Appearance harmony of materials using real objects and displayed images

Midori Tanaka and Takahiko Horiuchi

Graduate School of Advanced Integration Science, Chiba University, Japan
Email: horiuchi@faculty.chiba-u.jp

In the present study, we investigated the appearance harmony of various materials by conducting psychophysical experiments to collect quantitative data. We 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 judgment of harmony based on the reflectance properties of the actual surface, in addition to the surface appearance. In the second experiment, the samples were placed such that their surfaces and the viewing direction were perpendicular to the subject. In the third experiment, static sample images were displayed on a monitor. Our 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 affected significantly by the reactions of the subjects to the visual information regarding the samples with/without observing the monitor, rather than by the tilting of a sample. The results of a principal component analysis indicated that the harmony among categories of glossy or transparent materials was more likely to change, especially when the materials were displayed as images.

Received 07 August 2015; revised 04 November 2015; accepted 08 November 2015

Published online: 15 March 2016

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 colour, shape, texture, and material.

Since long, colour harmony has interested researchers involved in colour design studies based on various objects [1-3]. Although there are many theories related to colour 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 colour harmony theory and extended the harmony prediction theory from two-colour combinations [4] to three-colour combinations [5].

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 [6-8], these studies investigated the preference for a single shape, such as a kettle [7] or a chair [8], but they did not consider two-shape combinations. Chen et al. investigated the relationship between preferences for colour-pairs and shapes [9], 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 [10].

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 [11-13], translucency [14-16], or roughness [17-19]. 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, we 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, we investigated the harmony across material categories. In our experiments, we used 435 round-robin pairs of 30 samples made from 10 actual materials. We 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.

Experimental stimuli

Materials dataset

To investigate the subjects’ perception of materials without being influenced by shape, we produced a dataset of 30 exemplars (size = 50 × 50 mm). 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 Figure 1 and Table 1, respectively.

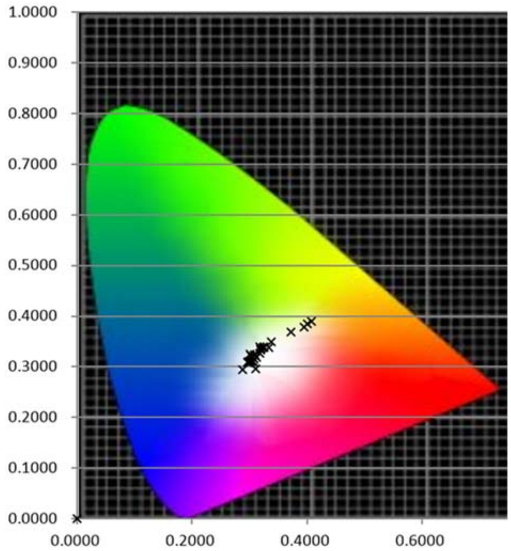


Figure 1 (left): The dataset of 30 exemplar materials.

Figure 2 (right): Chrominance of the material samples.

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 grey	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 grey	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

Table 1: Specifications of the material dataset.

In Table 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 and Hurlbert colour has a strong influence on perceptual qualities [20]. However, it is difficult to collect exemplars of various hues with uniform material; therefore, we only collected exemplars with low saturation. In Figure 2, 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.

Image dataset

We hypothesised that when materials are reproduced on a monitor, the following factors might strongly affect the perceptual harmony: intensity, colour reproduction, and resolution. Thus, we 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 quantisation level of 14 bits. We then prepared a colour 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. We used the following procedure to reproduce the actual display scene. Let $[R_C\ G_C\ B_C]^{\mathrm{T}}$ be a colour signal vector of a pixel captured by the RGB system, where T indicates the transpose operator of the vector. The three-dimensional vector was converted into CIE-XYZ tristimulus values $[X_C\ Y_C\ Z_C]^{\mathrm{T}}$ by multiplying a 3 × 3 matrix $\mathbf{M}_1$, which was developed by approximating the CIE1931 (2 deg) colour matching function using the accumulative camera sensitivity function. The tristimulus values were then converted into linear RGB values $[R_L\ G_L\ B_L]^{\mathrm{T}}$ by multiplying by a 3 × 3 matrix $\mathbf{M}_2$ which was developed in the display calibration process. Finally, the RGB vector $[R_M\ G_M\ B_M]^{\mathrm{T}}$ used for transmitting to the display was obtained by using the gamma operator φ. This procedure can be summarised by the following equation:

