Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance, pp. 52-55, Edinburgh, UK, ISBN 978-1-4716-6869-2, April 2012.

Integration of Multiple Cues for Visual Gloss Evaluation

Frédéric B. Leloup, Peter Hanselaer
Light & Lighting Laboratory, Catholic University College Sint-Lieven Gebroeders De Smetstraat 1, B-9000 Gent, Belgium
frederic.leloup@kahosl.be
peter.hanselaer@kahosl.be

Michael R. Pointer
Department of Colour Science, University of Leeds
Leeds LS2 9JT, UK
mrpointer@btinternet.com

Philip Dutré
Computer Graphics Group, KU Leuven
Celestijnenlaan 200A, B-3001 Heverlee, Belgium
philip.dutre@cs.kuleuven.be

ABSTRACT
This study reports on a psychophysical experiment with real stimuli that differ in multiple visual gloss criteria. Four samples were presented to 15 observers under different conditions of illumination, resulting in a series of 16 stimuli. Through pairwise comparisons, a gloss scale was derived and the observers' strategy to evaluate gloss was investigated. The preference probability matrix **P** indicated a dichotomy among observers. A first group of observers used the distinctness-of-image as a principal cue to glossiness, while a second group evaluated gloss primarily from differences in brightness. It could therefore be questioned if surface gloss can be characterized by one single quantity, or that a set of quantities is necessary to describe differences in gloss.

Keywords
Gloss perception, distinctness-of-image, brightness.

1. INTRODUCTION
The understanding of how human observers evaluate and interpret surface gloss has received increased attention during the last decades. Hunter was first to investigate the multidimensional nature of gloss perception, and defined six criteria that could possibly be used to rank surfaces based on their gloss: specular gloss, contrast gloss, dinstinctness-of-image (DOI) gloss, sheen, absence-of-bloom gloss, and surface-uniformity gloss [1].

Several studies questioned the perceptual gloss dimensions since then [2-4]. However, less has been reported on how these multiple cues interact and are integrated for the perception of surface gloss [5]. In this study, the multidimensionality of gloss perception and the integration of multiple cues for the visual evaluation of surface gloss are examined.

2. EXPERIMENTAL SETUP
Two identical sets of four black samples were selected for the study. Standard specular gloss measurements were performed on all samples at the three basic geometries (20°, 60°, 85° angle of incidence) with a Byk-Gardner micro-TRI-gloss-S glossmeter. Mean values and variances from 5 consecutive measurements are reported in Table 1.

Predicting Perceptions: The 3rd International Conference on Appearance, 17-19 April, 2012, Edinburgh, UK. Conference Proceedings Publication ISBN: 978-1-4716-6869-2, Pages: 52-55

Spatial reflection characteristics were determined with a home- built measurement instrument, capable of determining the absolute bidirectional reflectance distribution function (BRDF) [6]. To mimic the geometric conditions applied in both the specular glossmeter and the light booth (see below), the angle of incidence with respect to the sample normal was fixed at 60°. The viewing angle ranged from 0° to 80° in the opposite half-plane, 60° corresponding to the specular reflection direction. Absolute spectral BRDF values $q_{e,\lambda}$ of all four samples are presented in Fig. 1. The BRDF functions can be clearly discerned from each other, with both differences in specular peak values and in the width of the curves (related to the DOI).

Table 1. Average specular gloss values obtained in three basic geometries (20°, 60°, 85°), expressed in specular gloss units (SGU).

Sample	20° (SGU)	60° (SGU)	85° (SGU)
A	85.0 ± 0.4	91.0 ± 0.2	99.2 ± 0.1
B	52.1 ± 1.3	88.3 ± 0.4	94.2 ± 0.5
C	31.6 ± 0.7	70.7 ± 0.5	88.0 ± 0.3
D	16.4 ± 0.2	53.8 ± 0.5	78.1 ± 0.5

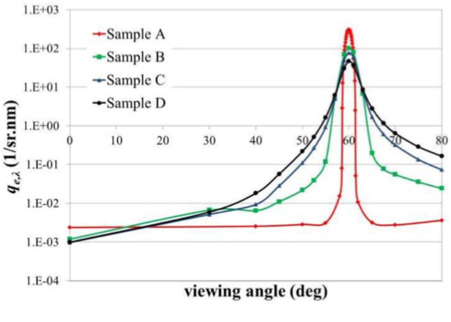

Figure 1. BRDF functions at an angle of incidence of 60° and at wavelength 589 nm. The viewing angle ranges from 0° to 80° in the opposite half-plane.

Psychophysical experiments were conducted in a light booth presented in Fig. 2. Two uniform rectangular light sources are positioned 60 cm from the sample holder, with an incidence angle of 60° toward the sample holder normal. The luminance of both sources can be regulated separately, by adjusting the current through two current-voltage sourcemeters. An additional ambient light source is positioned perpendicular to the sample holder, at a distance of 60 cm. This source generates a constant illuminance of approximately 500 lux on the samples.

Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance, pp. 52-55, Edinburgh, UK, ISBN 978-1-4716-6869-2, April 2012.

By application of different illumination settings to each of the 4 samples, a series of 16 stimuli was created. The luminance of the specular light sources was regulated such that for each sample approximately the same luminance of the specular highlight was obtained at four distinct levels; (1) low (±300 $cd.m^{-2}$), (2) medium-low (±700 cd.m), (3) medium-high (±1000 cd.m), and (4) high (±1300 $cd.m^{-2}$). Thus, 4 different stimuli were created with each sample. Each stimulus is further denoted by a letter/number combination, applying to the sample and the highlight luminance level, respectively.

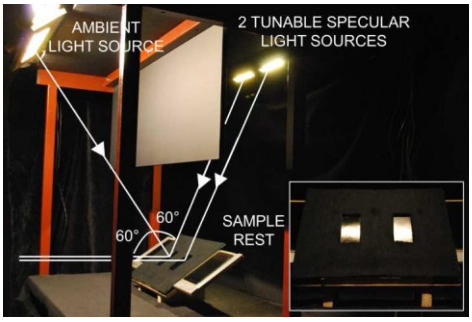

Figure 2. Side-view of the light booth, together with a close-up of 2 stimuli employed in the study.

3. VISUAL EXPERIMENTS

Stimuli were presented in pairs to the observers, who viewed both samples under test, each in the mirror reflection direction of one of the two specular light sources. Observers were told that they would be presented the samples under different conditions of illumination, but not that these conditions would differ between both samples under test. 15 observers (7 female, 8 male; age range 22-56 years) rated each possible pair of all 16 stimuli in both orders of presentation (left vs. right). On the exception of one author, all observers were naïve as to the purpose of the experiment. All observers had normal or corrected to normal visual acuity.

An overall gloss scale of the 16 stimuli was derived, based on the method of analyzing paired comparisons as proposed by Scheffé [7]. All observers rated the glossiness of the left sample i as compared to the right sample j, by use of a preference scale ranging from -2 to 2. When observers rated both alternatives to be equal, they were still asked to indicate a preference. As such, all responses could be converted to wins or losses of a two- alternatives forced choice (2AFC) task. The converted responses were used to analyze the preference probability matrix **P**, in which is reported the average number of times pij that stimulus i is rated to be glossier than stimulus j, based on 30 observations.

4. RESULTS

The derived overall gloss scale is presented in Fig. 3. As gloss differences between stimuli are reported when only one of both cues differs, it can be concluded that both differences in DOI and differences in brightness affect gloss appraisal. However, DOI seems to be the most preponderant cue. Indeed, the 4 stimuli incorporating sample A are rated to be glossier than all other stimuli. On the exception of stimulus B1, all stimuli incorporating sample B are rated to be glossier than stimuli that show an inferior DOI, i.e., stimuli incorporating sample C or sample D. When differences in DOI become smaller (e.g., a stimulus incorporating sample C in comparison with a stimulus incorporating sample D), the relative importance of brightness differences increases. Both cues seem now to be simultaneously taken into account, and the order of increasing DOI and increasing brightness with surface gloss is maintained in the scale. In analogy with Sakano and Ando [5], these general results may be explained by two simple models of interactions and integration of the glossiness cues, i.e., the disambiguation model and the linear combination model.

The preference probability matrix **P** is presented in Fig. 4. According to Dixon and Mood [8], at most 9 out of the 30 observations may indicate the less preferred stimulus in order to conclude that the two stimuli under test are different at the 5% level of significance. The elements indicating a significant difference between two stimuli are highlighted in green. Significant differences are always encountered, except when stimuli are assessed that show conflicting cues. In the situation of conflicting cues, one stimulus has a higher DOI yet lower highlight brightness than the other stimulus, or vice versa. It seems now, that a part of the observers base their judgment primarily on the differences in DOI, while another part of the observers base their judgment on the differences in brightness. This could suggest that observers did not have a consistent idea of what surface appears glossier in the presence of cue conflicts. This was however not the case, as each observer's internal consistency was checked to be satisfactory high.

Therefore, it seems that two groups of observers can be derived, which use a different strategy to assess gloss. This dichotomy is corroborated by calculation of the Spearmann rank correlation coefficient between every two observers. Positive correlations are found between each two observers of a first group of 10 observers, and between each two observers of a second group of 5 observers ($p < 0.05$).

It could therefore be questioned if surface gloss can be characterized by one single quantity, or that a set of quantities is necessary to describe visual gloss differences. If more studies indicate a dissension between observers, future psychophysical studies could probably better concentrate on the characterization of single cues to glossiness. Indeed, while it is complicated to determine the overall visual gloss impression from physical measurements, it is probable that single cues to glossiness can be quantified more accurately.

