Texture Affects Color Emotion

Marcel P. Lucassen,* Theo Gevers, Arjan Gijsenij

Intelligent System Laboratory Amsterdam, University of Amsterdam, Amsterdam, The Netherlands

Received 11 March 2010; revised 27 April 2010; accepted 8 June 2010

Abstract: Several studies have recorded color emotions in subjects viewing uniform color (UC) samples. We conduct an experiment to measure and model how these color emotions change when texture is added to the color samples. Using a computer monitor, our subjects arrange samples along four scales: warm–cool, masculine–feminine, hard–soft, and heavy–light. Three sample types of increasing visual complexity are used: UC, grayscale textures, and color textures (CTs). To assess the intraobserver variability, the experiment is repeated after 1 week. Our results show that texture fully determines the responses on the Hard-Soft scale, and plays a role of decreasing weight for the masculine–feminine, heavy–light, and warm–cool scales. Using some 25,000 observer responses, we derive color emotion functions that predict the group-averaged scale responses from the samples' color and texture parameters. For UC samples, the accuracy of our functions is significantly higher (average $R^2 = 0.88$) than that of previously reported functions applied to our data. The functions derived for CT samples have an accuracy of $R^2 = 0.80$. We conclude that when textured samples are used in color emotion studies, the psychological responses may be strongly affected by texture. © 2010 Wiley Periodicals, Inc. Col Res Appl, 36, 426–436, 2011; Published online 12 November 2010 in Wiley Online Library (wileyonlinelibrary.com). DOI 10.1002/col.20647

Key words: color; texture; color emotion; observer variability; ranking

INTRODUCTION

There is growing interest in the understanding of human feelings in response to seeing colors and colored objects. The so called "color emotions" (i.e., psychological responses to color), involved in published studies, do usually not refer to basic human emotions, such as happiness, surprise, or fear. Rather, they capture an observers' response on an associated affective dimension specified by the investigators, such as warm–cool and hard–soft. Color emotion studies recently published[1–3] focus on the selection of emotional scales and investigate how these scales are related by means of factor analysis. Then, regression analysis is usually applied to reveal the relationships of human responses on these scales with the underlying color appearance attributes, such as lightness, chroma, and hue. Roughly summarizing these studies, the common finding is that the color emotions are reasonably well described by a small number of semantic factors, such as the colour weight, colour activity, and colour heat found by Ou *et al*.[1] or valence, arousal, and dominance by Suk and Irtel.[4] Of the perceptual attributes that characterize the samples, lightness and chroma are most frequently reported as being the relevant parameters for quantitative prediction of the color emotions, although hue cannot be ignored in scales, such as warm–cool.

Several studies investigate whether color emotions can be regarded as culture specific or universal.[1,3,5–8] In most of them, it is found that the influence of cultural background is limited.

Additionally, the effect of media type (paper vs. CRT display) upon the emotional responses to color is studied.[4] No effect of media type is measured.

Many color vision studies regard color as the main experimental variable, as if it is an isolated object property. However, real life objects are seldom uniformly colored. Nonuniformity of object colors (texture) and their environment seems to be the rule rather than the exception.[9–11] Therefore, a logical next step in color emotion studies is the extension from uniform color (UC) toward color texture (CT). So far, the role of texture in color emotion has received only little attention. An early study by Tinker[12] shows that surface texture, as represented by coated paper or cloth, has little or no effect upon apparent warmth or

*Correspondence to: Dr. M. P. Lucassen (e-mail: m.p.lucassen@uva.nl).

Contract grant sponsor: Dutch Organization for Scientific Research; contract grant number: VICI NWO grant 639.023.705.

affective value of colors. Kim *et al.*[13] use color and texture features to predict human emotions based on textile images. Erhart and Irtel[14] indicate that surface structure can change the emotional effect of colored textile samples, depending upon the color. More recently, Simmons and Russell[15] report that the addition of texture can significantly change the perceived unpleasantness of colors, depending on the texture class. This finding, however, is confined to a single emotion, namely unpleasantness. So, although a handful of studies exist, what is still lacking is a systematic approach to color emotion in which the complexity of the color stimulus is gradually increased.

This article investigates the effect upon color emotion of adding texture to color, using a selection of four color emotion scales. Our experiments build upon, but differ in a number of ways from previous studies. Most importantly, instead of only studying uniformly colored samples, we also use samples with grayscale textures (GTs) and samples with CTs. These textures are primitive (no semantics) and synthesized to prevent strong associations, such as reported in Simmons and Russell,[15] and can be fully parameterized. Second, we introduce a method in which all samples (shown on a computer display) remain visible during experimental trials. The advantage is that they can be ordered conveniently along an emotion scale. Third, our subjects perform the full experiment twice, with at least 1 week in between the first and second measurement. This allows quantification of the intraobserver variability over time, on which we are the first to report. We believe that repeatability information is at least as important as the information obtained from more observers. Finally, we systematically sample the available color gamut of our color monitor to optimally cover the lightness, chroma, and hue domain.

We analyze our data in terms of rank correlations within subjects and between subjects and provide quantitative descriptions. We derive color and texture emotion formulae that predict the group-averaged responses on the emotion scales from the samples' color and texture descriptors. Using these models, we present visualizations of the arrangement of the samples used in the experiments.

METHODS

One of the problems we encountered in a pilot experiment is that when samples are shown one after the other, subjects tend to forget what responses they gave on the emotion scales for similar samples shown earlier in the trial. This leads to an unnecessary increase in variability in the subjects' responses, and therefore lower intra- and interobserver correlations. We therefore design our experiment in such a way that all samples remain visible during a trial. We ask our subjects to order 105 square samples horizontally along an emotion scale labeled with opposite word pairs (e.g., warm-cool). They use the computer mouse to drag samples from their initial location on the top of the screen. Samples can be dragged to any position on the screen to keep an overview of the arrangement of the samples. Subjects know that only the horizontal position of the samples on the scale will be analyzed.

Four emotion scales are used: warm–cool, masculine–feminine, hard–soft, and heavy-light. These four scales are tested in separate experimental trials. The section "Selection of Emotion Scales" motivates the selection of these four scales. There are three conditions differing only in the type (complexity) of samples used. In the UC condition, uniformly colored samples are used that were systematically selected from the sRGB color gamut of our color monitor. In the GT condition, grayscale samples have a texture created in luminance, but not in the chromatic domain. Textures are generated using Perlin noise.[16] The samples in the CT condition are basically blended from the UC and GT samples, thus showing the GTs applied to a single color.

Sample Selection

All samples were square patches of 100 $\times$ 100 pixels. Below we discuss the selection of the three types of samples used, in order of increasing visual complexity: UC, GT, and CT.

Uniform Color. Our goal is a systematic sampling of the available color gamut. The color monitor that we use to display the samples is calibrated to the sRGB color space.[17] Details hereof are presented in the section "Monitor." Within the sRGB color gamut, we select 100 chromatic samples and five achromatic samples. The chromatic samples are selected at five lightness levels ($L^* = 10, 30, 50, 70, 90$). For each level in L^*, 10 hue angles are selected at 36 degree interval ($h = 0, 36, 72, \ldots, 288, 324$). Finally, for each of these hue angles two levels in C^* are selected, being the maximum value (C^*_{max}) within the sRGB gamut and half the maximum value ($C^*_{max}/2$). Figure 1 shows the positions of the samples in the a^* and b^* plane of CIELAB color space. Note that different C^*_{max} values are obtained for the different hue angles, typical for any color gamut. Five additional achromatic samples are selected at $L^* = 20, 40, 60, 80, 100$. A specification of these samples in terms of CIE L^*, C^*, and h_{ab} is presented in Table AI in Appendix A. With the above sample selection, we cover the lightness, chroma, and hue domain of our monitor's color gamut.

