Examining Perceptual Luminance Uniformity of Simulated Luminaire Patterns

Belal Abboushi
Eduardo Rodriguez-Feo Bermudez
Lia Irvin
Benjamin Feagin Jr.
Michael Royer
Robert Davis

Pacific Northwest National Laboratory
620 SW 5th Avenue, Suite 810
Portland, OR 97204
Belal.Abboushi@pnnl.gov

This is an archival copy of an article published in *Light-Emitting Devices, Materials, and Applications XXV*. Please cite as:

Belal Abboushi, Eduardo Rodriguez-Feo Bermudez, Lia Irvin, Benjamin Feagin, Michael Royer, Robert Davis, "Examining perceptual luminance uniformity of simulated luminaire patterns," Proc. SPIE 11706, *Light-Emitting Devices, Materials, and Applications XXV*, 117061E (8 March 2021); https://doi.org/10.1117/12.2577901

Examining Perceptual Luminance Uniformity of Simulated Luminaire Patterns

Belal Abboushi*[a], Eduardo Rodriguez-Feo Bermudez[a], Lia Irvin[a], Benjamin Feagin Jr. [a], Michael Royer[a], Robert Davis[a]

[a] Pacific Northwest National Laboratory. 620 SW 5th Ave, Suite 810, Portland, OR 97204

ABSTRACT

Luminaire luminance uniformity is an important aspect that can affect perceived lighting quality, discomfort glare, and efficacy. While several metrics have been proposed to characterize luminance uniformity, previous studies have shown that current metrics such as Max:Min or Avg:Min luminance ratios can be insensitive to important differences in luminance gradient that may affect perceived uniformity. In an attempt to resolve this issue, previous studies incorporated a contrast sensitivity function for the human eye based on spatial frequency, such as in the recently proposed U_{HVS} metric; however, this metric has not been comprehensively studied in relation to perceived uniformity ratings.

The study presented in this paper aimed to examine the relationship between U_{HVS} and perceived uniformity ratings. Specifically, simulated luminance patterns were presented, and participants were asked to assess uniformity using a two-alternative forced-choice procedure. The results of 94 participants' evaluations showed a significant correlation between U_{HVS} and perceived uniformity. However, comparisons between patterns that had similar U_{HVS} sometimes resulted in statistically different ratings and comparisons between patterns that had larger differences in U_{HVS} sometimes did not result in a statistically significant difference in ratings. These results suggest that U_{HVS} might be used for general guidance but may warrant further studies to better understand its sensitivity and improve its alignment with perceived uniformity ratings.

Keywords: perceived uniformity, luminaire luminance uniformity, simulated patterns, uniformity metrics.

1. INTRODUCTION

1.1 Luminaire Luminance Uniformity

Luminaire Luminance Uniformity (LLU) describes variations in luminance distribution across the face of the luminaire aperture and is influenced by many design variables. These variables include the luminous intensity and photometric distribution of LEDs, distance between LED array and optical material, spacing between LEDs, optical material properties, LED shape (including optical lens) and size, luminaire geometry and finish, viewing angle, and distance between observer and the aperture.

Previous studies showed that uniform luminaires were perceived less glary compared to non-uniform luminaires at the same average luminance or illuminance at the eye.[1,2] While increasing LLU is generally desired,[3] it may reduce luminaire efficacy. This reduction occurs because improving the uniformity typically requires smoothing high luminance spots using more diffusive materials with lower transmittance. Overall, given the importance of LLU for visual comfort and efficacy, it is important to utilize a luminance uniformity metric that accurately characterizes human perception. Ultimately, a LLU metric can help lighting manufacturers and specifiers to make informed decisions that balance LLU and efficacy. Examples include selecting a luminaire with higher efficacy while delivering the same level of uniformity or selecting a luminaire that delivers a higher level of uniformity without compromising efficacy.

1.2 Uniformity Metrics

Most luminance uniformity metrics were developed with the goal of examining luminance uniformity within a space, such as for walls or outdoor areas,[4,5] and have not been applied to luminaire apertures. Currently, a commonly used metric for luminance uniformity is maximum-to-minimum luminance ratio (Max:Min). Since Max:Min only relies on two points, it may not describe the perceived uniformity of complex spatial patterns. Another metric is the ratio of the average luminance to minimum luminance (Avg:Min), which was adopted in the ANSI/IES-RP-8-18 document.[6] For both Max:Min and Avg:Min, a lower ratio implies a more uniform pattern. A third metric that is sometimes considered is the coefficient of variation (CV), which is the ratio of the standard deviation to the mean as shown in equation (1). This means that the entire luminous area is sampled, producing a more stable metric that is less likely to be affected by photometric measurement errors or other anomalies. For CV, a lower number implies a more uniform pattern.

