment 3), low-resolution color images (Experiment 4), and low-resolution gray images (Experiment 5).

The results are again broadly intuitive. Our results from Experiment 1 are in accordance with Fleming et al.’s results. For example, most materials classes received low scores in the perceptual quality of “Transparency,” whereas glass received high average ratings. However, the signatures of “Glossiness” were generally lower than Fleming et al.’s results, because the author used flat exemplars without specular highlights. The signatures of “Colorfulness” were also lower, because the author used de-saturated exemplars. Since “Fragility” often correlates with 3D shape, our results were expected to be different from Fleming et al.’s. However, as our results did not differ greatly from their results, the author assumes that shape did not influence “Fragility.”

The signature of each quality was similar between the five experiments in our study, but there were some notable differences. First, as the author expected, the rating of “Glossiness” and “Transparency” for half of the materials decreased when ratings were insensitive to the chromatic information and the image resolution. This suggests that gloss reproduction is an important factor for the realistic material perception on a display device. In another example, the ratings of “Hardness” and “Coldness” tended to increase when reproducing the material samples as low-quality images on the display device. This suggests that the poor information obtained from the images provided qualities of “Hardness” and “Coldness” to the material perception.

3.4.4 Correlations between Perceptual Qualities

As mentioned above, different qualities have different distributions across material classes, suggesting that these qualities provide a means of distinguishing between material types. The correlation matrix relating the perceptual qualities to one another is shown in Fig. 3.7.

The most positively correlated qualities were "Coldness" and "Hardness" $(r = 0.6397)$ in Experiment 1. The second most positively correlated qualities were "Coldness" and "Glossiness" $(r = 0.5308)$. These results support the results by Fleming et al. However, the most positively correlated qualities in their experiment, "Transparency" and "Glossiness," were absent in our experiments' $(r = 0.0193)$. The fact that solid metals in our material set are by default not transparent could explain why the author obtained these results.

The most negatively correlated qualities were "Hardness" and "Fragility" $(r = -0.6421)$ in Experiment 1, which further supports the results by Fleming et al. However, the second most negatively correlated qualities "Coldness" and "Fragility" $(r = -0.4832)$ did not match Fleming et al.'s results. This result could be explained by having only the surface property but no shape information available.

There are some notable differences between the real objects and the representative images. Our second major finding was that the strongly positively correlated qualities depended on the experimental condition. In Experiment 1, "Glossiness" and "Roughness" were strongly negatively correlated qualities, which support the results of Fleming et al. However, some of these correlations were not significant in Experiments 2 through 5, most likely because correlations with "Glossiness" decreased for various qualities in the reproduced images. On the other hand, the correlations between

“Prettiness” and various qualities increased the case of the reproduced images.

Different characteristics were obtained between the color and gray information. “Glossiness” and “Transparency” obtained higher positive correlations for gray materials $(0.2653 \le r \le 0.2723)$, but lower correlation for color materials $(0.1917 \le r \le 0.2068)$. Similarly, “Naturalness” and “Fragility” obtained higher positive correlations for gray materials $(0.2136 \le r \le 0.2618)$, but lower correlations for color materials $(0.1584 \le r \le 0.1981)$. According to these results, color information is important to distinguish perceptual qualities such as “Glossiness” and “Transparency,” and “Naturalness” and “Fragility.”

Different characteristics were obtained between the high- and low-resolution information. “Glossiness” and “Prettiness” obtained higher positive correlations for low-resolution images $(0.3351 \le r \le 0.3815)$, but lower correlations for high-resolution images $(0.2768 \le r \le 0.2822)$. Similarly, “Roughness” and “Fragility” obtained higher positive correlations for low-resolution images $(0.4166 \le r \le 0.4430)$, but lower correlations for high-resolution images $(0.3192 \le r \le 0.3871)$. According to the detailed analysis of the rating scores, for both “Roughness” and “Fragility” the ratings decreased when the image resolution decreased. Based on these results, high resolution is required to distinguish perceptual qualities such as “Glossiness” and “Prettiness,” and “Roughness” and “Fragility.”

Overall, the different qualities were only weakly correlated with one another, as Fleming et al. founded in their study. Half of the correlation coefficients in Experiment 1 had an unsigned magnitude of less than 0.14, and about 80% had an unsigned magnitude of less than 0.28. However, throughout Experiments 2 to 5, these correlations became slightly stronger. In the case of the low-resolution gray images in Experiment 5, 39% had an unsigned magnitude of less than 0.14, and 78% had an unsigned magnitude of

less than 0.28. This indicates that the perceptual qualities are weakly correlated with one another, but their correlations gain strength images.

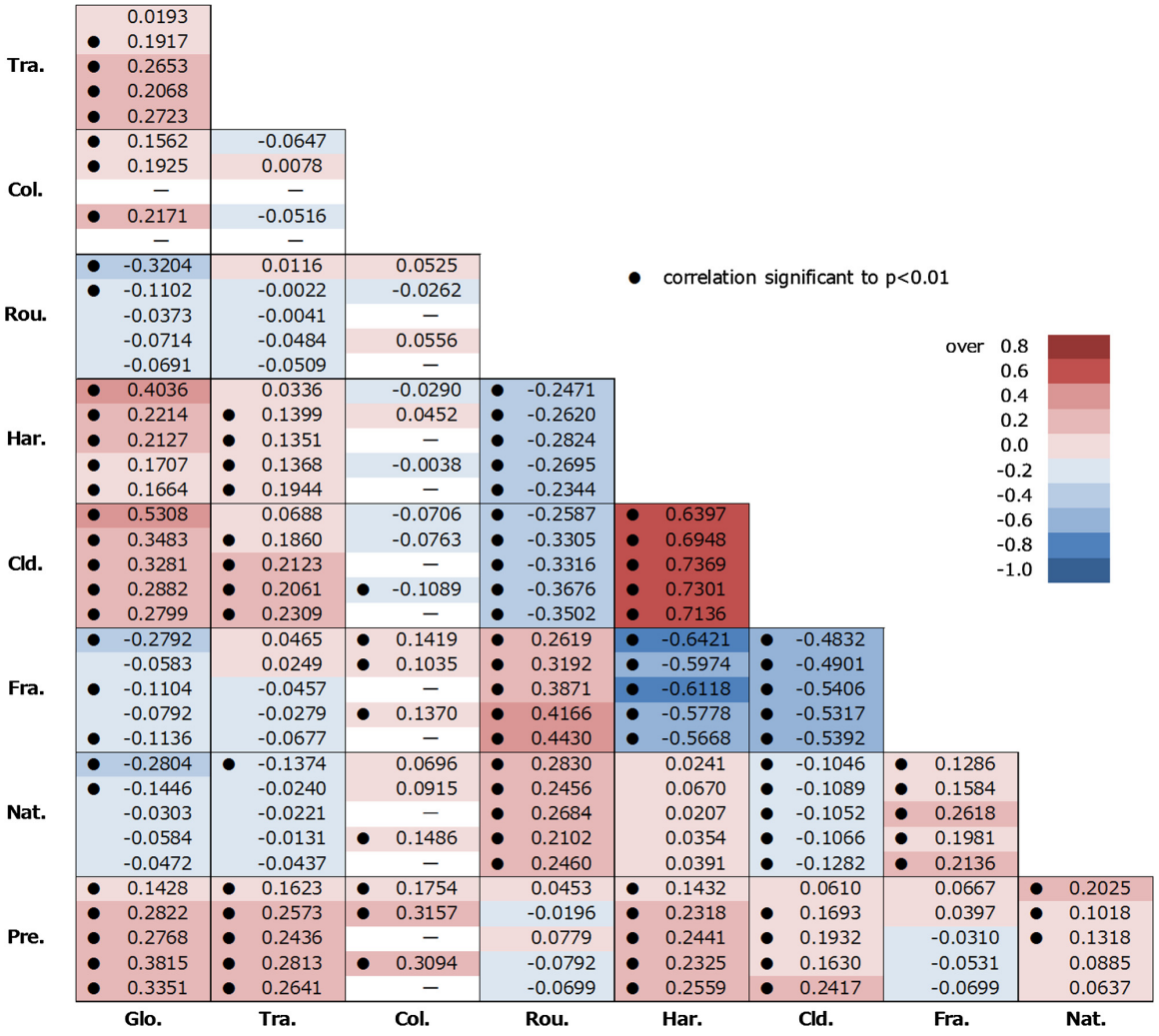


Figure 3.7. Correlation matrix relating the nine different perceptual qualities to one another. Colors indicate correlation coefficients as specified by the color bar. Reds indicate positive correlations, while blues indicate negative correlations. The correlation coefficient values are listed in each cell. Dots indicate that the correlation in the corresponding cell was statistically significant at the $p < 0.01$ level.

3.4.5 Correlations between Material Classes

The previous section addressed correlations between different perceptual qualities. In Fig. 3.8, the author plots the correlation matrix between the various material classes. Most of the correlation coefficients were strongly positive and higher than those reported by Fleming et al. The most positively correlated material classes in Experiment 1 were plastic and rubber $(r = 0.9113)$, followed by ceramic and metal $(r = 0.8845)$. These results are not surprising. Fabric and leather had similar signatures which might be caused by the flat surface representations without shape information that the author used. According to Fleming et al., the most positively correlated material classes were stone and wood $(r = 0.5815)$. The author also obtained a relatively strong positive correlation for this combination $(r = 0.4050)$.

Different characteristics were obtained for images with color and gray information. The most remarkable result is that all correlations were positive in the case of gray images (Experiments 3 and 5). For example, fabric and several materials (rubber, glass, metal, plastic, and ceramic) had higher negative correlations for the color images $(-0.5065 \le r \le -0.1383)$, but positive correlations for the gray images $(0.0251 \le r \le 0.2810)$. Glass and wood had higher negative correlations for the color gray images $(0.8619 \le r \le 0.8659)$, but lower correlations for the color images $(0.3951 \le r \le 0.4246)$. Most of the correlations increased from the color to the gray images. These results indicate that these material qualities were strongly influenced by color information, and that the information in the low-resolution gray images was insufficient for discriminating the materials.

Interestingly, the image resolution did not substantially affect the correlations between material classes.

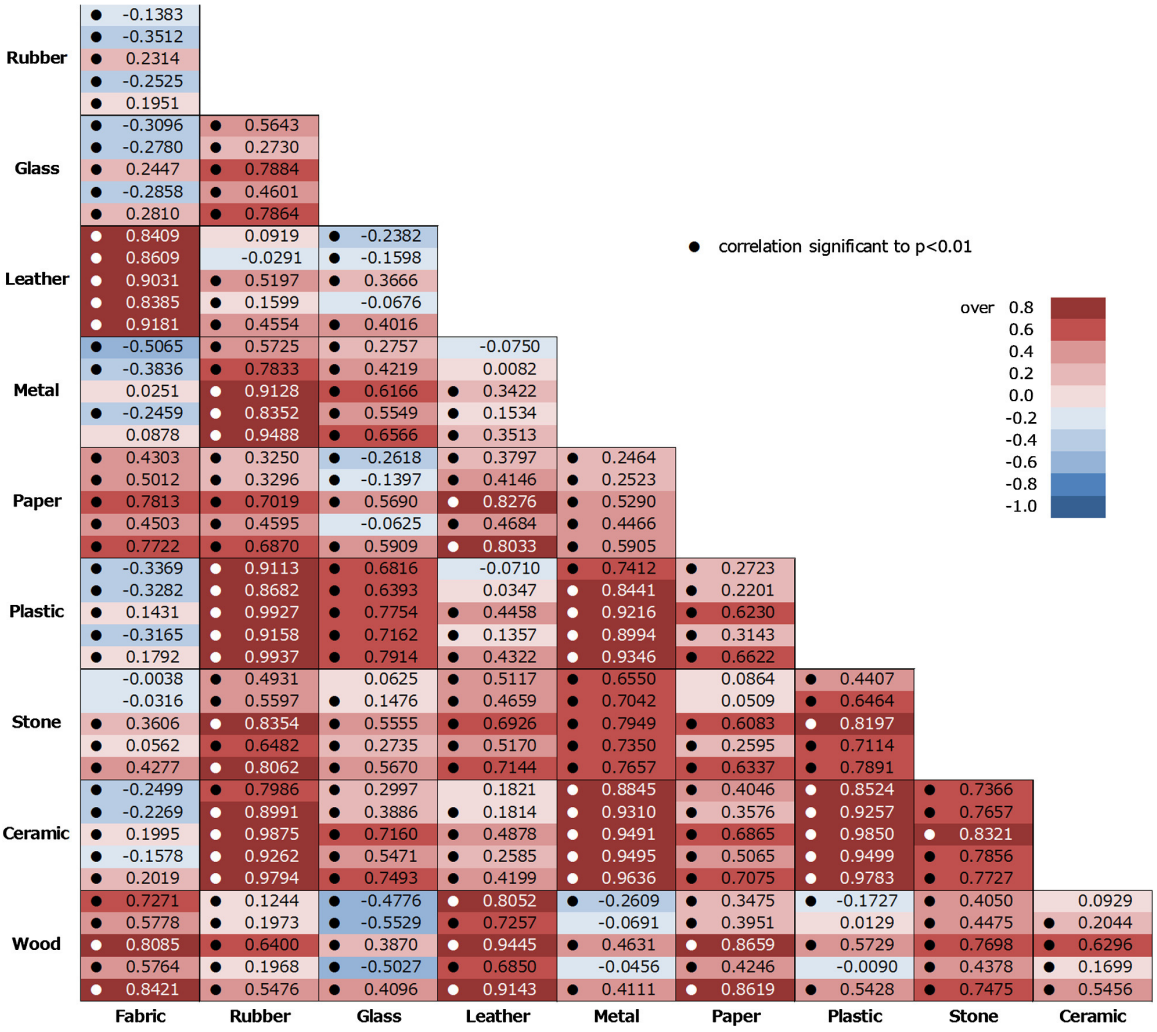


Figure 3.8. Correlation matrix relating the 10 different materials classes to one another. Colors indicate correlation coefficients as specified by the color bar. Reds indicate positive correlation, while blues indicate negative correlations. Correlation coefficients are presented in each cell. Dots indicate that the correlation in the corresponding cell was statistically significant at the $p < 0.01$ level.

3.4.6 Distributions of Material Classes in the Space of Perceptual Qualities

The author performed PCA on all ratings across participants to aid visualizing the distribution of material classes in the 9-dimensional feature space of perceptual qualities. Table 3.3 shows the percent variance of the first three principal components (PCs). By

showing the gray images in Experiments 3 and 5, higher percent variances were obtained. This result supports the notion that material qualities were harder to discriminate in the low-resolution gray images. Interestingly, the other three conditions (real object, high-resolution color images, and low-resolution color images,) had very similar percent variances. Throughout all experiments, the first three PCs accounted for more than 91% of the variance. It is important to note that the residual 9% of the variance in the distribution falls along the other six dimensions. This means that regardless of the representation method to show the materials, we can get an approximate impression of the overall distribution using only the first few PCs. This property can be illustrated by the scree plot in Fig. 3.9 which shows that 3 of those factors explain most of the variability because the line starts to straighten after factor 3. Figure 3.10 shows the PCs for perceptual qualities. From Figs. 3.10 (a) to (e) show nine PCs for each experiment. From Figs. 3.10 (f) to (n) show each PC through five experiments. According to these figures, the 1st PC with the high contribution ratio did not have undulations through all perceptual quality evaluations. Interestingly, through the 1st PC to the 3rd PC were almost the same for each perceptual quality through all experiments as shown in Figs. 3.10 (f)-(h). The 2nd PC related to the hardness, coldness and fragility under any conditions for different material categories and display methods. The 3rd PC also involved in glossiness, fragility and naturalness. As shown Figs. 3.10 (i)-(n), other PCs from the 4th PC to the 9th PC indicated characteristics for each perceptual quality under each condition of different display methods. For example, in Experiment 1 as shown in Figs. 3.10 (a), (f)-(n), the 4th PC was related to transparency, and the 5th PC involved in roughness. In Fig. 3.11, the author plotted the ratings for each material class projected onto the first three PCs in Experiment 1 and color-coded each image by its true class membership. Figures 3.11 (a), (b), and (c) show the spaces of the

1st – 2nd PCs, 1st – 3rd PCs, and 2nd – 3rd PCs, respectively. In each space, the stimuli within each class are generally closely clustered. Some clusters overlap, but they appear clearly localized within the space. The distributions of the other four experiments were similar to those of Experiment 1. Throughout all experiments, we observed that the space of the 2nd – 3rd PCs separated the material clusters the best. The interpretation of these principal component vectors is an interesting challenge for future studies.

Table 3.3 Percent variance of first three PCs.

	PC1	PC2	PC3
Experiment 1	0.802	0.877	0.912
Experiment 2	0.807	0.874	0.910
Experiment 3	0.823	0.895	0.927
Experiment 4	0.813	0.878	0.912
Experiment 5	0.822	0.891	0.924

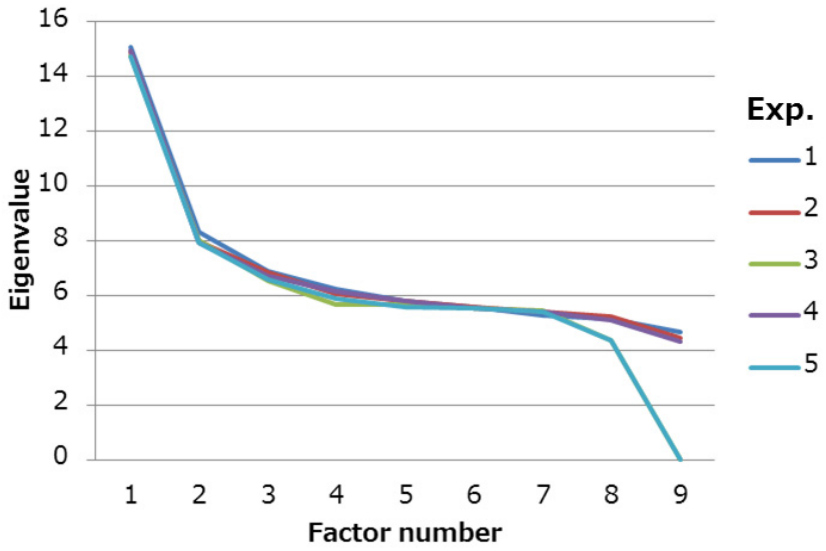
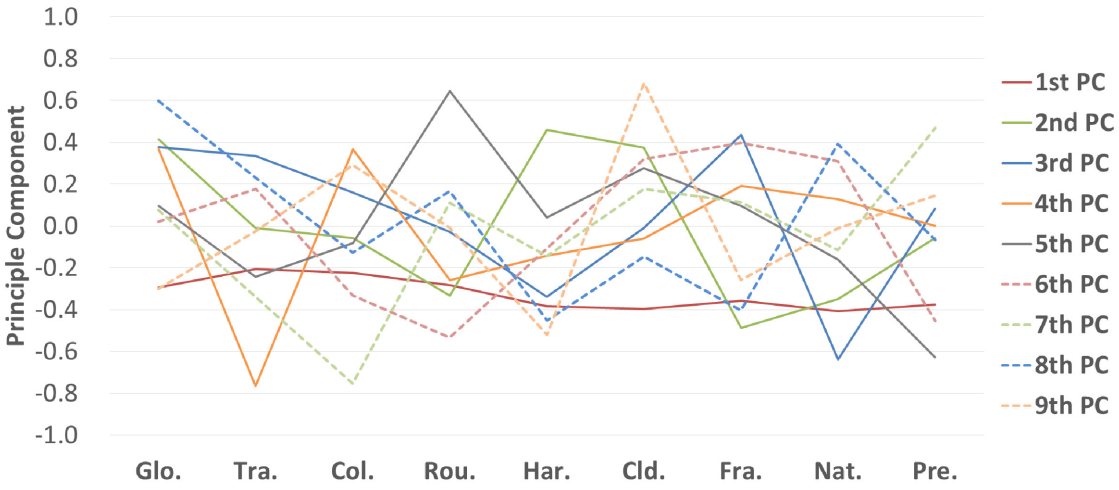
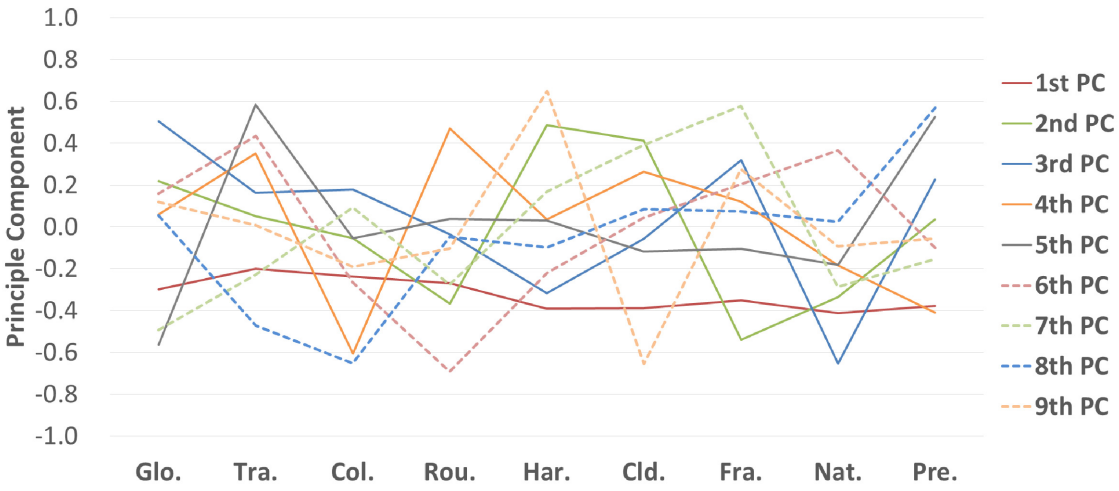


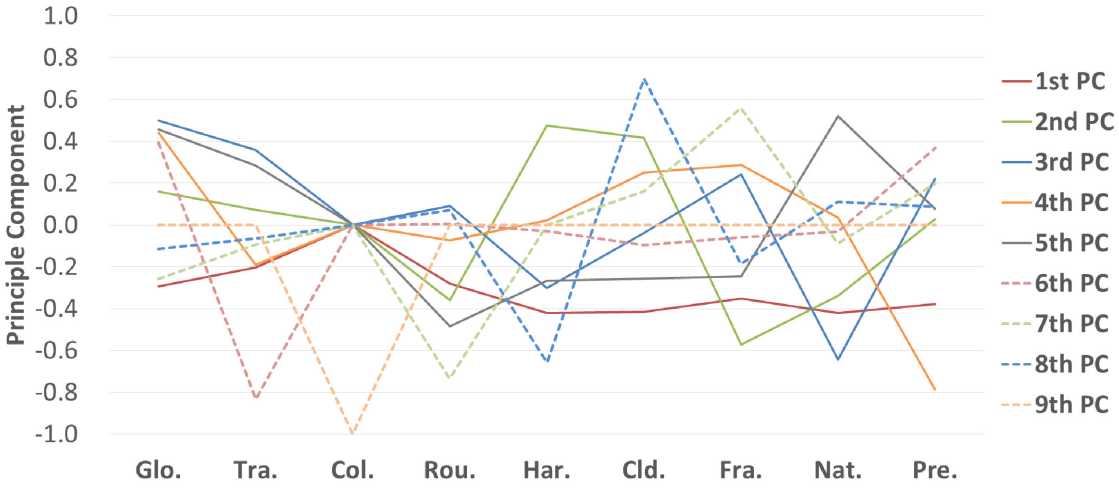
Figure 3.9. Scree plot for PCs in five experiments.



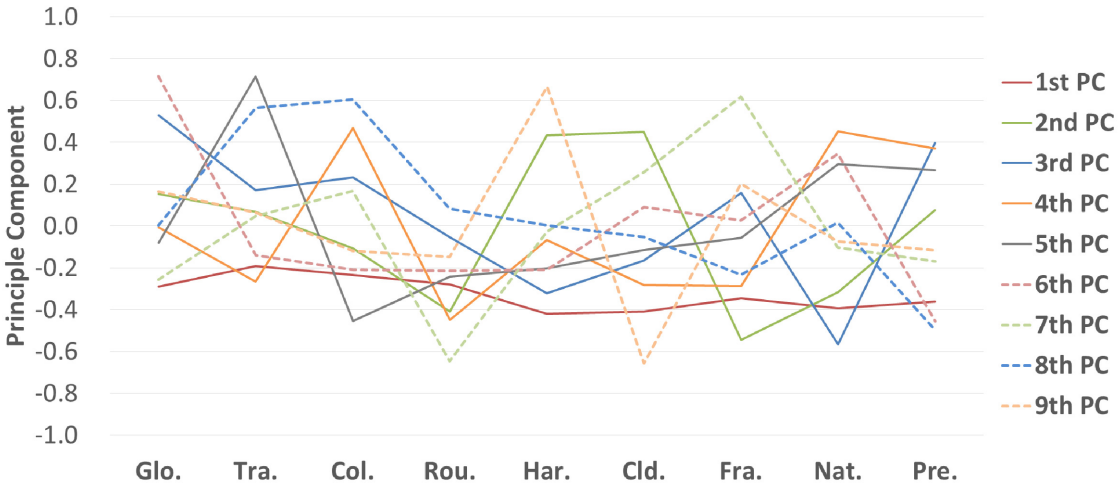
(a) Experiment 1.



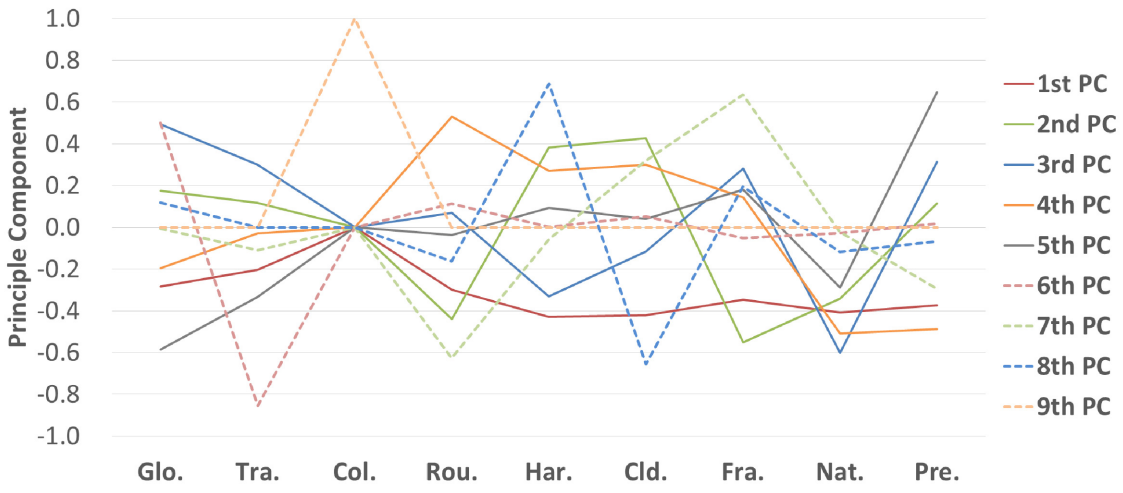
(b) Experiment 2.