Grayscale Texture. Being aware of the fact that texture is one of the characteristics that may alter surface perception,[18,19] at this point in our research we do not want to use natural textures to avoid the possibility of strong inherent emotional associations. Therefore, we use textures that are synthesized on the basis of Perlin noise[16] using the open-source libnoise library (http://libnoise.sourceforge.net/). Perlin noise is a primitive structure used

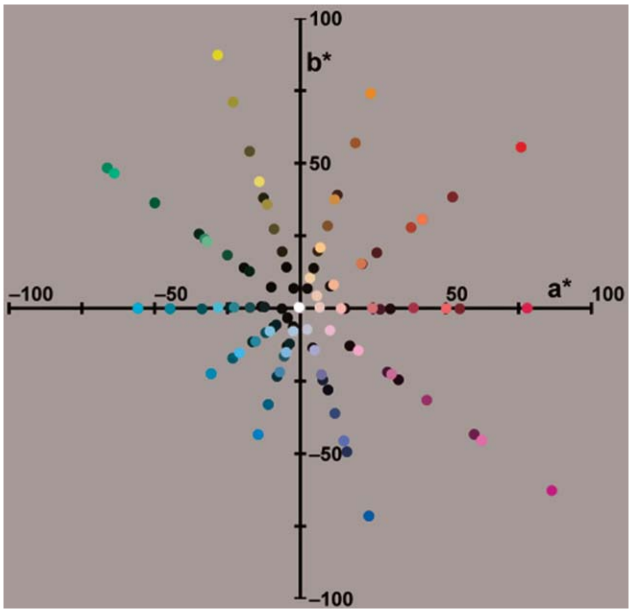

FIG. 1. Positions of the uniform color samples in CIELab color space, covering the sRGB gamut of our color monitor. One hundred chromatic samples are selected at 5 lightness levels (L^*), 10 hue angles (h_{ab}), and 2 chroma levels (C^*). Another five achromatic samples are selected at intermediate lightness levels. A specification of these samples is presented in Table AI in Appendix A.

in procedural texture generation, and is pseudo-random in appearance. All visual details in Perlin noise are the same size, which means that theoretically such an image can be said to truly represent a single texture. Perlin noise can be fully parameterized implying that we can reliably generate a random sample of textures by randomly sampling from the Perlin parameter space. Through controlling the number of octaves, the frequency of each octave and the amplitude of each octave we can respectively control the level of detail, the granularity and the contrast of the resulting texture. For a more detailed description of Perlin noise-based textures, we refer to the aforementioned libnoise library. The GTs are achromatic, showing only spatial variations in lightness. Figure 2 shows an example of how changes in the individual Perlin parameters (number of octaves, frequency, persistence, lacunarity) affect the visual appearance. The extent to which changes in these parameters result in changes in visual appearance depends on the actual position in this parameter space. Summarizing, we have created textures that have no semantics, which can be systematically controlled by the Perlin noise parameters, form a subset of all possible textures and have a natural appearance.

Color Texture. Our CT samples are colored versions of the GTs. They are not multicolored, but consist of lightness variations in a singe color. Although more complex (multicolor) textures exist in reality, we used the simpler

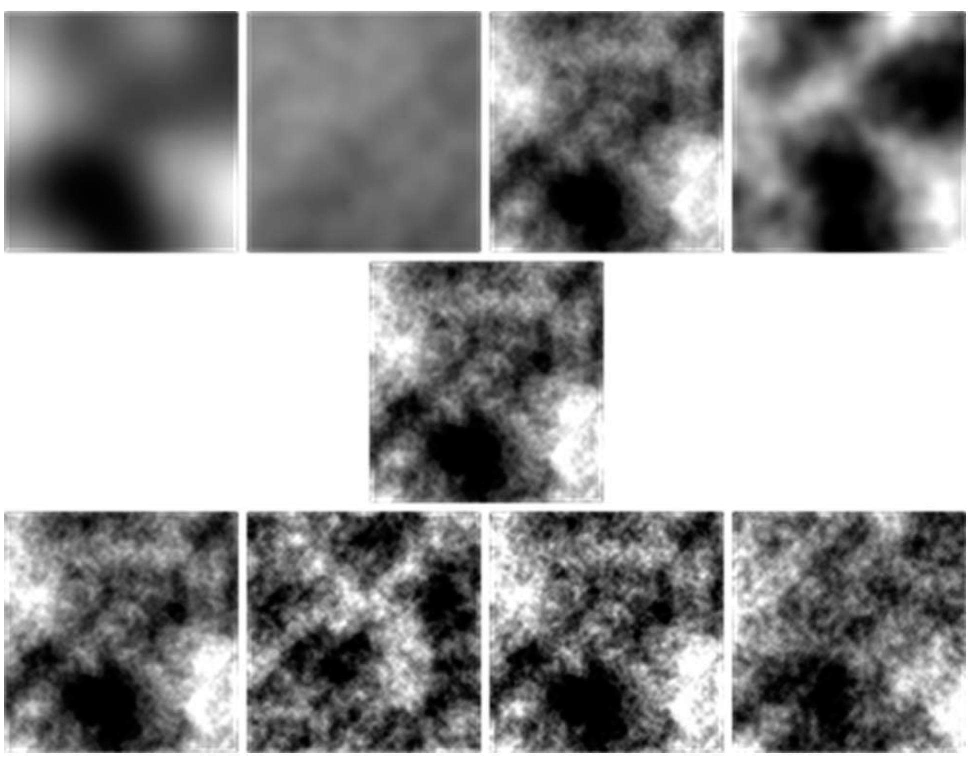

FIG. 2. The effect of varying the Perlin noise parameters is demonstrated here. From left to right: number of octaves, frequency, persistence, and lacunarity. The middle row shows identical samples at parameter values (6, 0.5, 0.6, and 2.5). The top row shows lower values for the parameter in question, while keeping the other parameters fixed. For example, the top left sample has values (1, 0.5, 0.6, and 2.5). The bottom row shows higher values for the parameter in question, while keeping the others fixed. For example, the bottom left sample has values (12, 0.5, 0.6, and 2.5). Note that there is some overlap in visual effect for the higher parameter values (bottom row), but this depends on the position in parameter space. For this particular example, the upper row shows clear differences in visual effect.

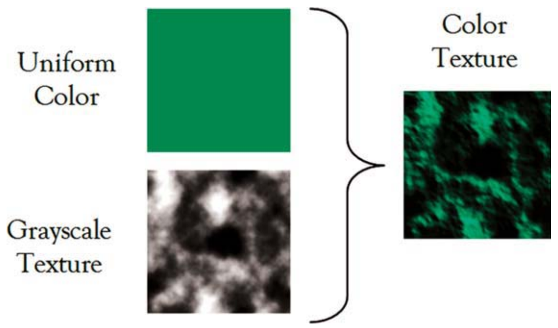

FIG. 3. Color Textures are created by combining uniform color patches with the grayscale textures.

monochrome textures as a first step. The advantage is that this way, the effect of adding texture to color can be studied in stages involving an increasing visual complexity. Figure 3 illustrates how the colored textures are created.

