Towards a Novel Prediction Model for Visual Interest in Daylit Renderings

Siobhan Rockcastle, María Lovísa Ámundadóttir, and Marilyne Andersen

Interdisciplinary Laboratory of Performance-Integrated Design (LIPID)
School of Architecture, Civil and Environmental Engineering (ENAC)
Ecole Polytechnique Fédérale de Lausanne (EPFL)
Lausanne, Switzerland
siobhan.rockcastle@epfl.ch

ABSTRACT

In spaces where daylight is a primary source of illumination, our visual perception of architecture is largely influenced by the ephemeral composition of sunlight and shadow. To evaluate these perceptual effects, the authors will apply quantitative contrast measures to HDR renderings for a series of existing contemporary architectural spaces under variable sunlight conditions. These measures will then be compared to subjective ratings of visual interest, collected through an online survey designed to test the influence of spatial and temporal parameters. The objectives of this study assess the impact of sunlight dynamics on subjective ratings of daylit architectural renderings and compare the relationship between these subjective ratings and existing quantitative metrics. The results show that one modified contrast metric can be used to predict factors of visual interest in daylit renderings. When applied through an annual simulation-based approach, this novel metric reveals human perceptual responses to dynamic daylight conditions.

Author Keywords

Daylight; visual interest; renderings; online survey; contrast

1 INTRODUCTION

The compositional effects of shadow, contrast, and light directionality are essential to the visual performance of architecture, and yet their effects are most often defined as qualitative, and research that seeks to measure the impacts on human perception has been limited. To complicate this issue, variable sunlight and climate-driven sky conditions produce diverse compositions of light and shadow. While electric light can be fine-tuned to achieve a specific visual appearance, the ephemerality of natural lighting conditions can produce un-anticipated and even surprising visual affects over time.

Over the last several decades, daylighting research has gravitated toward the development of task-based illumination metrics to assess general illumination thresholds [1]. Visual comfort metrics, specifically those pertaining to glare, have also gained momentum as daylight integration as an energy efficient alternative to electric light has led to an increase in glazing and shading systems that can trigger occupant discomfort in workplaces [2]. Performance indicators for the visual appearance of daylight in architecture, such as those presented in this paper, have only gained momentum in recent years due to concerns that existing illumination-based metrics are not evaluating light perceived from an occupant's field-of-view [3].

In some ways, the idea of evaluating perceptual lighting quality through quantitative measures is somewhat superfluous. Why would we need to quantify the performance of something that we can readily evaluate using qualitative judgment? Although people can observe and assess the visual effects of daylight in a single moment of time, they cannot intuitively comprehend or predict the range of effect that might be experienced over time. As daylight is a highly dynamic source, the complexity of predicting performance necessitates a method that can evaluate a space over time and across diverse sun positions to communicate the variable impacts of light and shadow. Simulation is a powerful tool for evaluating performance dynamics as we can assess a range of temporally-induced effects. Existing tools assess illumination and glare risk, yet there are no dynamic simulation-based methods for evaluating the positive perceptual aspects of daylight composition or its impact on architectural design.

As discussed in Section 2.1, while there are studies linking global contrast measures to perceived impressions of visual interest, more sophisticated local contrast measures exist in vision science and psychology but have not been used to evaluate the perceptual performance of daylit architecture. If we use image processing to quantify contrast-based visual effects within a single rendering and successfully link these values to impressions of spatial composition and visual interest, then we can apply that measure to a series of hourly and daily instances and predict these effects over time. This would help designers to understand where (within a defined view) and when (across hourly and daily moments) the effects of contrast, light, and shadow are likely to produce specific perceptual responses.

In this paper, the authors will apply existing contrast metrics from vision science and psychology to high dynamic range (HDR) renderings for a series of nine contemporary

SimAUD 2016 May 16-18 London, UK

architectural spaces under 3 different sunny sky conditions that vary in daylight composition. These measures will then be compared to subjective ratings for contrast, uniformity, variation, direction, complexity, excitement, and stimulation that have been gathered through an online survey.