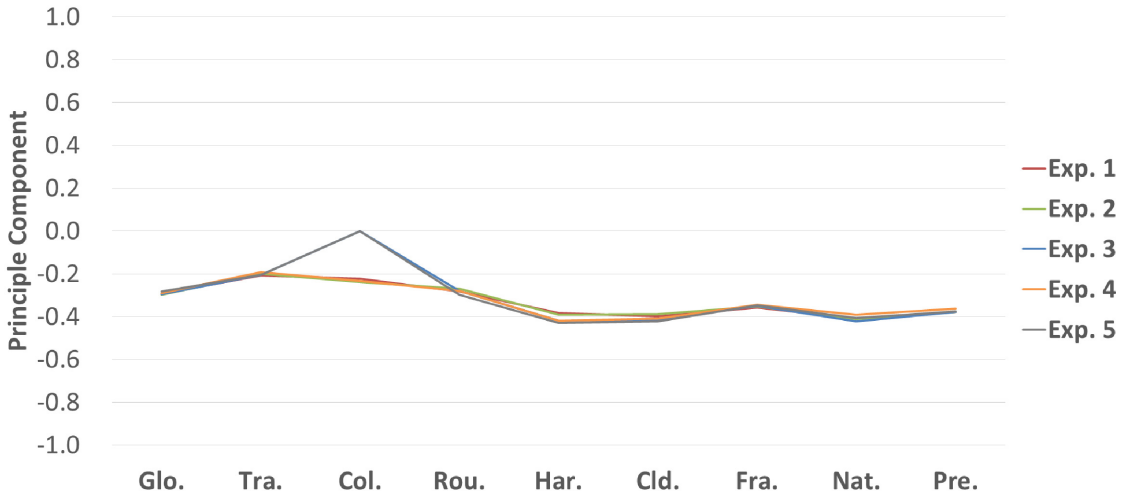
(c) Experiment 3.



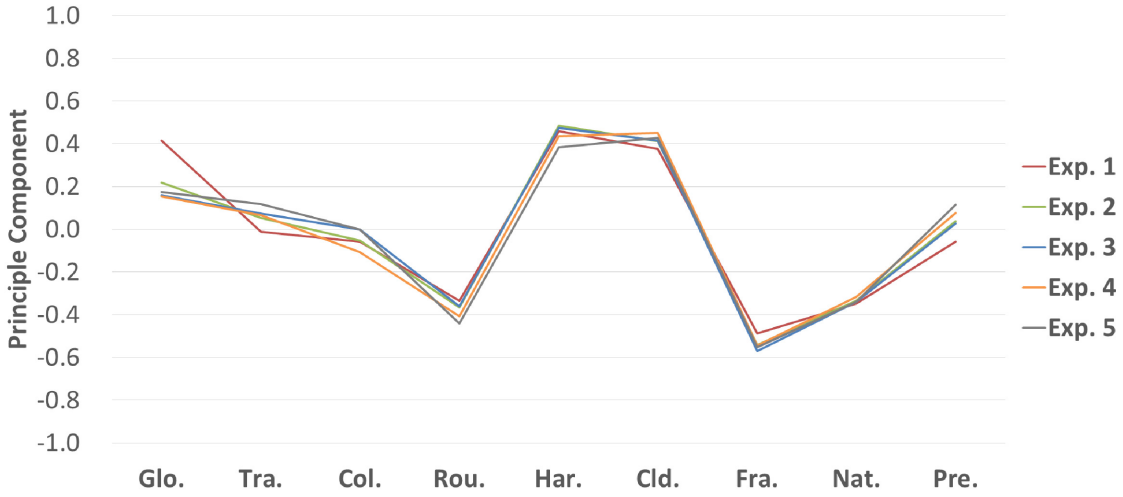
(d) Experiment 4.



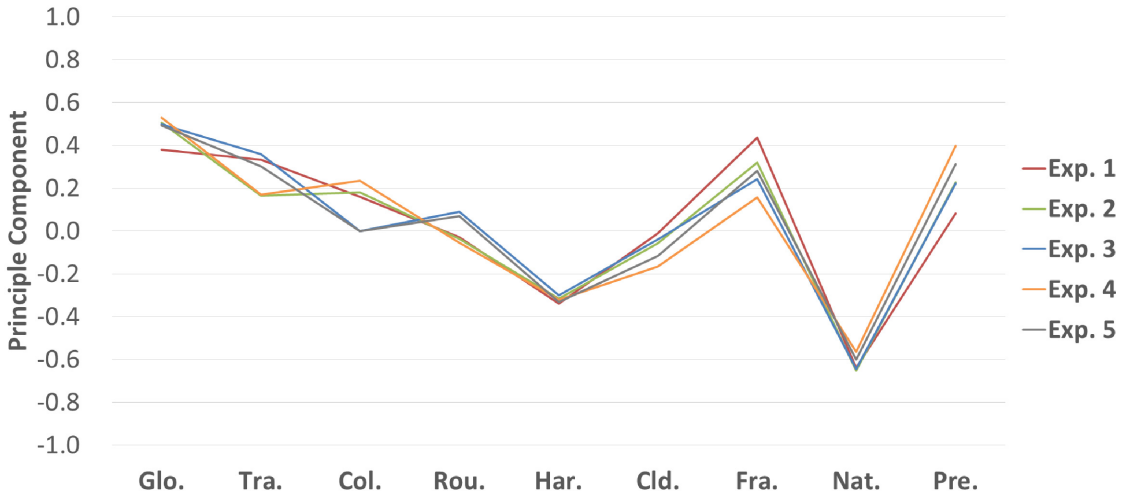
(e) Experiment 5.



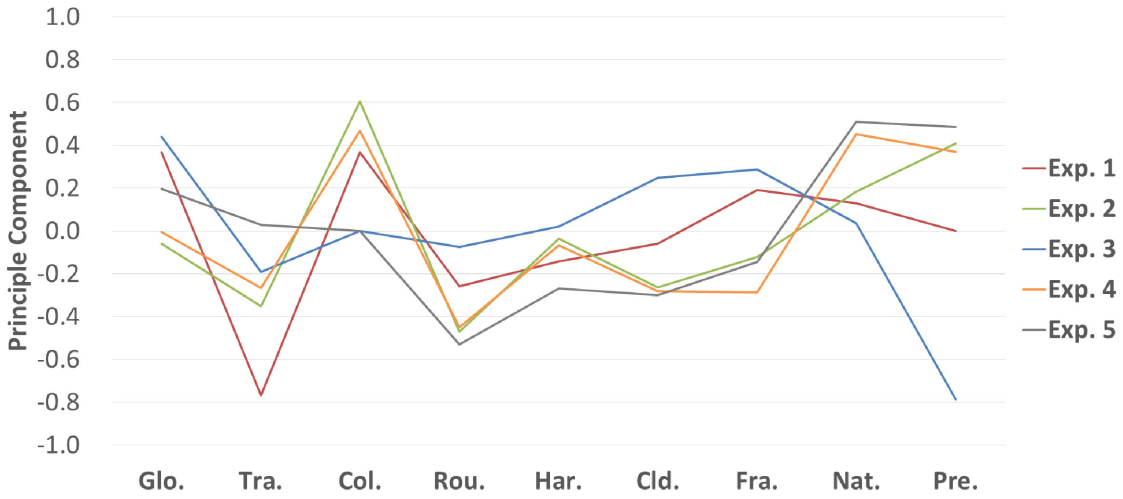
(f) 1st Principle Component.



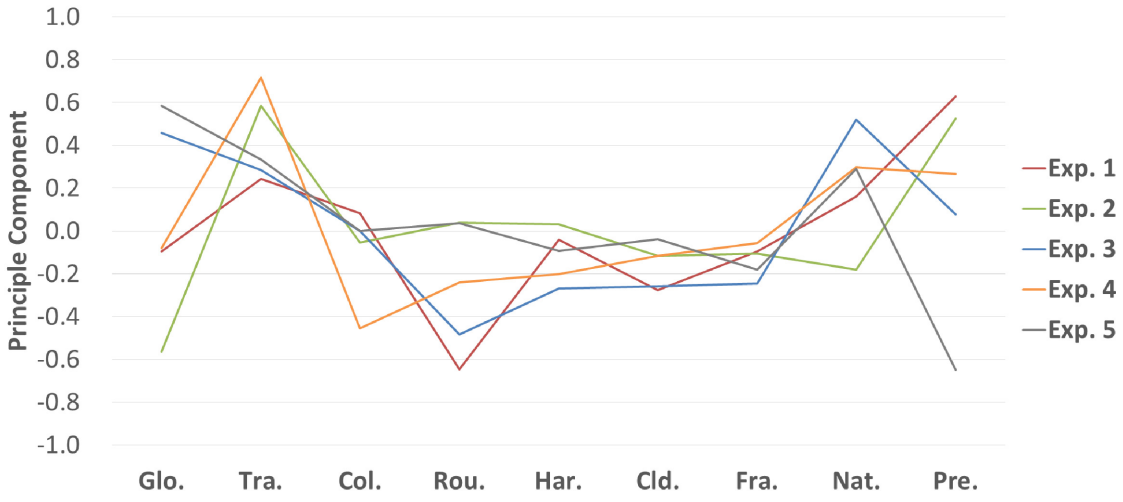
(g) 2nd Principle Component.



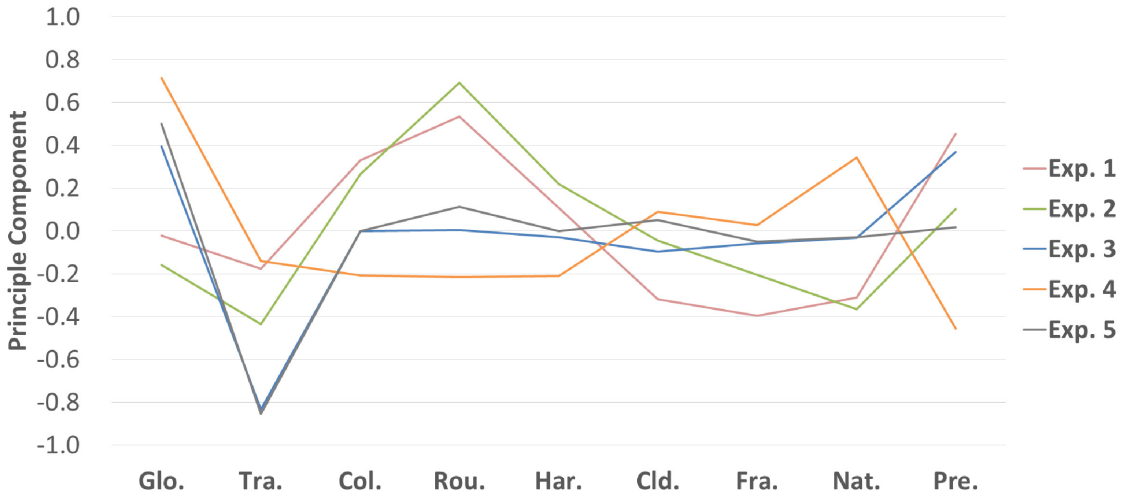
(h) 3rd Principle Component.



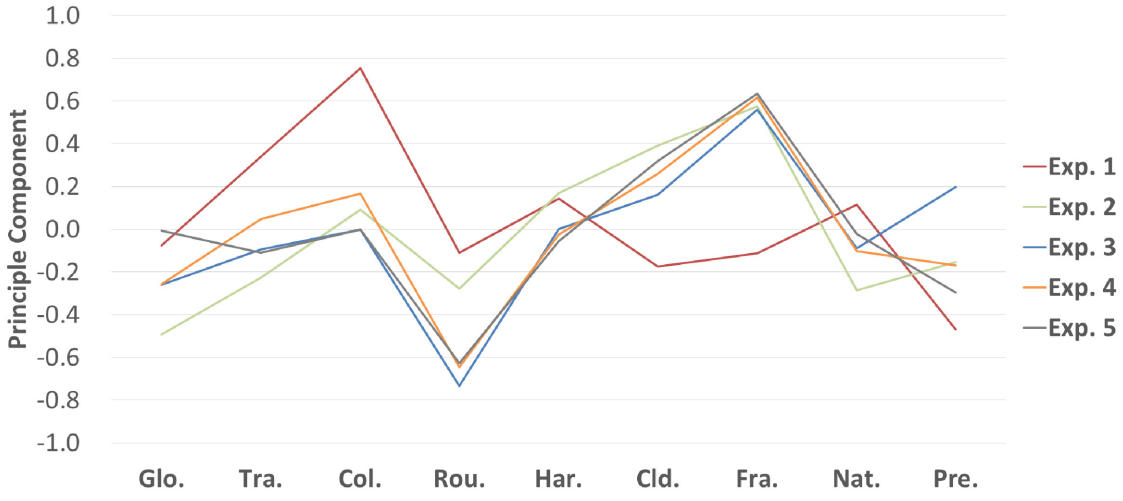
(i) 4th Principle Component.



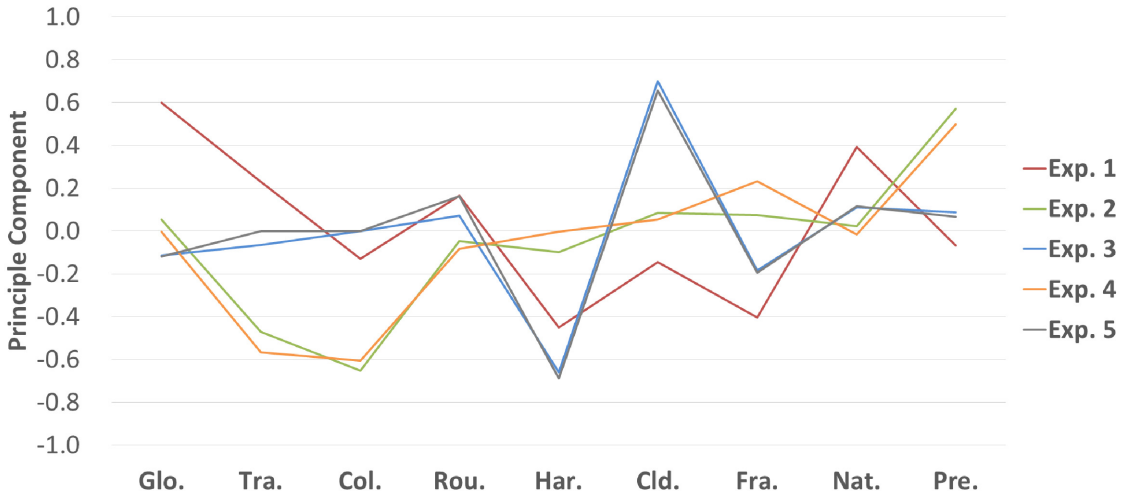
(j) 5th Principle Component.



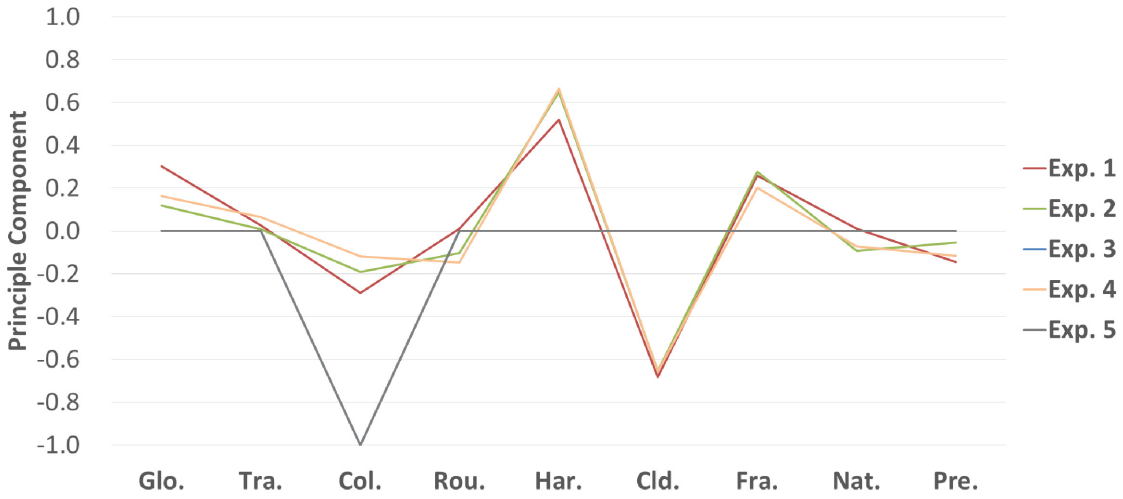
(k) 6th Principle Component.



(l) 7th Principle Component.

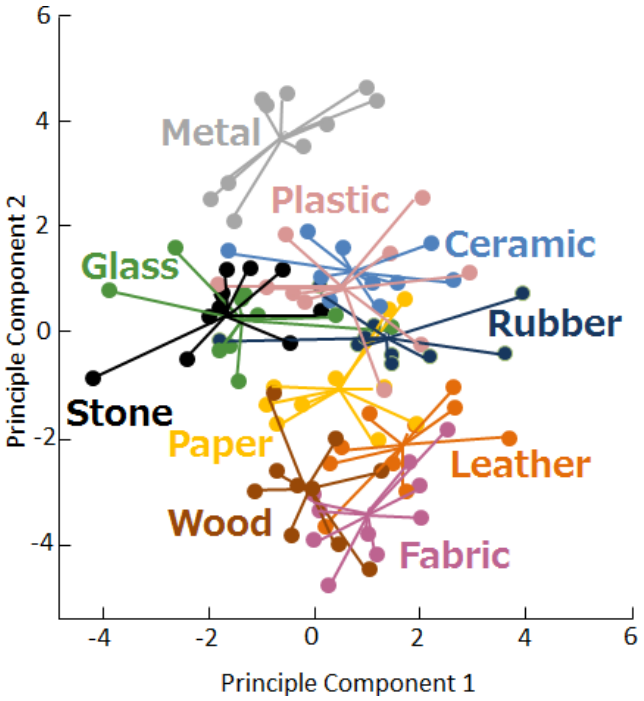


(m) 8th Principle Component.

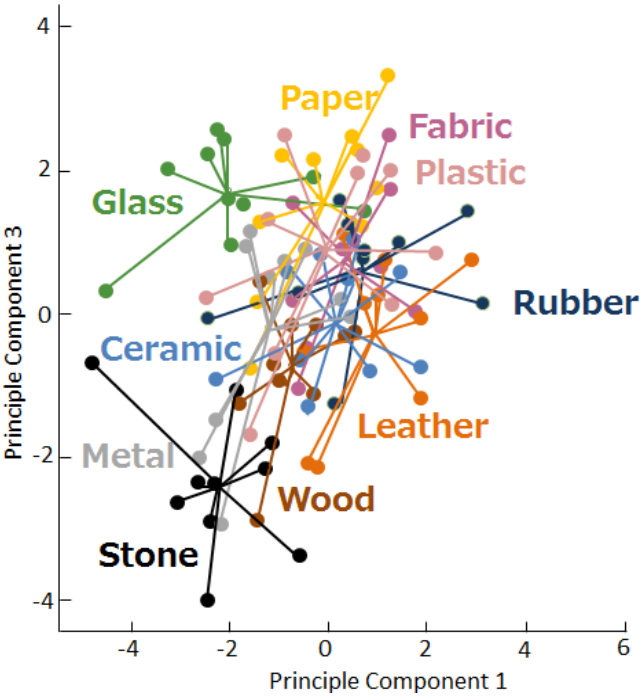


(n) 9th Principle Component.

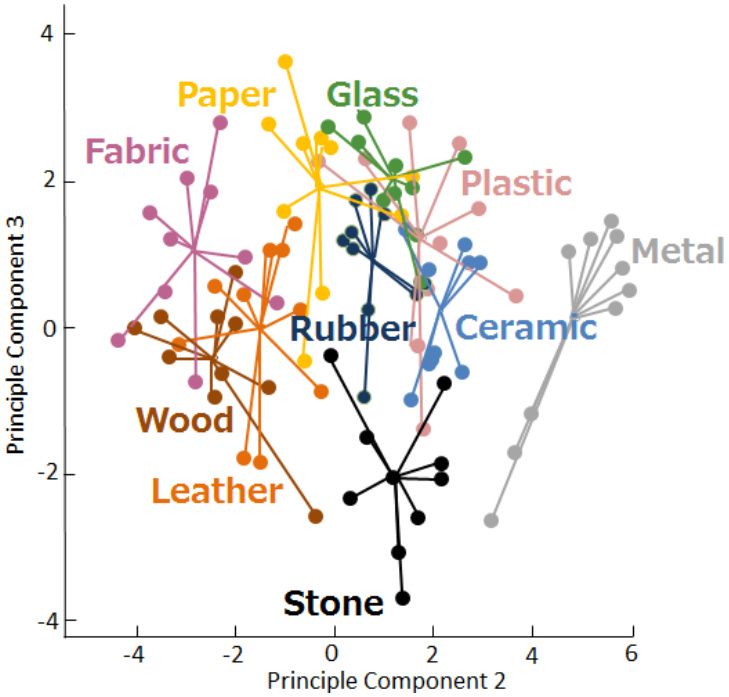
Figure 3.10. Principle components for perceptual qualities.



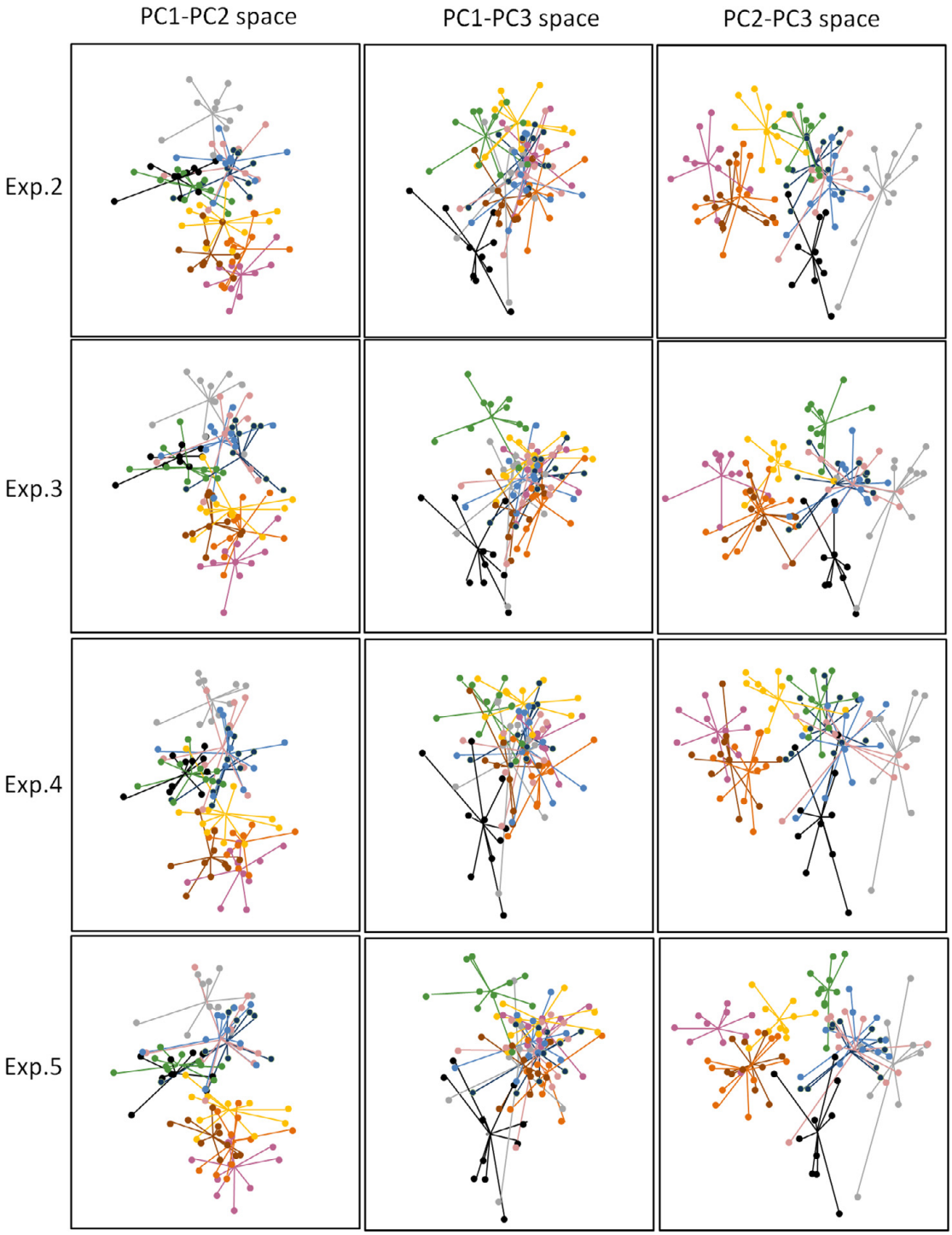
(a) The space of the 1st – 2nd PCs in Experiment 1.



(b) The space of the 1st – 3rd PCs in Experiment 1.



(c) The space of the 2nd – 3rd PCs in Experiment 1.

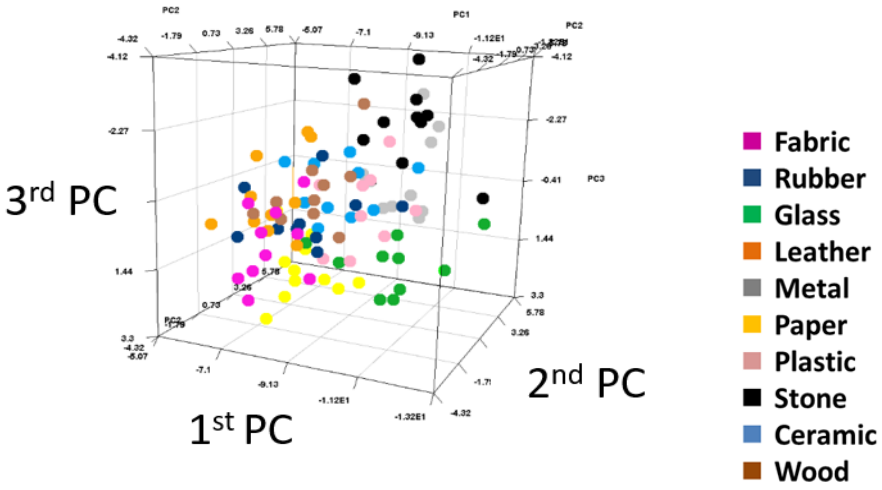


(d) The spaces of the PCs in Experiments 2-5.

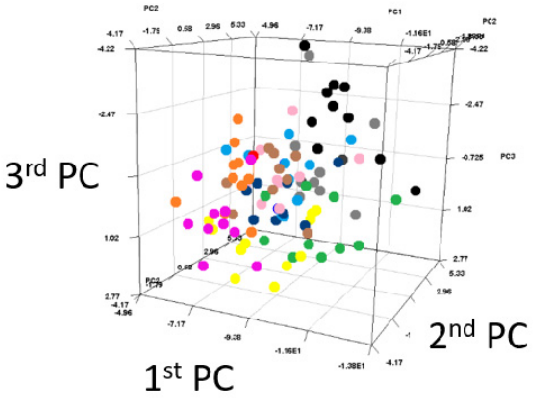
Figure 3.11. Distribution of the samples for the first three PCs in Experiment 1. Ten circles for each material represent projected positions of averaged rating scores (two sessions per participant) for all ten participants; lines join each sample to the projected mean location of each cluster. Color coding is based on true class membership, which was not revealed to the participants.