Selection of Emotion Scales

Contrary to a number of preceding studies,[1–3] our primary aim is not to find out which scales are most appropriate to capture color emotions, but rather to explore the effects of adding texture to color samples. We therefore select four scales with opposite word-pairs that have been frequently used in previous studies[1–3,7,13,20,21] and for which we also gained experimental confidence in our pilot studies. These four scales are warm–cool, masculine–feminine, hard–soft, and heavy–light. The warm–cool scale is not used for the GT samples, because our subjects found this combination very hard, if not impossible. With the exception of the masculine–feminine scale, quantitative descriptions of the scales on the basis of CIELAB parameters are available from previous studies, which enable us to compare our results with that of other investigators.

Subjects

Ten subjects participated in the experiments, six men and four women. Their ages range from 26 to 53, with an average of 31.9. Subjects are from seven different nationalities: Dutch (4), Chinese (1), Russian (1), Italian (1), Spanish (1), Polish (1), and German (1). All subjects have normal color vision and normal or corrected to normal visual acuity. Subjects are screened for color vision deficiencies with the HRR pseudo-isochromatic plates (4th edition), allowing color vision testing along both the red-green and yellow-blue axes of color space.[22] The HRR test is viewed under prescribed lighting (CIE illuminant C) using the True Daylight Illuminator (Richmond Products), whereas illumination by other light sources is reduced to a minimum. The first author also participates as a subject in the experiment; the other subjects are unaware of the purposes of the experiment. Subjects participate on voluntary basis and do not receive a financial reward; they are all employed or studying at the institute where the experiment is carried out.

Monitor and Calibration

Stimuli are presented on a high-resolution (1600 $\times$ 1200 pixels, 0.27 mm dot pitch) calibrated LCD monitor, an Eizo ColorEdge CG211. The monitor is driven by a computer system having a 24-bit (RGB) color graphics card operating at a 60 Hz refresh rate. Before each experimental session, a colorimetric calibration of the LCD is performed using a spectrophotometer (Eye-one, Gretag-Macbeth [now X-Rite]). The monitor is calibrated to a D65 white point of 80 cd/m^2, with gamma 2.2 for each of the three color primaries. The CIE 1931 x,y chromaticities coordinates of the primaries were $(x,y) = (0.638, 0.322)$ for red, (0.299, 0.611) for green and (0.145, 0.058) for blue, respectively. With these settings of our monitor, we closely approximate the sRGB standard monitor profile.[17] Spatial uniformity of the display, measured relative to the center of the monitor, was $\Delta E^*_{ab} < 1.5$, according to the manufacturer's calibration certificates.

Procedure

Subjects were seated in front of the monitor at a viewing distance of about 60 cm. The screen size extended $39.6° \times 30.2°$ of visual angle, and a sample (square patch of 100×100 pixels) $2.6° \times 2.6°$. Samples were initially displayed in random order at the top of the screen. Subjects dragged the samples away from their initial position to give them a relative ordering along the horizontal emotion scale. Subjects knew that only the horizontal position would be analyzed, the vertical space could be used to keep an overview of the samples. After ordering the first group of 50 samples, subjects pressed a button after which the second group of 55 samples was shown (the first 50 samples remained visible).

FIG. 4. Experimental result (data from a single observer) for the uniform color samples, ordered horizontally along the masculine–feminine scale. Only the horizontal position matters. At a viewing distance of 60 cm, the screen size extends $39.6° \times 30.2°$ visual angle, and one sample $2.6° \times 2.6°$. The 100 chromatic patches systematically sample the sRGB color gamut at 5 lightness levels, 10 hue levels, and 2 chroma levels. Additionally, five achromatic samples are used.

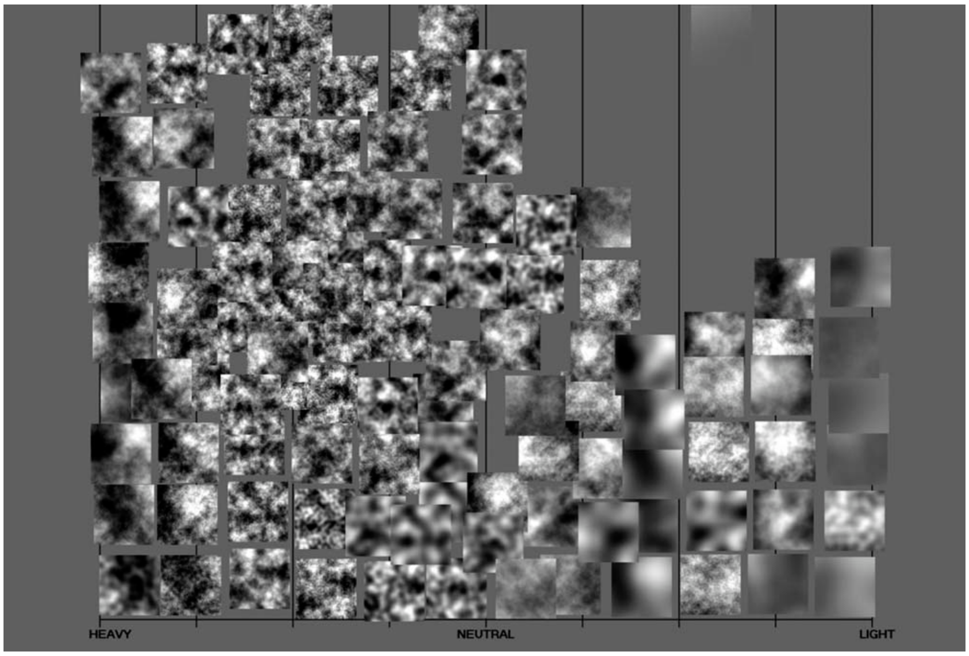

FIG. 5. Experimental result (data from a single observer) for the grayscale texture samples, ordered horizontally along the heavy–light scale. Heavy extends to the left from the center, Light to the right side from the center. Neutral (neither Heavy nor Light) is at the center of the horizontal scale. Textures are made using Perlin noise.

During a trial, all samples could be reordered if desired. One trial of 105 samples took about 5–10 min. All subjects repeated the experiment with at least 1 week in between the first and the second measurement.

RESULTS

Examples of the results for a single observer on the three sample types UC, GT, and CT are shown in Figs. 4, 5, and 6, respectively. The emotion scale arbitrarily extends from -4 (outer left) to $+4$ (outer right), with value zero being neutral (center). Actual scale values for the samples are calculated from their horizontal midpoints. Throughout this article we use ranks (i.e., a relative order from the left side to the right side of the scale) and rank correlations rather than the absolute scale values, because the scales are not expected to be linear. An additional advantage of using ranks is that it corrects for individual differ-