The group of local contrast measures selected for this online survey were identified after an initial proof-of-concept experiment conducted with a small subject sample size revealed a stronger correlation between local contrast measures and ratings of contrast, excitement, and stimulation [20] compared to the global measures also tested. This paper builds upon those findings with a larger subject pool and expanded group of local contrast measures to extract a new model for predicting factors of visual interest.

2 BACKGROUND

Those studies that have assessed the perceptual impacts of contrast on daylit space have relied primarily on subjective surveys to explore the relationship between simple photometric measurements and perceived impressions of interior space [4-7]. Existing research has identified two factors that impact subject impressions of daylit space: average luminance and luminance variation [8]. While average luminance has been associated with impressions of brightness, luminance variation has been linked to visual interest [9]. Studies into subject preference have found that mean luminance and luminance variation (distribution and strength of variation) within an office environment contribute to occupant impressions of preference [6-7, 10-13].

2.1 Existing Contrast Measures

Studies that rely on simple photometric measures such as average luminance and luminance variation do not address the *spatial* diversity of luminance values within an occupants' field-of-view. The definition of luminance variation or contrast in these studies is most commonly defined by a global measure, such as Michelson or Root Mean Square (RMS) contrast. Where Michelson computes a ratio from two single points of extreme brightness [14], RMS measures the root mean square of pixel intensities [15] (Appendix A.1). These global contrast measures provide a single value, that existing studies in daylight perception have utilized due to the ease of comparing this value to subjective rankings [5]. Global measures cannot, however predict perceived contrast between two images that vary in the distribution of luminance values [16].

To overcome this limitation, more sophisticated contrast measures have been developed in the fields of image analysis and vision research. The current state of the art in these fields would define two types of measures that are commonly used to quantify contrast: those that rely on global measures (such as Michelson and RMS) and those that rely on local measures [16]. Local contrast measures overcome the limitations associated with global measures by quantifying the effect of *composition* on contrasting areas of brightness and darkness. The authors have focused on neighborhood metrics for their ability to quantify the local contrast values between pixels within a neighborhood or sub-region within an image and assign a singular measure that represents the strength of local variation across all pixels. This led them to define Spatial Contrast (SC) measures as the sum of local pixel variations across a single image resolution [17]:

$$\mathrm{SC} = \frac{1}{WH}\sum_{i=1}^{W} \sum_{j=1}^{H} \overline{\Delta p}_{i,j} \tag{1}$$

where $\overline{\Delta p}_{i,j}$ is the average difference between the four pixels orthogonally surrounding the central pixel $p_{i,j}$ or

$$\overline{\Delta p}_{i,j} = \frac{1}{4}\left(\left|p_{i,j} - p_{i+1,j}\right| + \left|p_{i,j} - p_{i-1,j}\right| + \cdots \right.$$
$$\left.\left|p_{i,j} - p_{i,j+1}\right| + \left|p_{i,j} - p_{i,j-1}\right|\right). \tag{2}$$

RAMMG, a contrast algorithm developed by Rizzi et al, [17] applies a multi-level approach to compute mean local pixel variations across a subsampled pyramid structure, taking into account perceived differences in brightness across multiple image resolutions:

$$\mathrm{RAMMG} = \frac{1}{N}\sum_{l=1}^{N} \bar{c}_l, \tag{3}$$

where N is the number of levels (image resolutions) and $\bar{c}_l$ is the mean contrast in the level l. The image resolution is halved in each subsequent level, where $W_l = W_{l-1}/2$ and $H_l = H_{l-1}/2$ are the width and height of the image at level l and $c_{i,j}$ is the contrast of each pixel, calculated as:

$$c_{i,j} = \sum_{k \epsilon K_8} \alpha\left|p_{i,j} - p_k\right|, \tag{4}$$

where pixels p_k are the 8 neighbouring pixels of $p_{i,j}$ and the weight α applied to each of the 8 surrounding pixels k is:

$$\alpha = \frac{1}{4+2\sqrt{2}}\begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix}. \tag{5}$$

Multi-level metrics like RAMMG were developed to assess both small and large pixel. Where large image resolutions (>100,000 pixels) provide the detail to compute small, localized contrast valued between pixel neighbors, small image resolutions (<25,000) provide the opportunity to measure the difference between larger areas of brightness (i.e. larger neighborhoods).