Figures 3.12 and 3.13 show plotted the ratings for each material class projected onto the 3-dimentional distributions of the space of 1^{st}-2^{nd}-3^{rd} PCs and 2^{nd}-3^{rd}-4^{th} PCs in all experiments, respectively. As described above, in Fig. 3.10 (f), 1^{st} PC gave a flat component without any remarkable characteristics for specific perceptual quality because of showing a common component. Therefore, in Fig. 3.12, it was difficult to distinguish material categories and most of them overlapped with each other. On the other hand, in Fig. 3.13(a), glass appears clearly localized, because 4^{th} PC was related to transparency.

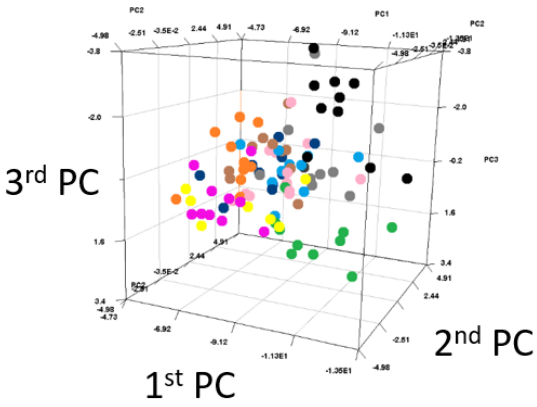
The author further applied k-means clustering to the data, which derives clusters solely on the proximity of different ratings in the 9-dimensional principal component space. Thus, by comparing the true clusters to those extracted by the k-means, we can measure the extent to which data of a given category are clumped together in feature space. The results of the k-means clustering depend on the initial setting of seeds. The author distributed initial seeds at random ten times.



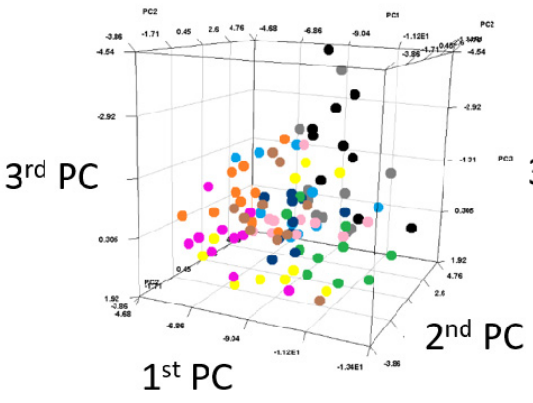
(a) Experiment 1.



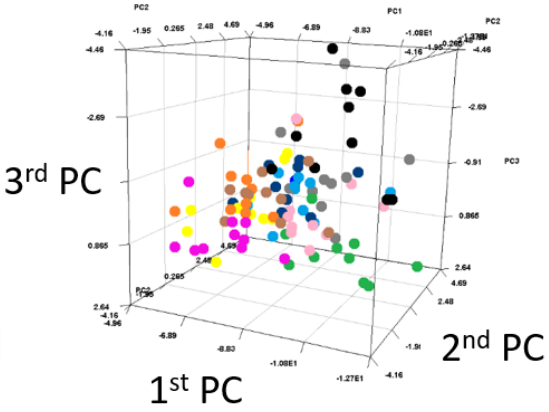
(b) Experiment 2.



(c) Experiment 3.

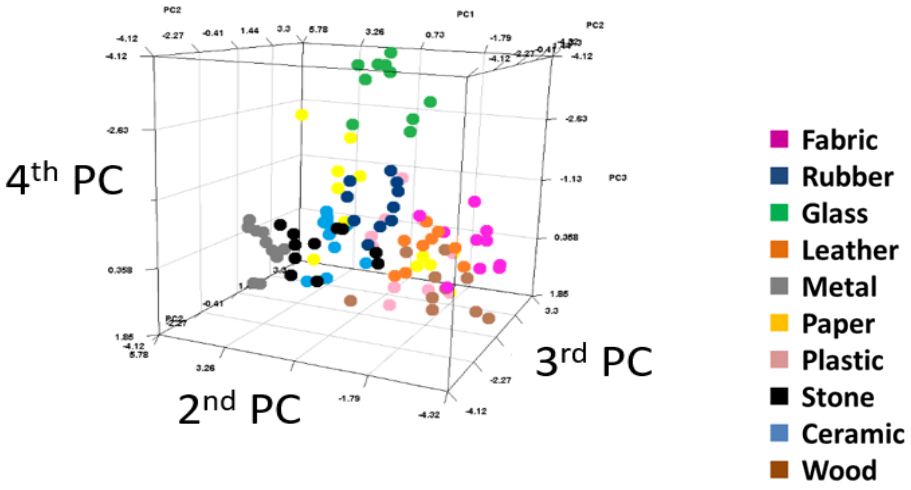


(d) Experiment 4.

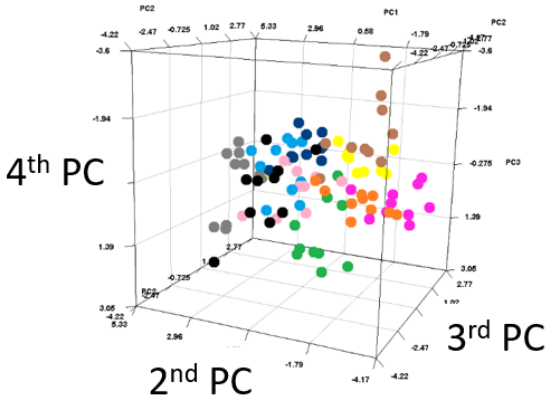


(e) Experiment 5.

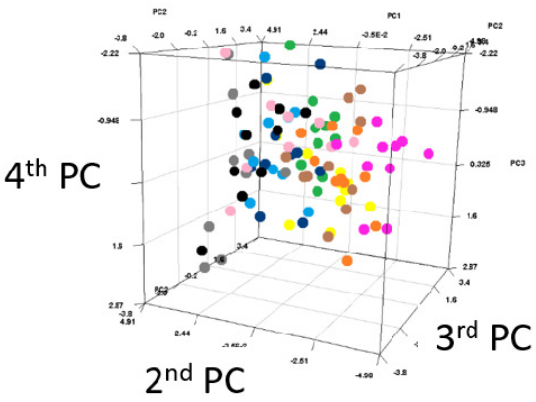
Figure 3.12. 3-Dimentional distribution in the 1st-2nd-3rd PCs' space.



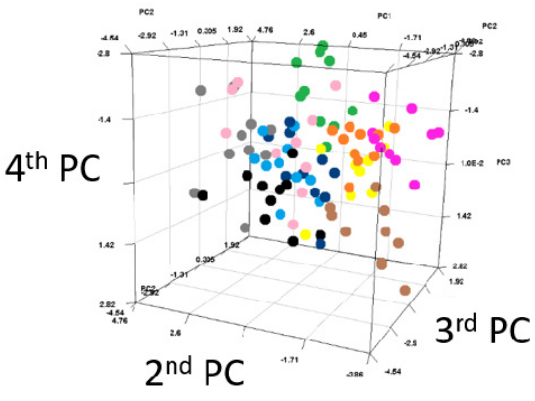
(a) Experiment 1.



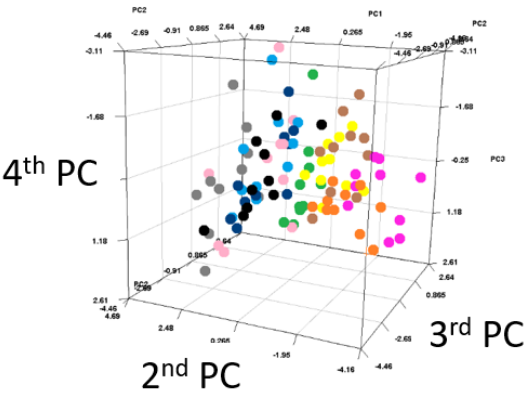
(b) Experiment 2.



(c) Experiment 3.



(d) Experiment 4.



(e) Experiment 5.

Figure 3.13. 3-Dimentional distribution in the 2nd-3rd-4th PCs' space.

Figure 3.14 shows the average of the similarity ratio according to which the quality data mapped onto the PC space was classified into the same material category by the k-means clustering. The horizontal and vertical axes show the experimental condition and the similarity ratio, respectively. As Fig. 3.14 illustrates, the k-means clustering algorithm clustered over 66% of the samples the same way as humans did (i.e., the same mutual class membership) based on the human quality ratings in Experiment 1. The similarity ratio decreased when the materials were displayed as images, and was lowest for Experiments 3 and 5. This suggests that the participants confused material categories more easily when they were represented as gray images.

Figure 3.15 shows the same data of Fig. 3.14 separately for each material class. Glass and stone were more often classified into the same material category. In contrary, the ratios of ceramic, paper, plastic and rubber were relatively low. An interesting characteristic was observed for the ratios in Experiment 2. The correlations of the similarity ratio between the different experiments are shown in Table 3.4. As shown in the table, the correlations with the other experiments were low $(0.198 < r < 0.352)$. In particular, the ratios in Experiment 2 fluctuated greatly between different material classes (e.g., glass and metal), as compared with other experimental results. By investigating the distribution in the PC space, the form of these clusters resembled more an ellipse than a circle. Therefore, the data of glass and metal could be separated into several material categories by the property of k-means algorithm. A more detailed analysis is required.

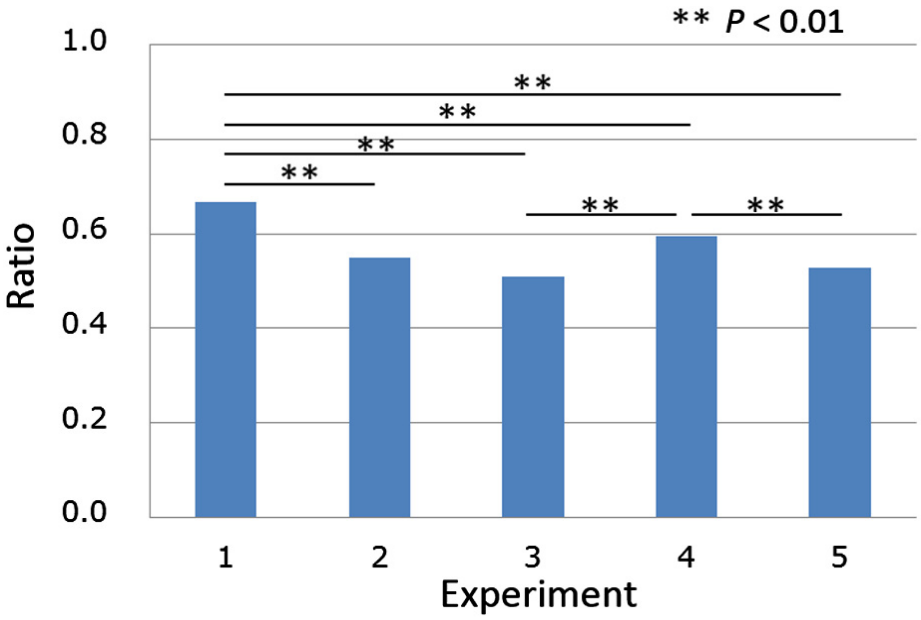


Figure 3.14. Average ratios of the same material samples in each cluster by the k-means algorithm. From left to right, each bar shows scores for real objects (Experiment 1), color images (Experiment 2), gray images (Experiment 3), low-resolution color images (Experiment 4), and low-resolution gray images (Experiment 5).

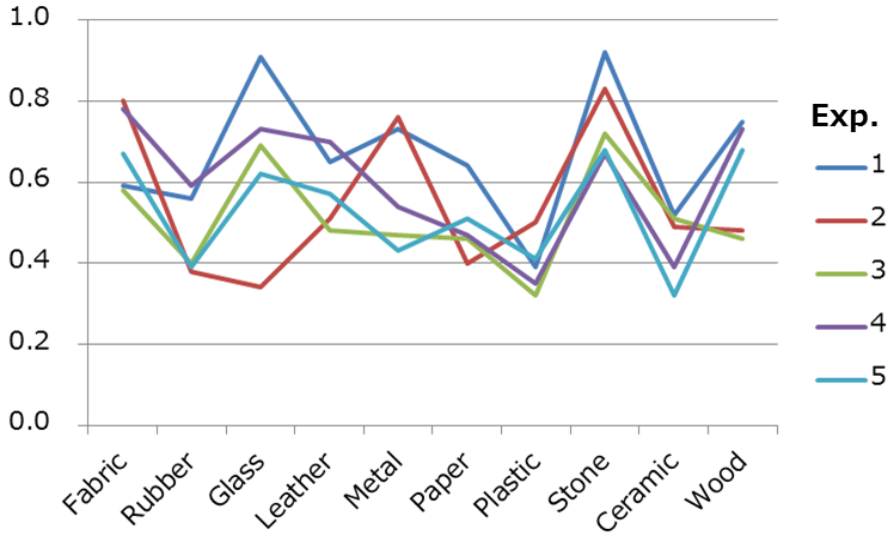


Figure 3.15. The ratio for each material of Fig. 3.14.

Table 3.4. Correlation of the similarity ratio in Fig. 3.15 between the different experiments.

Experiment	1	2	3	4	5
1	1.000	0.198	0.815	0.637	0.664
2		1.000	0.352	0.234	0.322
3			1.000	0.584	0.616
4				1.000	0.842
5					1.000

The author also applied k-means clustering to the perceptual qualities. Figure 3.16 shows the clustering results when k = 10 that is equivalent to the number of material categories in our dataset. The k-means clustering was performed using the average of 9-dimensional perceptual quality scores for all observers. In Fig. 3.16, Materials within the same color bar under material images means that they were classified into the same cluster. Here, the same colored bars among different experiments have no meanings. Each image is color coded by its true class membership.

As shown in the figure, material and category are not entirely consistent. However, metals were classified into the same category through all experiments. The result of the experiment 5 was significantly different from the other experimental results.

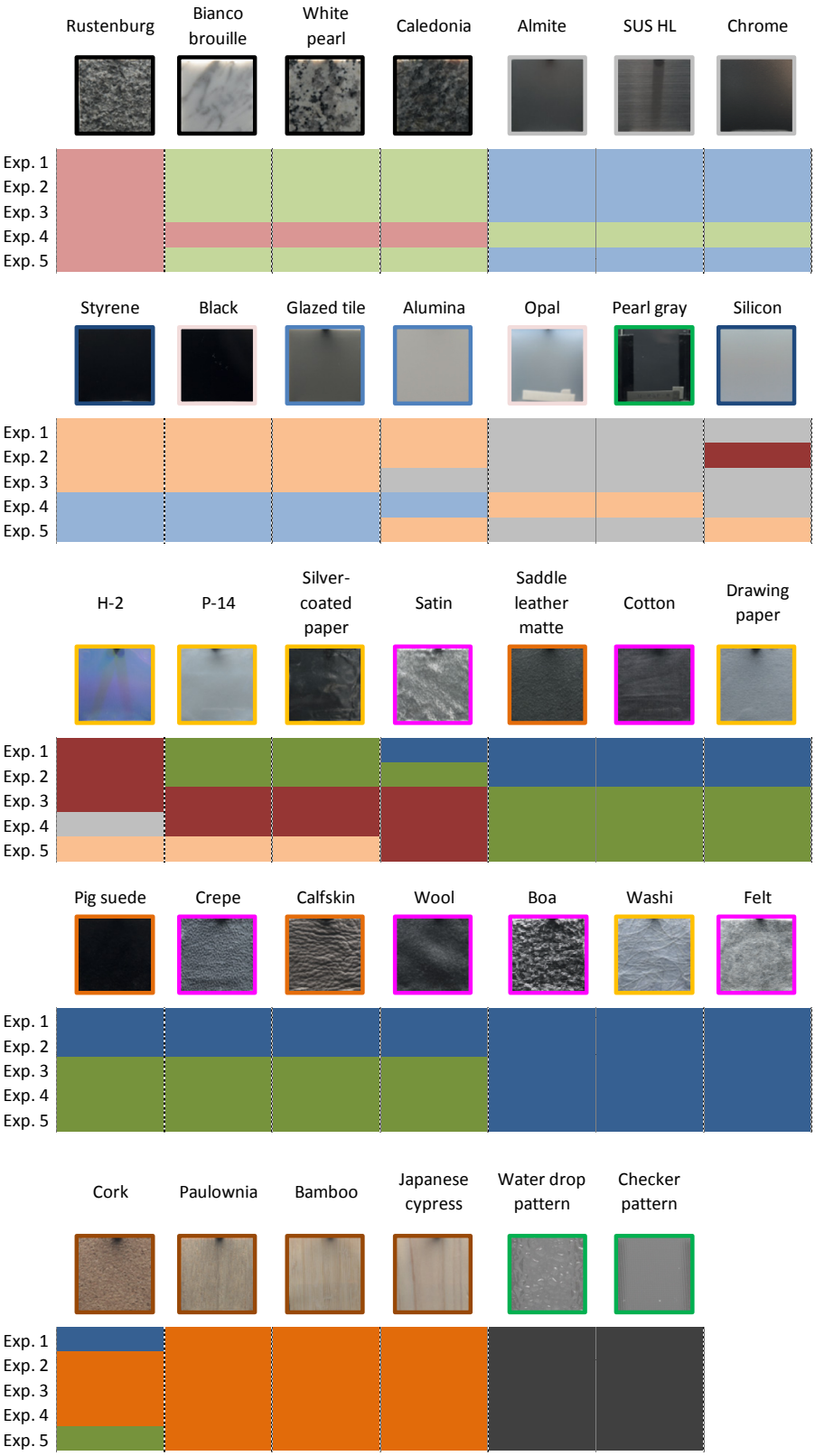


Figure 3.16. Category transition with similar perceptual qualities in each experiment using k-means algorithm (k = 10).

3.5 The Effect of Binocular Parallax for Material Perception

As described in Sec. 3.4.2, the perceptual qualities of materials differed between real objects and reproduced images. Figure 3.5 showed that the rating of "Glossiness" of metal and plastic materials decreased from Experiment 1 to the remaining experiments, and the rating of "Transparency" of glass materials decreased to the rendered images. In this section, the author further investigates the effect of binocular parallax for perceptual qualities of materials.

In the case of viewing real objects, human observes objects using binocular parallax caused from right and left eyes. On the other hand, in the case of rendered images, human observes the images as monocular images obtained from camera with single lens. This means that even the brightness and color between real objects and rendered images are equivalent by using our rendering system, the observation process is different between them.

The author conducted an additional experiment to verify the effect of the binocular parallax. Eight real objects selected from three material categories (metal, glass and plastic) were used as the experimental stimuli to investigate the effect of binocular parallax for glossiness and transparency perception. These samples are a partial set of our material dataset in Table 3.1. The image dataset was newly prepared with three different positions of monocular vision. Figure 3.17 shows a set of images in the image dataset captured at each eye position. The images for right and left eyes were captured 35mm away from center position, because the average distance of both eyes is typically 70mm.

Six kinds of stimuli were prepared for each material sample as follows:

(Pattern 1) A rendered image for right eye with monocular observation.

(Pattern 2) A rendered image for left eye with monocular observation.

(Pattern 3) A rendered image for both eyes with binocular observation.

(Pattern 4) A real object for right eye with monocular observation.

(Pattern 5) A real object for left eye with monocular observation.

(Pattern 6) A real object for both eyes with binocular observation.

(a) Left (b) Center (c) Right

Figure 3.17. A set of images captured at each eye position.

The experimental stimuli were evaluated using the same method in Sec. 3.3. Each stimulus was showed once to be in sequence from the pattern 1 to 6. Ten participants with normal color vision participated in this experiment.

Figure 3.18 shows the mean quality scores of glossiness and transparency for each observation pattern. The author found that the difference of display method between the real objects and rendered images has bigger effects for evaluating perceptual glossiness and transparency than the difference between the binocular and monocular observation. This means that the binocular parallax is little effect on the result. The rating score of binocular observation is almost same as our previous experiment in Fig. 3.5. However, the rating for transparency of plastic differed from the previous result. The reason might be due to the order of stimuli. In the previous experiment, the participants evaluated perceptual qualities by fixing the display method and changing the material samples in

order. In contrast, in this experiment, the participants evaluated perceptual qualities by fixing the material sample and changing the display methods. Therefore, this experiment became easy to compare and evaluate the difference of perceptual qualities than the previous experiment. As the introspection report obtained from participants, the reflected glare sometimes existed on the real material surfaces and the effects enhanced the glossiness of the real materials. We need to consider their effects carefully.

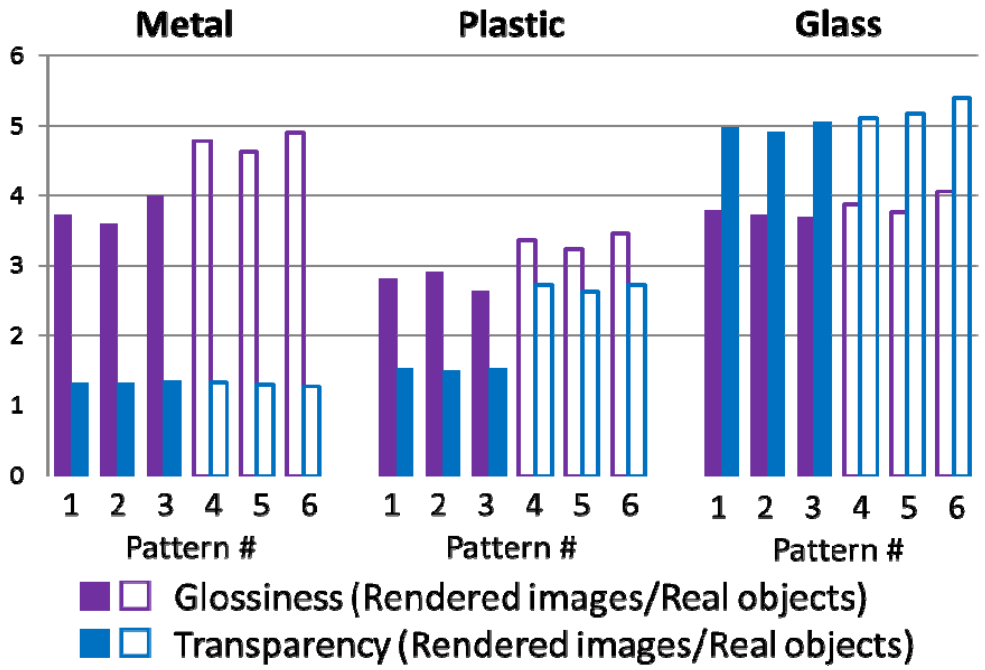


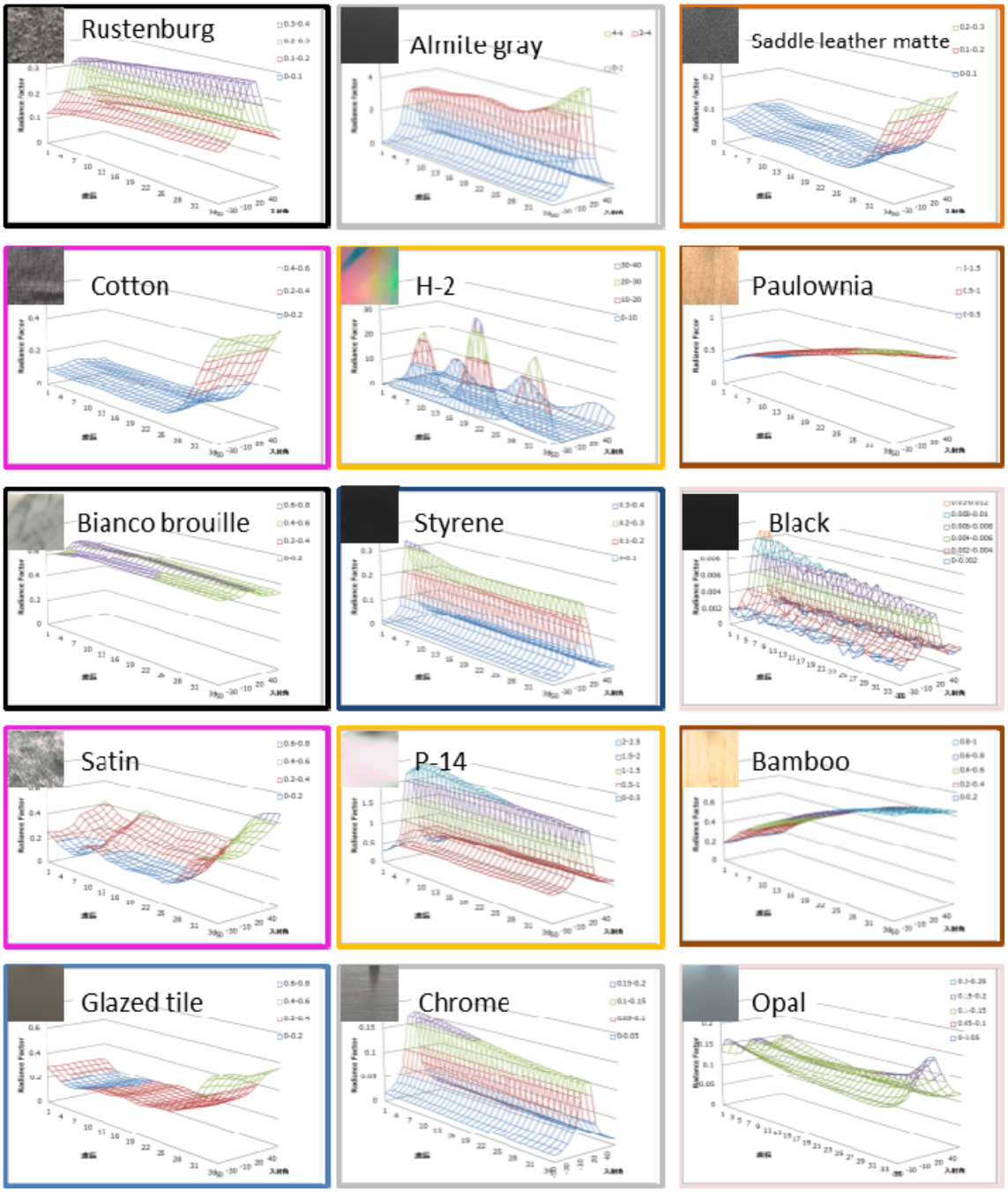
Figure 3.18. Mean quality scores for each observation pattern. From left to right, each bar shows score of the rendered image for right side eye (pattern 1), the rendered image for left side eye (pattern 2), the rendered image for eyes (pattern 3), the real object for right side eye (pattern 4), the real object for left side eye (pattern 5) and the real object for eyes (pattern 6).