$$[R_M\ G_M\ B_M]^{\mathrm{T}} = \varphi\,(\mathbf{M}_2\mathbf{M}_1[R_C\ G_C\ B_C]^{\mathrm{T}}) \qquad (1)$$

Using this calibration process, we verified that the intensity and chromaticity of the real materials and their images reproduced on the display were almost equivalent.

Experimental method

We 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 (see Figure 3).

(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 colour, 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 3 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 colour vision.

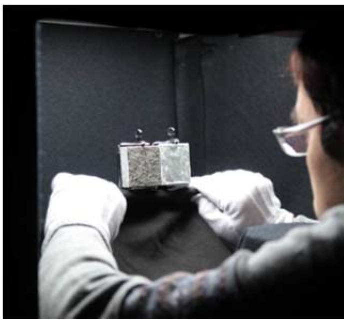


Figure 3: Snapshot of Experiment A.

Experimental results

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.

Intra- and inter-participant 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 3. The intra-participant variance was defined as:

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

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

$$\sigma_{\text{inter}}^2(i) = \frac{1}{20 \times 435} \sum_{k=1}^{20} \sum_{l=1}^{435} \left(\overline{b}_l(i) - b_{k,l}(i) \right)^2 , \qquad (3)$$

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 2 summarises 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 2 shows the average variance of the ratings among the 435 samples. As shown in Table 2, we 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 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.

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

Table 2: Variances in the Inter- and intra-participant ratings.

Perceptual harmony ratings within and across the categories of materials

Table 3 summarises 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 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 categorised 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.

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

Table 3: Averaged harmony ratings for the categories of materials.

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 harmonised across material categories.

Figure 4 shows the averaged harmony ratings for each sample pair. The colours indicate the averaged ratings as specified by the colour bar in Figure 4(d). Red indicates harmony whereas blue indicates disharmony. As shown in Figure 4, 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 harmonised among material categories, as indicated by the red dotted lines in Figure 4. 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 (grey calfskin, leather).

Changes in harmony between experiments

As shown in Table 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 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.

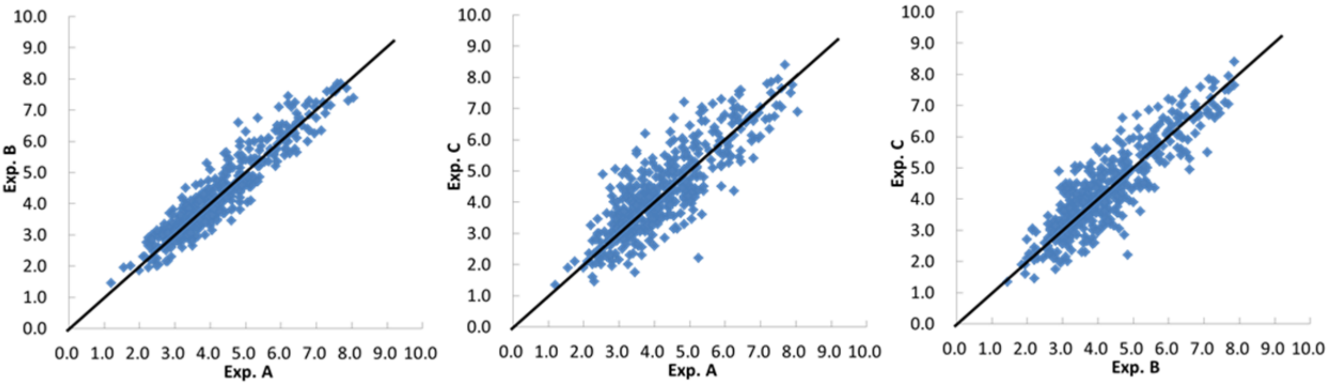


Figure 4: Scattered graphs of averaged harmony ratings between experiments: (a) correlation between Experiments A and B (R^2=0.89) (left); (b) correlation between Experiments A and C (R^2=0.71) (middle) and (c) correlation between Experiments B and C (R^2=0.78) (right).