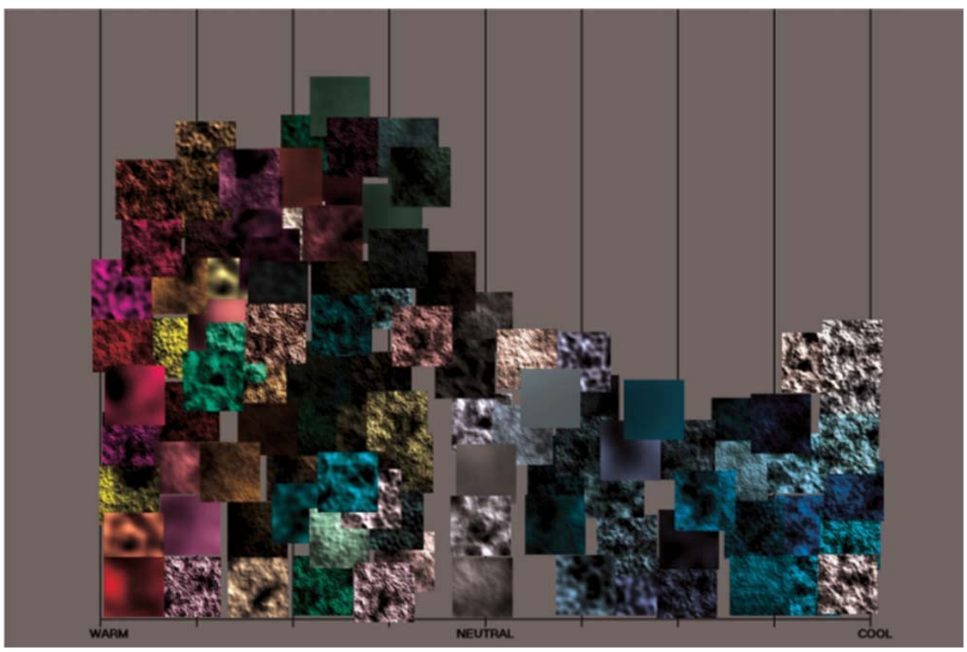

FIG. 6. Experimental result (data from a single observer) for the color texture samples, ordered horizontally along the warm–cool scale. The neutral point is in the center, warm extends to the left side of the scale, cool to the right side of the scale. The color textures are made from blending the uniform color samples with the grayscale textures.

TABLE I. Intraobserver agreement.

Sample type	Emotion scale	Subject 1	2	3	4	5	6	7	8	9	10	Average
Uniform color	WC	0.73	0.83	0.78	0.74	0.81	0.82	0.76	0.78	0.79	0.79	0.78
	MF	0.82	0.72	0.77	0.79	0.74	0.92	0.76	0.79	0.40	0.65	0.74
	HS	0.86	0.67	0.42	0.32	0.66	0.66	0.33	0.77	0.86	0.56	0.61
	HL	0.83	0.87	0.87	0.90	0.85	0.89	0.54	0.75	0.93	0.60	0.80
	Average	0.81	0.77	0.71	0.69	0.77	0.82	0.60	0.78	0.74	0.65	0.73
Grayscale texture	WC	–	–	–	–	–	–	–	–	–	–	–
	MF	0.87	0.79	0.76	0.57	0.61	0.67	0.67	0.68	0.85	0.38	0.68
	HS	0.73	0.82	0.71	0.78	0.72	0.76	0.80	0.68	0.87	0.58	0.74
	HL	0.66	0.54	0.69	0.62	0.76	0.10	0.68	0.42	0.50	0.47	0.54
	Average	0.75	0.71	0.72	0.66	0.70	0.51	0.72	0.59	0.74	0.47	0.66
Color texture	WC	0.82	0.67	0.85	0.74	0.88	0.66	0.87	0.50	0.76	0.66	0.74
	MF	0.75	0.58	0.73	0.76	0.38	0.63	0.29	0.43	0.78	0.61	0.60
	HS	0.54	0.66	0.78	0.75	0.59	0.67	0.29	0.59	0.87	0.68	0.64
	HL	0.76	0.74	0.67	0.84	0.76	0.70	0.52	0.37	0.68	0.35	0.64
	Average	0.72	0.66	0.76	0.77	0.65	0.67	0.50	0.48	0.77	0.57	0.65

Shown are the correlation coefficients between rank orders of the first and second (after 1 week) measurement. WC, Warm–Cool; MF, Masculine–Feminine; HS, Hard–Soft; HL, Heavy–Light.

ences in the used scale range. For instance, one subject may use the full scale range to position the samples, whereas another subject may use only 75% of that range. Statistical analyses are performed with the Statgraphics Centurion XV software package.

Quantitative Analysis: Observer Variability

Intraobserver Agreement. How well do observers agree with themselves? For each observer, sample type and emotion scale, we determine the rank correlation between the first and second measurement (Table I). This correlation is a measure for the intraobserver agreement, or in other words, the repeatability. For 105 samples, the critical value of the correlation coefficient is about 0.195 at the 95% confidence level. Table I shows that the correlation between the first and second measurement is highly significant, for all subjects and all conditions, except for subject 6 on the Heavy-Light scale for the GT samples. For the UC samples, the correlation averaged over subjects and emotion scales is 0.73, which is higher than the corresponding values for the GT samples ($r = 0.66$) and the CT samples ($r = 0.65$). A paired t-test on the UC and the CT data shows that the difference is significant at the 95% confidence level ($P = 0.015$). The same test on the UC and GT data reveals that the difference is not significant ($P = 0.23$), but this is based on less data because the warm–cool scale was not measured for the GT samples. Apparently, subjects reproduce their color emotional responses on UC samples better than on the CT samples. Averaged over the three sample types, the highest intraobserver agreement is found for the warm–cool scale ($r = 0.74$), followed by heavy–light ($r = 0.70$), masculine–feminine ($r = 0.69$), and hard-soft ($r = 0.60$). Considering that the second measurement is made about 1 week after the first measurement, these intraobserver values

TABLE II. Interobserver agreement.

Sample type	Emotion scale	Subject 1	2	3	4	5	6	7	8	9	10	Average
Uniform color	WC	0.41	0.83	0.90	0.20	0.82	0.27	0.59	0.51	0.47	0.73	0.57
	MF	0.91	0.81	0.79	0.76	0.86	0.85	0.74	0.83	0.73	0.54	0.78
	HS	0.77	−0.04	0.56	0.25	0.81	0.86	0.77	0.49	0.02	0.06	0.46
	HL	0.91	0.95	0.97	0.95	0.84	0.68	0.90	0.88	0.96	0.87	0.89
	Average	0.75	0.64	0.81	0.54	0.83	0.67	0.75	0.68	0.55	0.55	0.68
Grayscale texture	WC	–	–	–	–	–	–	–	–	–	–	–
	MF	0.94	0.89	0.89	0.73	0.81	0.75	0.74	0.87	0.93	0.75	0.83
	HS	0.88	0.94	0.84	0.81	0.88	0.85	0.83	0.44	0.92	0.79	0.82
	HL	0.85	0.77	0.79	0.55	0.87	0.63	0.79	0.30	0.44	0.70	0.67
	Average	0.89	0.87	0.84	0.70	0.85	0.74	0.79	0.54	0.77	0.75	0.77
Color texture	WC	0.63	0.87	0.79	0.11	0.81	0.63	0.82	0.78	0.63	0.60	0.67
	MF	0.79	0.80	0.78	0.45	0.71	0.73	0.59	0.60	0.33	0.22	0.60
	HS	0.53	0.88	0.76	0.79	0.54	0.83	0.67	0.64	0.78	0.69	0.71
	HL	0.77	0.83	0.28	0.80	0.79	0.56	0.84	0.06	0.78	0.70	0.64
	Average	0.68	0.85	0.65	0.54	0.71	0.69	0.73	0.52	0.63	0.55	0.65

Shown are the correlation coefficients between rank orders of a single observer with the average rank orders of the nine other observers. WC, Warm–Cool; MF, Masculine–Feminine; HS, Hard–Soft; HL, Heavy–Light.

TABLE III. Color and texture emotion formulae and percentages of explained variance.