3.6 Analysis between Physical Properties and Psychophysical Evaluations

In this section, the author conducts an analysis between physical properties and the psychophysical evaluations obtained from our experiments explained in the previous sections.

First, the author measured surface reflectance properties of real materials as the Bidirectional Reflectance Distribution Function (BRDF) using a Gonio spectrophotometer (Murakami Color Research Lab. GSP-2). Figure 3.19 shows the measured BRDF data of observation angle at 0 deg. for thirty real materials in our dataset. Four materials, which were difficult to measure the BRDF due to their transparency property, were excluded. The boundary color of each graph corresponds to the material categories as shown in Fig. 3.22. Figure 3.20 shows the histogram calculated from the BRDF data (λ = 550 nm, incident angle = (- 50 deg., +50 deg.), and observation angle = 0 deg) for the 30 materials. Statistical values such as kurtosis, averaged reflectance and reflectance deviation were calculated for each graph in Fig. 3.20 as physical properties.

Second, the author calculated surface texture properties of rendered images as the frequency histogram shown in Fig. 3.21. Statistical values such as lightness deviation, skewness and kurtosis were calculated for each graph in Fig. 3.21 as physical properties.


Rustenburg
Almile gray
Saddle leather matte
Cotton
H-2
Paulownia
Bianco brouille
Styrene
Black
Satin
P-14
Bamboo
Glazed tile
Chrome
Opal

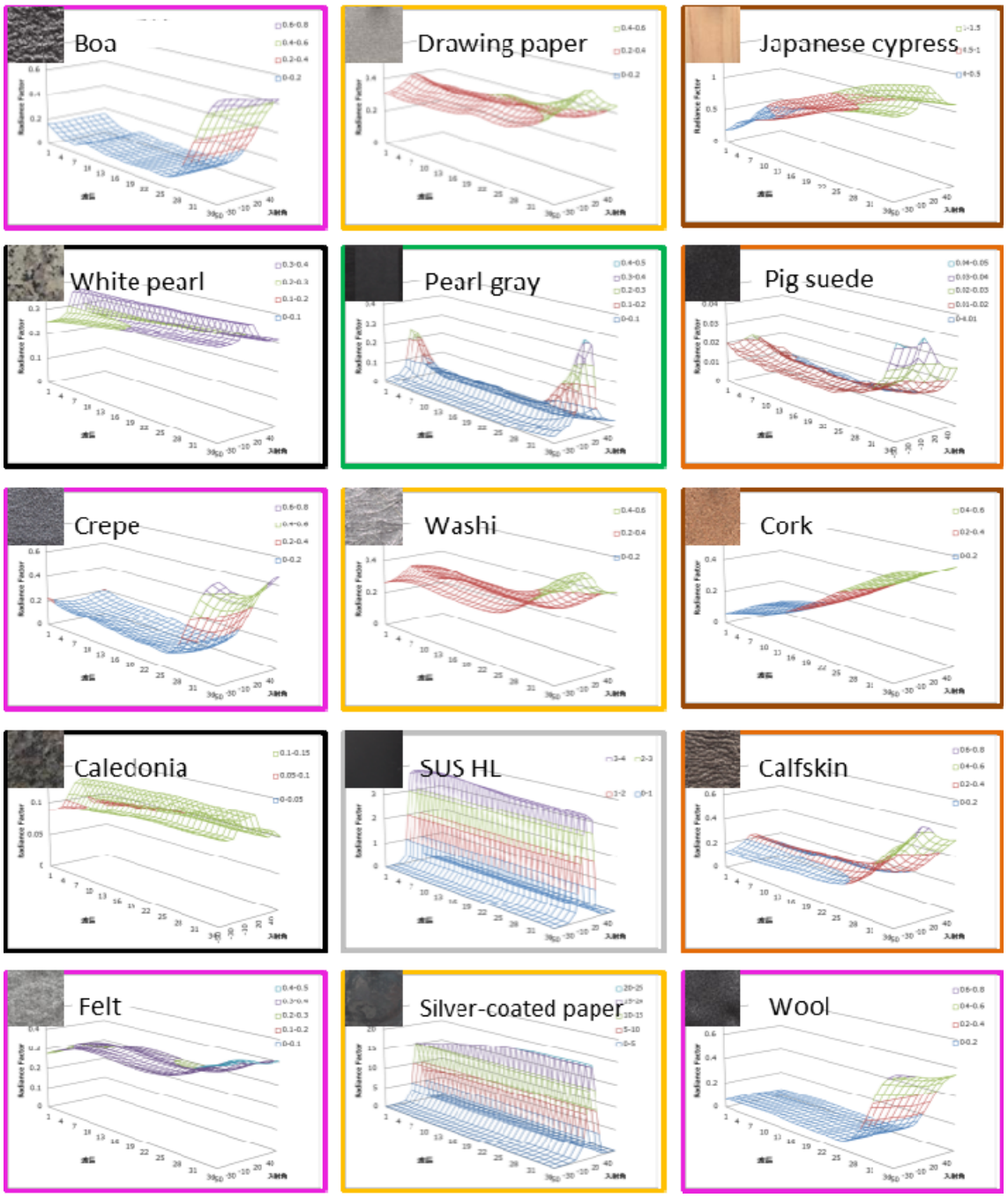
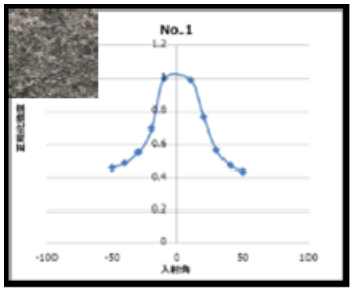
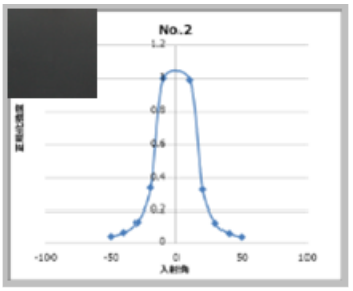


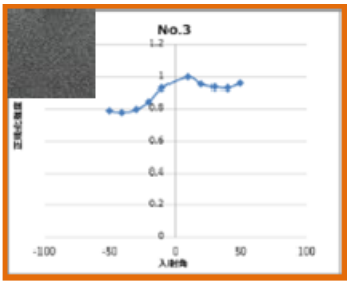
Figure 3.19. BRDF data for 30 samples.



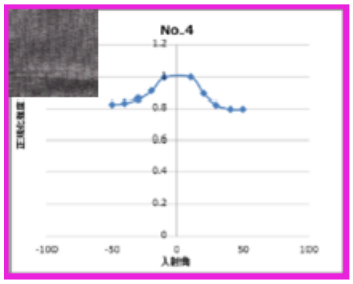
Rustenburg



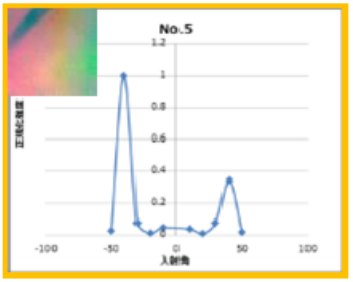
Almite gray



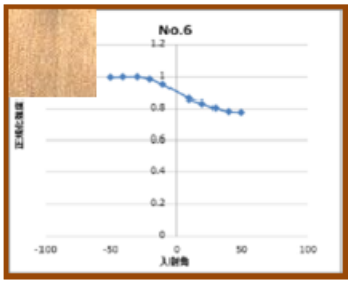
Saddle leather matte



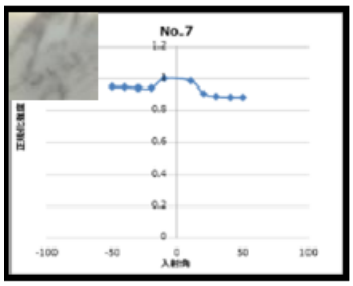
Cotton



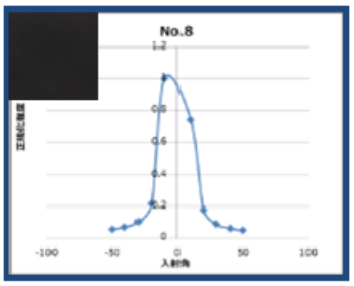
H-2



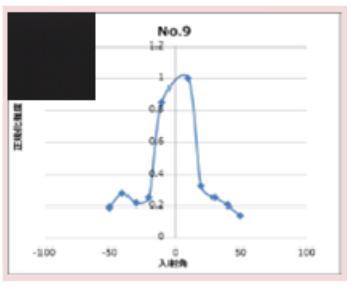
Paulownia



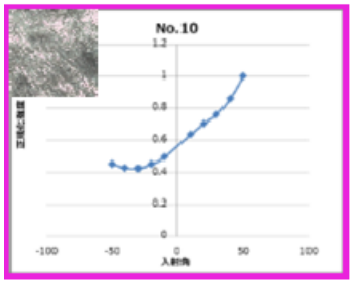
Bianco brouille



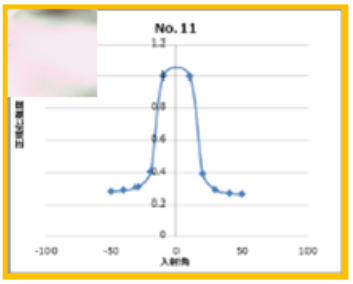
Styrene



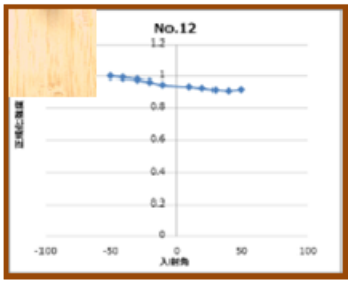
Black



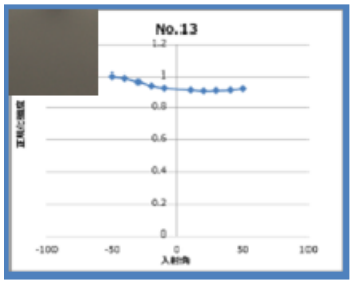
Satin



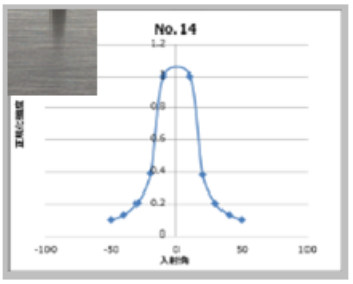
P-14



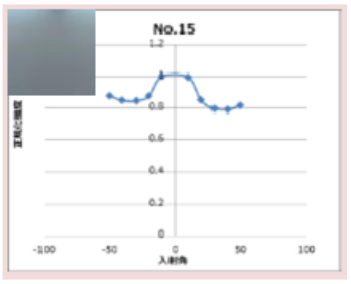
Bamboo



Glazed tile



Chrome



Opal

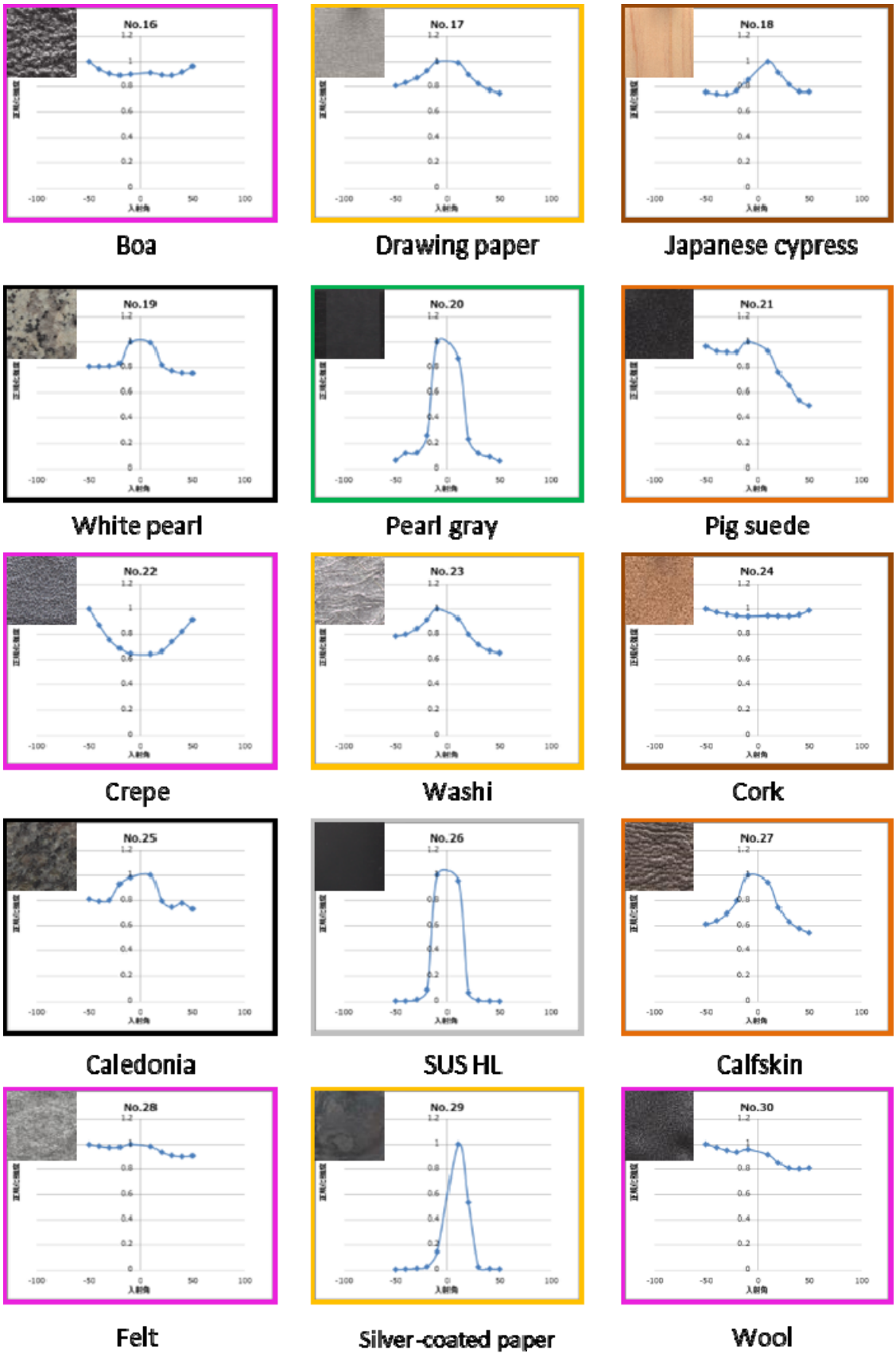
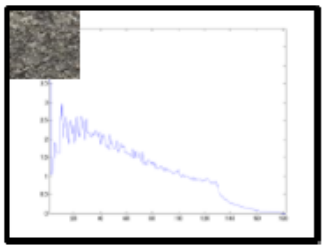
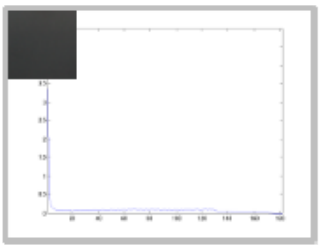


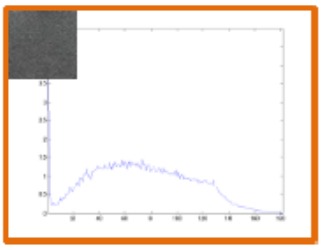
Figure 3.20. 2D Histogram extracted from BRDF data for 30 samples.



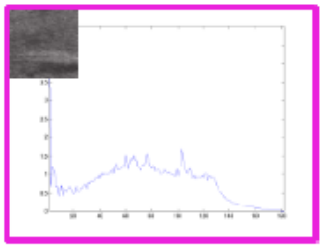
Rustenburg



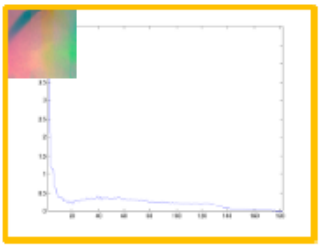
Almite gray



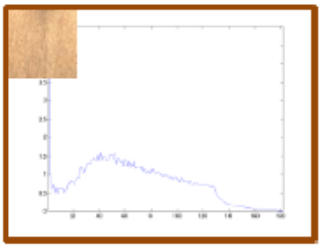
Saddle leather matte



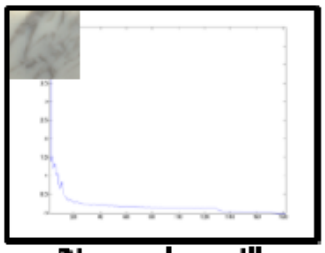
Cotton



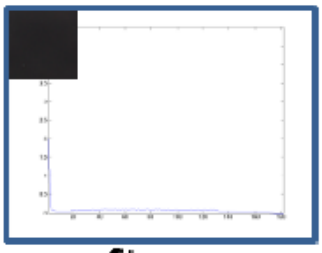
H-2



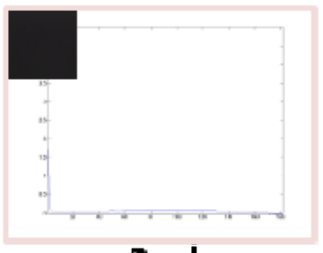
Paulownia



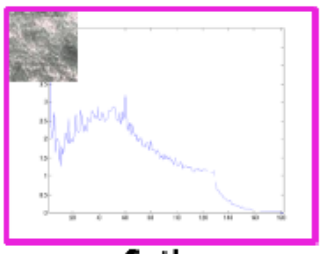
Blanco brouille



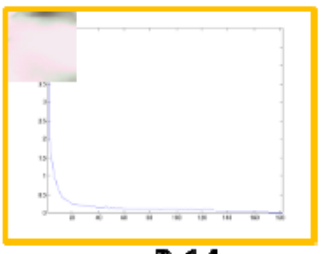
Styrene



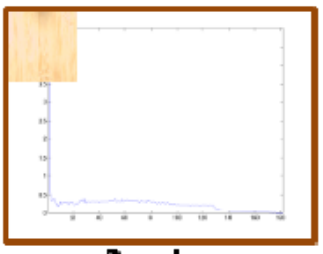
B ack



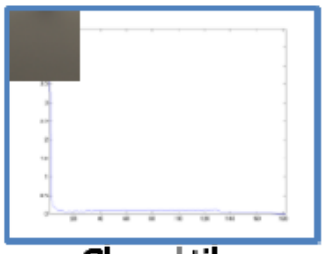
Satin



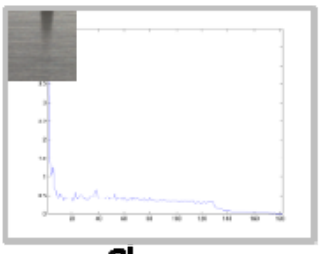
P-14



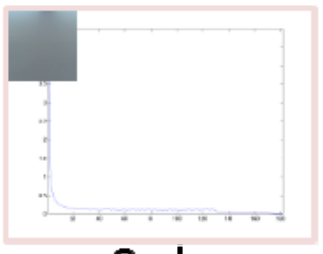
Bamboo



Glazed tile



Chrome



Opal

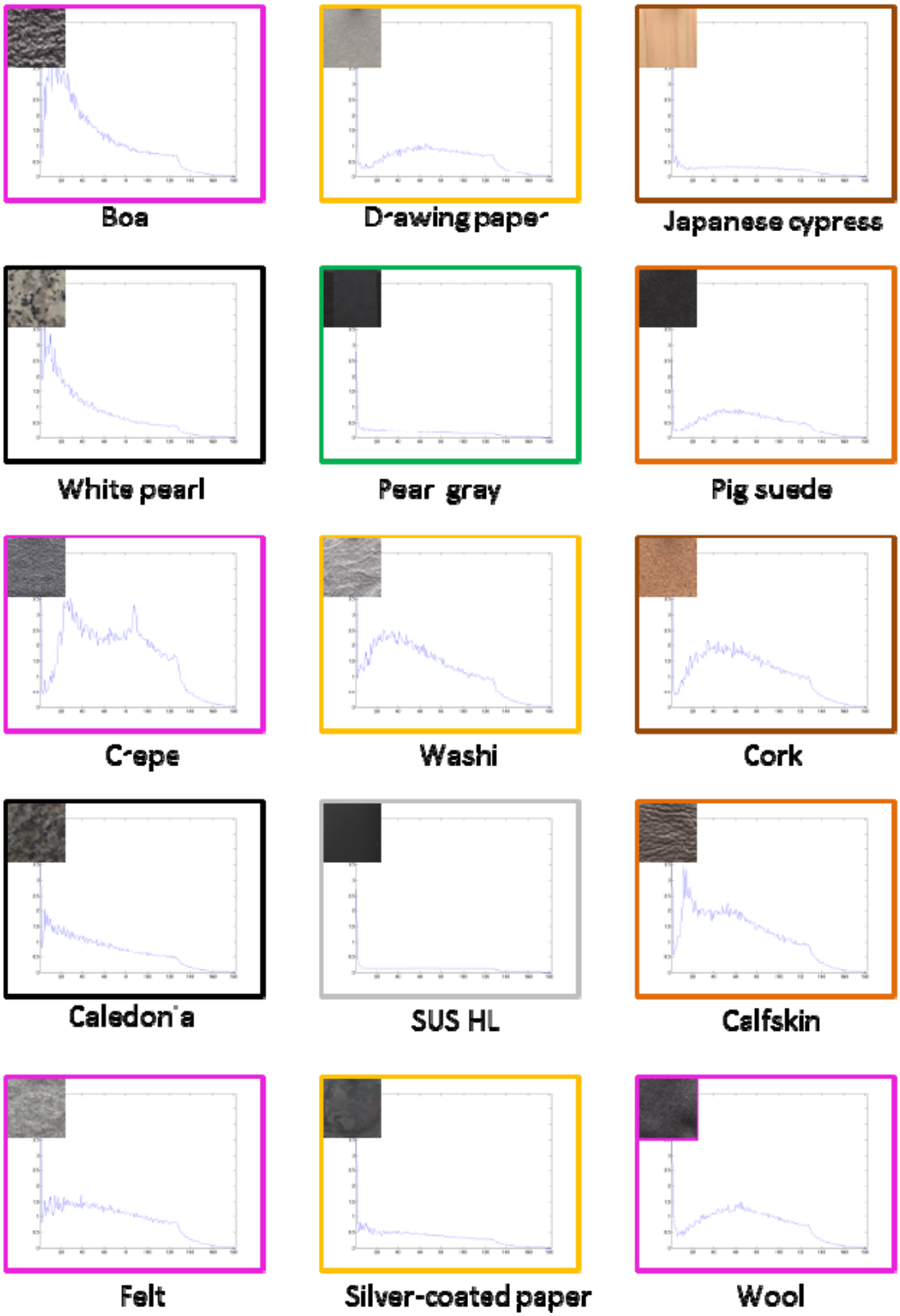


Figure3.21. Frequency histogram for 30 rendered images.

Figure 3.22. Color index.

Table 3.5 shows the correlations between physical properties and psychophysical evaluations explained in Sec. 3.4. Upper table shows the correlation with physical properties and obtained shitsukan perceptual evaluations for real-world objects. Lower table shows that for rendered color images. As shown in Table 3.5, the averaged reflectance calculated from the BRDF data and glossiness obtained from subjective assessment was the highest correlation (0.7485). Glossiness had also high correlation with kurtosis and reflectance deviation. These results indicated that human might evaluate the glossiness based on physical properties of the BRDF information as a clue. In addition, naturalness had high correlation with averaged reflectance and reflectance deviation. This suggests that human might evaluate the naturalness to real-world objects and rendered images based on the reflectance characteristics. Furthermore, roughness had high correlation with lightness deviation of rendered images. This result indicated that the roughness of rendered images could be controlled by regulating the lightness deviation.

Table 3.5. Correlations between physical properties and psychophysical evaluations.

(a) Correlations to evaluations for real-world objects.

	Real-world objects	Glo.	Tra.	Col.	Rou.	Har.	Cld.	Fla.	Nat.	Pre.
BRDF data	Kurtosis	0.6442	0.0249	0.3874	-0.3257	0.1352	0.3748	-0.1344	-0.4688	0.0565
	Averaged reflectance	-0.7458	-0.2366	-0.0086	0.3466	-0.3087	-0.5962	0.3919	0.6384	0.0059
	Reflectance deviation	0.6878	0.2749	-0.1128	-0.2955	0.3417	0.5933	-0.4457	-0.6198	0.0041
Frequency data	Lightness deviation	-0.2327	-0.1479	0.2572	0.6666	-0.3659	-0.3811	0.3784	0.4532	0.1201
	Skewness	-0.0409	0.2392	-0.3738	0.1573	0.0058	-0.0034	-0.1447	-0.1122	0.1573
	Kurtosis	0.1620	0.5075	-0.1847	-0.0707	0.0922	0.2135	-0.0464	-0.2333	0.0768

Over 0.7 / 0.6 / 0.5

(b) Correlations to evaluations for rendered images.

	Color images	Glo.	Tra.	Col.	Rou.	Har.	Cld.	Fla.	Nat.	Pre.
BRDF data	Kurtosis	0.5244	0.0392	0.3606	-0.3013	0.1439	0.3403	-0.1868	-0.4574	0.2060
	Averaged reflectance	-0.6175	-0.2253	-0.0084	0.2947	-0.3397	-0.5794	0.3896	0.6447	-0.1658
	Reflectance deviation	0.5689	0.2614	-0.0979	-0.2354	0.3682	0.5834	-0.4141	-0.6250	0.1422
Frequency data	Lightness deviation	-0.0128	-0.2032	0.1410	0.6589	-0.3896	-0.4080	0.4009	0.4216	-0.1009
	Skewness	-0.2034	0.1898	-0.3132	0.2146	0.0135	0.0490	-0.0705	-0.0551	-0.0043
	Kurtosis	0.2965	0.5262	-0.1818	-0.0143	0.0815	0.2287	-0.0538	-0.2149	0.1063

Over 0.7 / 0.6 / 0.5

3.7 Conclusions

This study investigated the perceptual qualities of static surfaces using real materials and four types of image reproductions. In order to study material perception that was not influenced by shape and saturated color, the author produced a material dataset that consisted of 34 exemplars selected from 10 material categories.

By considering intra- and inter-participant variance, the author found that the quality evaluation obtained from the richer information available from real materials was almost equivalent between the individual participants. On the other hand, the evaluation obtained from gray or low-resolution images differed between the individual participants due to the diminished information available in the reproduced images.

Through subjective assessments, the author confirmed that the representation method of materials affected perceptual qualities only in certain cases that can be classified into the following three types: (1) when reproducing images of materials, perceptions of

qualities such as “Glossiness” decreased for metal and plastic materials, whereas perceptions of the “Transparency” decreased for glass materials; (2) for gray images, perceptions of qualities such as “Glossiness” of paper and “Prettiness” of rubber materials decreased; and (3) when materials are reproduced in low-resolution images, perceptions of “Hardness” and “Coldness” tended to increase.

Correlations between perceptual qualities were observed, but the majority of these correlations were weak. However, when the materials were represented and displayed as images, their correlations gained strength. The correlations between material classes were presented as gray images. These results indicate that the qualities of these materials were strongly influenced by their color information, and in the low-resolution gray image conditions, the materials could not be discriminated as well as in the color conditions. Interestingly, the difference in image resolution hardly influenced the correlations between material classes.