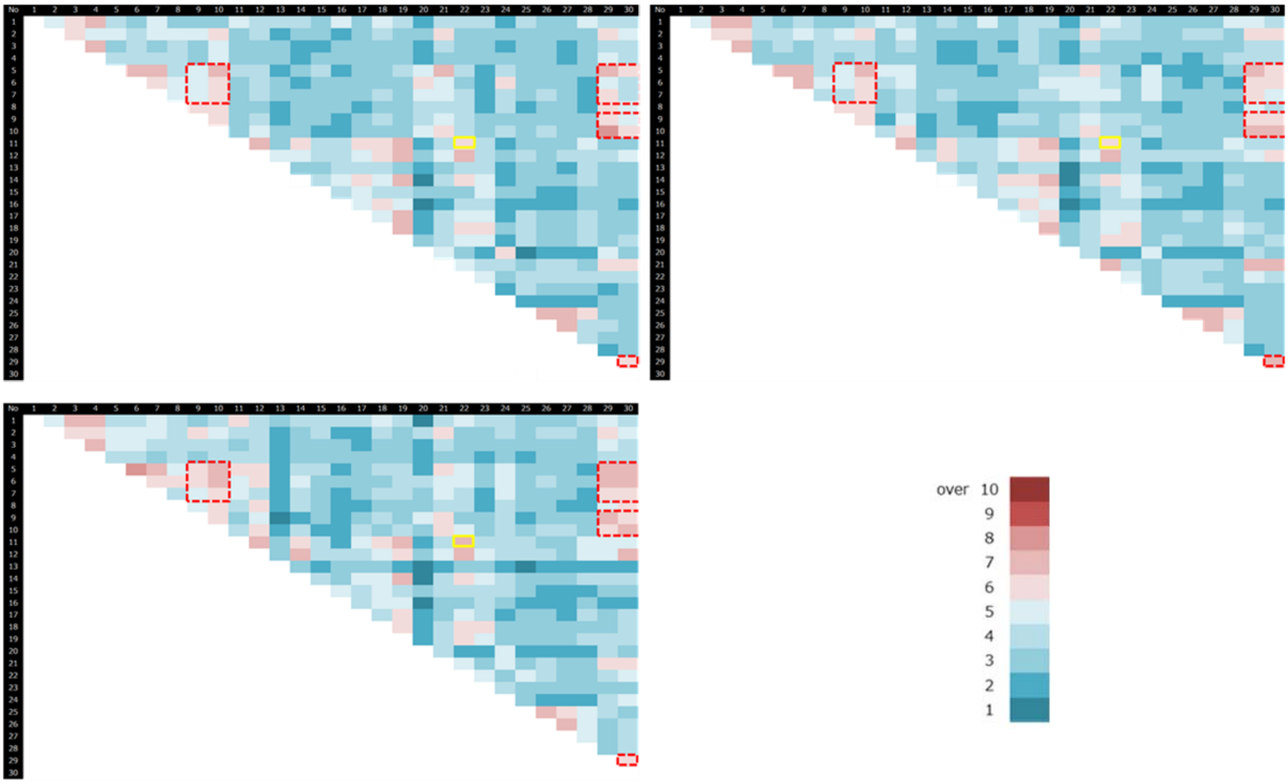


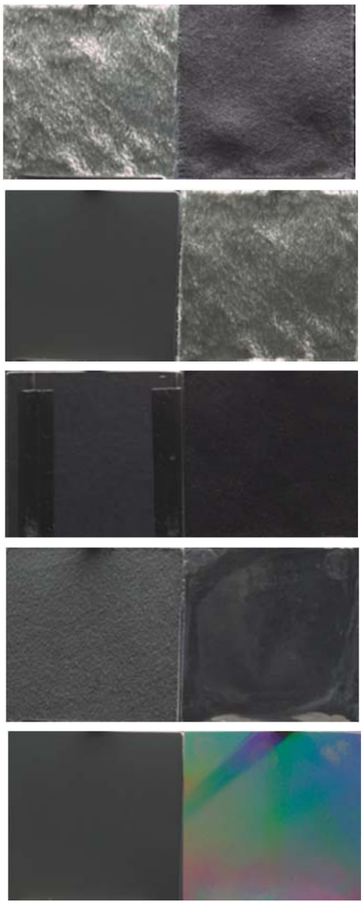
Figure 5: Harmony ratings provided by the twenty subjects for Experiment A (top left), Experiment B (top right) and Experiment C (bottom left). The index of the ratings is shown in 10 levels (bottom right). The vertical and horizontal numbers correspond to the material numbers for the samples on the left and right, respectively, as presented in Table 1.

As shown in Figure 5, some of the ratings for each pair differed in the experiments. Table 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 6(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 grey, metal) and 15 (satin, fabric), as shown in Figure

6(b), had a high rating in Experiment A because these glossy surfaces with grey colour 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 colour. In these cases, the appearance harmony was highly sensitive to the reflectance properties.

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)

Table 4: Harmony changes between experiments.



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

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

(c) Pair 8 (pearl grey, 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 grey, metal) and 20 (H-2, paper). A: 5.25, B: 4.85, C: 2.20.

Figure 6: Pairs that changed significantly between experiments.

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 grey, glass) and 12 (pig suede, leather), and Pair 11 (saddle leather matte, leather) and 24 (silver coated paper, paper) as shown in Figures 6(c) and 6(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 grey, metal) and 20 (H-2, paper), as shown in Figure 6(e).

Pair 5 (almite grey, 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 colour displayed on the monitor, as shown in Figure 6(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 6(d). In these cases, the material appearance may have differed between the real objects and the displayed images.