The Difference of Gaussian (DOG) measure, developed by Tadmor & Tolhurst [18], measures local differences between two bi-dimensional Gaussian components with a center radius and a surround radius. In 2009, Simone et al. combined the multilevel approach developed for RAMMG and the DOG measure to create a multi-level measure called Retinal-like Subsampling Contrast (RSC) [19]. These metrics are described in more detail in Appendix A.2.

2.2 Existing Experimental Studies

Existing research into qualitative lighting performance has seen studies which apply subjective rating methods to HDR photographs [7, 21-22] or rendered images, usually of a simulated office environment [6]. These experiments have asked participants to rate images for pleasantness, contrast, brightness, spaciousness, and/or distribution which [4] are then compared to photometric measurements taken from the digital images.

When using renderings to collect qualitative impressions of daylight related to brightness and contrast, it is essential that tone-mapping algorithms are used to provide the broadest possible luminance range. In controlled laboratory experiments, tone-mapped HDR images have been displayed to subjects using 2D or 3D projection, HDR displays, and conventional low dynamic range (LDR) displays. While there are now backlit HDR screens which can display luminance values up to 4,000 cd/m^2 [23], a study by Cauwerts in 2013 found that conventional LDR displays of 200 cd/m^2 (with images tone mapped to 256 distinct luminance levels) could be used as a surrogate for real world spaces to conduct subjective assessments involving contrast and brightness [22]. In 2012, Villa and Labayrade developed a protocol for lighting quality research using digital images distributed through online survey methods. Their study found that 40 subjects were sufficient to measure significant effects despite systematic error due to uncontrolled conditions (variations in display, background, ambient illumination) [24].

In this paper, the authors use an online survey with tone-mapped images, accepting the limitations of conventional displays in order to reach a broader range of test subjects using the method introduced in Section 3.

3 METHODS

The experimental objectives presented in this paper are two-fold: 1) To measure the impact of sky conditions and architectural composition on subjective ratings of contrast-related characteristics in rendered images, and 2) to compare the relationship between these subjective ratings and existing quantitative contrast measurements. The first objective is to test whether subjects agree on ratings of contrast-based visual effects in architectural spaces and whether these ratings are sensitive to sunlight dynamics (sky types). The second objective is to compare existing contrast measurements and subjective ratings in search of a quantitative model for predicting perceptual responses to daylight.

3.1 Architectural Spaces

For this experiment, the authors modeled nine contemporary architectural spaces that display a range of contrast-based visual effects. On the high contrast side of the spectrum, the authors selected the Arab World Institute by Jean Nouvel (*arab*), the Zolleverein School by SANAA (*zoll*), and the Serpentine Pavilion by Toyo Ito (*serp*). The middle of the spectrum contains the Neughebauer House by Richard Meier (*neug*), the Toldeo Glass Museum by SANAA (*toledo*), and the First Unitarian Church by Louis Kahn (*first*). Finally, the low contrast holds the Poli House by Pezo Von Ellrichshausen (*poli*), the Thermal Baths at Vals by Peter Zumthor (*vals*), and the Menil Gallery by Renzo Piano (*menil*) (Figure 1).

Each of the selected spaces was modelled in Rhinoceros version 5 sr6 and exported to Radiance using the Diva 3.0 toolbar to produce HDR daylight renderings. The authors did not model temporary artifacts (furniture, people) in order to limit visual obstructions and minimize biases toward space use. The PCOND mapping algorithm [25] was used to compress HDR images down to conventional computer screens (0.5 to 200 cd/m^2) as all images in this experiment are displayed on personal tablet, laptop, and desktop screens. The authors acknowledge the limitations associated with a compressed range of values and will use screen technologies with an expanded luminance range in a forthcoming laboratory-based experiment.