The findings from the PCA and k-means clustering indicate that material categories are more likely to get confused when the materials are displayed as images, especially as gray images. The additional analysis indicated the possibility to explain the relationship between physical properties and psychophysical assessments.

Further investigations on how shape and motion influence perceptual information need to be conducted in the future to add depth and validity to the present study.

Chapter 4. Investigating Appearance Harmony of Materials

4.1 Introduction

In visual design, harmony refers to the similarity among components or objects that look like they belong together. Harmony is often related to the body, mind, and emotions in our living space, which means that the harmony of real objects is an important characteristic. Indeed, harmony might be affected by the shared traits between objects, such as their color, shape, texture, and material.

Since long, color harmony has interested researchers involved in color design studies based on various objects (Judd & Wyszecki, 1975 ; Hård & Sivik, 2001 ; Burchett, 2002). Although there are many theories related to color harmony, there appear to be a number of common shared 'principles', such as complementary hue, equal hue, equal chroma, and equal lightness. Recently, Ou et al. examined the color harmony theory and extended the harmony prediction theory from two-color combinations (Ou & Luo, 2006) to three-color combinations (Ou et al., 2010).

In contrast, other traits related to harmony have not been investigated deeply in previous studies. Though the relationship between product identity and shape has been discussed (Bar & Neta, 2006 ; Nasser & Marjan, 2010 ; Ye et al., 2014), these studies investigated the preference for a single shape, such as a kettle (Nasser & Marjan, 2010) or a chair (Ye et al., 2014), but they did not consider two-shape combinations. Chen et al. investigated the relationship between preferences for color-pairs and shapes (Chen et al., 2014), but they did not discuss two-shape combinations. In the field of texture analysis, a single texture has been used in preference analyses. In 2014, Qiao et al. began the study of texture harmony (Qiao et al., 2014).

When sensing harmony among actual objects, both harmony among colors and the appearance of harmony among materials are important considerations. Figure 4.1 shows

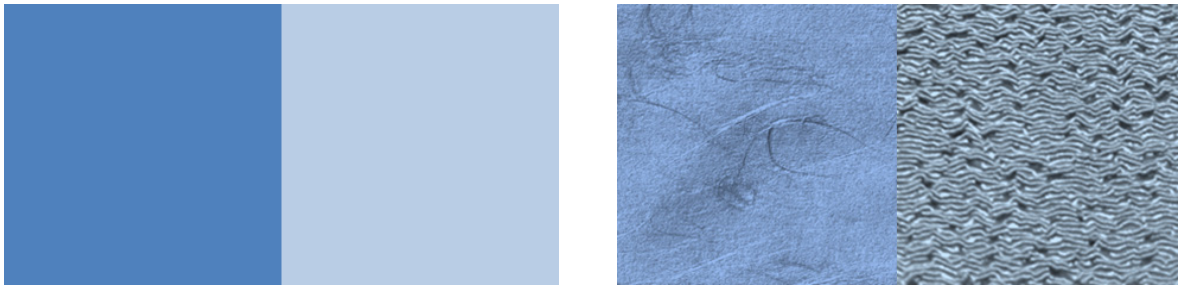
two examples of a color harmony pair. The pair in Figure 4.1(a) shows color patches typically used in color harmony studies. Figure 4.1(b) shows real materials: a *washi* (handmade Japanese paper) and a crape fabric. Basic colors shown in Figures 4.1(a) and (b) are nearly the same. According to color harmony theory, the pair in Figure 4.1 harmonizes. However, as shown in Figure 4.1(b), the appearance of materials with harmonized color does not necessarily harmonize. Although the appearance of harmony in relation to specific materials has been investigated, harmony among different materials has not received adequate attention.

Recently, the analysis of material appearance has been studied actively. Most of these studies have focused on visual estimates of specific properties of materials, such as glossiness (Fleming et al., 2003 ; Motoyoshi & Matoba, 2012 ; Nishida & Shinya, 1998), translucency (Fleming & Bülthoff), 2005 ; Bülthoff et al., 2011; Motoyoshi, 2010), or roughness (Padilla et al., 2008 ; Pont & Koenderink, 2005,2008). According to experimental studies of material harmony, most of our empirical knowledge of harmony is based on specific material clusters in the actual field of industrial design, such as combinations of wood or stone used in architecture, or combinations of metals used in car production. However, to the best of our knowledge, there have been no previous studies on the appearance harmony of materials.

Thus, in the present study, the author investigated the appearance harmony of materials based on psychophysical experiments. Although the real world comprises numerous materials, the harmony among different materials has received little attention. In this study, the author investigated the harmony across material categories. In our experiments, the author used 435 round-robin pairs of 30 samples made from 10 actual materials. The author conducted three experiments because the appearance of materials

can change greatly depending on the observation conditions. In Experiment A, the subjects were allowed to tilt the sample pairs to obtain a comprehensive assessment of harmony, which was based on the reflectance properties of the actual surface as well as the surface appearance. In Experiment B, the samples were placed such that their surfaces and viewing direction were perpendicular to the subject. Furthermore, to reflect engineering applications, static sample images were displayed on a monitor in Experiment C, and the harmony of the displayed samples was investigated.

In these experiments, subjects assessed the appearance harmony or disharmony of each sample pair based on their surface appearance. Overall, these three experiments investigated the appearance harmony of various materials.



(a) A color harmony pair without textures.

(b) A color harmony pair with textures from different materials.

Figure 4.1. Examples of color harmony pairs with/without textures.

4.2 Experimental Stimuli

4.2.1 Material Dataset

To investigate the subjects' perception of materials without being influenced by shape, the author produced a dataset of 30 exemplars (size = 50 × 50 mm). The exemplars were a partial dataset of material dataset described in Sec.2.2.1. The

individual exemplars were selected from 10 material categories, i.e., stone, metal, glass, plastic, leather, fabric, paper, wood, ceramic, and rubber, thereby covering a wide range of material appearances. The materials and their specifications have been shown in Fig. 4.2 and Table 4.1, respectively. In Table 4.1, the glossiness was measured using a glossiness checker (HORIBA IG-410) that could calculate gloss in 100% specular reflectance as 1000. It should be noted that the measured glossiness might assign approximate values for uneven surfaces or transparent materials. As noted by (Albertazzi & Hurlbert, 2013), color has a strong influence on perceptual qualities. However, it is difficult to collect exemplars of various hues with uniform material; therefore, the author only collected exemplars with low saturation. In Fig. 4.3, the symbol "×" represents the location of each exemplar on the CIE xy chromaticity diagram. The number of exemplars in each material category was unequal because they were collected according to the differences in their surface properties. The material samples were used to generate 435 round-robin pairs, which were coupled arbitrarily and presented to the subjects, i.e., two samples each time.

Figure 4.2. The dataset of 30 exemplar materials.

(The samples form a line from the upper left.)

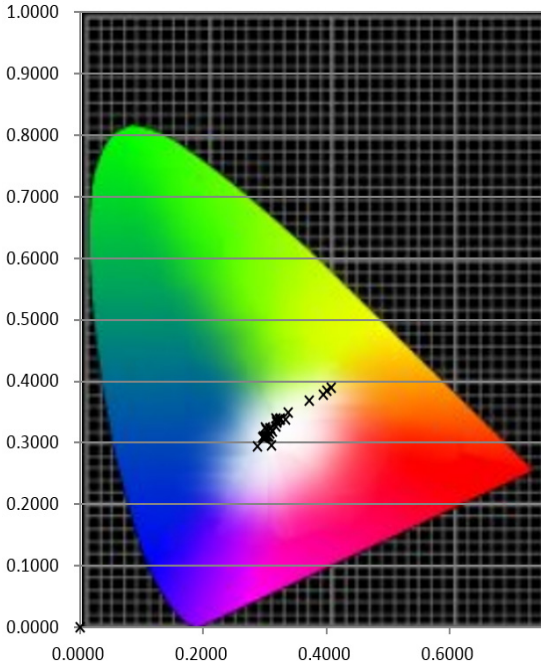


Figure 4.3. Chrominance of the material samples.

Table 4.1. Specifications of the material dataset.

Category	Remarks	L*	C^*_{ab}	h_{ab} (deg.)	Glossiness	#
Stone	Rustenburg	29.4	2.21	71.9	3.2	1
	Bianco brouille	34.7	1.56	69.2	93.3	2
	White pearl	38.3	2.74	68.8	80.0	3
	Caledonia	29.5	1.27	54.9	71.2	4
Metal	Almite gray	14.5	1.52	39.3	63.1	5
	Chrome	11.8	0.74	-82.8	241.7	6
	SUS HL	23.8	2.18	76.2	173.7	7
Glass	Pearl gray	18.5	3.42	-70.8	92.9	8
Plastic	Opal	60.4	7.22	-82.0	78.7	9
	Black	9.5	0.07	0.0	83.9	10
Leather	Saddle leather matte	24.6	1.74	75.3	1.0	11
	Pig suede	6.2	1.69	-61.2	1.2	12
	Calfskin	40.0	4.81	47.4	4.8	13
Fabric	Cotton	22.7	3.44	-53.2	2.4	14
	Satin	41.1	2.68	-62.5	1.4	15
	Boa	29.8	2.76	-36.2	1.9	16
	Crepe	31.0	7.62	-75.6	2.1	17
	Felt	47.6	1.60	-61.9	2.1	18
	Wool	26.6	3.83	-70.3	2.4	19
Paper	H-2	39.4	10.43	-39.0	340.0	20
	P-14	49.3	2.24	87.8	39.7	21
	Drawing paper	65.7	2.89	-2.7	4.0	22
	Washi (handmade)	59.2	5.25	-54.3	3.5	23
	Silver-coated paper	26.2	3.66	-81.3	230.3	24
Wood	Paulownia	29.9	17.09	68.6	6.1	25
	Bamboo	39.9	22.52	70.0	3.1	26
	Japanese cypress	60.0	19.17	68.9	5.6	27
	Cork	35.1	17.47	66.5	1.8	28
Ceramic	Glazed tile	24.4	4.37	75.1	93.9	29
Rubber	Styrene	9.9	0.18	0.0	33.6	30

4.2.2 Image Dataset

The author hypothesized that when materials are reproduced on a monitor, the following factors might strongly affect the perceptual harmony: intensity, color reproduction, and resolution. Thus, the author developed an imaging system to facilitate the accurate reproduction of the real-world display materials, where the camera system comprised an RGB camera and a standard lens. The camera used to obtain a linear output was a Canon EOS 5D Mark II, with a *sRAW2* image size of 2784 × 1856 pixels and a quantization level of 14 bits. The author then prepared a color image dataset by capturing the materials placed in a viewing booth.

The output monitor was an Apple 15.4" MacBook Pro with Retina display, where the widescreen, LED-backlit IPS screen had a glossy finish, with a native resolution of 2880 × 1800 pixels and 220 pixels per inch. The author used the same procedure as Sec. 2.2.2 to reproduce the actual display scene. Using the calibration process, the author verified that the intensity and chromaticity of the real materials and their images reproduced on the display were almost equivalent.

4.3 Experimental Methods

The author conducted three different experiments, as follows:

(1) Experiment A:

Subjects were allowed to tilt the sample pairs to obtain a comprehensive judgment of harmony based on the reflectance properties of the actual surface, as well as the surface appearance.

(2) Experiment B:

Sample pairs were placed such that their surfaces and viewing directions were perpendicular to the subject. In this experiment, subjects assessed the harmony or disharmony of each sample pair based on their two-dimensional surface appearance.

(3) Experiment C:

The static sample pairs used in Experiment B were photographed using a digital camera. Subjects assessed the appearance harmony or disharmony of each sample pair that appeared on the images displayed on a calibrated monitor.

All experiments were conducted according to the principles laid down in the Helsinki Declaration. Written informed consent was obtained from all participants. After dark adaptation for two min, the subjects evaluated the pairs according to each experimental method using a forced-choice, 10-point scale to rate harmony-disharmony. The subjects determined the appropriate rating for each combination from 1 (disharmony) to 10 (harmony) and recorded them on answer sheets. 'Harmony' was defined as a pleasing combination based on color, texture and reflectance properties obtained from the objects' surface. In each experiment, 435 pairs were evaluated and over 30 pairs were then selected from the 435 pairs, to confirm the reproducibility of the experimental results.

Figure 4.4 shows a snapshot of the evaluation of the appearance harmony during Experiment A. In Experiments A and B, each pair of samples was placed in a viewing booth (Macbeth Judge II) with a D65 ceiling light. Therefore, specular reflection did not occur on the surface. The viewing booth was set in a dark room, and the inside wall was covered with black felt. The subjects were asked to wear gloves in order to avoid the possibility of tactile effects confounding their assessments. Therefore, participants could not acquire tactile information, such as temperature or roughness from the materials

while setting up the experiment. The participants set up the materials because our experiment required 465 repetitions during the evaluation, and it was not realistic that the experimenters would place all the stimuli. In Experiment C, participants rated the harmony ratings for each pair of materials displayed on the retina display in the dark room.

In each experiment, the subjects conducted the evaluation in a specified order and they changed the evaluation samples themselves. Twenty subjects participated in this experiment. All the subjects were native Japanese with normal color vision.

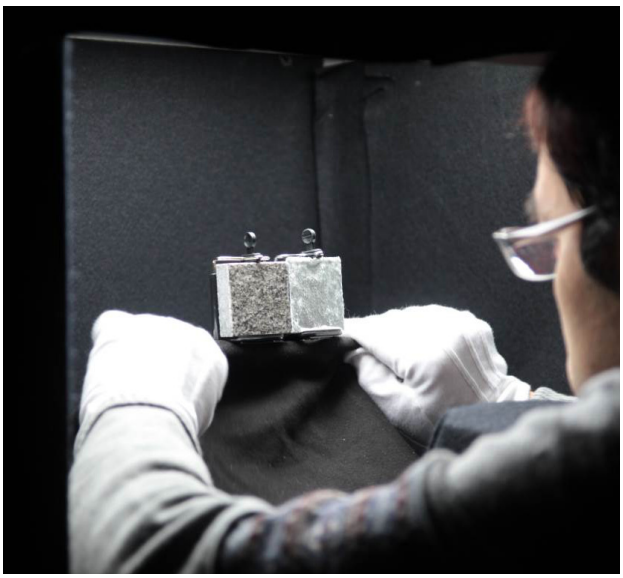


Figure 4.4. Snapshot of Experiment A.

4.4 Results and Discussion

On an average, each session required 198 min, 170 min, and 67 min for Experiments A, B, and C, respectively. Therefore, on an average, 435 min were required to complete all the three experiments.

4.4.1 Intra- and Inter- Subject Variances

The intra-participant variance was calculated as the average variance in the ratings

between the two trials for the thirty pairs, to confirm the reproducibility within each participant, as mentioned in Section 4. The intra-participant variance was defined as

$$\sigma^2_{\text{intra}}(i) = \frac{1}{20 \times 30 \times 2} \sum_{k=1}^{20} \sum_{l=1}^{30} \sum_{m=1}^{2} \left(\overline{a}_{k,l}(i) - a_{k,l,m}(i) \right)^2 , \quad (4.1)$$

where $a_{k,l,m}(i)$ is the rating for the *l*-th pairs in the *m*-th trials by the *k*-th participant, and $\overline{a}_{k,l}(i)$ is the average rating for the two trials.

The inter-participant variance $\sigma^2_{\text{inter}}(i)$ was calculated as the averaged ratings for each of the 435 pairs by the twenty participants, as follows:

$$\sigma^2_{\text{inter}}(i) = \frac{1}{20 \times 435} \sum_{k=1}^{20} \sum_{l=1}^{435} \left(\overline{b}_l(i) - b_{k,l}(i) \right)^2 , \quad (4.2)$$

where $b_{k,l}(i)$ is the rating for the *l*-th pair by the *k*-th participant, and $\overline{b}_l(i)$ is the average rating by the twenty participants.

Table 4.2 summarizes the intra- and inter-participant variances in the ratings. The intra-subject variance was based on 30 samples, which were presented twice. The inter-subject variance in the right row of Table 4.2 shows the average variance of the ratings among the 435 samples. As shown in Table 4.2, the author confirmed that the variance in the intra-subject ratings was remarkably less than the variance in the inter-subject ratings.

The observed variance had one notable feature, as shown in Table 4.2, i.e. the ratings in Experiments A and B varied among subjects, whereas the ratings in Experiment C were stable. This suggests that the richness of the real-world information was sensitive to the perceptual harmony ratings among subjects. In contrast, intra-subject variances were almost constant through all three experiments.

Table 4.2. Variances in the Inter- and intra-participant ratings.

Experiment	Intra-variance	Inter-variance
A	0.36	4.53
B	0.35	4.49
C	0.39	4.12

4.4.2 Perceptual Harmony Ratings within and across the Categories of Materials

Table 4.3 summarizes the average perceptual harmony ratings for all subjects within the same material category and across different material categories in each experiment. As shown in Table 4.3, the ratings for sample pairs within the same material category were higher than those across material categories. In all the experiments, the 'Metal-Metal' pair had the highest harmony ratings (Exp. A: 7.70, Exp. B: 7.85, Exp. C: 8.40). By contrast, the harmony ratings for the 'Paper-Paper' (Exp. A: 2.30, Exp. B: 2.60, Exp. C: 2.35) and the 'Leather-Leather' (Exp. A: 3.75, Exp. B: 3.60, Exp. C: 3.05) pairs were categorized as showing perceptual disharmony (< 5.5, i.e. the boundary score between harmony and disharmony). These results suggest that the perceptual harmony ratings depended on the materials, where two samples within the same material category could be perceived as having appearance disharmony. Interestingly, the harmony rating for the 'Paper-Metal' pair (Exp. A: 5.80, Exp. B: 5.55, Exp. C: 5.80) was higher than that for the 'Paper-Paper' pair. This indicates that the perceptual harmony of the pairs in different material categories could be higher than that of pairs within the same material category.

Table 4.3. Averaged harmony ratings for the categories of materials.

Experiment	Within category	Across categories
A	5.71	4.04
B	5.89	4.07
C	5.58	4.15

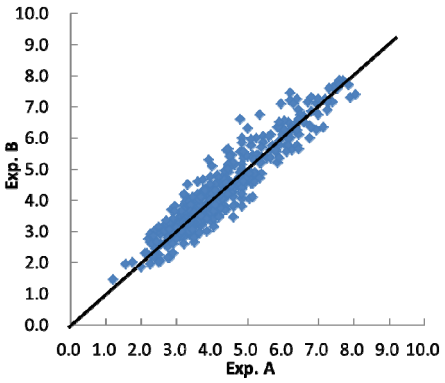
Metal, plastic, ceramic and rubber shared the most harmony with each other across material categories as follows: 'Metal-Plastic' (Exp. A: 5.93, Exp. B: 5.80, Exp. C: 6.49), 'Metal-Ceramic' (Exp. A: 6.60, Exp. B: 6.61, Exp. C: 7.08), 'Metal-Rubber' (Exp. A: 5.54, Exp. B: 6.06, Exp. C: 6.81), 'Plastic-Ceramic' (Exp. A: 7.20, Exp. B: 6.88, Exp. C: 6.98), 'Plastic-Rubber' (Exp. A: 6.08, Exp. B: 6.83, Exp. C: 6.68) and 'Ceramic-Rubber' (Exp. A: 6.20, Exp. B: 7.45, Exp. C: 6.75). The 'Glass-Plastic' pair also exhibited a high perceptual harmony rating (Exp. A: 6.40, Exp. B: 6.28, Exp. C: 6.03). These results indicate that materials can be harmonized across material categories.

Figure 4.6 shows the averaged harmony ratings for each sample pair. The colors indicate the averaged ratings as specified by the color bar in Fig. 4.6(d). Red indicates harmony whereas blue indicates disharmony. As shown in Fig. 4.6, the harmony ratings close to the diagonal, which indicate harmony within the same material category, were generally high. However, as described earlier, metal, plastic, ceramic, and rubber were harmonized among material categories, as indicated by the red dotted lines in Fig. 4.6. Moreover, high harmony ratings were obtained between different material samples such as Pair 11 (saddle leather matte, leather) and 22 (drawing paper, paper), as indicated by the solid yellow line. Regardless of whether the materials in the pair belonged to the same category, some materials were in disharmony with other material samples, such as those in Pair 13 (gray calfskin, leather).

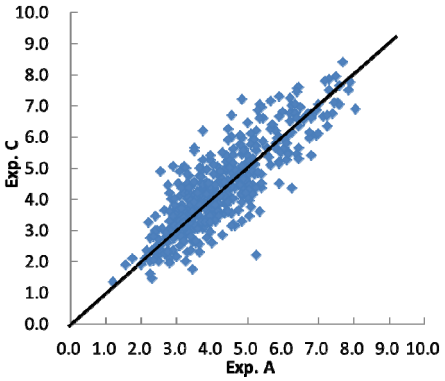
4.4.3 Changes in Harmony among Experiments

As shown in Table 4.3, the harmony ratings were lowest in Experiment C, for the combination within material category, but the opposite result was obtained for the combination across material categories. This result suggests that the harmony was not sensitive of the material categories obtained from the rendered images. The average ratings in Experiments A and B did not differ significantly.

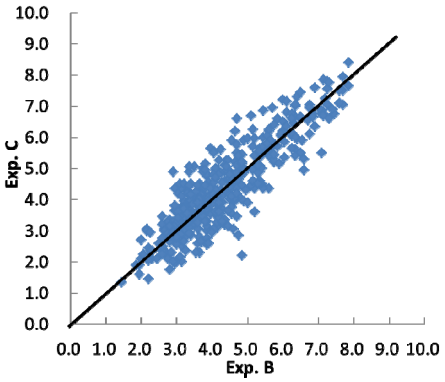
Figure 4.5 shows scatter graphs of harmony ratings between experiments. As shown in the graphs, the correlation using real materials between Experiments A and B, was higher than that between real materials (Experiments A and B) and rendered images (Experiment C). However, harmony ratings had generally high correlation between experiments. Here, there were some notable differences between the experiments. For the 'Leather-Rubber' pair, the average rating changed from disharmony in Experiment A (4.28), to harmony in Experiments B (4.64) and C (5.08). In this case, the reflectance property was very sensitive to the appearance harmony. For the 'Metal-Leather' pair, the average rating changed from disharmony in Experiments A (3.86) and B (4.36), to harmony in Experiment C (5.03). In this case, the appearance of the material may have differed between the real objects and the displayed images.



(a) Correlation between Experiments A and B (R^2=0.89).



(b) Correlation between Experiments A and C (R^2=0.71).

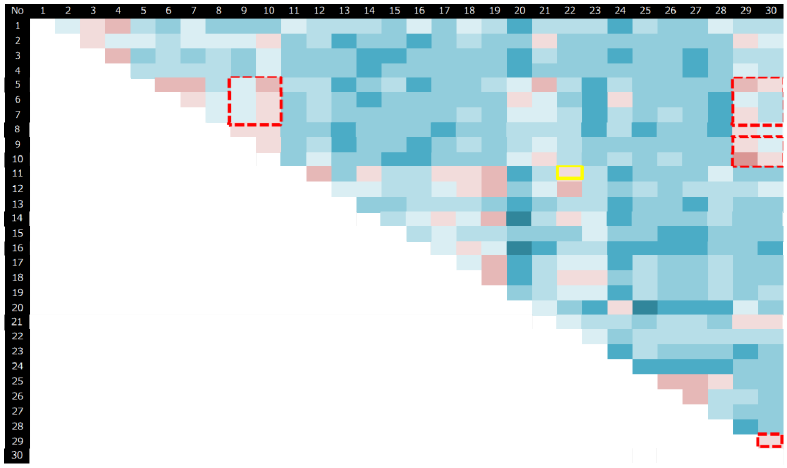


(c) Correlation between Experiments B and C (R^2=0.78).

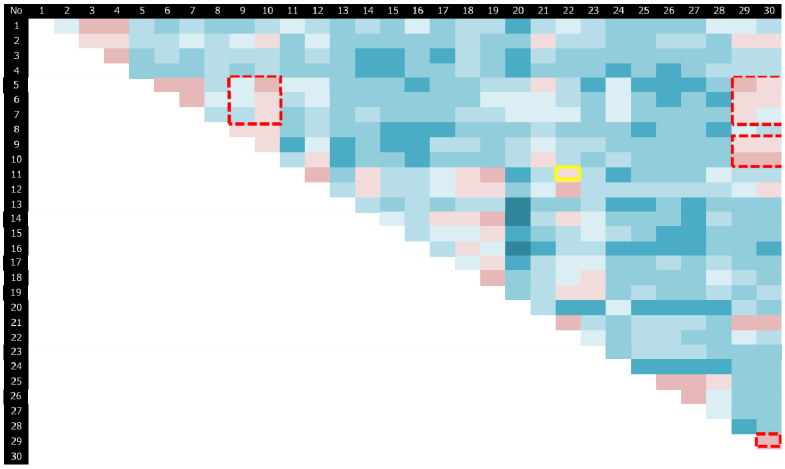
Figure 4.5. Scatter graphs of averaged harmony ratings between experiments.

As shown in Fig. 4.6, some of the ratings for each pair differed in the experiments. Table 4.4 shows the ratings and pairs that changed greatly between experiments. The average ratings across all subjects changed by a maximum of + 1.8 (Pair 15 and 19) and by a minimum of - 1.15 (Pair 5 and 15) between Experiments A and B. In particular, Pair 15 (satin, fabric) and 19 (wool, fabric), as shown in Figure 4.7 (a), had a low rating in Experiment A because the reflective properties made their appearances differ greatly. By contrast, their rating was high in Experiment B because both surfaces resembled the same rough fabric. Pair 5 (almite gray, metal) and 15 (satin, fabric), as shown in Figure 4.7 (b), had a high rating in Experiment A because these glossy surfaces with gray color had similar reflective properties. In contrast, a low rating was obtained in Experiment B due to the differences in the surface information such as roughness and color. In these cases, the appearance harmony was highly sensitive to the reflectance properties.

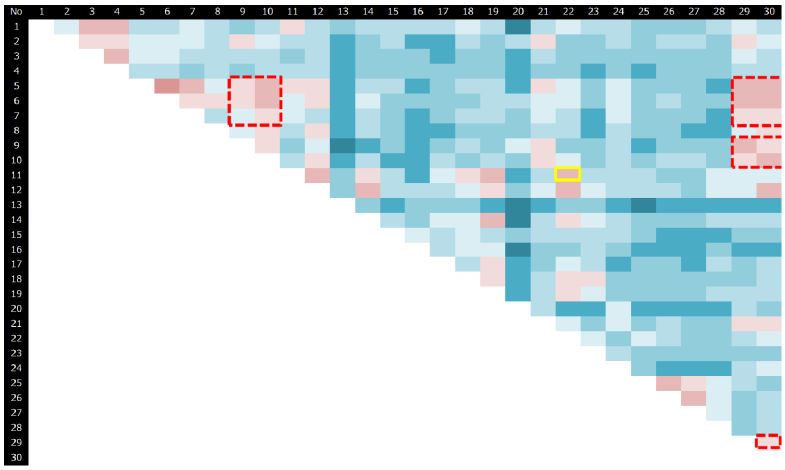
In Experiment C, the average ratings across all subjects changed by a maximum of +2.45 (Pair 8 and 12) from Experiment A to C and by a maximum of +2.00 (Pair 11 and 24) from Experiment B to C. Pair 8 (pearl gray, glass) and 12 (pig suede, leather), and Pair 11 (saddle leather matte, leather) and 24 (silver coated paper, paper) as shown in Figures 4.7 (c) and 4.7 (d), had low ratings in Experiments A and B because their surface properties differed greatly in appearance. However, this difference could not be perceived when this pair was displayed on the monitor in Experiment C. In contrast, the average rating across all subjects changed by a minimum of - 3.05 (Pair 5 and 20) from Experiment A to C and by a minimum of - 2.65 (pair 5 and 20) from Experiment B to C. Pair 5 (almite gray, metal) and 20 (H-2, paper), as shown in Figure 4.7 (e).