Distributions of samples in the appearance harmony space

We performed a principal component analysis (PCA) of all the ratings across materials to facilitate the visualisation of the distributions of material classes in the appearance harmony feature space. We 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 we assumed symmetry among the rating. The diagonal components were postulated to be the maximum ratings, because we considered that a combination of the same stimuli should be harmonised and the maximum rating was a reasonable assumption. We derived 30 dimensions and 30 principal components (PCs) from the matrix. Therefore, materials with the same harmony properties had the same PC coefficients.

Table 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 7 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, we can obtain an approximation of the overall distribution by simply using 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

Table 5: Percentage variance explained by the first three PCs.

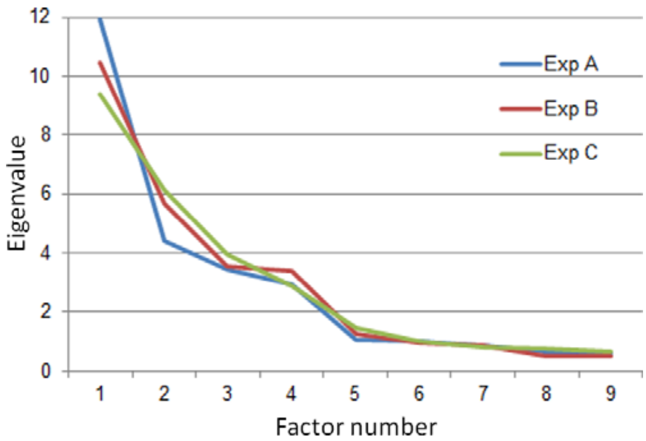


Figure 7: Scree plot for the PCs in all three experiments.

Figure 8 shows the ratings for each sample projected onto the first two PCs where each image is colour coded by its true class membership. The open circles indicate the average for each material class and the same colour corresponds to the same material category. The observed distribution in the PC space has several key features. As shown in Figure 8(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 8(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 8(c). This suggests that the degree of harmony did not depend on the material clusters in the displayed images.