3.2 Experimental Design

The experimental design selected for this online study is a repetitive 3 x 3 Semi-Latin-Square which allows for the comparison of three factors – space, subject group, and sky - while limiting experimental fatigue by showing each subject 9 images, rather than the 27 which are required by a full factorial design. The Semi-Latin-Square allows for repetition (in the case of multiple subjects within a given group) and nesting (with three architectural examples per sub category of high, medium, and low contrast – nine spaces in total). Each subject within a group is shown a single rendering for of each of the 9 spaces, under one of the 3 sun positions (Figure 1). This methodology was tested in a proof-of-concept experiment using a small subject sample size and limited range of contrast measures to verify the approach [20]. This paper expands that subject pool and range of metrics through an online survey.

To select the dates and times for each rendering within the study, the authors divided half the year (from the winter to summer solstice) into 28 moments which represent symmetrical daily and monthly instances. Each of the nine architectural spaces was then rendered for each of the 28 moments and analyzed in MATLAB R2012b using the RAMMG contrast metric (eq. 3) [17], which was selected to represent the broader group of neighborhood metrics introduced in Section 2.1. From the assessment of RAMMG contrast across these 28 renderings, three images were then selected: the highest, lowest, and mean contrast composition for each space. Based on the mean RAMMG contrast for each architectural space, the 9 spaces were then ordered and divided into three architectural sub-groups: high, medium, and low.

Table 1 shows the contrast measures applied to the 27 renderings selected for this study: both global (Michelson and RMS) and local contrast metrics (SC, RAMMG, DOG and RSC). As DOG measurements are dependent on the center and surround radii of Gaussian components, the authors applied a selection of radii (r_c = 1-4 to r_s = 2-8) based on past experiments [18,26]. Local measurements such as

RAMMG and RSC are dependent on multiple levels within the image, therefore the authors looked at each resolution level independently. In this study, the original images were 1488 x 1024 pixels and as each subsequent level is halved, we looked at 9 independent image levels for RAMMG (RAMM1, RAMM2,…,RAMM9), and 5-6 levels for RSC, depending on the r_c and r_s.

Table 1 List of contrast measures considered in study.

Global Measures			
•	Michelson	*Michelson, 1927*	*A.1*
•	RMS	*Pavel et. Al, 1987*	*A.1*
Local Measures			
•	SC	*Rockcastle & Andersen, 2014*	*Eq.1*
•	RAMMG	*Rizzi et al, 2004*	*Eq.3*
•	DOG	*Tadmor & Tolhurst, 2000*	*A.2*
•	RSC	*Simone et al., 2009*	*A.2*

3.3 Experimental Procedure

The online survey designed for this experiment was created using Survey Gizmo (http://www.surveygizmo.com/) with a branch logic which allowed for random group assignment upon subject initiation of the survey link. The survey was distributed using multiple diffusion methods: email, Facebook, LinkedIn, and Twitter over a duration of 10 days. Each subject group was asked to respond to some basic demographic questions regarding geographic location and profession and then shown the nine architectural spaces at random, under one of the three possible sky conditions. For example, group 1 (Figure 1) was shown three high contrast spaces under sky 1, three medium contrast spaces under sky 3, and the three low contrast spaces under sky 2. While smartphones were forbidden, we allowed tablet, laptop, and desktop computers. Subjects were asked to turn the brightness on their device to maximum, to ensure the maximum possible pixel range was observed.

For each image, subjects were asked to rate the daylight composition using the following seven point semantic differential scales: low contrast – high contrast, uniform – non-uniform, unvaried – varied, diffuse – direct, simple – complex, calming - exciting, sedating – stimulating (Figure 1). Flynn introduced the use of semantic differential scales to gather subjective assessments of daylight quality in terms of visual clarity, spaciousness, evaluation, relaxation, social prominence, complexity, modifying influence, and spatial modifiers [4]. For the proposed study, the authors have focused on scales associated with complexity and spatial modifiers as well as visual interest.