$$CV = \frac{\sigma}{\bar{x}} \tag{1}$$

The recently proposed entropy uniformity (EU) was shown to be exponentially related to perceived uniformity[7]. EU is equal to one when the luminous surface is entirely uniform, and zero when entirely non-uniform. In equation (2), n is the number of luminance points measured and p_i is the ratio of the i^{th} luminance value to total luminance from all points.

$$EU = \frac{1}{n} \cdot exp\,(-\sum p_i\, ln(p_i)) \tag{2}$$

Metrics such as luminance ratios and the CV characterize the photometric conditions that ultimately affect a perception of uniformity. But the perception of luminance uniformity is also affected by the ways in which the human visual system processes the photometric stimulus. Ashdown[8] discussed luminance gradients in relation to contrast sensitivity and highlighted the need to measure the human visual system's ability to decipher luminance patterns. Another article showed that incorporating the contrast sensitivity function improved correlations with subjective preference of MR16 lamp beams[9]. One uniformity metric that combines photometric conditions with assumptions about visual processing is U_{HVS}[10], which is uniformity based upon the human visual system (HVS). In equations (3) and (4), the variables κ, α, β, and C are constants, and NU_{HVS} is the non-uniformity based on the human visual system. NU_{HVS} weights the summation of the Fourier transform of the luminance pattern $F(\omega_n)$ by the human visual contrast sensitivity function $CSF(\omega_n)$. It is then divided by the addition of a constant C added to the sum of the Fourier transform (the magnitudes of all spatial frequencies present in the data). A U_{HVS} value closer to one implies a more uniform pattern.

$$U_{HVS} = \frac{1}{1 + k \cdot CV^{\alpha} \cdot NU_{HVS}^{\beta}} \tag{3}$$

$$NU_{HVS} = \frac{\sum_n F(\omega_n) CSF(\omega_n)}{C + \sum_n F(\omega_n)} \tag{4}$$

$CSF(\omega_n)$ describes the human eye's sensitivity to luminance contrast as a function of spatial frequency[11]. Sensitivity increases up to about three cycles (spatial wavelengths) per degree in the visual field, and then decreases slowly to where ten cycles per degree is barely visible[10]. This means that although the presence of high frequencies indicates a less uniform pattern with more drastic differences in luminance, past a certain frequency, the human eye is less sensitive and therefore less able to detect these photometric differences.

In summary, previous studies used different uniformity metrics to examine and predict perceived uniformity. Accounting for CSF in uniformity metrics has likely improved these correlations[10] because CSF addresses the human visual system ability to distinguish bright from dim areas, compared to traditional metrics solely based on a statistical analysis. However, the U_{HVS} performance as a metric has not been tested using priori hypotheses. This study aimed to examine perceived uniformity in relation to existing uniformity metrics with a focus on U_{HVS}. It used an online questionnaire to present simulated luminance patterns to a group of participants who assessed perceived luminance uniformity. We hypothesized that: 1) patterns that had similar U_{HVS} value ($\pm$ 0.01) would receive similar perceived uniformity ratings; 2) patterns that had U_{HVS} values that differed by more than 0.01 would receive different uniformity ratings; 3) U_{HVS} would correlate with perceived uniformity ratings (spearman's coefficient $r > 0.7$, $p < 0.05$).

2. METHODS

The dependent variable in this study was perceived uniformity, as indicated by participants through a choice between two stimuli—these choices were later converted to ratings, as subsequently described. Given that this study was conducted online, there were inherent uncontrolled variables such as illumination conditions in participant's room, computer screen settings such as contrast and brightness, the amount of time spent viewing the patterns, and any individual variability in contrast sensitivity. Nonetheless, the next sections discuss some measures implemented to improve the quality of data collected.

2.1 Stimuli

Eight theoretical grayscale patterns were created, using formulas subsequently described, to represent possible luminance patterns. Variation in the patterns was created by manipulating the number of modelled point sources, the distance between these sources, and the distance between the point sources and diffuser. The patterns were generated such that: 1) all patterns had the same average luminance when viewed on a computer screen; and 2) of the eight patterns, six were pairs having a similar U_{HVS} value. Figure 1 shows the eight simulated patterns and Table 1 lists corresponding uniformity metric values. For U_{HVS} calculations, default constant values of $k = 5$, $\alpha = 1$, $\beta = 0.5$, and $C = 1\text{x}10^{-7}$ were used.[10]