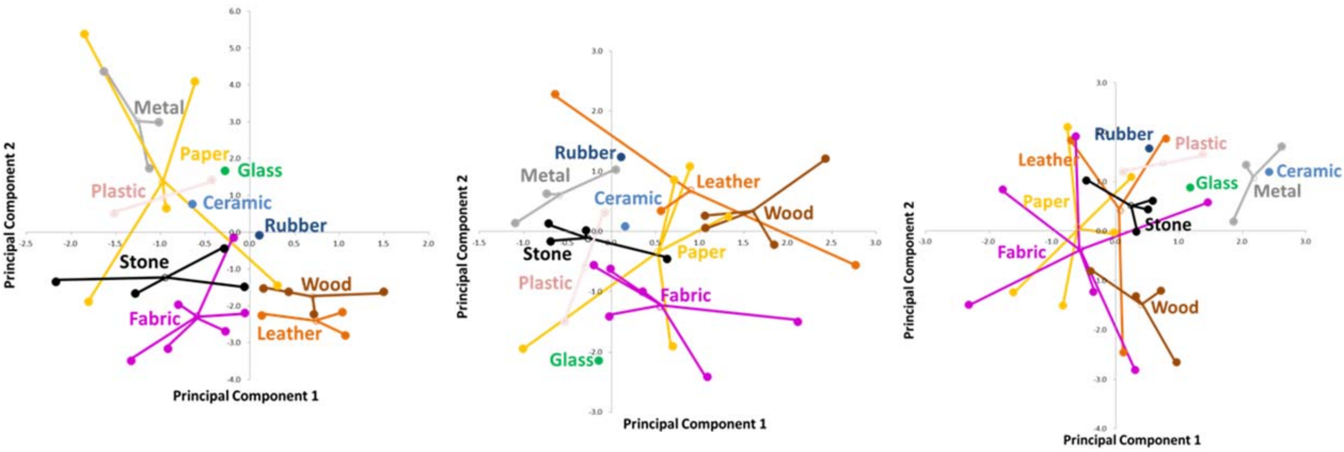


Figure 8: Distribution of samples in the first two PCs in Experiment A (left), Experiment B (middle) and Experiment C (right).