3.4 Data Management

In total, there were 334 subjects who initiated the survey with 200 complete responses and 134 partially completed, which were discarded. Interestingly, we did see a significant effect on responses from those subjects using tablets. These subjects (4.5%) were discarded as this effect could be due to the smaller screen size (which forced subjects to manually zoom in to view each image) or the default button format which was automatically adjusted in Survey Gizmo on the tablet version. There was no significant effect observed between subjects using a laptop or a desktop computer. Of the remaining 175 subjects, 96% selected their English language capacity as professional, bilingual, or native, with the remaining 4% responding with elementary or limited working proficiency. These subjects were also discarded.

From the remaining 168 subjects, 64% were composed of designers (architecture, landscape, urban, or interior), 36 % non-designers, with 55% reporting their expertise in lighting design as competent, proficient, or expert, and the remaining 45% claiming novice or beginner expertise. There was no significant effect observed between subjects with a design background or expertise in lighting. One subject was excluded from the analysis because 73% of responses were neutral. We normalized the responses (from 1 to 7) for five other subjects, as they did not use either extreme on the rating scale. The remaining 167 subjects were evenly distributed among the three groups (G1: 55, G2: 56, G3: 56).

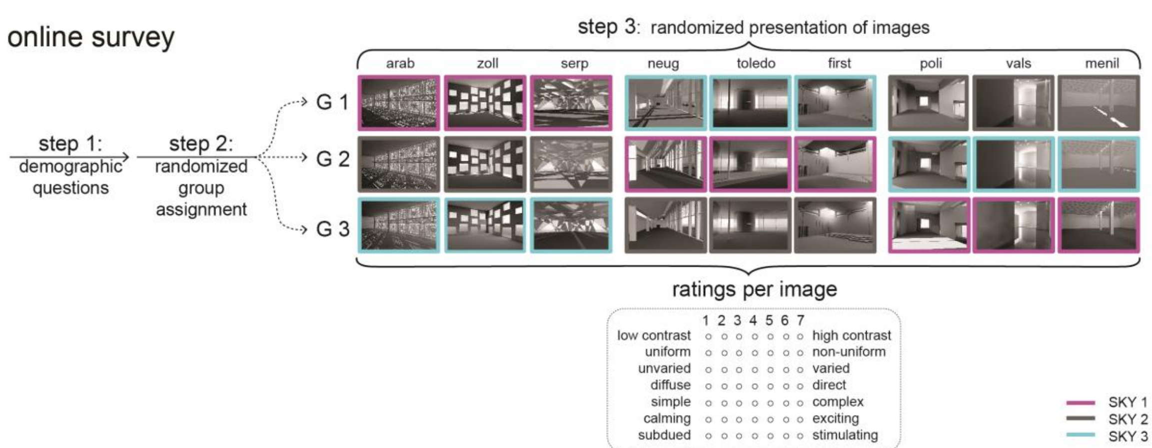

Figure 1 Subjects are first introduced to basic demographic questions, after which they are randomly sorted into one of three groups (G1, G2, or G3) and asked to rate the selected images as they are presented in fully randomized order.

3.5 Data Analysis

To test the significance of experimental factors on the data from each rating pair collected in the experiment, a 3-way ANOVA was used to test the effects of sky, space, and subject group. As the residuals for each rating pair was not normally distributed, a post-hoc analysis was conducted using Kruskal-Wallis to determine the significance of each group within the factor under consideration. To analyze the relationship between subject ratings and existing contrast measures, the authors calculated the Spearman's rank correlation coefficient. Using Spearman's correlation, the authors then selected those combinations of rating-pair and contrast measurement with $\rho_s \geq 0.70$ ($p < 0.0001$). A cumulative logistic model was then applied to fit the subject ratings to selected contrast measures, as the subjective ratings are ordinal response scales.