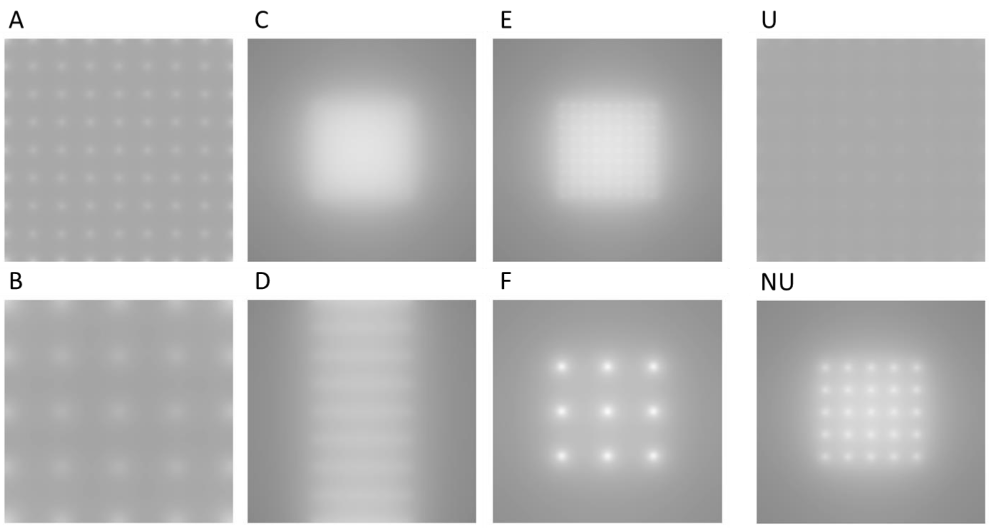

Figure 1: The eight patterns used in the experiment.

Table 1: Uniformity metrics of the eight simulated patterns. Pairs of interest are highlighted (pairs similar in U_{HVS}, or different in U_{HVS} but have visual similarity).

Pattern	**U_{HVS}**	**Pairs similar in U_{HVS}**	**Pairs different in U_{HVS}**	**Max:Min**	**Avg:Min**	**EU**	**CV**
A	0.95	A-B		1.31	1.02	1.00	0.02
B	0.96		A-C	1.31	1.04	1.00	0.03
C	0.88	C-D		1.78	1.25	0.99	0.17
D	0.89		C-E	1.56	1.24	0.99	0.13
E	0.82	E-F		1.83	1.25	0.99	0.17
F	0.81			1.98	1.17	0.99	0.11
U	0.99		F-NU	1.06	1.01	1.00	0.01
NU	0.79			1.86	1.23	0.99	0.15

The patterns were simulated in Python3, primarily using the NumPy and matplotlib libraries. In the simulation, two two-dimensional arrays were created representing the point sources and diffusing material. These two planes were parallel to each other with a distance (D) between them (Figure 2). Assuming a cosine distribution from a theoretical point source, its vector intensity (N) was distributed on the diffusing plane using equations (5) and (6):

$$N = \cos(\theta) \quad (5)$$

$$\theta = \arctan\left(\frac{\left(\left[X_{LED} - X_{Dif}\right]^2 + \left[Y_{LED} - Y_{Dif}\right]^2\right)^{1/2}}{D}\right) \quad (6)$$

where X and Y represent the coordinates of LED and receiving point on the diffusing plane and N is the vector intensity for θ range (-90° to 90°). Vector intensities that landed outside the diffusing plane, such as those reaching side surfaces, were recalculated to reflect inwards towards the diffusing plane assuming 80% Lambertian reflectance of these surfaces. Lastly, a linear grayscale was applied to the values of the diffusing plane to generate the patterns.

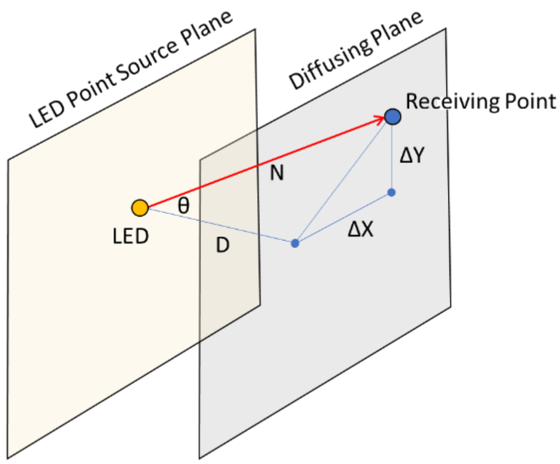

Figure 2: A diagram of the theoretical luminaire used to establish the luminance patterns, showing the emitting point and the receiver point.

2.2 Participants

Since computer screens from different manufacturers might have differences in their display capabilities, the sampling frame for this study was restricted to office employees working in one firm to improve the homogeneity of computer screens and laptop make. Participants were recruited using internal social-media and information exchange sites. No compensation was provided for participation. This study was approved by the institutional review board at the Pacific Northwest National Laboratory (No. IRB00011131).

To determine an appropriate sample size, calculations were conducted using G*Power software[12]. Assuming a medium Cohen's *D* effect size of 0.3, a power of 0.8, and a two-tailed test, a sample size of 84 was required to examine bivariate correlations, and a sample of 90 was required to examine differences using a paired *t*-test. Hence, for all anticipated statistical tests, a sample of 90 was determined to be enough to detect medium size effects.