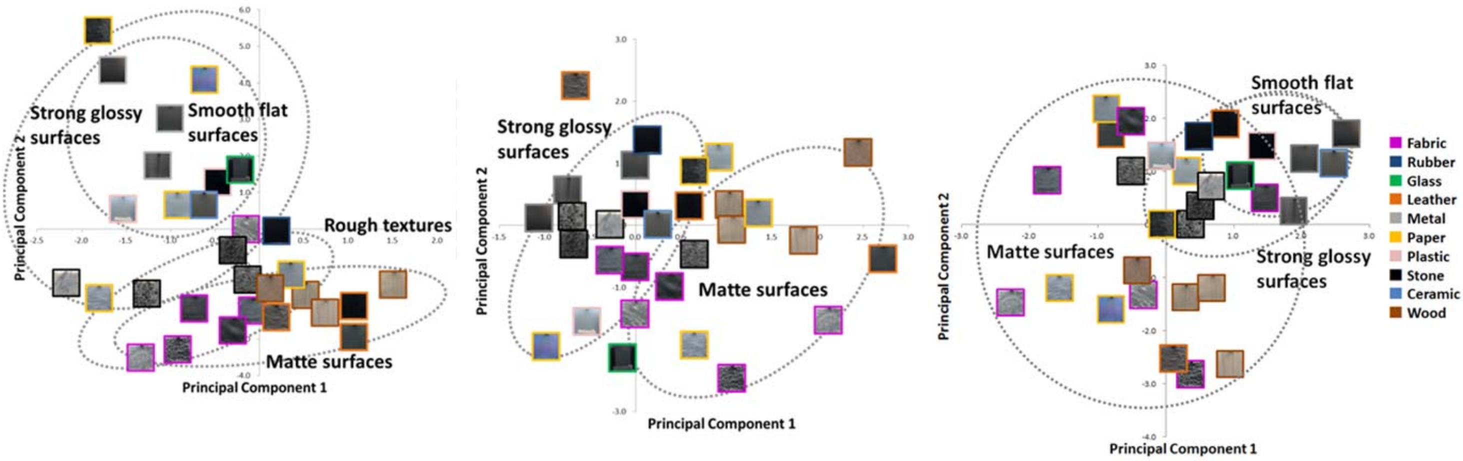


Figure 9: Samples categorised according to the similarities in appearance harmony in Experiment A (left), Experiment B (middle) and Experiment C (right).

Figure 9 shows the material samples projected onto the components in the first two PC spaces in Figure 8. The boundary colour of each component corresponds to the material clusters. We classified the samples with common properties, which have been encircled using a broken line. The property of glossiness has been presented in Table 1. Figure 10 shows the glossiness map of each sample. It was difficult to measure roughness of all materials under the same condition; therefore, we judged the roughness property subjectively.

In Experiment A, we found that the reflectance properties of material surfaces comprised an important factor that affected the appearance harmony. As shown in Figure 9(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.

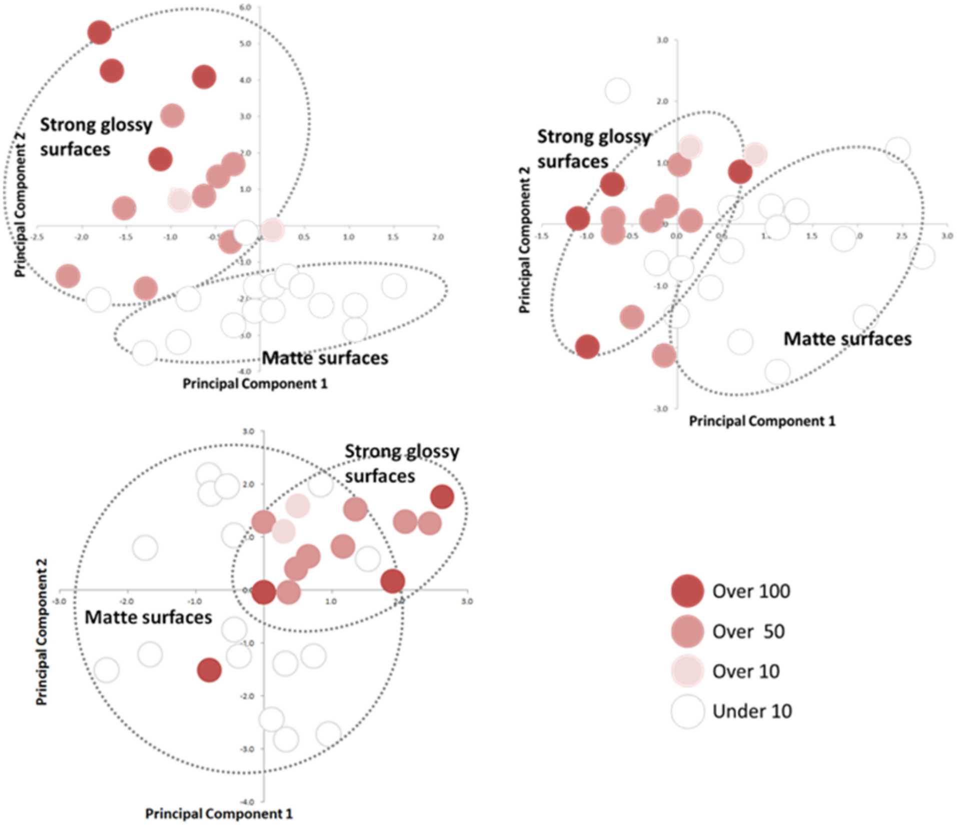


Figure 10: Glossiness maps for Experiment A (top left), Experiment B (top right) and Experiment C (bottom left). The index is represented in 4 levels (as shown in bottom right).

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.