Sample type	Emotion scale	Function predicting absolute scale values	Adjusted R^2	Average adjusted R^2
Uniform color	WC	$-0.59 + 0.017\,L - 0.21\,C^{0.6}\cos(h - 45)$	0.90	0.88
	MF	$-2.47 + 0.035\,L + 0.80\,C^{0.3} - 0.018\,h - 0.000021\,h^2 + 0.00000023\,h^3$	0.83	
	HS	$-10.26 + 7.35\,L^{0.1} + 0.053\,C - 0.0019\,C^2 + 0.000011\,C^3 + 0.42\cos(h - 30)$	0.82	
	HL	$-4.41 + 0.30\,L^{0.7} - 0.26\cos(h - 130)$	0.98	
Grayscale texture	WC	–	–	0.82
	MF	$101.36 + 9.27\,L^{0.1} - 30.06\,\mathrm{oct}^{0.05} - 6.06\,\mathrm{freq}^{0.3} - 53.38\,\mathrm{pers}^{0.1} - 25.15\,\mathrm{lac}^{0.1}$	0.83	
	HS	$116.12 + 6.10L^{0.1} - 32.30\,\mathrm{oct}^{0.05} - 13.13\,\mathrm{freq}^{0.1} - 48.81\,\mathrm{pers}^{0.1} - 29.33\,\mathrm{lac}^{0.1}$	0.84	
	HL	$42.67 + 0.064\,L - 12.46\,\mathrm{oct}^{0.05} - 11.35\,\mathrm{freq}^{0.1} - 5.84\,\mathrm{pers}^{0.5} - 17.23\,\mathrm{lac}^{0.05}$	0.80	
Color texture	WC	$-0.80 + 0.015\,L - 0.2\,C^{0.65}\cos(h - 40) + 0.056\,\mathrm{oct}$	0.84	0.80
	MF	$0.84\,L^{0.25} + 0.022\,C - 0.017\,h + 0.00000014\,h^3 - 0.57\,\mathrm{oct}^{0.5} - 0.70\,\mathrm{freq}^{0.5}$	0.76	
	HS	$586.33 - 178.78\,\mathrm{oct}^{0.01} - 84.20\,\mathrm{freq}^{0.01} - 106.83\,\mathrm{pers}^{0.02} - 213.89\,\mathrm{lac}^{0.01}$	0.73	
	HL	$0.33\,L^{0.6} + 0.020\,C^{0.8} - 2.57\,\mathrm{oct}^{0.1} - 1.41\,\mathrm{freq}^{0.1} + 0.015\,\mathrm{lac}^3$	0.86	

The adjusted R^2 measure accounts for the number of free parameters in the formulae. The functions predict the activity on the emotion scales based on the CIELAB color parameters L^*, C^*, h, and/or the Perlin noise texture parameters (oct, number of octaves; freq, frequency; pers, persistence; lac = lacunarity). WC, warm–cool; MF, masculine–feminine; HS, hard–soft; HL, heavy–light.

seem satisfactory. It is impossible to compare this result with other studies because previous color emotion studies did not repeat experiments to assess the level of intra-observer variability.

Interobserver Agreement. How well do observers agree with each other? We calculate the rank correlation between each observer (averaged rank from the first and second measurement) and the average of all other observers. This data is shown in Table II. From the data in Table II, we note that the average interobserver correlation is $r = 0.68$ for the UC samples, $r = 0.77$ for the GT samples and $r = 0.65$ for the CT samples, respectively. Apparently observers agree best on the GT samples. One salient result on the UC samples is that subjects 2, 4, 9, and 10 have low correlations with the group average on the hard–soft scale. This is partly attributable to the positioning of the dark samples along the scale. Further analysis shows that the standard deviation in the subject responses shows a minimum at $L^* + C^* = 100$ and a more than two-fold increase at lower and higher values. Obviously, dark colors and saturated colors lead to lower agreement among subjects. This is found to apply to both the warm–cool and hard–soft scale. We do not consider the four observers as outliers. Their correlation coefficients calculated between the first and second measurement ($r = 0.67$, 0.32, 0.86, and 0.56, respectively) indicate that three of the four observers are able to replicate their results fairly well.

Before discussing the results of adding texture to the color samples, we first present the results of regression analysis. This provides color emotion formulae with which we can more easily explain the effects of texture.

Quantitative Analysis: Modeling

The goal of this section is to derive quantitative formulae that describe the color and texture emotions as a function of the samples' color and texture parameters. As a first step, one-way ANOVA's are performed to find out which of the parameters are significantly connected to the emotion scales. Using both the results of the one-way ANOVAs and formulas derived in previous studies[1] as a starting point, we search for the analytical functions giving the highest amounts of variance explained on the color emotion scales. This is done using our statistical software that indicates the significance of each parameter in the nonlinear regression. The resulting functions are shown in Table III. These functions predict the activity on the emotion scales, based on the color parameters L^*, C^*, h, and/or the texture parameters number of octaves, frequency, persistence, and lacunarity. Before the functions for the CTs are derived, we first recalculate the L^*, C^*, and h values as obtained from averaging over each samples' 100 $\times$ 100 pixels. This is done because the blending procedure used to create the CTs as sketched in Fig. 3 results in somewhat darker samples compared with the UC samples.

The models are derived on group averaged scale values, that is, averaged over 10 observers. A negative scale value indicates a response toward the left word of the opposite word-pair (e.g., warm on the warm–cool scale), a positive value indicates a response toward the right word (e.g., cool on the warm–cool scale). A value of zero, corresponding to the scale center, indicates neutral response, that is, neither warm nor cool on the warm–cool example.

Table III reports the adjusted R^2 as a goodness-of-fit measure for the regression functions. This measure corrects R^2 (variance explained) for the number of free parameters in the regression models. The table shows that for the UC samples, the functions based on the CIELAB parameters L^*, C^*, and h give rise to high values of adjusted R^2, with an average of 0.88. For the Grayscale and CT samples, the average adjusted R^2 is 0.82 and 0.80, respectively.

All in all, the color and texture emotion functions provide a reasonably accurate description of the average observer response on the emotion scales. UC samples are best described, followed by GT and CT. In Figs. B1, B2, and B3, we show visualizations of the samples used in our experiments, ranked along the emotion scales as predicted from the functions in Table III.

In the following section, we return to our main research question: what is the effect of texture on the color emotion scales?

The Effects of Texture on Color Emotion. We already noted that the intraobserver agreement for the UC samples is higher than for the textured samples. At the same time, the interobserver agreement is better for the grayscale samples (average $R^2 = 0.77$) than for the uniform samples ($R^2 = 0.68$) and the CT samples ($R^2 = 0.65$). This may be due to the fact that the GT samples have no variations in hue or chroma, and so the observers have to deal with less color dimensions as they arrange the samples along the scales.

The analytical functions presented in Table III reflect the dependencies on the samples' color and texture parameters. The color parameters L^*, C^*, and h play an important role in all functions for UC samples, with the exception that C^* does not appear in the function for the heavy–light scale. With respect to the functions for the CT samples, three things are noted. First, all color parameters L^*, C^*, and h appear in the warm–cool and masculine–feminine scales. Second, only L^* and C^* appear in the heavy–light scale, and third, no color parameters appear in the function for the hard–soft scale. So, when texture is added to the UC samples, only the hard–soft scale loses its dependency on color parameters. In other words, hard–soft is fully dominated by texture. Warm–cool, masculine–feminine, and heavy–light are dominated by color parameters (in order of descending dominance), adding the texture parameters explains for another 2.9, 36.2, and 27.5%-point of the variance in the data, respectively, as shown in Table IV. This Table presents a comparison of model performances on the CT samples. For example, when the function for warm–cool derived on the UC samples is applied to the CT data, already 82% of the data variance is explained. Adding texture parameters to this function increases the model performance by 2.9%-point. Likewise, for the hard–soft scale, the model derived on the UC samples has no explanatory power at all ($R^2 = 0$) on the CT samples and adding texture parameters results in $R^2 = 0.73$. The last column shows the "added value" of texture parameters, calculated as the difference between the adjusted R^2 obtained with- and without texture parameters.