2.3 Procedure

While conducting experiments online has important benefits, it also comes with limitations. For example, different computer screens and internet browsers might have different contrast and brightness settings, ambient illumination may vary among participants' rooms, and computer screen size and resolution cannot be controlled. The procedure used in this study included steps aimed to help, to some extent, address and document this variability.

The responses were collected using the online platform SurveyMonkey. On average, it took about eight minutes to complete the questionnaire. Duplicate responses from the same participant were prevented without collecting any personally identifiable information. After completing the consent form, participants were asked to: 1) view the questionnaire on the native laptop or PC screen and not to view the questionnaire using phones or tablets; and 2) to sit an arm's length away from the computer screen in a comfortable position.

An introduction consisted of four parts:

1. A set of questions asked about the participant's laptop make, internet browser, age, and vision condition (*e.g.*, if they needed lenses and whether they were wearing them or not).
2. The participants were shown two gradients (Figure 3) – one with black background and another with white background – and asked participants to click on the darkest/brightest bar that they could distinguish from the black/white background. Participants were not instructed to adjust screen contrast or brightness; instead, a procedure was used to ensure that participants could discern between different gradient levels, similar to the procedure used in previous studies[13,14]. Gradient discernment was checked because it was thought that contrast levels might affect perceived uniformity.
3. To ensure a consistent viewing size of the patterns across participants, participants were asked to adjust the viewing size (zoom) settings of their internet browser. The questionnaire showed a picture of a driver's license card and asked participants to hold their own license card against the screen while adjusting their browser viewing size to match the size of their actual card.
4. To help explain the questionnaire and patterns, participants were shown a picture of an office space with a luminaire, provided with a definition of uniformity as "the consistency/evenness of color across the face of the fixture" and viewed examples of a very uniform and a very non-uniform pattern.

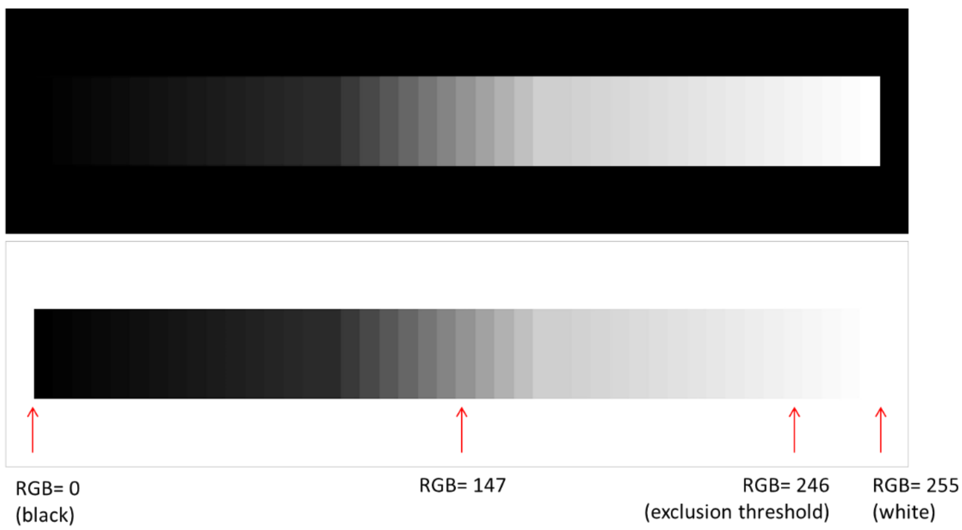

Figure 3: The gray bars with a black background (top image) and white background (bottom image) that were used to check gradient discernment. The red arrows highlight reference RGB values.

After the introduction, the two-alternative forced choice method (2AFC) was used such that each pattern was paired with every other pattern for uniformity assessment. Additionally, null conditions were created by pairing each pattern with itself. Participants were asked to assess the uniformity of the resultant 36 combinations/comparisons, responding to the prompt: "Please look at the two light patterns and click on the one that looks more uniform." The order of pairs was randomized to address order bias, and the left/right position of patterns was counterbalanced across participants to account for potential left/right bias.

3. RESULTS AND ANALYSIS

A total of 118 responses were collected. Incomplete responses (n = 8), those that needed corrective lenses but were not wearing them (n = 8), those with a visual disability that could not be corrected (n = 1), those that were not able to adjust their screen setting (n = 3), and those that could not distinguish at least the bar with RGB = 246 from the white background (n = 4, see Figure 3) were excluded. The reason the white background differentiation test was used for exclusion is because patterns generally did not include dark areas with RGB smaller than 147. RGB = 246 was used as an exclusion criterion because responses below that were determined to be outliers (*i.e.*, values that lie beyond the whisker: 75th percentile + 1.5 x interquartile range). These criteria resulted in 94 responses that were included in the analyses. Of the 94 participants whose data was included, 66 needed corrective lenses and were wearing them while completing the questionnaire. Regarding computer screen makes and internet browsers used, 62 participants used a Dell screen, and 83 used Google Chrome. The rest of participants used HP (n = 14), Acer (n = 1), Mac (n = 8), and other screen makes (n = 9). There were nine participants that used Firefox, and two that used Internet Explorer. Participants' ages were distributed across different age groups such that 17 participants were within the 18-29 age group, 21 were within 30-39, 22 were within 40-49, 20 were within 50-59, 12 were within 60-69, and two were within 70-79. This paper did not explore potential effects of computer screen make, internet browser, corrective lens use, or age.