We also applied k-means clustering to the harmony rating data. By comparing the true clusters with those extracted by k-means clustering, we 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, we distributed initial seeds, at random 10 times. Figure 11 shows the clustering results when k = 5. Experiments A and B obtained the same clustering results indicated as red line in Figure 11, which shared similar harmony properties. The clustering result in Experiment C has been indicated using a blue line in Figure 11.

It should be noted that *k*-means clustering was performed using 30-dimensional harmony rating data, whereas the plot shown in Figure 8 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, colour 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 Experiment C, which supports the results shown in Figure 8. The other materials were separated into different clusters. These results indicate that some sample pairs were viewed as harmonious, although the materials were different.

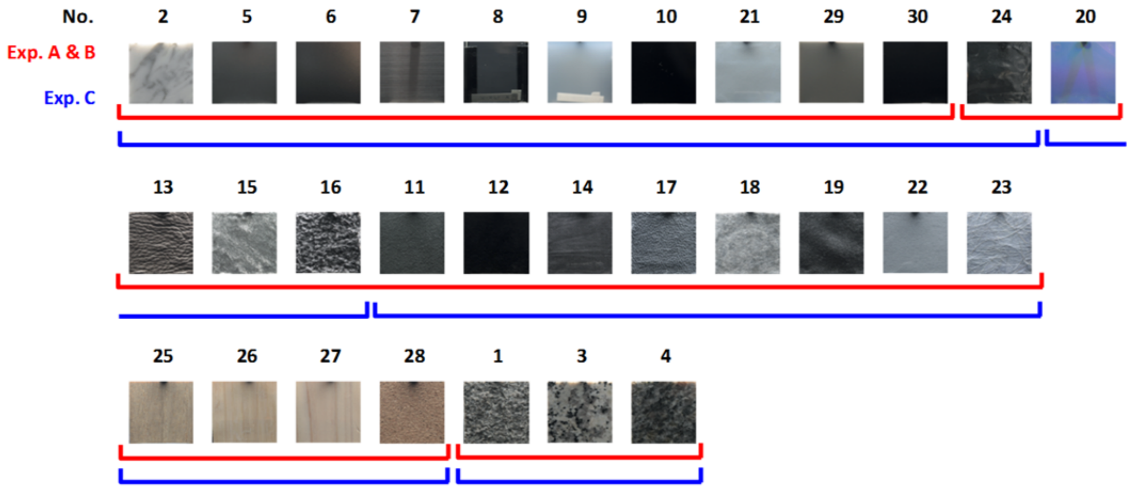


Figure 11: Category transition among experiments in harmony rating data using k-means algorithm (k = 5).

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.

Consideration for the colour effect

As described in the Experimental Stimuli section, colour has a strong influence on perceptual qualities. Therefore, we used exemplars with low saturation. In this section, we analyse the effect of colour to harmony in our experiments by conducting an additional experiment (Experiment D). In Experiment D, the colour images used in Experiment C were converted to grey images. We carefully chose 87 grey images that had similar distribution with all 435 pairs in regard to the CIE xy distance between two materials shown in Figure 12. 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 Figure 13, colour and harmony rating differences between two materials were independent. These results suggest that the colour effect in our experiments was weak using our samples. However, the correlation for all pairs excluding brownish coloured wood samples increased 0.88. Figure 14 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 colours were different each other. In contrast, in Experiment D, the pair had a high rating of 7.1, because colour difference disappeared and their surface textures were resembled. In this case, colour might influence for evaluating appearance harmony.

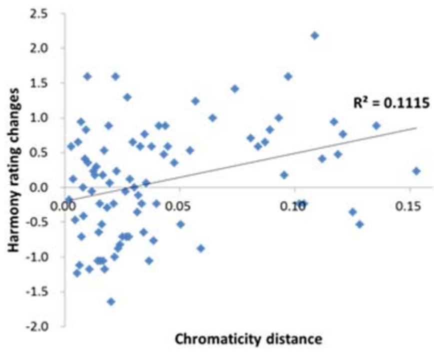
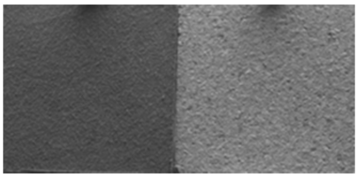


Figure 12 (left): A histogram of the CIE xy distance for each pair in Experiment D.