4 RESULTS

4.1 Distribution of Subject Responses

Figure 2 shows stacked bar plots with the distribution of subject responses for each level of the seven-point rating scale for a selection of 3 spaces (*arab*, *neug*, and *menil*). Subject responses are clustered into gradients by color, with responses that fall on the left side of the scale (1-3) in cyan and responses that fall on the right (5-7) shown in magenta. White is used for neutral ratings (4) and the dotted line shows where the median responses fall for each rating pair. The most frequent responses (summed by color) are shown as a percentage of the total number of responses. There is a substantial effect of sky type in some, but not all spaces – specifically those that see the most obvious variation in daylight composition due to sunlight penetration. The space with strongest subject consensus toward the cyan end of the rating scale (low contrast, uniform, unvaried, diffuse, simple, calming, subdued) was *menil*, while the magenta side of the rating scale (high contrast, non-uniform, varied, direct, complex, exciting, stimulating) was dominated by *arab*. While all rating scales were found to be significantly correlated, subject responses for ratings of excitement and stimulation were the most highly correlated ($\rho_s = 0.75$, Spearman's correlation).

4.2 Effects of Experiment

The significance of experimental factors was evaluated using a 3-way Anova to test the effects of subject group, space, and sky type on each rating scale. While the ANOVA revealed a significant effect of both space and sky factors for all rating scales (p<0.01), the residuals were not normally distributed. A post-hoc analysis was conducted using Kruskal-Wallis, a non-parametric test, to study pair-wise comparisons between each group between the factors under consideration within each rating. This test was run for both sky type and space group on each of the semantic scales. This test revealed the effect of sky was significant on subject responses to all rating scales (p<0.01), except unvaried-varied.

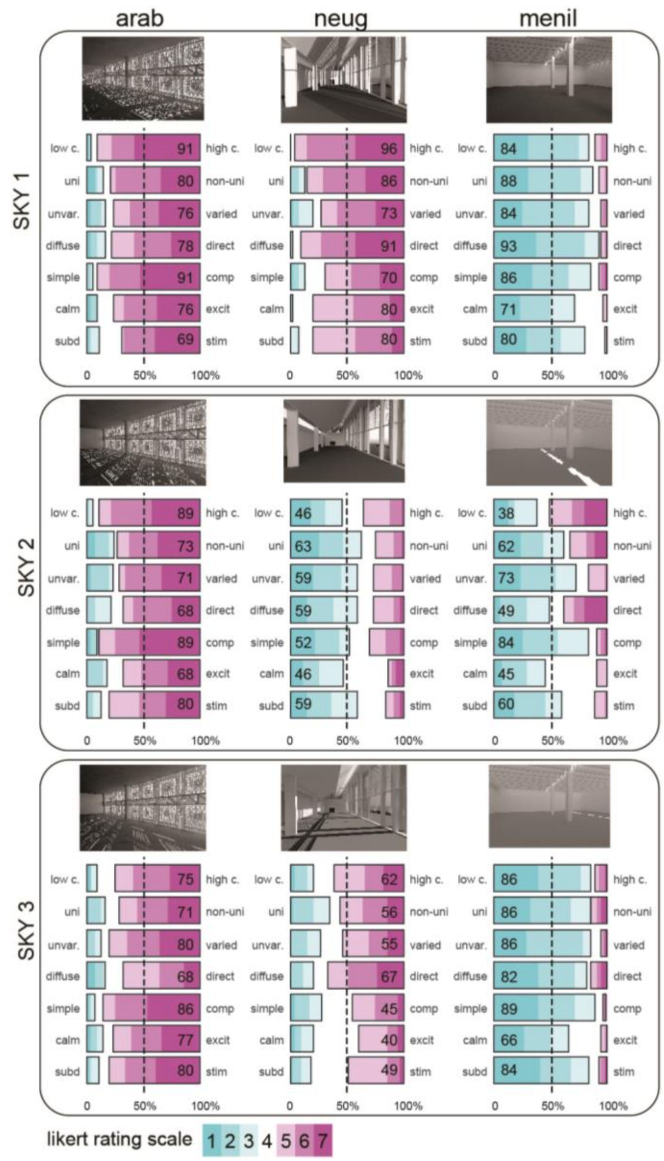

Figure 2 Shows subject ratings for *arab*, *neug*, and *menil* under all three sky types. Ratings are clustered into cyan (ratings 1,2,3) and magenta (ratings 5,6,7).