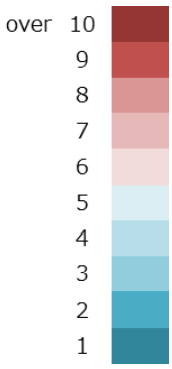
(a) Experiment A.



(b) Experiment B.



(c) Experiment C.



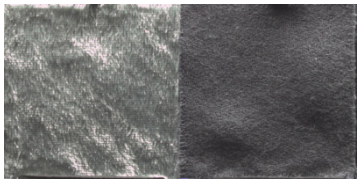
(d) Index.

Figure 4.6. Harmony ratings provided by the twenty subjects. The vertical and horizontal numbers correspond to the material numbers for the samples on the left and right, respectively, as presented in Table 4.1.

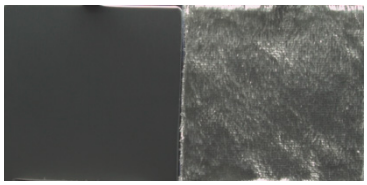
Table 4.4. Harmony changes between experiments.

Experiment	Up (pair)	Down (pair)
A→B	+1.80 (15–19)	-1.15 (5–15)
A→C	+2.45 (8–12)	-3.05 (5–20)
B→C	+2.00 (11–24)	-2.65 (5–20)

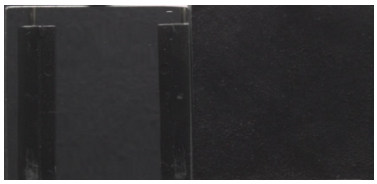
Pair 5 (almite gray, metal) and 20 (H2 hologram, paper) had a high rating in Experiments A and B because, although they were different material types, they had similar reflective properties. However, in Experiment C, a low rating was obtained for the holographic color displayed on the monitor, as shown in Figure 4.7 (e). Pair 11 (saddle leather matte, leather) and 24 (silver coated paper, paper) had a low rating in Experiment B because their surface properties and textures were differ. However, in Experiment C, a high rating was obtained due to the low resolution of the monitor, as shown in Figure 4.7 (d). In these cases, the material appearance may have differed between the real objects and the displayed images.



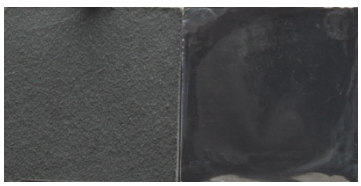
(a) Pair 15 (satin, fabric) and 19 (wool, fabric). Rating A: 4.80, B: 6.60.



(b) Pair 5 (almite gray, metal) and 15 (satin, fabric). Rating A: 4.60, B: 3.45.



(c) Pair 8 (pearl gray, glass) and 12 (pig suede, leather). Rating A: 3.75, C: 6.20.



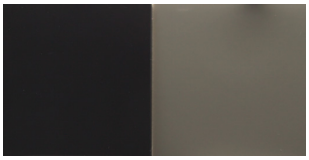
(d) Pair 11 (saddle leather matte, leather) and 24 (silver coated paper, paper). Rating B: 2.90, C: 4.90.

(e) Pair 5 (almite gray, metal) and 20 (H-2, paper). A: 5.25, B: 4.85, C: 2.20.

Figure 4.7. Pairs that changed significantly between experiments.

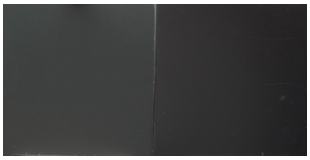
Figures 4.8 (a) and (b) show the five pairs with the highest and lowest harmony rating for each experiment, respectively. As shown in Fig. 4.8 (a), the pairs in the same material category, with similar texture, color and reflectance property had high harmony

ratings. On the other hand, the pairs of sample number 20 (H-2, paper) had low harmony ratings as shown in Fig. 4.8 (b). Figure 4.9 presents the pairs evaluated as harmony rating '10' for each experiment from over five participants. In every experiment, pair 5 (almite gray, metal) and 6 (chrome, metal) obtained the maximum harmony rating '10' even though both materials were not equivalent. Four participants did not give harmony rating '10' through all experiments. This result shows that 80 % participants were not simply to evaluate the similarity of the materials as the appearance harmony.



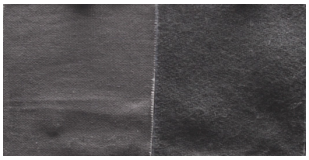
Pair 10 (black, plastic) and 29 (glazed tile, ceramic). Rating: 8.05.

Pair 5 (almite gray, metal) and 6 (chrome, metal). Rating: 7.85.



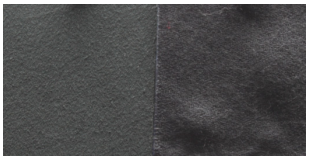
Pair 5 (almite gray, metal) and 6 (chrome, metal). Rating: 8.40.

Pair 11 (saddle leather matte, leather) and 12 (pig suede, leather). Rating: 7.90.



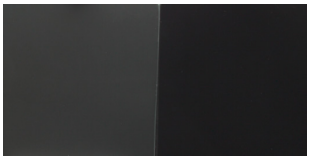
Pair 14 (cotton, fabric) and 19 (wool, fabric). Rating: 7.85.

Pair 1 (rustenburg, stone) and 4 (Caledonia, stone). Rating: 7.95.



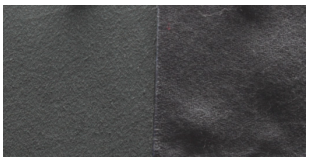
Pair 11 (saddle leather matte, leather) and 19 (wool, fabric). Rating: 7.85.

Pair 25 (paulownia, wood) and 26 (bamboo, wood). Rating: 7.70.

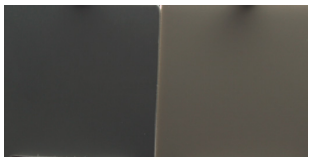


Pair 5 (almite gray, metal) and 10 (black, plastic). Rating: 7.85.

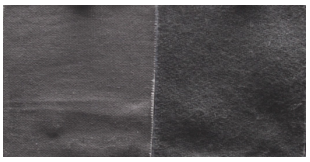
Pair 5 (almite gray, metal) and 6 (chrome, metal). Rating A: 7.70.



Pair 11 (saddle leather matte, leather) and 19 (wool, fabric). Rating B: 7.70.

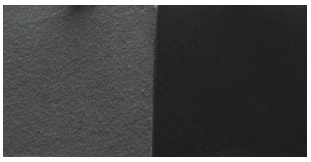


Pair 5 (almite gray, metal) and 29 (grazed tile, ceramic). Rating C: 7.80.



Pair 14 (cotton, fabric) and 19 (wool, fabric). Rating A: 7.60.

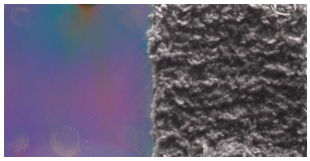
Pair 1 (rustenburg, stone) and 4 (Caledonia, stone). Rating B: 7.70.



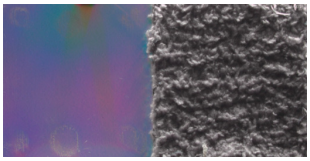
Pair 11 (saddle leather matte, leather) and 12 (pig suede, leather). Rating C: 7.75.

Experiment A. Experiment B. Experiment C.

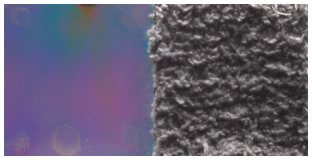
(a) Pairs with the best five harmony ratings for each experiment.



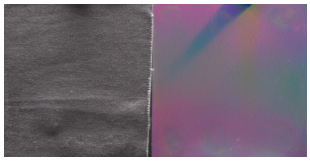
Pair 20 (H-2, paper) and 16 (boa, fabric). Rating: 1.20.



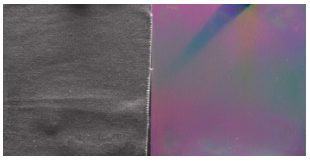
Pair 20 (H-2, paper) and 16 (boa, fabric). Rating: 1.45.



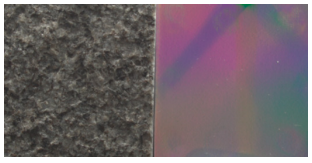
Pair 20 (H-2, paper) and 16 (boa, fabric). Rating: 1.35.



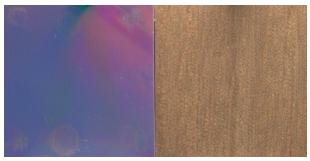
Pair 14 (cotton, fabric) and 20 (H-2, paper). Rating: 1.55.



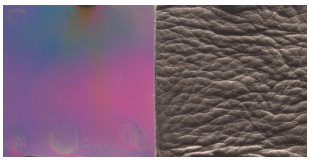
Pair 14 (cotton, fabric) and 20 (H-2, paper). Rating: 1.85.



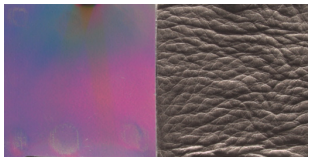
Pair 1 (rustenburg, stone) and 20 (H-2, paper). Rating: 1.45.



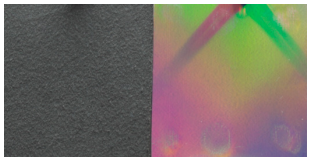
Pair 20 (H-2, paper) and 25 (paulownia, wood). Rating: 1.75.



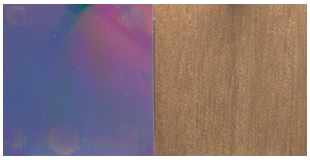
Pair 20 (H-2, paper) and 13 (calfskin, wood). Rating: 1.95.



Pair 20 (H-2, paper) and 13 (calfskin, wood). Rating: 1.60.



Pair 11 (saddle leather matte, leather) and 20 (H-2, paper). Rating: 2.10.

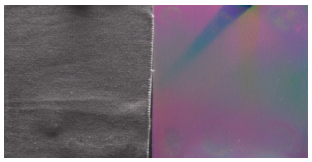


Pair 20 (H-2, paper) and 25 (paulownia, wood). Rating: 2.00.

Pair 25 (paulownia, wood) and 13 (calfskin, wood). Rating: 1.75.

Pair 20 (H-2, paper) and 26 (bamboo, wood). Rating: 2.15.

Pair 26 (bamboo, wood) and 24 (silver-coated paper, paper). Rating: 2.00.



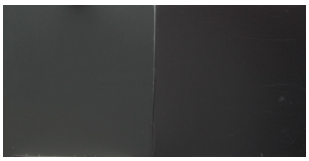
Pair 14 (cotton, fabric) and 20 (H-2, paper). Rating: 1.90.

Experiment A. Experiment B. Experiment C.

(b) Pairs with the worst five harmony ratings for each experiment.

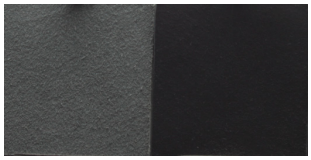
Figure 4.8. Pairs with the highest and lowest harmony ratings.

Pair 10 (black, plastic) and 29 (glazed tile, ceramic).

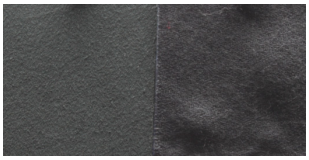


Pair 5 (almite gray, metal) and 6 (chrome, metal).

Pair 5 (almite gray, metal) and 6 (chrome, metal).

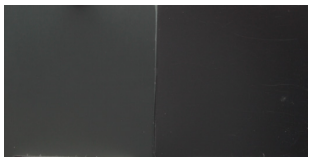


Pair 11 (saddle leather matte, leather) and 12 (pig suede, leather).

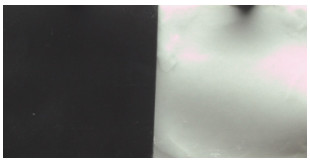


Pair 11 (saddle leather matte, leather) and 19 (wool, fabric).

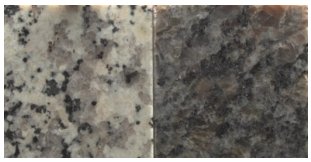
Pair 26 (bamboo, wood) and 27 (Japanese cypress, wood).



Pair 5 (almite gray, metal) and 6 (chrome, metal).



Pair 30 (styrene, rubber) and 21 (P-14, paper).



Pair 3 (white pearl, stone) and 4 (Caledonia, stone).

Pair 5 (almite gray, metal) and 29 (grazed tile, ceramic).

Experiment A. Experiment B. Experiment C.

Figure 4.9. Pairs with the maximum harmony rating ‘10’ for each experiment.

4.4.4 Distributions of Samples in the Appearance Harmony Space

The author performed a PCA of all the ratings across materials to facilitate the visualization of the distributions of material classes in the appearance harmony feature space. The author created a 30 × 30 diagonal matrix where the lines and columns represented the 30 material samples. Each element in the matrix showed the average rating for a pair, where the author assumed symmetry among the rating. The diagonal components were postulated to be the maximum ratings, because the author considered that a combination of the same stimuli should be harmonized and the maximum rating was a reasonable assumption. The author derived 30 dimensions and 30 PCs from the matrix. Therefore, materials with the same harmony properties had the same PC coefficients.

Table 4.5 shows the percentage of variance explained by the first three PCs, the first five PCs, and first ten PCs. For the first PC, the amount of variance was difference among experiment. However, for the first three PCs, a similar amount of variance was explained in all of the experiments. Figure 4.10 presents the scree plot, which shows that the three factors explained variability because the tilt becomes smooth after the three factors are presented, and five factors explained most of the variability because the line starts to straighten after the five factors are presented. Thus, regardless of the methods used to determine the appearance harmony among materials, the author can obtain an approximation of the overall distribution by simply using the first three PCs.

Table 4.5. Percentage variance explained by the first three PCs.

	PC1	PC2	PC3	PC5	PC10
Experiment A	0.398	0.546	0.660	0.795	0.915
Experiment B	0.348	0.538	0.656	0.812	0.920
Experiment C	0.313	0.518	0.656	0.794	0.918

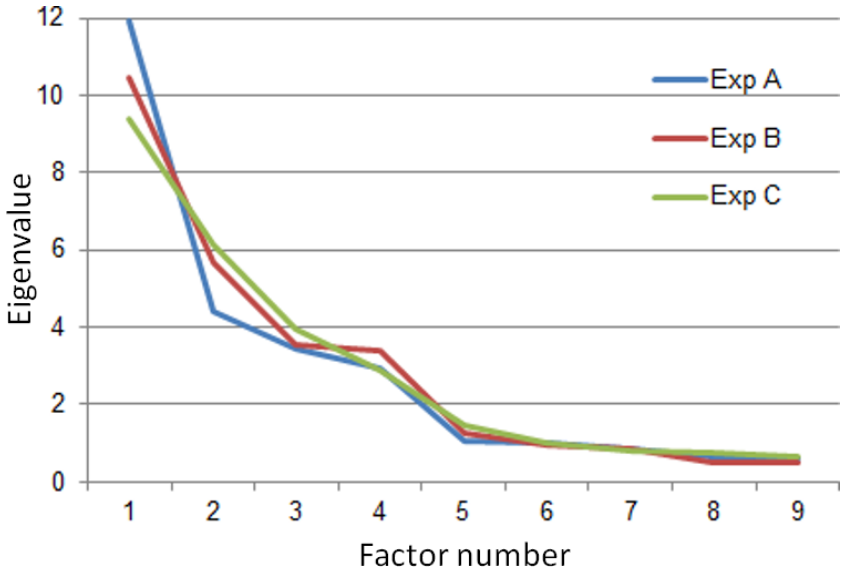


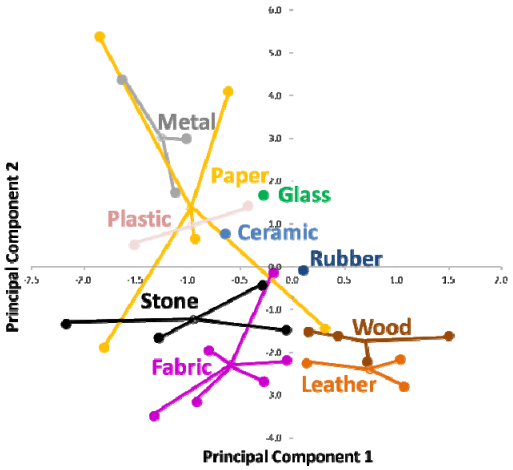
Figure 4.10. Scree plot for the PCs in all three experiments.

Figure 4.11 shows the ratings for each sample projected onto the first two PCs where each image is color coded by its true class membership. The open circles indicate the average for each material class and the same color corresponds to the same material category. The observed distribution in the PC space has several key features. As shown in Figure 4.11 (a), most of the material categories were isolated except for papers in a two-dimensional space in Experiment A. This indicates that the degree of harmony depended on the material clusters formed by real objects that could be moved. In contrast, with the exceptions of metal samples, the material categories overlapped with each other in Experiment B, as shown in Figure 4.11 (b). This indicates that the degree of harmony did not depend on the material clusters in a stationary state, unlike the

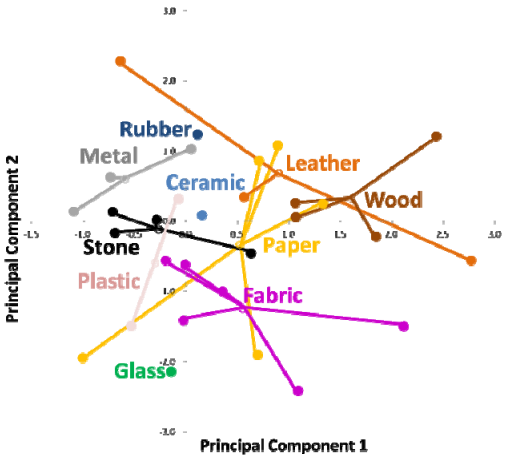
objects that could be moved. Furthermore, most of the material categories overlapped with each other in Experiment C, as shown in Figure 4.11 (c). This suggests that the degree of harmony did not depend on the material clusters in the displayed images.

Figure 4.12 shows the material samples projected onto the components in the first two PC spaces in Figure 4.11. The boundary color of each component corresponds to the material clusters. The author classified the samples with common properties, which have been encircled using a broken line. The property of glossiness has been presented in Table 4.1. Figure 4.13 shows the glossiness map of each sample. It was difficult to measure roughness of all materials under the same condition; therefore, the author judged the roughness property subjectively.

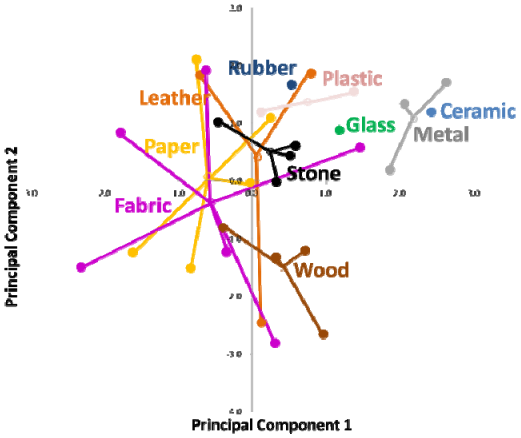
In Experiment A, the author found that the reflectance properties of material surfaces comprised an important factor that affected the appearance harmony. As shown in Fig. 4.12 (a), glossy and matte surfaces categories were clearly separated, and smooth flat and rough textured surfaces were also separated. In Experiment B, the smooth flat and rough textured surfaces categories in Experiments A and B disappeared in the two-dimensional PC space. Thus, the roughness properties might have been reduced by not tilting the sample pairs. In Experiment C, glossy and matte surfaces categories were overlapped. Glossiness properties might have been reduced by displaying the materials on the monitor.



(a) Experiment A.

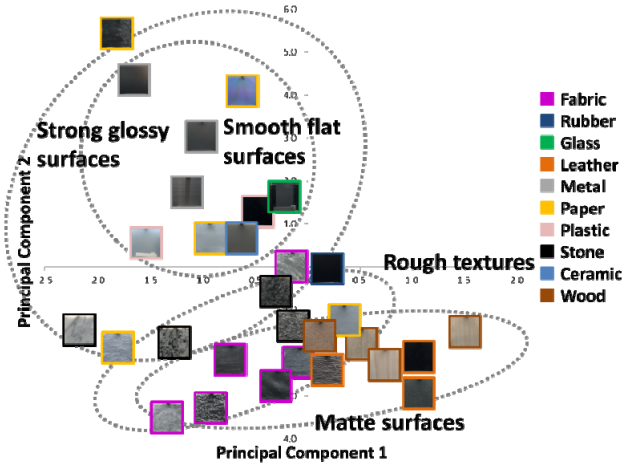


(b) Experiment B.

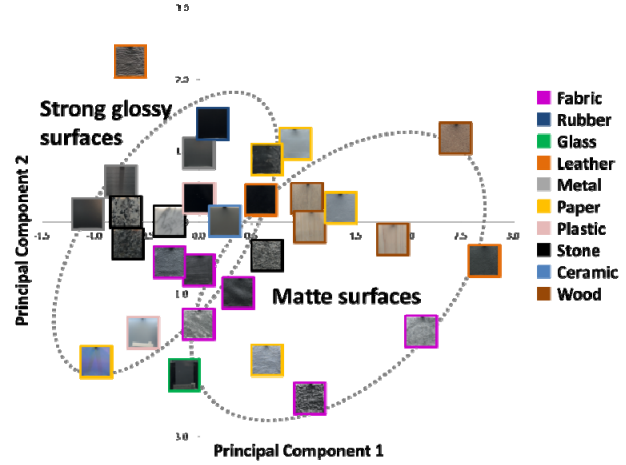


(C) Experiment C.

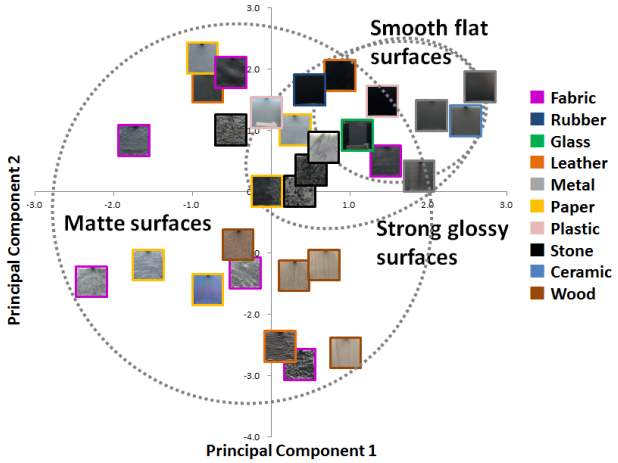
Figure 4.11. Distribution of samples in the first two PCs.



(a) Experiment A.

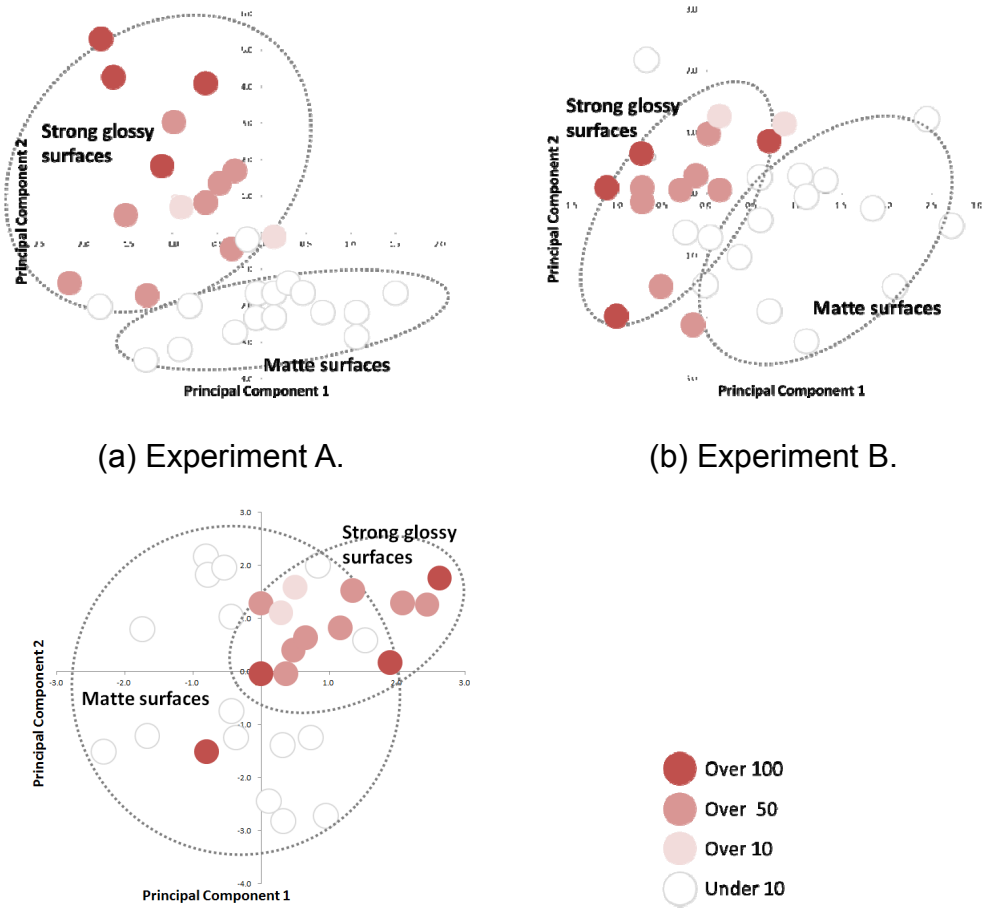


(b) Experiment B.



(c) Experiment C.

Figure 4.12. Samples categorized according to the similarities in appearance harmony.



(a) Experiment A.

(b) Experiment B.

(c) Experiment C.

(d) Index.

Figure 4.13. Glossiness maps on PC1-PC2 for each experiment.

The relationship between the paper and fabric groups is a good example. In Experiment A, the paper and fabric samples were plotted in different areas of the two-dimensional PC space. However, some materials from the paper and fabric groups were plotted close together because their glossy appearances were similar. These groups overlapped in Experiment C, but these materials were not plotted close together.