Figure 13 (right): Relationships between harmony rating changes and chromaticity distance.

(a) A colour image used in Experiment C.



(b) A grey image used in Experiment D.

Figure 14: The pairs showing maximum change between Experiments C and D [Pair 11 (saddle leather matte, leather) and 28 (cork, wood)].

Conclusions

We 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, we 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, we 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.

Acknowledgement

This work was supported by Grant-in-Aid for Scientific Research on Innovative Areas "Shitsukan" (No. 25135706) and (15H05914) from MEXT, Japan.

References

1. Judd DB and Wyszecki G (1975), *Color in Business, Science and Industry*, 3rd edition. New York: Wiley, 390-396.
2. Hård A and Sivik L (2001), A theory of colors in combination: A descriptive model related to the NCS color-order system. *Color Research and Application*, **26** (1), 4-28.
3. Burchett KE (2002), Color harmony, *Color Research and Application*, **27** (1), 28-31.
4. Ou L and Luo MR (2006), A colour harmony model for two-colour combinations, *Color Research and Application*, **31** (3), 191-204.
5. Ou L, Chong P, Luo MR and Minchew C (2010), Additivity of colour harmony, *Color Research and Application*, **36** (5), 335-372.
6. Bar M and Neta M (2006), Humans prefer curved visual objects, *Psychological Science*, **17**, 645-648.
7. Nasser KM and Marjan T (2010), Design with emotional approach by implementing Kansei engineering—Case study: Design of kettle, *Proceedings of the International Conference on Kansei Engineering and Emotion Research*, 625-632, Paris, France.
8. Ye Y, Zhang Z and He R (2014), Study on design of chair shaping based on Kansei engineering, *International Journal of Scientific and Engineering Research*, **5**, 273-276.
9. Chen N, Tanaka K, Matsuyoshi D and Watanabe K (2013), Cross preference for color combination and shape, *Proceedings of The First Conference of Asia Color Association*, 98-101, Thanyaburi, Thailand.
10. 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.
11. Fleming RW, Dror RO and Adelson EH (2003), Real-world illumination and the perception of surface reflectance properties, *Journal of Vision*, **3**, 347-368.
12. Motoyoshi I and Matoba H (2012), Variability in constancy of the perceived surface reflectance across different illumination statistics, *Vision Research*, **53**, 30-39.
13. 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** (12), 2951-2965.

14. Fleming RW and Bülthoff HH (2005), Low-level image cues in the perception of translucent materials, *ACM Transactions on Applied Perception*, **2** (3), 346-382.
15. Fleming RW, Jäkel F and Maloney LT (2011), Visual perception of thick transparent materials, *Psychological Science*, **22** (6), 812-820.
16. Motoyoshi I (2010), Highlight–shading relationship as a cue for the perception of translucent and transparent materials, *Journal of Vision*, **10** (9), 1-11.
17. Padilla S, Drbohlav O, Green PR, Spence AD and Chantler MJ (2008), Perceived roughness of 1/f noise surfaces, *Vision Research*, **48**, 1791-1797.
18. Pont SC and Koenderink JJ (2005), Bidirectional texture contrast function, *International Journal of Computer Vision: Special Issue on Texture Synthesis and Analysis*, **62** (1-2), 17-34.
19. Pont SC and Koenderink JJ (2008), Shape, surface roughness, and human perception, in Handbook of Texture Analysis, Mirmehdi M, Xie X and Suri J (eds.), Imperial College Press, 197–222.
20. Albertazzi L and Hurlbert A (2013), The perceptual quality of color, in *Handbook of Experimental Phenomenology: Visual Perception of Shape, Space and Appearance*, Mirmehdi M and Albertazzi L (eds.), 369-394, John Wiley & Sons.