3.1 Statistical Analysis

The mean number of times each pattern was selected as being more uniform is provided in Figure 4. Given that data from the two-alternative forced-choice procedure are ordinal related data, the non-parametric Friedman Rank Sum test was used for analyzing complete block designs where there were k = 8 experimental treatments (patterns) and b = 94 blocks (participants). Assumptions of ordinal data and randomized presentation order were evaluated and confirmed to be met. This test was conducted using R stats package and confirmed a significant difference in perceived uniformity among the eight patterns $X^2(7) = 427.95$, $p < 0.01$. After the Friedman test confirmed a significant difference, analyses of the previously identified pairs of interest were conducted using the Wilcoxon Signed Rank test. To address the first hypothesis, three pairs similar in U_{HVS} (A-B, C-D, and E-F) were tested. For the second hypothesis, there were several pattern combinations that could be compared, but we focused on three pairs that had visual similarity in terms of the geometrical arrangement of point sources and had different U_{HVS} values (A-C, C-E, and F-NU). Testing six comparisons required adjusting alpha, using the Bonferroni correction, to 0.05/6 comparisons = 0.0083 at the 5% level and 0.01/6 comparisons = 0.0016 at the 1% level. In the results below we use these corrected alpha levels and report the effect size.

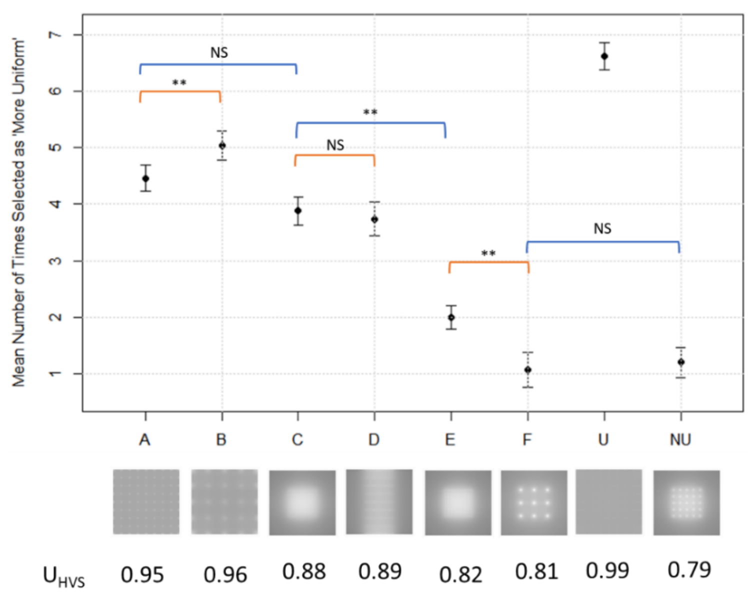

Figure 4: Mean number of times selected as more uniform for each pattern. Self-pairs (null conditions) were not included in this analysis, hence the maximum number of times a pattern can be selected as more uniform was seven. The bars show 95% confidence intervals. The symbol ** refers to the adjusted 1% significance level ($p < 0.0016$); and NS indicates not significant. The orange brackets are for comparisons between patterns with U_{HVS} within ± 0.01, whereas blue brackets are for comparisons between patterns with $U_{HVS} > 0.01$.

The Wilcoxon Signed Rank test showed that uniformity ratings for pattern B were significantly higher than A ($p < 0.0016$, $r = 0.39$) and ratings for E were significantly higher than F ($p < 0.0016$, $r = 0.46$). No significant difference was found between patterns C and D. While uniformity ratings for C and D were not significantly different, the results for A-B and E-F suggest rejection of the first hypothesis, which expected patterns that had similar U_{HVS} values (± 0.01) to receive similar uniformity ratings.

Regarding patterns that had different U_{HVS} values, uniformity ratings for patterns A and C ($p = 0.012$) as well as F and NU ($p = 0.27$) were not significantly different. Uniformity ratings for pattern C were significantly higher than E ($p < 0.0016$, $r = 0.77$). Thus, the results suggest rejection of the second hypothesis, that a difference in U_{HVS} greater than 0.01 would lead to differences in perceived uniformity.