The author also applied k-means clustering to the harmony rating data. By comparing the true clusters with those extracted by k-means clustering, the author could measure the extent to which the data from a given category were clumped together in the feature

space. The results obtained by k-means clustering depend on the initial settings of the seeds. Therefore, the author distributed initial seeds, at random 10 times. Figure 4.14 shows the clustering results when k = 5. Experiments A and B obtained the same clustering results indicated as red line in Fig. 4.14, which shared similar harmony properties. The clustering result in Experiment C has been indicated using a blue line in Fig. 4.14.

It should be noted that k-means clustering was performed using 30-dimensional harmony rating data, whereas the plot shown in Fig. 4.11 is based on two-dimensional PCs. The materials surrounded by black bold lines were classified into the same cluster, which shared similar harmony properties. In all of the experiments, wood samples were isolated from the other materials. As described above, color might be a strong feature that affects the appearance harmony of materials. Metals and plastics were classified into the same cluster. In Experiments A and B, leather and fabric were also classified into the same cluster, whereas they were classified into different clusters in Experiments C, which supports the results shown in Fig. 4.11. The other materials were separated into different clusters. These results indicate that some sample pairs were viewed as harmonious, although the materials were different.

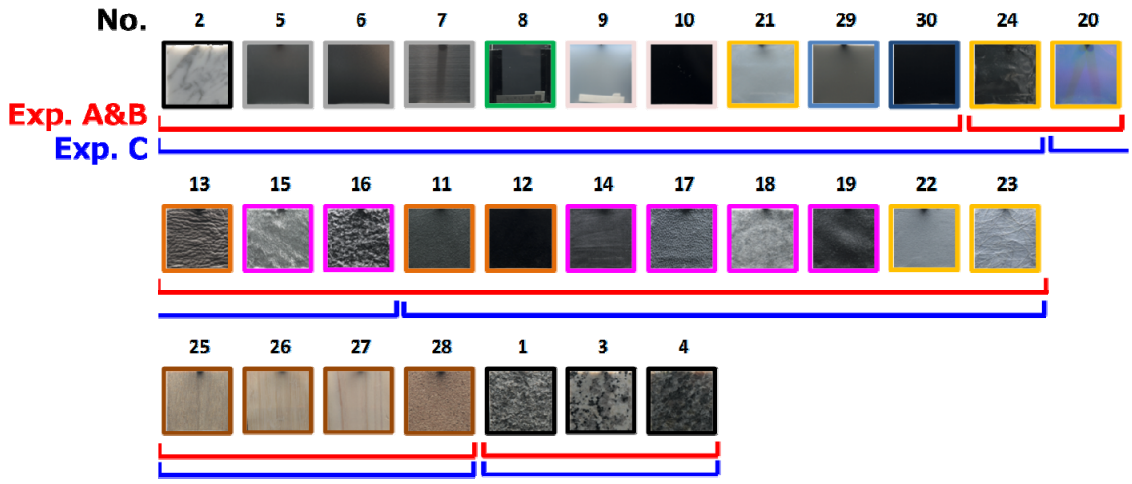


Figure 4.14. Category transition among experiments in harmony rating data using k-means algorithm (*k* = *5*).

As explained before, clustered results between Experiments A and B were equivalent. Interestingly, the one-dimensional ordering in Experiment C was also equivalent, and only the partitions between clusters were shifted. This result suggests that the overall structure of the appearance harmony was equivalent in all experiments, and the appearance harmony of some materials was affected significantly by the reactions of the subjects to the visual information obtained from the samples viewed on the monitor in Experiment C.

4.4.5 Consideration for the Color Effect

As described in Section 4.2, color has a strong influence on perceptual qualities. Therefore, the author used exemplars with low saturation. In this section, the author analyzes the effect of color to harmony in our experiments by conducting an additional experiment (Experiment D). In Experiment D, the color images used in Experiment C were converted to gray images. The author carefully chose 87 gray images that had similar distribution with all 435 pairs in regard to the CIE xy distance between two

materials shown in Fig. 4.15. The experimental condition was the same as that used in Experiment C.

By comparing the results of Experiment D and Experiment C, the correlation of harmony rating was 0.85 and the inclination of the regression line was 1.00. Furthermore, as shown in Fig. 4.16, color and harmony rating differences between two materials were independent. These results suggest that the color effect in our experiments was weak using our samples. However, the correlation for all pairs excluding brownish colored wood samples increased 0.88. Figure 4.17 shows the maximum changed pair of harmony rating between experiments (11: saddle leather matte, leather and 28: cork, wood). The pair had a low rating 4.9 in Experiment C, because their surface colors were different each other. In contrast, in Experiment D, the pair had a high rating of 7.1, because color difference disappeared and their surface textures were resembled. In this case, color might influence for evaluating appearance harmony.

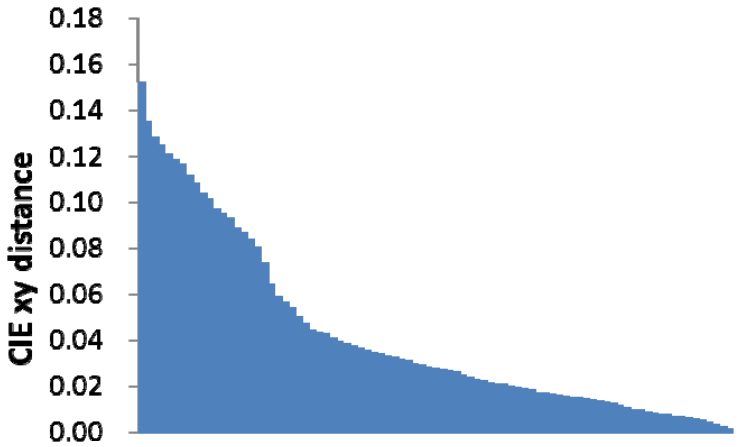


Figure 4.15. Histogram of the CIE xy distance for each pair in Experiment D.

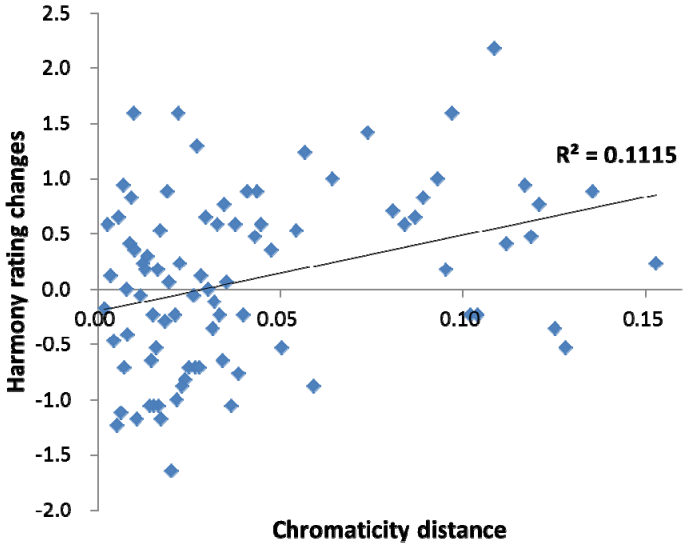
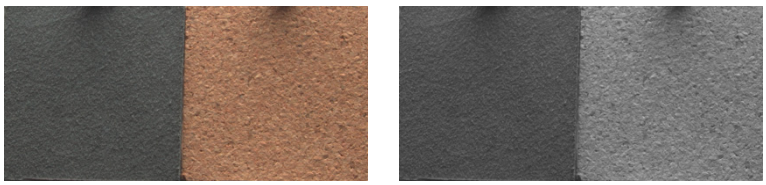


Figure 4.16. Relationships between harmony rating changes and chromaticity distance.



(a) A color image used in experiment C. (b) A gray image used in experiment D.

Figure 4.17. The pairs showing maximum change between Experiments C and D [Pair 11 (saddle leather matte, leather) and 28 (cork, wood)].

4.5 Analysis between Physical Properties and Psychophysical Evaluations

In this section, the author conducts an analysis between physical properties and the psychophysical evaluations obtained from our experiments explained in the previous sections. First, the author calculated the anisotropy for rendered images of 30 samples using following equations.

$$anisotropy = \frac{s^2(f_r)}{P_r^2(f_r)} \tag{4. 3}$$

$$P_r(f_r) = \frac{1}{N_r(f_r)} \sum_{i=1}^{N_r(f_r)} P(f) \tag{4. 4}$$

$$s^2(f_r) = \frac{1}{N_r(f_r) - 1} \sum_{i=1}^{N_r(f_r)} (P(f) - P_r(f_r))^2 \tag{4. 5}$$

where f means the frequency vector of an rendered image, $P(f)$ means the power vector of f, f_r means the same radius r from center in power spectral coordinate and $N_r(f_r)$ means the number of factor obtained from annulus.

Figure 4.18 shows the rendered color image dataset of 30 samples used in our experiment. The author calculated the anisotropy for these images. Figure 4.19 shows the calculated anisotropy histogram. The boundary color of each image corresponds to the material categories the same as in Fig. 4.12.

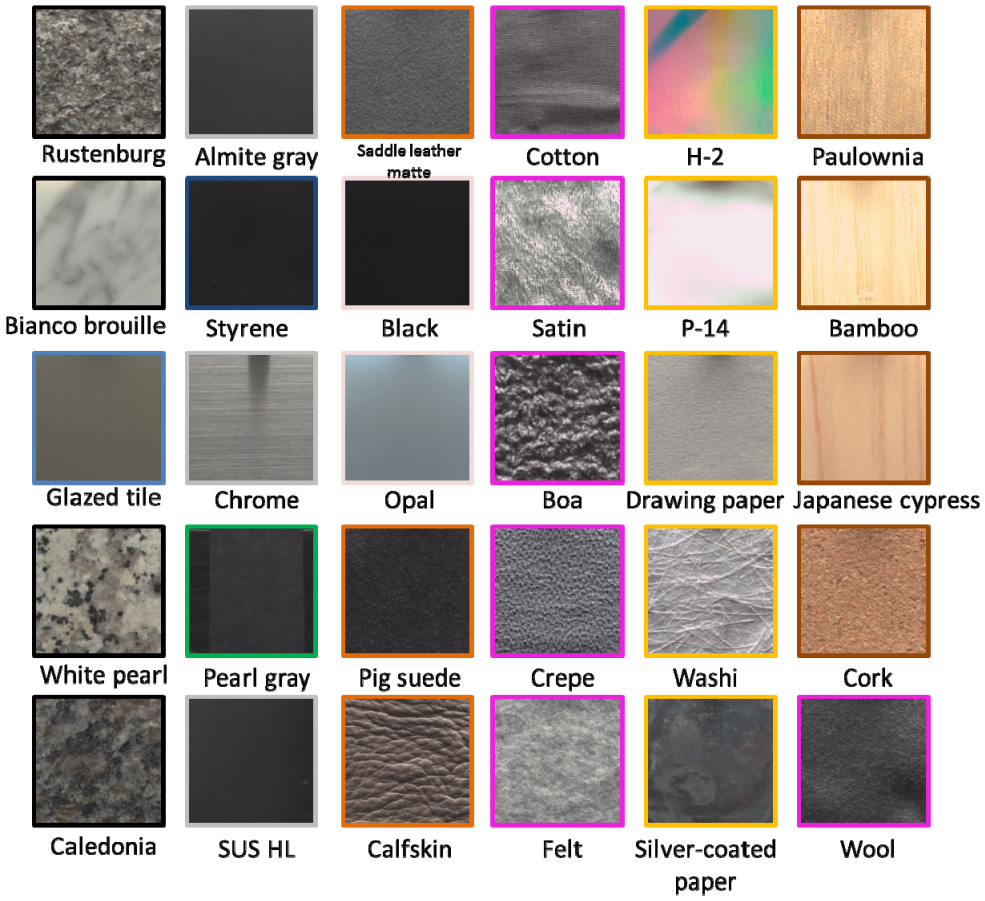
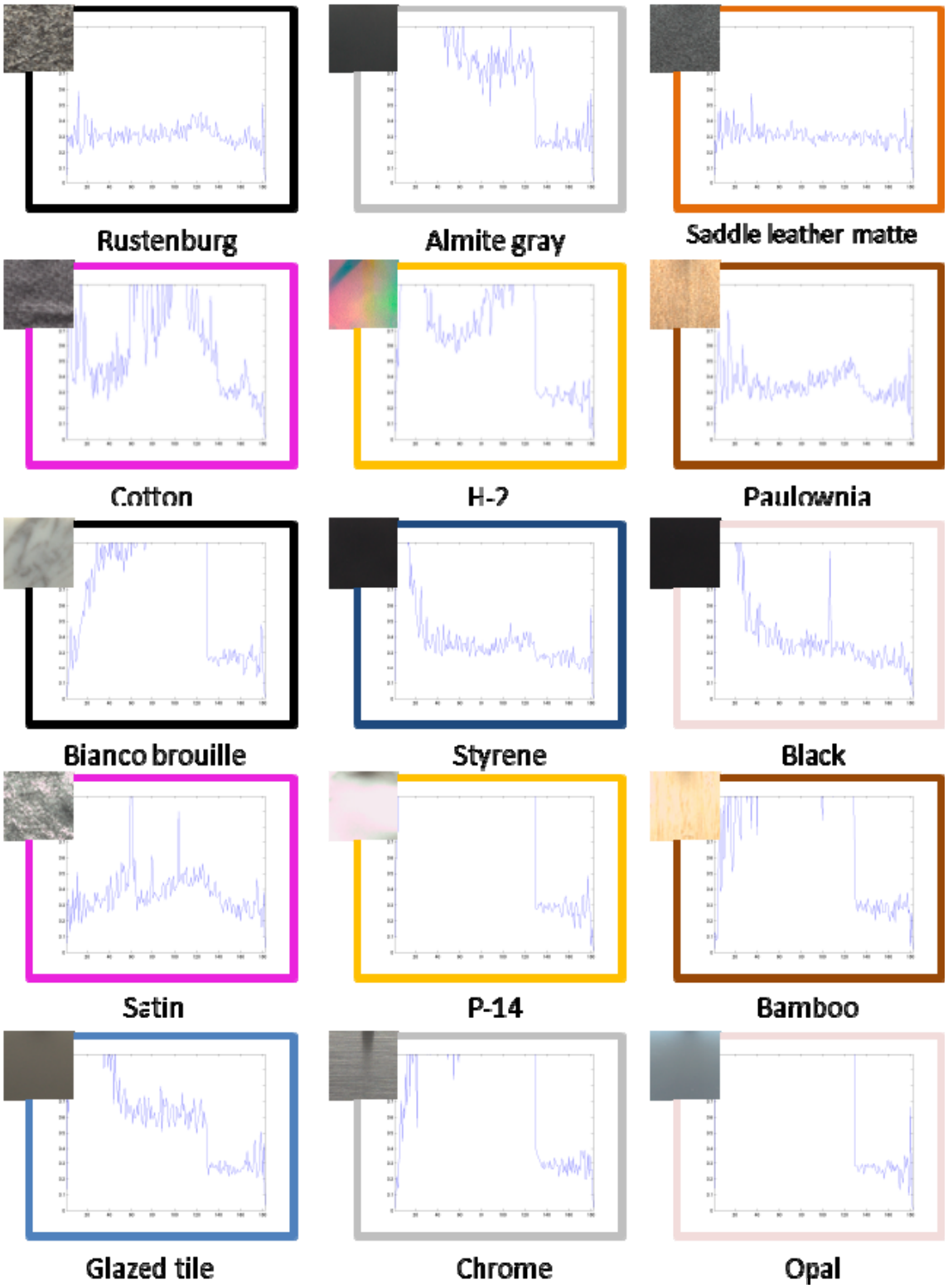


Figure 4.18. Rendered images of 30 samples.


Rustenburg
Almite gray
Saddle leather matte
Cotton
H-2
Paulownia
Bianco brouille
Styrene
Black
Satin
P-14
Bamboo
Glazed tile
Chrome
Opal

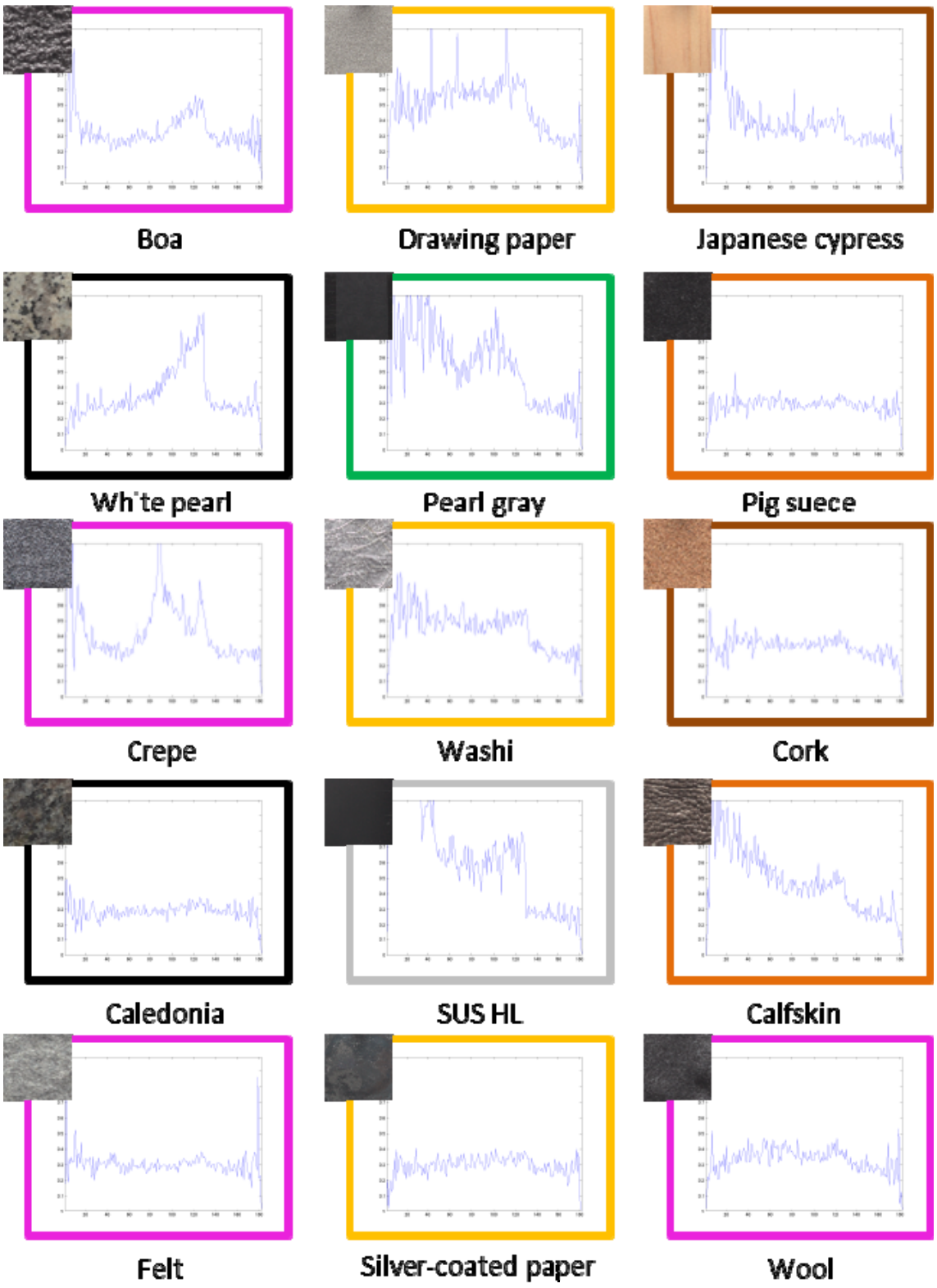
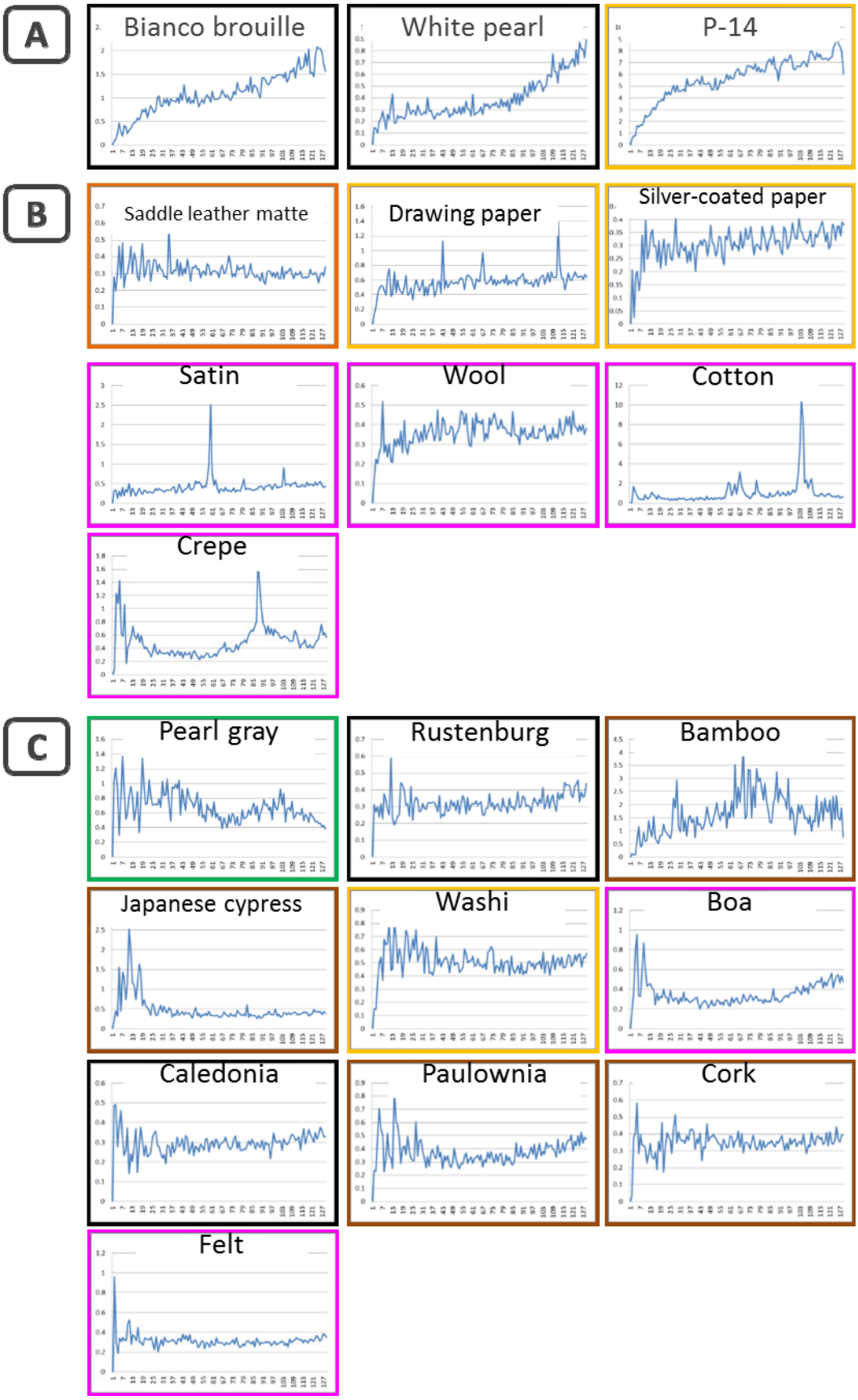


Figure 4.19. Anisotropy hisogram for 30 samples.

Second, the author subjectively classified the material samples based on the shape of the anisotropy histogram. The clusters might be reflected by the surface texture properties of rendered images. The classified result is shown in Fig. 4.20. Group A consists of three samples with a graph rising to the right side of high frequency. Group B consists of seven different samples that have similar properties with comparative even and some spikes. Group C consists of ten samples that have flat histgram for all frequencies. Group D consists of nine samples with bumps in low frequency. One specific sample (H-2, paper) is separated from other samples and named group E.

For each classified group, the averaged harmony ratings were calculated when the materials belonged to the same or not group. Table 4.6 shows the result. As shown in the table, the averaged ratings within the same group (Table 4.6(a)) were higher than one of different group (Table 4.6(b)). This result suggest that the texture property represented by the anisotropy histogram is an important index to judge the harmonious of materials.


A
Bianco brouille
White pearl
P-14
B
Saddle leather matte
Drawing paper
Silver-coated paper
Satin
Wool
Cotton
Crepe
C
Pearl gray
Rustenburg
Bamboo
Japanese cypress
Washi
Boa
Caledonia
Paulownia
Cork
Felt

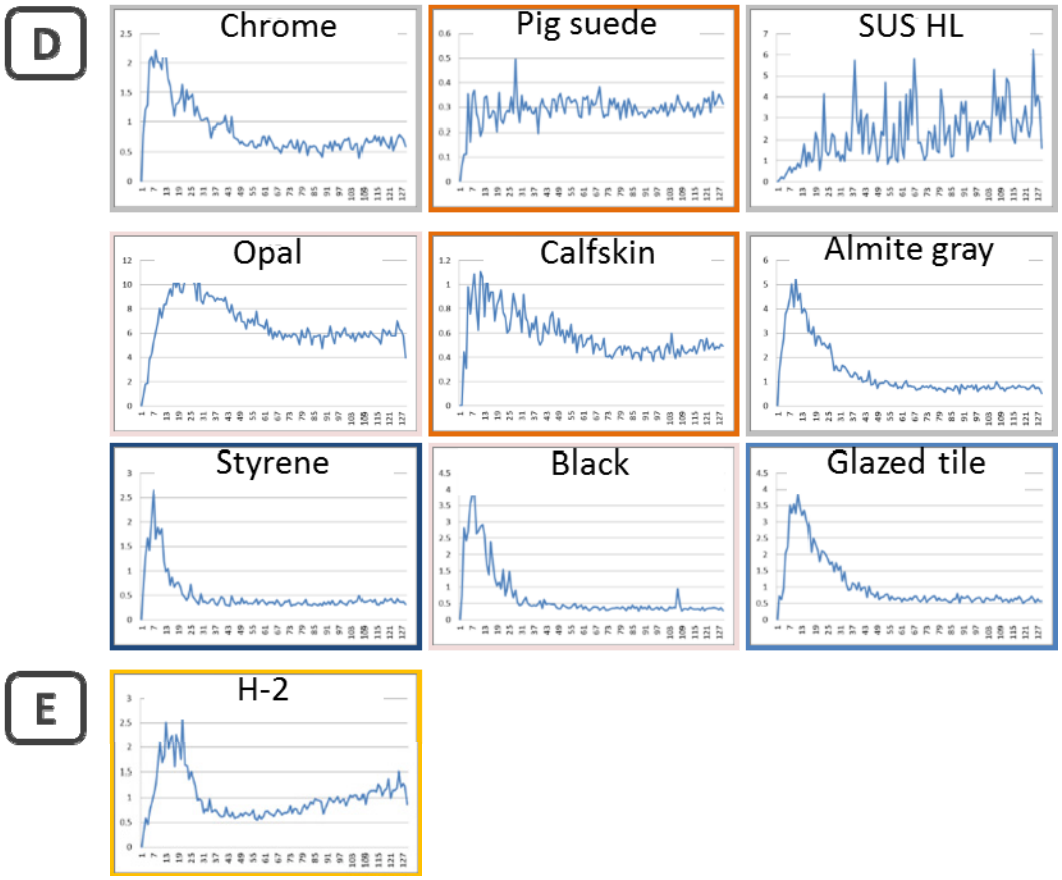


Figure 4.20. Categorized anisotropy hisogram to 5 groups.