In conclusion, when texture is added to UC samples, color emotions change. Responses along the hard–soft scale are fully determined by texture, and in decreasing extent for the masculine–feminine, heavy–light, and warm–cool scales. The impact of this is that when textured samples are involved in color emotion studies, texture cannot be ignored.

TABLE IV. Comparison of the performance (adjusted R^2) of color emotion models on our data for the Color Texture samples, when using the Uniform Color functions (derived on the Uniform Color samples) or the Color Texture functions including texture parameters.

	Adjusted R^2		
Emotion scale	Uniform color function	Color texture function	Added value of texture parameters
Warm–Cool	0.82	0.84	0.029
Masculine–Feminine	0.40	0.76	0.36
Hard–Soft	0	0.73	0.73
Heavy–Light	0.59	0.86	0.28

TABLE V. Performance (adjusted R^2) of color emotion models by different investigators on our experimental data for the Uniform Color samples.

	Adjusted R^2		
Color emotion scale	Our study	Ou et al. (2004)	Xin and Cheng (2000)
Warm–Cool	0.90	0.70	0.14
Hard–Soft	0.82	0.16[a]	0.36
Heavy–Light	0.98	0.96	0.96

[a] Excluding the five achromatic samples. When including these samples (having $C^* = 0$), there is no correlation between our data and the model prediction by Ou et al. (2004).

Comparison with Other Studies. We can evaluate the performance of color emotion functions derived by others on our experimental data, but only for the UC samples. Functions for grayscale or CT samples have not been published previously. For the scales warm–cool, hard–soft, and heavy–light, we determine the adjusted R^2 for models derived in studies by Xin and Cheng[21] and Ou *et al.*,[1] see Table V. The results show that our experimental data for the heavy–light scale (which strongly depend on lightness L^*) is very well described by all three models. For the warm–cool scale, the model by Ou *et al.*[1] is reasonably good ($R^2 = 0.70$), but the model by Xin and Cheng[21] completely fails. For the hard–soft scale, both models by Ou *et al.*[1] and Xin and Cheng[21] fail. An explanation for this may be the different methods used for obtaining the observer scores. In our experiments, the subjects put the samples in relative order along the scale, whereas the other investigators only record the preference for one of the scale directions (for instance warm or cool). In the latter case, a final scale value is obtained by performing some sort of averaging over the scores of the observers, and therefore many observers are necessary.

DISCUSSION

We demonstrate a systematic approach to the study of color emotions and the effect thereupon of adding texture to the color samples. A limited number of scales (four) are used, because we are mainly interested in the specific effect of adding texture, and not so much in factor analysis that reveals how different scales may combine into new descriptors. Nevertheless, we have gathered a valuable set of experimental data using an improved method in which subjects order the samples along the scale while maintaining a view on all samples. Another methodological improvement in comparison with other studies is that our subjects repeat the experimental trials after 1 week, which provides us with an estimate of the intraobserver agreement. We derive analytical functions that predict the group-averaged scale responses, with a precision exceeding that reported in other studies. Also, our functions outperform the functions derived by Xin and Cheng[21] and Ou *et al*.[1] when applied to our data, which is probably explained by the differences in methods. We note that the adjusted R^2 measure is the preferred measure to report, because that one corrects for the number of free parameters in the functions.

Our subjects are from seven different nationalities. Testing on cross-cultural effects, as done in other studies[1,3,5,6,8] is not performed as that would require more subjects. Neither do we test on possible gender differences. Again, our focus is on the effect of adding texture, not on other issues. In the experimental design we adopt the minimum number of observers (10) as discussed in Engeldrum.[21] As long as the desired scale precision is unknown it is impossible to make precise estimates on the required number of observers. All that can be said is that the use of more subjects leads to lower standard deviations in the estimates. Scale accuracy increases with about the square root of the number of observers. Other studies have used more subjects (e.g., Ou *et al*.[1] used 31 observers, Gao and Xin[2] used 70 observers, Gao *et al*.,[3] used 50–70 observers per cultural group) but we prefer to perform a repetition of the full experiment, which we regard equally important. In this respect, an interesting question is what the subjects' long term repeatability on the color and texture emotion scales is. That kind of information would greatly help to assess the validity and applicability of the color and texture formulae derived here. From our study it is clear, though, that whenever textured samples are used, texture may play an important role in color emotion.

As for future experiments, there are several routes to go. In addition to the four scales we study, other scales used in color emotion studies may be selected. Moreover, scales from other studies (not necessarily color studies) may be adopted to better capture the responses for texture classes. Although we already enhance the complexity of our stimuli by adding lightness textures to UCs, the textures that we use are still rather primitive. Using chromatic textures, having chromatic variations around the average, would be a logical next step. When the chromatic distribution is mainly in one direction in color space, discrimination thresholds for natural and synthesized textures were found to be identical,[23] which allows us to continue working with synthetic textures. It may be expected that when using natural textures, certain color–texture combinations will fit prototypical templates like green grass, and initiate strong associations. Recent findings from neuroimaging studies suggest that the cerebral processing of form, texture, and color may be independent.[24,25] Yet, these studies provide no answer to the question how those object features interact when subjects have to respond on emotion scales.

CONCLUSIONS

When texture is added to UC samples, color emotions change. Texture fully determines the responses on the hard–soft scale, and plays a role of decreasing weight for the masculine–feminine, heavy–light, and warm-cool scales. We conclude that when textured samples are used in color emotion studies, the psychological responses may be strongly affected by texture.

APPENDIX A

TABLE AI. CIELAB L^*, C^*, and h_{ab} specification of the uniform color samples.

Sample	L^*	C^*	h_{ab}	Sample	L^*	C^*	h_{ab}
1	10	31	0	51	10	13	180
2	10	15.5	0	52	10	6.5	180
3	30	55	0	53	30	24	180
4	30	27.5	0	54	30	12	180
5	50	78	0	55	50	34	180
6	50	39	0	56	50	17	180
7	70	50	0	57	70	45	180
8	70	25	0	58	70	22.5	180
9	90	14	0	59	90	56	180
10	90	7	0	60	90	28	180
11	10	26	36	61		11	216
12	10	13	36	62	10	5.5	216
13	30	65	36	63	30	20	216
14	30	32.5	36	64	30	10	216
15	50	94	36	65	50	29	216
16	50	47	36	66	50	14.5	216
17	70	52	36	67	70	38	216
18	70	26	36	68	70	19	216
19	90	14	36	69	90	26	216
20	90	7	36	70	90	13	216
21	10	15	72	71	10	14	252
22	10	7.5	72	72	10	7	252
23	30	41	72	73	30	25	252
24	30	20.5	72	74	30	12.5	252
25	50	60	72	75	50	35	252
26	50	30	72	76	50	17.5	252
27	70	78	72	77	70	46	252
28	70	39	72	78	70	23	252
29	90	22	72	79	90	16	252
30	90	11	72	80	90	8	252
31	10	15	108	81	10	29	288
32	10	7.5	108	82	10	14.5	288
33	30	40	108	83	30	52	288
34	30	20	108	84	30	26	288
35	50	57	108	85	50	75	288
36	50	28.5	108	86	50	37.5	288
37	70	75	108	87	70	48	288
38	70	37.5	108	88	70	24	288
39	90	92	108	89	90	15	288
40	90	46	108	90	90	7.5	288
41	10	24	144	91	10	42	324
42	10	12	144	92	10	21	324
43	30	43	144	93	30	74	324
44	30	21.5	144	94	30	37	324
45	50	62	144	95	50	107	324
46	50	31	144	96	50	53.5	324
47	70	82	144	97	70	77	324
48	70	41	144	98	70	38.5	324
49	90	79	144	99	90	25	324
50	90	39.5	144	100	90	12.5	324
				101	20	0	–
				102	40	0	–
				103	60	0	–
				104	80	0	–
				105	100	0	–

The first 100 samples are chromatic, and the last five samples are achromatic.