3.2 Null conditions

The analysis of null condition pairs (*i.e.*, each pattern paired with itself) examined the percentage of times the pattern on the left and right were selected (Table 2). Wilcoxon Signed Rank tests showed no significant differences in any of the null comparisons, as well as no significant differences in overall left/right choices, indicating no significant left/right bias in the responses.

Table 2: The percentage of times left and right patterns (null conditions) were selected as more uniform.

Pairs	Left pattern (%)	Right Pattern (%)
A-A	50	50
B-B	44	56
C-C	50	50
D-D	41	59
E-E	48	52
F-F	46	54
U-U	50	50
NU-NU	55	45
Overall	48	52

3.3 Thurstone Model for Paired Comparisons (Case V)

Thurstone's Case V model is a popular method that can be used to investigate responses from paired comparisons where a participant is asked to select one of two stimuli[15]. This model assumes that 1) each participant has a continuous preference for each stimulus; 2) these continuous preferences are normally distributed; and 3) continuous preferences are uncorrelated and have a common variance.[16] The number of times each pattern was selected over another pattern was formulated as a proportion matrix. The R package 'psych'[17] was then used to calculate scaled Thurstone's values for all patterns, with a goodness of fit of 0.98.

The scaled values for all patterns are shown in Table 3. These scaled values are at the interval level and can be used in parametric statistical tests. Table 4 shows Pearson's correlation results. In this analysis, five metrics were examined, which required adjusting the significance level using Bonferroni correction 0.05/5 = 0.01, and 0.01/5 = 0.002. The EU metric was shown in a previous article[7] to be exponentially related to perceived uniformity; hence a log transformation was applied to the scaled Thurstone values. A constant of one was added to these values to avoid a zero value for pattern F. Both the U_{HVS} and the Max:Min metric values were significantly associated with scaled Thurstone values ($p < 0.002$), though the percent of variance explained using U_{HVS} was higher ($r^2 = 0.94$). This result supports the third hypothesis expecting U_{HVS} to be significantly correlated with perceived uniformity ratings.

Table 3: Scaled Thurstone values represent the perceived uniformity of patterns using the paired forced-choice responses. A higher value represents higher uniformity.

	Patterns							
	A	**B**	**C**	**D**	**E**	**F**	**U**	**NU**
Scaled Thurstone value	2.47	2.60	2.19	2.09	1.35	1	3.45	1.07

Table 4: Linear regression models of Thurstone scaled values+1 as predicted by each metric. P values adjusted to 5% to 0.01 (represented with *) and 1% to 0.002 (represented with **). nl: The log of Thurstone scaled values were used in the regression with EU to transform the exponential relationship to a linear relationship.

Statistical test	**Parameter**	**Uniformity Metric**				
		U_{HVS}	**Max:Min**	**Avg:Min**	**EU ^{nl}**	**CV**
Pearson's correlations	Coefficient	0.97**	-0.94**	-0.7	0.61	-0.76
	p value	<0.002	<0.002	0.05	0.11	0.03
Linear regression models	r^2	0.94**	0.88**	0.50	0.37	0.57
	p value	<0.002	<0.002	0.05	0.11	0.03
	Estimate	11.05	-2.42	-5.48	46.84	-9.28