5. REFERENCES

[1] Hunter, R. S. 1937. Methods of determining gloss. *Natl. Bur. Stand. Research Paper RP958, J. Res. Natl. Bur. Stand.* 18 (1937), 19-39.

[2] Billmeyer, F. W. and O'Donnell, F. X. D. 1987. Visual gloss scaling and multidimensional scaling analysis of painted specimens. *Color Res. Appl.* 12 (Dec. 1987), 315-326.

[3] Pellacini, F., Ferwerda, J. A., and Greenberg, D. P. 2000. Toward a psychophysically-based light reflection model for image synthesis. In *Proceedings of the ACM Transactions on Graphics (SIGGRAPH'00)*, (New Orleans, Louisiana, USA, 23-28 July 2000), ACM, New York, NY, 55–64.

[4] Ged, G., Obein, G., Silvestri, Z., Le Rohellec, J., and Viénot, F. 2010. Recognizing real materials from their glossy appearance. *Journal of Vision* 10, 9 (Dec. 2010), 18:1-17.

[5] Sakano, Y. and Ando, H. Effects of head motion and stereo viewing on perceived glossiness. *Journal of Vision* 10, 9 (Nov. 2010), 15: 1-14.

[6] Leloup, F. B., Forment, S., Dutré, P., Pointer, M. R., and Hanselaer, P. 2008. Design of an instrument for measuring the spectral bidirectional scatter distribution function. *Appl. Opt.* 47 (Oct. 2008), 5454-5467.

[7] Scheffé, H. 1952. An analysis of variance for paired comparisons. *Journal of the American Statistical Association* 47 (Sept. 1952), 381-400.

[8] Dixon, W. J. and Mood, A. M. 1946. The statistical sign test. *Journal of the American Statistical Association* 41 (Dec. 1946), 557-566.

Predicting Perceptions: Proceedings of the 3rd International Conference on Appearance, pp. 52-55, Edinburgh, UK, ISBN 978-1-4716-6869-2, April 2012.

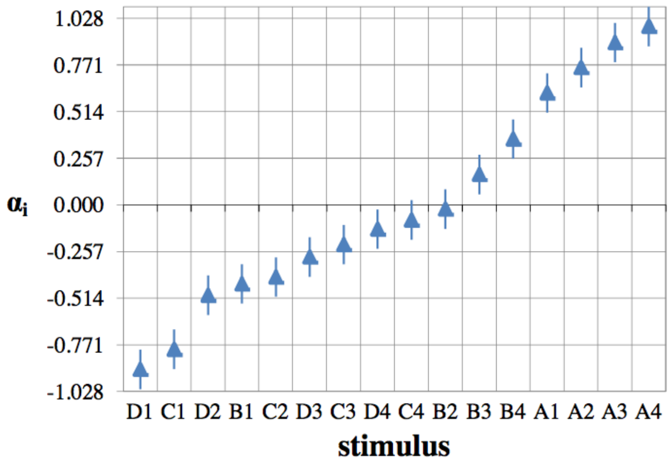

Figure 3. Overall gloss scale of the 16 stimuli.

D	4																
	3																0.03
	2															0.00	0.00
	1														0.00	0.00	0.00
C	4													1.00	1.00	0.93	0.67
	3												0.27	1.00	0.93	0.83	0.33
	2											0.20	0.03	1.00	0.87	0.33	0.23
	1										0.07	0.03	0.03	0.77	0.17	0.13	0.13
B	4									1.00	0.97	0.93	0.77	0.97	0.97	0.90	0.80
	3								0.17	1.00	0.90	0.83	0.70	0.97	0.90	0.80	0.73
	2							0.13	0.07	1.00	0.73	0.63	0.53	1.00	0.80	0.70	0.63
	1						0.03	0.00	0.00	0.80	0.53	0.43	0.43	0.83	0.50	0.60	0.57
A	4					0.97	0.80	0.80	0.73	0.93	0.83	0.77	0.73	0.93	0.77	0.73	0.77
	3				0.20	0.93	0.70	0.77	0.67	0.83	0.73	0.73	0.70	0.93	0.77	0.73	0.67
	2			0.23	0.03	0.93	0.80	0.67	0.67	0.87	0.80	0.67	0.67	0.87	0.77	0.73	0.70
	1		0.17	0.00	0.03	0.77	0.73	0.67	0.63	0.80	0.70	0.67	0.63	0.80	0.70	0.70	0.70
Stimulus *i*		1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4
Stimulus *j*		A				B				C				D			

Figure 4. Preference probability matrix P, in which is indicated the average number of times p_{ij} that stimulus *i* is rated to be glossier than stimulus *j*.