APPENDIX B

Here, we show arrangements of the samples used in our experiments. For each emotion scale (except for the warm–cool scale for the GTs), the samples are ranked on scale values as calculated using the functions given in Table 3. The UC samples are displayed as vertical bars to save some space. For the same reason, we left out half the samples in the arrangements of the GTs and the CTs. The arrangements are illustrative; accuracy of color reproduction is limited.

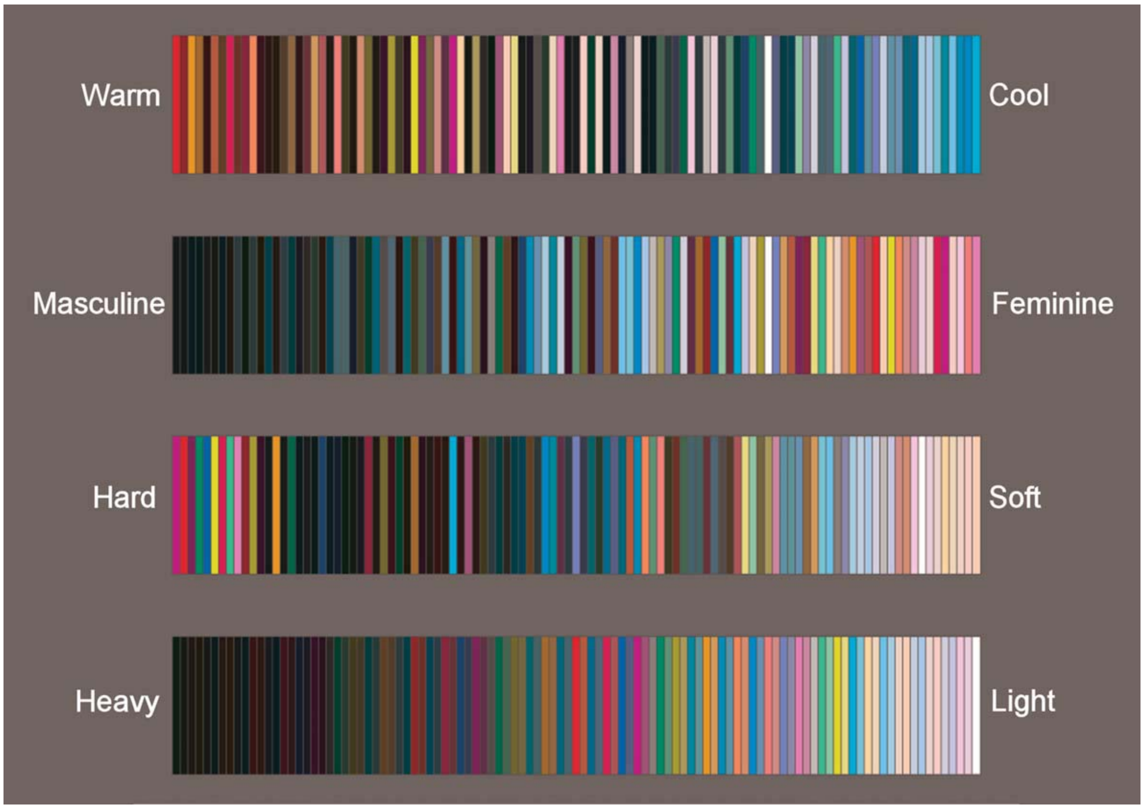

FIG. B1. Arrangement of the uniform color samples used in the experiments, based on the scale values predicted from the functions for uniform colors in Table III.

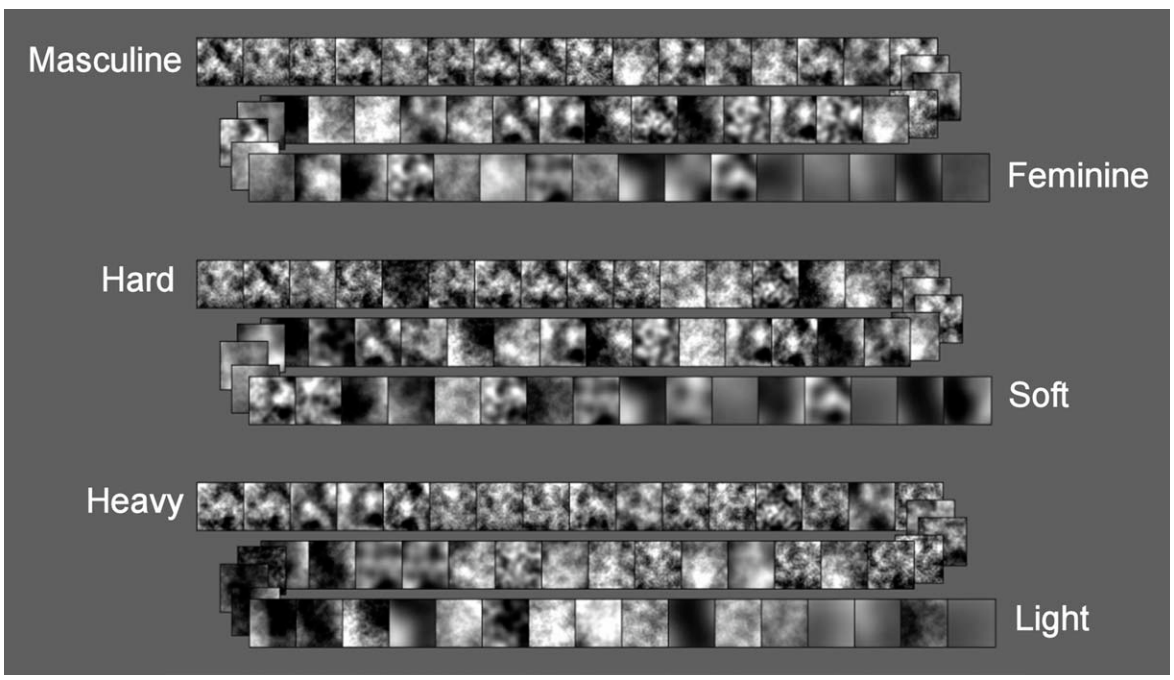

FIG. B2. Arrangement of the grayscale texture samples used in the experiments, based on the scale values predicted from the functions for grayscale texture in Table III.