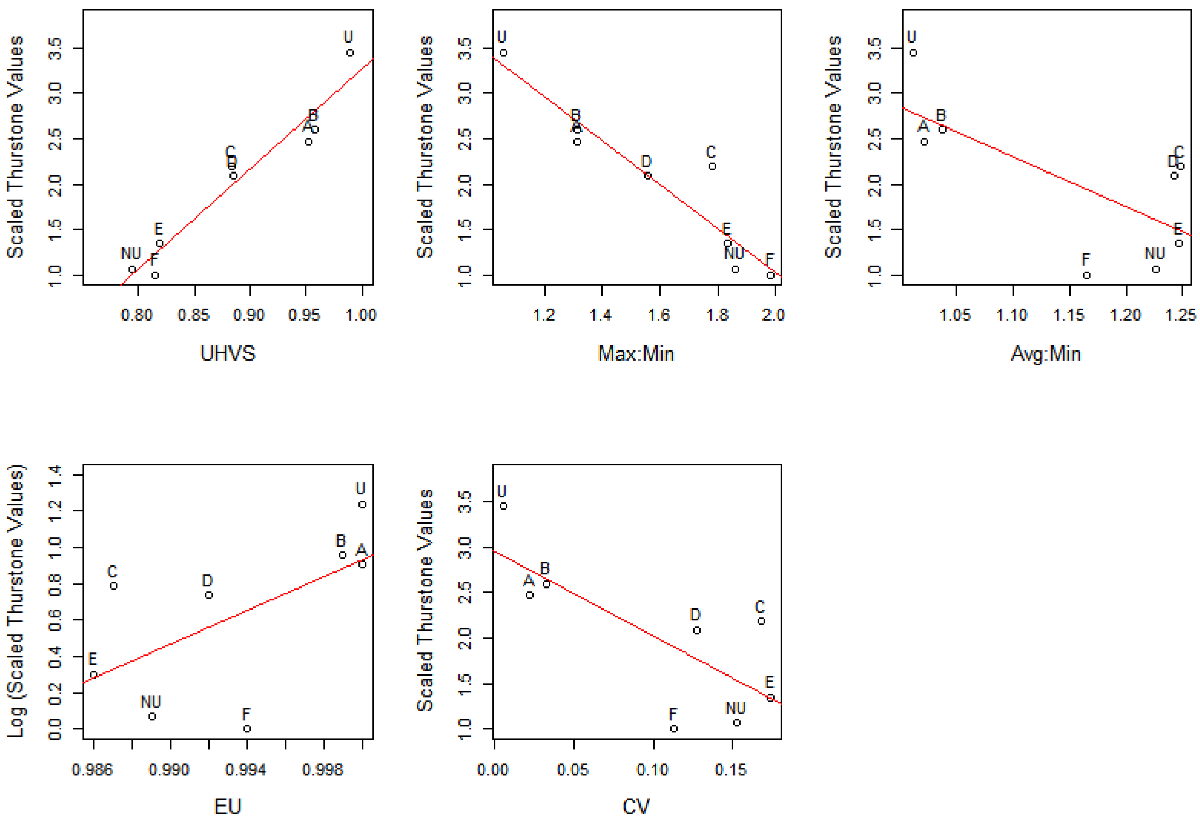

Figure 5: Scatterplots showing the relationship between different metrics and scaled Thurstone values. The red line is a linear regression fit. The letters refer to the patterns.

4. DISCUSSION

Although pattern pairs like A-B and E-F were similar in their U_{HVS} values (within 0.01 U_{HVS} units), there was a significant difference in their perceived uniformity. Pattern B, for example, was perceived more uniform than A. This could be due to assumptions underlying exponents in the U_{HVS} equation like α, β, and κ. For example, adjusting alpha from 1 to 0.7 increases regression model r^2 from 0.94 to 0.99 and results in a higher U_{HVS} for pattern B (0.891), compared to A (0.863). Based on the results of this study, there seems to be other factors that influenced uniformity ratings beyond those addressed in the current U_{HVS} formulation.

In contrast, patterns A and C had a difference of 0.07 in U_{HVS} but the results did not show a significant difference in ratings. Likewise, subjective uniformity ratings for patterns F and NU were not significantly different. It should be noted that the pair A-C had a p value = 0.012, which is close but did not achieve statistical significance using the Bonferroni corrected threshold of 0.0083 at the 5% level. On the other hand, the pair F-NU had a relatively small difference in U_{HVS} (0.02) and they both consisted of small bright light sources. These reasons warrant further exploration.

Overall, U_{HVS} showed better correlation with perceived uniformity than other metrics. Max:Min ranked second and performed better than Avg:Min, EU, and CV for this limited set of stimuli. Pattern C had a higher Max luminance value than D, though that did not seem to have affected perceived luminance ratings (Figure 5). The improved performance of U_{HVS} might be due to its accounting for contrast sensitivity and spatial frequency of patterns. One advantage of using U_{HVS} is that the range for U_{HVS} is known (between zero and one), compared to Max:Min, which has a much wider range. Technically, Max:Min does not address how abrupt or smooth the transition/gradient between the point with highest luminance and the point with lowest luminance is. It is important to note that the stimuli used were specifically chosen to examine U_{HVS}—and U_{HVS} and Max:Min are not always as well correlated as they are in this study—hence the current results on the performance of different metrics require further investigation.

It is important to interpret the results of this study considering several limitations. The decision to use a two-alternative forced-choice procedure might have made it easier for participants to judge the uniformity of patterns compared to a rating procedure where each pattern is individually presented on the screen. However, the use of this procedure limited the number of patterns that could be included. Therefore, the patterns examined in this study are unlikely to represent the wide range of luminance distributions seen in luminaire apertures.

Another limitation is related to viewing the patterns on a computer screen compared to viewing the luminaire aperture in an interior space. Luminaire apertures are most likely seen by occupants in perspective view from many different angles, such as while walking around in an office space or conducting different tasks. In addition to viewing angles, the luminance range of the aperture and its physical context might affect perceived uniformity. The impacts of these variables warrant further investigation in future studies.