Table 4.6. Averaged harmony ratings of materials in the same/different group.

(a) Two materials of the pair belongs to the same group.

Group	Exp. A	Exp. B	Exp. C	Average
A	5.77	5.90	5.72	5.79
B	5.07	5.32	5.32	5.24
C	4.18	4.26	4.00	4.15
D	5.34	5.60	5.69	5.54
E	-	-	-	-
Average	5.09	5.27	5.18	-

(b) Two materials of the pair belongs to the different groups.

	Exp. A	Exp. B	Exp. C	Average
Total	3.99	4.00	4.09	4.03

4.6 Conclusions

The author investigated the appearance harmony among various materials by conducting three psychophysical experiments using the following real materials and their displayed images: stone, metal, glass, plastic, leather, fabric, paper, wood, ceramic, and rubber. In Experiment A, the subjects were allowed to tilt the sample pairs to obtain a comprehensive assessment of harmony based on the reflective properties of the actual surfaces as well as their surface appearance. In Experiment B, the samples were placed such that their surfaces and viewing directions were perpendicular to the subject. In Experiment C, static sample images were displayed on a monitor.

Based on the intra- and inter-participant variances, the author found that the perceptual harmony ratings among subjects were sensitive to the richness of the information available in the real world. However, the perceptual harmony ratings within a subject were stable through all displayed methods. Based on subjective assessments, the author confirmed that sample pairs with similar surface properties were viewed as harmonious, although the materials were different. Indeed, the appearance harmony of the materials differed among static real samples, tilted samples, and static images. In particular, the appearance harmony of some materials was affected significantly by the reactions of subjects to the visual information related to samples with/without displaying them on the monitor, rather than tilting a sample. The results of the PCA indicated that the harmony among categories of glossy materials was more likely to change when the materials were displayed as images. According to the k-means clustering of the data, the overall structure of the appearance harmony was equivalent in all experiments, and the appearance harmony of some materials was affected significantly by the reactions of the subjects to the visual information obtained from the samples viewed on the monitor in

Experiment C.

Furthermore, the author investigated the relationship between physical properties and image features, the author found a fact that the texture property represented by the anisotropy histogram was an important index to judge the harmonious of materials.

Chapter 5. General Conclusions and Future Works

5.1 General Conclusions

In this dissertation, the author analyzed shitsukan using real-world objects and rendered images. After the fundamental investigation of shitsukan perception under different viewing conditions based on color naming experiments, the author conducted two different experiments to investigate the perceptual qualities, and appearance harmony using real materials and rendered images. In the first experiment for investigating the perceptual qualities of material appearance using real materials and rendered images, the author found that the representation method of some materials affected their perceptual qualities. By using methods such as PCA and k-means clustering, the author determined that material categories were more likely to be confused when materials were represented as images, especially gray images. In the second experiment for investigating the harmony of the material appearance, the author indicated that the appearance harmony of some materials was significantly affected by the reactions of subjects to the visual information about samples that are (are not) displayed on the monitor, rather than tilting a sample.

In Chapter 1, the author presented an overview of the motivations and purposes of the psychophysical experiments in this dissertation. Then he contents and the structure of the dissertation are introduced.

In Chapter 2, the author presented the fundamental experiment of shitsukan perception under different viewing conditions that investigates color naming for 2D and 3D rendered samples. Conventional color naming experiments using a priori clues generally involve 2D clues such as color patches. However, in real-world scenes, most objects have 3D shapes, and their colors are affected by illumination effects such as shadow and gloss. The author used 2D and 3D rendered samples as clues in the experiments, and

analyzed the relationship between color terms and object surfaces. First, the author developed a color term collection system that can produce 218 test colors. The author rendered the color images of a flat disk as a 2D sample and a sphere as a 3D sample on a calibrated display device. It is assumed that the 2D and 3D surfaces with the same object color are obtained under the same viewing and illumination conditions. The results of the color naming experiments show that for 2D and 3D samples, there are differences in the color terms. Important findings are as follows: (1) when observing achromatic colors, brighter color terms tend to be chosen as the 3D samples compared to the 2D samples, (2) achromatic color terms are chosen as 3D samples having low saturation, and (3) for chromatic colors, a darker color term is generally chosen relative to the corresponding 2D samples having the same color. By changing the illumination angle rom 0° to 45° to the surface normal, these properties become more prominent.

In Chapter 3, the author presented the first main topic to investigate the perceptual qualities of material appearance using real materials and rendered images. Recent experimental evidence supported the idea that human observers are good at recognizing and categorizing materials. Using projected images, Fleming et al. reported that perceptual qualities and material classes are closely related. In this experiment, the author further investigated these findings using real materials and degraded versions of images of the same materials. The author developed a real material dataset, as well as four image datasets by varying chromaticity (color vs. gray) and resolution (high vs. low) of the material images. To investigate the fundamental properties of the materials' static surface appearance, the author used stimuli that lacked shape and saturated color information. The author then investigated the relationship between these perceptual qualities and the various types of image representation by performing psychophysical experiments. The results showed that the representation method of some materials

affected their perceptual qualities. These cases could be classified into the following three types: (1) perceptual qualities decreased by reproducing the materials as images, (2) perceptual qualities decreased by creating gray images, and (3) perceptual qualities such as “Hardness” and “Coldness” tended to increase when the materials were reproduced as low-quality images. Using methods such as PCA and k-means clustering, the author found that material categories are more likely to be confused when materials are represented as images, especially gray images. Furthermore, the additional analysis showed the possibility of explaining the relationship between the physical properties and psychophysical assessments.

In Chapter 4, the second main topic was introduced. In this experiment, which was aimed at investigating the harmony of a material appearance, the author investigated the appearance harmony of various materials by conducting psychophysical experiments to collect quantitative data. The author conducted three experiments using 435 round-robin pairs of 30 samples made from 10 actual materials. In the first experiment, subjects were allowed to tilt the pair of samples to obtain a comprehensive assessment of the harmony, based on the reflectance properties of the actual surface and the surface appearance. In the second experiment, the samples were placed such that their surfaces and viewing directions were perpendicular to the subject. In the third experiment, static sample images were displayed on a monitor. The results indicated that the sample pairs with similar surface properties were viewed as harmonious, although their materials were different. Indeed, the appearance harmony of the materials differed among static real samples, tilted samples, and the displayed static images. In particular, the appearance harmony of some materials was significantly affected by the reactions of subjects to visual information about the samples with/without observation of the monitor, rather than by tilting a sample. The PCA results indicated that the harmony between categories

of glossy materials was more likely to change, when the materials were displayed as images. Further consideration for analyzing the relationship between physical properties and psychophysical assessments were conducted. The findings indicated that human evaluates the appearance harmony of materials using psychological properties obtained from physical information such as the texture and reflectance characteristics of a material surface.

5.2 Future Works

In future works, a definite goal of our study is to establish shitsukan management technology. In the color engineering field, color management systems are widely used to treat color correctly for different devices. As is the case with the color management system, shitsukan management technology is a very important and challenging issue when considering output devices. The management technology can control all of the perceptual qualities, and not only color and luminance. Currently, the author is aimed at investigating the relationship between the physical properties and psychological assessments for a material appearance using a few display methods. In future works, the author aims to complete the theoretical aspect, and realize applications to manufacturing systems that can control human sensibility based on a developed shitsukan management technology.

To realize the goal, the following five approaches are required:

1. Analysis of shitsukan perception characteristics by performing shitsukan assessment using movies.
2. Analysis of shitsukan perception characteristics depending on the effect of image quality deterioration such as color, resolution and dynamic range.

3. Analysis of shitsukan perception characteristics depending on the memory and learning for shitsukan.
4. Analysis of shitsukan perception characteristics depending on the reciprocal interaction such as the contrast effect of shitsukan, harmony and combinations.
5. Analysis of shitsukan perception characteristics depending on the effect of the viewing environment including illuminant conditions, viewing distance (material scaling), chromatic adaptation, and the different modes of color appearance.

Acknowledgements

The author wishes to express her gratitude to the persons who have directed her research and who have given her valuable comments, direction, and hospitality. First of all, a special debt of gratitude owes to my supervisor Professor Takahiko Horiuchi from the graduate school of Advanced Integration Science, Chiba University for the sincere supervision and hospitality. He gave me a lot of opportunities to study abroad for academic activity. He also guided me toward the right direction as a researcher. The author also deeply appreciates to Specially Appointed Researcher Dr. Shoji Tominaga, a member of the graduate committee, particular thanks for the valuable comments and direction and encouragements. He taught me how to behave as a researcher. The author would like to sincerely appreciate Professor Roland W. Fleming at Gießen University, Germany for his valuable suggestions and discussions. The author would like to appreciate the committee members, Professor Hirohisa Yaguchi, the chair of the committee, and Professor Yoshitsugu Manabe from Chiba University for his kind cooperation and valuable comments. With their considerable review comments, the author could complete this dissertation finally.

During her time in Tominaga-Horiuchi laboratory and Horiuchi-Hirai laboratory, the author had memorable days with wonderful people. She would like to express her gratitude the member of the laboratory. The author also would like to thank Assistant Professor Keita Hirai and Dr. Ryoichi Saito. She studied a researcher's attitude form their working. Additionally, the author would like to thank to all volunteers for taking part in the subjective assessments. The author would also like to appreciate Associate Professor Yoko Mizokami from Chiba University for their precious advices.

Finally, the author would like to give her special thanks to her family whose patient love enabled her to complete this work.

Without their support, the author could not live this colorful life.

With my sincere gratitude,

Midori TANAKA

References

Abe, T., Okatani, T. and Deguchi, K. (2012). Recognizing surface qualities from natural images based on learning to rank. Proc. International Conference on Pattern Recognition, 3712-3715.

Albertazzi, L. and Hurlbert, A. (2013). The Perceptual Quality of Color. In M. Mirmehdi, Albertazzi, L. (Eds.), Handbook of Experimental Phenomenology: Visual Perception of Shape. Space and Appearance (section 15). John Wiley & Sons.

Anderson, B. L. (2011). Visual perception of materials and surfaces. Current Biology, 21(24), R978-983.

Andreopoulos, A. and Tsotsos, J.K. (2014). 50 years of object recognition: directions forward. Computer vision and image understanding, 117(8) 827-891.

Bar, M. and Neta, M. (2006). Humans prefer curved visual objects. Psychological Science, 17, 645-648.

Baumgartner, E., Wiebel, C. B. and Gegenfurtner, K. R. (2013). Visual and Haptic Representations of Material Properties. Multisensory Research, 26, 429-455.

Berlin, B. and Kay, P. (1969). Basic Color Terms Their Universality and Evolution. University of California Press.

Borish I.M. and Benjamin W.J. (1998). Borish's Clinical Refraction. W.B. Saunders Co.

Boynton, R.M. and Olson, C.X. (1987). Locating Basic Colors in the OSA Space. Color Research and Application, 12(2), 94-105.

Beck, J. (1972). Surface Color Perception. Cornell University Press.

Berlin, B. and Kay, P. (1969). Basic Color Terms Their Universality and Evolution. University of California Press.

Boynton, R.M. and Olson, C.X. (1987). Locating Basic Colors in the OSA Space. Color Research and Application, 12(2), 94-105.

Bornstein M.H. (1973). Color Vision and Color Naming: A Psychophysiological Hypothesis of Cultural Difference. Psychological Bulletin, 80, 257-285.

Burchett, K.E. (2002). Color harmony. Color Res Appl., 27, 28-31.

Chen, N., Tanaka, K., Matsuyoshi, D. and Watanabe, K. (2013). Cross preference for color combination and shape. Proc. Asia Color Association Conference, 98-101.

Crowell, T. Y. (1953). OSA Committee on Colorimetry. The Science of Color.

Dalal, N. and Triggs, B. (2005). Histograms of oriented gradients for human detection. Proc. CVPR, 2, 886-893.

Dana, K. J., Van-Ginneken, B., Nayar, S. K. and Koenderink, J. J. (1999). Reflectance and texture of real world surfaces. ACM Trans. on Graphics, 18(1), 1-34.

Debevec, P., Hawkins, T., Tchou, C., Duiker, H. P., Sarokin, W. and Sagar, M. (2000). Acquiring the reflectance field of a human face. Proc.ACM SIGGRAPH, 145-156.

Doerschner, K., Fleming, R. W., Yilmaz, O., Schrater, P. R., Hartung, B., and Kersten, D. (2011). Visual motion and the perception of surface material. Current Biology, 21(23), 1-7.

Dror, R. O., Adelson, E. H. and Willsky, A. S. (2001). Recognition of surface reflectance properties from a single image under unknown real-world illumination. Proc. the Workshop on Identifying Objects Across Variations in Lighting at CVPR.

Eda, T., Ozaki, K., and Ayama, M. (2010). Relation between Blackness, Hue, and Saturation - Color Evaluation Experiment Using Surface Color Mode and Self-Luminous Color Mode -. Color Science Association of Japan, 34(1), 39-49 (in Japanese).

Fleming, R. W., Dror, R. O., and Adelson, E. H. (2003). Real-world illumination and the perception of surface reflectance properties. Journal of Vision, 3, 347-368.

Fleming, R. W., and Bülthoff, H. H. (2005). Low-level image cues in the perception of translucent materials. ACM Transactions on Applied Perception, 2(3), 346-382.

Fleming, R. W., Jäkel, F., and Maloney, L. T. (2011). Visual perception of thick transparent materials. Psychological Science, 22(6), 812-820.

Fleming, R. W., Wiebel, C., and Gegenfurtner, K. (2013). Perceptual qualities and material classes. Journal of Vision, 13(8), 9.

Grubbs, F. E. (1969). Procedures for Detecting Outlying Observations in Samples. Technometrics, 11(1), 1-21.

Hård, A. and Sivik, L. (2001). A theory of colors in combination: A descriptive model related to the NCS color-order system. Color Res Appl., 26, 4-28.

Hartline, H.K., Wagner, H.G. and Ratcliff, F. (1956). Inhibition in the Eye of Limulus. Journal of General Physiology, 39(5), 651-673.

Hiramatsu, C., Goda, N. and Komatsu, H. (2011). Transformation from image-based to perceptual representation of materials along the human ventral visual pathway. NeuroImage, 57(2), 482-494.

Jensen, H. W., Marschner, S., Levoy, M. and Hanrahan, P. (2001). A practical model for subsurface light transport. Proc. ACM SIGGRAPH, 511-518.

Judd, D.B. and Wyszecki, G. (1975). Color in Business. Science and Industry, 3rd edition. New York: Wiley; 390-396.

Katz, D., MacLeod, R. B., Fox, C.W. and Kegan, P. (1935) The World of Color.

Kay, P. and McDaniel, C.K. (1978). The Linguistic Significance of the Meanings of Basic Color Terms. Language, 54(3), 610-646.

Kay, P. and Regier, T. (2007). Color Naming Universals: The Case of Berinmo. Cognition, 102, 289-298.

Kay, P. (1999). Color. Paul Kay. Journal of Linguistic Anthropology, 1, 29-32.

Webster, M.A. and Kay, P. (2007). Individual and Population Differences in Focal Colors. Anthropology of Color, 29-53.

Koenderink, J. and van Doorn, A. (1987). Representation of local geometry in the visual system. Biological Cybernetics, 545, 367-375.

Land, E. H. (1964). The Retinex. Am. Scientist, 52, 247-264.

Liu, C., Sharan, L., Adelson, E. H., and Rosenholtz, R. (2010). Exploring features in a Bayesian framework for material recognition. Proc. CVPR, 239-246.

Lowe, D. G. (2004). Distinctive image-features from scale-invariant keypoints. IJCV, 60(2), 91-110.

Marschner, S., Westin, S. H., Arbree, A. and Moon, J. T. (2005). Measuring and modeling the appearance of finished wood. Proc. ACM SIGGRAPH, 727-734.

McCann, J. J., McKee, S. and Taylor, T. (1976). Quantitative Studies in Retinex Theory: A Comparison between Theoretical Predictions and Observer Responses to 'Color Mondrian' Experiments. Vision Research, 16, 445-458.

Mojsilovic, A. (2002) A Method for Color Naming and Description of Color Composition in Images. Proc. IEEE International Conference on Image Processing, 789-792.

Moroney, N. (2003). Unconstrained web-based color naming experiment. Proc. SPIE, 5008, 36-46.

Motomura, H. (2002). Analysis of Gamut Mapping Algorithms from the Viewpoint of Color Name Matching. Journal of the Society for Information Display, 10(3), 247-254.

Motoyoshi, I. (2010). Highlight-shading relationship as a cue for the perception of translucent and transparent materials. Journal of Vision, 10(9), 1-11.

Motoyoshi, I., Nishida, S., Sharan, L. and Adelson, E.H. (2010). Image statistics and the perception of surface qualities. Nature, 447(7141), 206-209.

Motoyoshi, I., and Matoba, H. (2012). Variability in constancy of the perceived surface reflectance across different illumination statistics. Vision Research, 53, 30-39.

Muryy, A., Welchman, A. E., Blake, A., and Fleming, R. W. (2013). Specular reflections and the estimation of shape from binocular disparity. Proceedings of the National Academy of Sciences, 110(6), 2413-2418.

Mylonas, D., Macdonald, L. and Wuerger, S. (2010). Towards an Online Color Naming Model. In CIC, 140-144.

Nagai, T., Matsushima, T., Koida, K., Tani, Y., Kitazaki, M. and Nakauchi, S. (2015). Temporal properties of material categorization and material rating: visual vs non-visual material features. Vision Research.

Nasser, K.M. and Marjan, T. (2010). Design with emotional approach by implementing Kansei engineering—Case study: Design of kettle. Proc. International Conference on Kansei Engineering and Emotion Research, 625-632.

Nishida, S., and Shinya, M. (1998). Use of image-based information in judgments of surface-reflectance properties. Journal of the Optical Society of America A, 15, 2951-2965.

Okajima, K., Ayama, M., Uchikawa, K. and Ikeda, M. (1988). Comparison of Luminous-efficiency for Brightness in a Light-source Color Mode and a Surface Color Mode. Kogaku, 17, 582-592.

Okajima, K. (2000). Mode of Color Appearance. Color Science Association of Japan, 24(1), 51-57 (in Japanese).

Okazawa, G., Goda, N. and Komatsu, H. (2012). Selective responses to specular surface in the macaque visual cortex revealed by fMRI. Neuroimage, 63, 1321-1333.

Olkkonen, M., Witzel, C., Hansen, T. and Gegenfurtner, K. R. (2010). Categorical color constancy for real surfaces. Journal of Vision, 10(9):16, 1-22.

Ou, L. and Luo, M.R. (2006). A colour harmony model for two-colour combinations. Color Res Appl. 31, 191-204.

Ou, L., Chong, P., Luo, M.R. and Minchew, C. (2010). Additivity of colour harmony. Color Res Appl., 36, 355-372.

Padilla, S., Drbohlav, O., Green, P. R., Spence, A. D., and Chantler, M. J. (2008). Perceived roughness of 1/f noise surfaces. Vision Research, 48, 1791-1797.

Phong, B. T. (1975). Illumination for Computer Generated Pictures. Communications of the ACM, 18(6), 311-317.

Pont, S. C. and Koenderink, J. J. (2005). Bidirectional texture contrast function. International Journal of Computer Vision, 62(1/2). April/May 2005, special issue on Texture Syndissertation and Analysis.

Pont, S. C., and Koenderink, J. J. (2008). Shape, surface roughness, and human perception. In M. Mirmehdi, X. Xie, and J. Suri (Eds.), Handbook of texture analysis (pp. 197-222). Imperial College Press.

Prasad, D. K. (2012). Survey of the problem of object detection in real images. International Journal of Image Processing, 6(6), 441-466.

Qiao, X., Wang, P., Li, Y. and Hu, Z. (2014). Study on a correlation model between the Kansei image and the texture harmony. International Journal of Signal Processing, Image Processing and Pattern Recognition, 7, 73-84

Regier, T., Kay, P., and Cook, R.S. (2005). Focal colors are universal after all. In the National Academy of Sciences, 102, 8386-8391.

Regier, T., Kay, P. and Khetarpal, N. (2007). Color Naming Reflects Optimal Partitions of Color Space. In the National Academy of Sciences. 104, 1436-1441.

Regier, T., Kay, P. and Khertapal, N. (2009). Color Naming and the Shape of Color Space. Language, 85, 884-892.

Roberson, D., Davies, I, and Davidoff, J. (2000). Color Categories are not Universal: Replications and New Evidence from a Stone-age Culture. Journal of Experimental Psychology, 129(3), 369-398.

Sarailidis, G. and Katsavounidis, I. (2012). A multiscale error diffusion technique for digital multitoning. IEEE Trans. On image processing, Vol.21, No.5, pp.2693-2705.

Sharan, L., Rosenholtz, R. and Adelson, E. (2009). Material perception: What can you see in a brief glance?. Journal of Vision, 9(8), 784.

Shinoda, H. (1991). Division of a color space by color terms in surface and aperture color mode. Color Science Association of Japan, 15(1), 53-54 (in Japanese).

Shinoda, H., Ushikawa, K., and Ikeda, M. (1993). Categorized Color Space on CRT in the Aperture and the Surface Color Mode. Color Research and Application, 326-333.

Shinoda, H. and Yamaguchi, H. (2009). Evaluation of Space-Brightness Using Border Luminance of Color Appearance Mode. Journal of the Illuminating Engineering Institute of Japan, 93(12).

Tangkijviwat, U., Rattanakasamsuk, K. and Shinoda, H. (2010). Color Preference Affected by Mode of Color Appearance. Color Res. Appl., 35(1), 50-61.

Thompson, W. B., Fleming, R. W., Creem-Regehr, S., and Stefanucci, J. (2011). Visual perception from a computer graphics perspective. Wellesley, MA, USA: CRC Press.

Tominaga, S. (1985). A Color-naming Method for Computer Color Vision. In IEEE Int. Conf. on Systems, Man, and Cybernetics. Proceedings. Osaka, 573-577.

Tominaga, S. (1987). A Computer Method for Specifying Colors by Means of Color Naming. Cognitive Engineering in the Design of Human-Computer Interaction and Expert Systems. Elsevier Science Publishers, 131-138.

Tominaga, S., Ono, A. and Horiuchi, T. (2010). Investigation and Analysis of Color Terms in Modern Japanese. In SPIE Electronic Imaging, Proceedings, 7528. San Jose, California, 752804-1 - 752804-8.

Toyama, H., Ptsuki, H., Sakakibara, N. and Ayama, M. (1997). Different of Color Appearances between Self-luminous Mode and Surface-color Mode. The Institute of Image Information and Television Engineers, Technical Report 21(28), 13-18 (in Japanese).

Uchikawa, H., Uchikawa, K. and Boynton, R. M. (1989). Influence of Achromatic Surrounds on Categorical Perception of Surface Colors. Vision Research, 29, 881-890.

Uchikawa, K., Kuriki, I. and Shinoda, H. (1996). Categorical Color-Name Regions of A Color Space in Aperture and Surface Color Modes. Journal of Light and Visual Environment, 20(1).

Varma, M. and Zisserman, A. (2009). A statistical approach to material classification using image patch exemplars. IEEE Trans. PAMI, 31(11), 2032-2047.

Wendt, G., Faul, F., and Mausfeld, R. (2008). Highlight disparity contributes to the authenticity and strength of perceived glossiness. Journal of Vision, 8, 14.

Weijer, J., Schmid, C. and Verbeek, J. J. (2007). Learning Color Names from Real-World Images. Proc. IEEE International Conference on Computer Vision and Pattern Recognition, 1-8.

Wiebel, C. B., Baumgartner, E. and Gegenfurtner, K. R. (2013a). Visual and Haptic Representations of Material Qualities. Journal of Vision, 13(9) 198.

Wiebel, C. B., Valsecchi, M. and Gegenfurtner, K. R. (2013b). The speed and accuracy of material recognition in natural images. Attention, Perception & Psychophysics, 75(5) 954-966.

Yamashita, Y. (1997). Change in Modes of Color Appearance and Perceived Colors. Color Science Association of Japan, 21(1), 15-24 (in Japanese).

Ye, Y., Zhang, Z. and He, R. (2014). Study on design of chair shaping based on Kansei engineering. International Journal of Scientific & Engineering Research, 5, 273-276.

Zaidi, Q. (2011). Visual inferences of material changes: Color as clue and distraction. Wiley Interdisciplinary Reviews: Cognitive Science, 2(6), 686-700.

Contributions

Journal Papers

1. M. Tanaka, T. Horiuchi and S. Tominaga, Color Naming Experiments using 2D and 3D Rendered Sample, Color Research and Application, Vol.40, Issue 3, pp.270-280, Jun., 2015 (DOI: 10.1002/col.21886, Article first published online: Apr., 2014).
2. M. Tanaka and T. Horiuchi, Investigating Perceptual Qualities of Static Surface Appearance using Real Materials and Displayed Image, Vision Research, Vol.115, Part B, pp.246-258, Oct., 2015 (DOI: 10.1016/j.visres.2014.11.016, Article first published online: Dec., 2014)
3. M. Tanaka and T. Horiuchi, Appearance Harmony of Materials using Real Objects and Displayed Images, Journal of the International Colour Association, (Accepted for publication).

Conferences

1. M. Tanaka, T. Horiuchi and S. Tominaga, Color Control of a Lighting System using RGBW LEDs, Proc. 23rd IS&T/SPIE Symposium on Electronic Imaging, 78660W (Jan., 2011)
2. M. Tanaka, S. Tominaga and T. Horiuchi, Color Naming Experiment using 2D and 3D Rendered Samples, Proc. Midterm Meeting of the International Colour Association, pp.90-93 (June, 2011)
3. M. Tanaka, T. Horiuchi and S. Tominaga, Color Naming Experiment using Plural Stimuli under Different Viewing Environments, Proc. Interim Meeting of the International Colour Association, pp.72-75 (Sep., 2012)
4. M. Tanaka, T. Horiuchi and S. Tominaga, Visual Perception of Fluorescent and Neon Colors on an LCD Monitor, Proc. Congress of the International Colour Association, pp.1053-1056 (July, 2013)
5. M. Tanaka and T. Horiuchi, An Investigation of the Appearance Harmony of Material, Proc. Interim Meeting of the International Colour Association, pp.449-457 (Oct., 2014)
6. M. Tanaka and T. Horiuchi, An Investigation of the Appearance Harmony using Real Materials and Displayed Image, Proc. Midterm Meeting of the International Colour Association, pp.314-319 (May, 2015)

Invited Talk

1. M. Tanaka, Perceptual Qualities and Harmony of Material Appearance, International Symposium on Foundations of Visual Information, pp.31-36 (Sep., 2015)