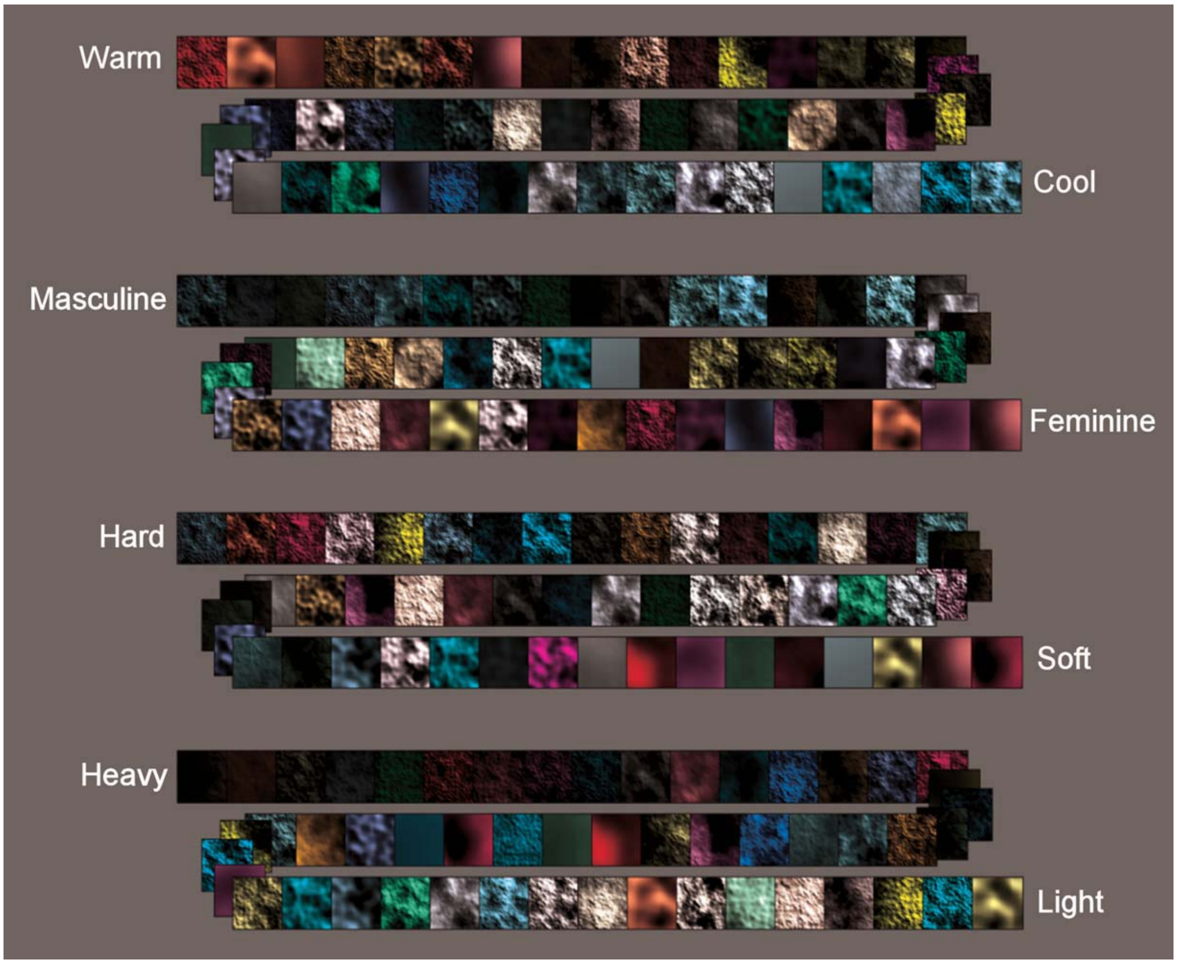

FIG. B3. Arrangement of the color texture samples used in the experiments, based on the scale values predicted from the functions for color texture in Table III.

1. Ou L-C, Luo MR, Woodcock A, Wright A. A study of colour emotion and colour preference. Part I. Colour emotions for single colours. Col Res Appl 2004;29:232–240.
2. Gao XP, Xin JH. Investigation of human's emotional responses on colors. Col Res Appl 2006;31:411–417.
3. Gao XP, Xin JH, Sato T, Hansuebsai A, Scalzo M, Kajiwara K, Guan S-S, Valldeperas J, Lis MJ, Billger M. Analysis of cross-cultural color emotion. Col Res Appl 2007;32:223–229.
4. Suk H-J, Irtel H. Emotional response to color across media. Col Res Appl 2010;35:64–77.
5. Adams FM, Osgood CE. A cross-cultural study of the affective meanings of color. J Cross Cult Psychol 1973;4:135–156.
6. D'Andrade R, Egan M. The colors of emotion. Am Ethnol 1974;1:49–63.
7. Xin JH, Cheng KM, Taylor G, Sato T, Hansuebsai A. Cross-regional comparison of colour emotions Part I: Quantitative analysis. Col Res Appl 2004;29:451–457.
8. Clarke T, Costall A. The emotional connotations of color: A qualitative investigation. Col Res Appl 2008;33:406–410.
9. Ekroll V, Faul F, Niederee R. The peculiar nature of simultaneous colour contrast in uniform surrounds. Vision Res 2004;44:1765–1786.
10. Nascimento SM, Foster DH, Amano K. Psychophysical estimates of the number of spectral-reflectance basis functions needed to reproduce natural scenes. J Opt Soc Am A 2005;22:1017–1022.
11. Hansen T, Giesel M, Gegenfurtner KR. Chromatic discrimination of natural objects. J Vision 2008;8:1–19.
12. Tinker MA. Effect of stimulus-texture upon apparent warmth and affective value of colors. Am J Psychol 1938;51:532–535.
13. Kim S-J, Kim EY, Jeong KJ, Kim J-I. Emotion-based textile indexing using colors, texture and patterns. Proc ISVC 2006;LNCS4292:9–18.
14. Erhart L, Irtel H. Der Einfluss von Oberflächeneigenschaften auf die emotionale Wirkung von Farbe. FarbInfo, Dresden; 2006.
15. Simmons DR, Russell CL. Visual texture affects the perceived unpleasantness of colours. Perception37 ECVP Abstr Suppl 2008:146.
16. Perlin K. An image synthesizer. Comp Graph 1985;19:287–296.
17. Stokes M, Anderson M, Chandrasekar S, Motta R. A standard default color space for the Internet–sRGB. Available at: http://www.w3.org/Graphics/Color/sRGB.html; November, 1996.
18. Anderson BL, Kim J. Image statistics do not explain the perception of gloss and lightness. J Vision 2009;9:10.1–17.
19. Kingdom FAA. Perceiving light versus material. Vision Res 2008;48:2090–2105.
20. Ou L-C, Luo MR, Woodcock A, Wright A. A study of colour emotion and colour preference. Part III: Colour preference modeling. Col Res Appl 2004:29:381–389.
21. Xin JH, Cheng K. Quantitative evaluation of colour emotion. Japan Soc Promotion of Sci Fund Meet 2000:71–86.
22. Bailey JE, Neitz M, Tait D, Neitz J. Evaluation of an updated HRR color vision test. Visual Neurosci 2004;22:431–436.
23. Engeldrum P. Psychometric Scaling: A Toolkit for Imaging Systems Development, Winchester, Massachusetts, USA: Imcotek Press; 2001.
24. Giesel M, Hansen T, Gegenfurtner KR. The discrimination of chromatic textures, J Vision 2009;9:1–28.
25. Cant JS, Goodale MA. Attention to form or surface properties modulates different regions of human occipitotemporal cortex. Cereb Cortex 2007;17:713–731.
26. Cavina-Pratesi C, Kentridge RW, Heywood CA, Milner AD. Separate channels for processing form, texture, and color: Evidence from FMRI adaptation and visual object agnosia. Cereb Cortex 2010;20:433–446.