5. CONCLUSION

In this study, the ability of U_{HVS} to consistently discern differences between patterns was examined. Specifically, we investigated whether patterns that had a small difference in U_{HVS} (± 0.01) would yield similar perceived uniformity ratings (first hypothesis), and whether a larger difference in U_{HVS} (> 0.01) would yield different ratings (second hypothesis). In both cases, the results were mixed, indicating that the U_{HVS} does not match perceived uniformity in all situations. Assumptions underlying U_{HVS} and other factors might have contributed to these results. Nonetheless, U_{HVS} had a statistically significant correlation with perceived uniformity, confirming the third hypothesis. Overall, U_{HVS} might be used for general guidance such as to rank a set of patterns from highest to lowest in uniformity but might not be able to consistently predict similarities or differences in perceived uniformity ratings. Further studies are needed to examine U_{HVS} in laboratory settings under different luminance ranges and patterns. Further refinement of U_{HVS} might make it effective for use by lighting manufacturers or in other situations where greater specificity is needed.

REFERENCES

[1] T. Tashiro, S. Kawanobe, T. Kimura-Minoda, S. Kohko, T. Ishikawa, and M. Ayama, "Discomfort glare for white LED light sources with different spatial arrangements," *Light. Res. Technol.*, 2015, doi: 10.1177/1477153514532122.

[2] Y. Yang, M. Ronnier, S. N. Ma, and X. Y. Liu, "Assessing glare. Part 1: Comparing uniform and non-uniform LED luminaires," *Light. Res. Technol.*, vol. 49, no. 2, pp. 195–210, 2017, doi: 10.1177/1477153515607396.

[3] L. M. Geerdinck, J. R. Van Gheluwe, and M. C. J. M. Vissenberg, "Discomfort glare perception of non-uniform light sources in an office setting," *J. Environ. Psychol.*, 2014, doi: 10.1016/j.jenvp.2014.04.002.

[4] Y. Hu, M. R. Luo, and Y. Yang, "A study on lighting uniformity for LED smart lighting system," *2015 12th China Int. Forum Solid State Light. SSLCHINA 2015*, pp. 127–130, 2015, doi: 10.1109/SSLCHINA.2015.7360705.

[5] K. W. Houser, M. Wei, and M. P. Royer, "Illuminance uniformity of outdoor sports lighting," *LEUKOS - J. Illum. Eng. Soc. North Am.*, vol. 7, no. 4, pp. 221–235, 2011, doi: 10.1080/15502724.2011.10732148.

[6] ANSI/IES, “American National Standard Practice for Design and Maintenance of Roadway and Parking Facility Lighting (ANSI/IES RP-8-18),” 2018.
[7] Q. Yao, B. Zhong, Y. Shi, and J. Ju, “Evaluation of Several Different Types of Uniformity Metrics and Their Correlation with Subjective Perceptions,” *LEUKOS - J. Illum. Eng. Soc. North Am.*, 2017, doi: 10.1080/15502724.2016.1172488.
[8] I. Ashdown, “Luminance gradients: Photometric analysis and perceptual reproduction,” *J. Illum. Eng. Soc.*, vol. 25, no. 1, pp. 69–82, 1996, doi: 10.1080/00994480.1996.10748135.
[9] K. Simonson, N. Narendran, P. Boyce, and A. Bierman, “Development of a metric to quantify beam quality of reflectorized lamps,” *J. Illum. Eng. Soc.*, vol. 32, no. 1, pp. 63–72, 2003, doi: 10.1080/00994480.2003.10748405.
[10] I. Moreno, “Illumination uniformity assessment based on human vision,” *Opt. Lett.*, 2010, doi: 10.1364/ol.35.004030.
[11] P. G. J. Barten, “Physical model for the contrast sensitivity of the human eye,” *Hum. Vision, Vis. Process. Digit. Disp. III*, vol. 1666, no. August 1992, p. 57, 1992, doi: 10.1117/12.135956.
[12] F. Faul, E. Erdfelder, A. G. Lang, and A. Buchner, “G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences,” *Behav. Res. Methods*, vol. 39, no. 2, pp. 175–191, 2007, doi: 10.3758/BF03193146.
[13] C. Villa and R. Labayrade, “Validation of an online protocol for assessing the luminous environment,” *Light. Res. Technol.*, 2013, doi: 10.1177/1477153512450452.
[14] I. Sprow, P. Zolliker, Z. Baranczuk, T. Stamm, and P. Zolliker, “Web-based psychometric evaluation of image quality,” in *Image Quality and System Performance VI*, 2009, vol. 7242, pp. 95–106, doi: 10.1117/12.805313.
[15] L. L. Thurstone, “A law of comparative judgment.,” *Psychol. Rev.*, vol. 34, no. 4, pp. 273–286, 1927, doi: 10.1037/h0070288.
[16] A. Maydeu-Olivares, “Thurstone’s Case V Model: A Structural Equations Modeling Perspective,” in *Recent Developments on Structural Equation Models: Theory and Applications*, K. van Montfort, J. Oud, and A. Satorra, Eds. Dordrecht: Springer Netherlands, 2004, pp. 41–67.
[17] W. Revelle, “An Introduction to Psychometric Theory with Applications in R,” *Personal. Proj.*, 2012.