A pair-wise comparison between sky 1 and sky 3 showed a significant effect (p<0.01) on ratings of contrast, uniformity, direct, complexity, excitement, and stimulation. Ratings of excitement and stimulation also showed a significant effect (p<0.01) between sky 1 and sky 2, which suggest that these ratings were more sensitive to the range of sky types presented in this experiment.

To test the effect of space, we grouped the examples into high, medium, and low based on the percentage of subject responses for all 7 rating pairs magenta cluster 5-7. In this test, there was a significant effect of space between all groups in the factor (p<0.001) for all rating pairs.

4.3 Subject Ratings vs. Quantitative Measures

To relate median subject responses for each rating pair as a function of the contrast metrics introduced in Section 2.1, a Spearman's correlation analysis was conducted. Although a range of center and surround radii were considered for the metrics that rely on Gaussian components (DOG and RSC), only the results for $r_c = 1$ to $r_s = 2$ are listed here. No radii combinations tested in this study were found to have particularly significant correlation to subject responses.

Table 2 shows that RAMM5 (the 5th resolution level in RAMMG - 64 x 93 pixels) achieved the strongest statistical dependence to median ratings of diffuse-direct ($\rho_s = 0.77$), calming-exciting ($\rho_s = 0.78$), and subdued-stimulating ($\rho_s = 0.77$), while RAMMG had the strongest dependence with ratings for low contrast – high contrast ($\rho_s = 0.74$). Using Spearman's correlation to pre-select contrast metrics as possible predictors of visual interest, we selected RAMM5, hereafter referred to as 'Modified Spatial Contrast.'

The authors then applied an ordered logit model to fit the Modified Spatial Contrast (RAMM5) to subjective ratings for diffuse - direct, calming - exciting, subdued - stimulating using ordered logistic regression. The deviance of these fits was 8.78, 9.36, and 9.21, respectively. Figure 3 shows the application to a proportional odds model to predict subject ratings of calming – exciting.

When we group ratings, such as we did in the cyan and magenta gradient plots in Figure 2, we can say that a Modified Spatial Contrast of 13 (or more) triggers responses of excitement (ratings of 5, 6, or 7) for 63% of subjects, whereas a Modified Spatial Contrast of 5 (or less) produces responses of calming (ratings of 1, 2, or 3) in 59% of subjects. This probabilistic model provides the first *ever* objective predictor for visual interest in daylit architecture.

Contrary to those metrics which address task-plane illuminance, autonomy from electric energy sources, and discomfort glare, Modified Spatial Contrast allows designers to compute the probability of achieving specific perceptual responses to daylight across the day and year.

Table 2: Spearman's correlation coefficients between median subject responses for each rating pair and contrast measure.

	Michelson	RMS	SC	RAMM5	RAMMG	DOG $r_c=1$ $r_s=2$	RSC $r_c=1$ $r_s=2$
contrast	0.10	0.62	0.52	**0.72***	**0.74***	0.30	0.17
uniformity	0.06	0.55	0.44	0.67	0.66	0.32	0.19
variation	0.08	0.42	0.40	0.58	0.55	0.30	0.19
direct	0.17	0.59	0.63	**0.77***	**0.75***	0.50	0.00
complex	0.14	0.53	0.48	0.65	0.62	0.36	0.14
exciting	0.06	**0.70***	**0.70***	**0.78***	**0.74***	0.38	0.26
stimulating	0.16	0.61	0.60	**0.77***	**0.75***	0.31	0.17

*Rating pair and contrast measurements with $\rho \geq 0.70$ ($p < 0.0001$) were considered most significant.

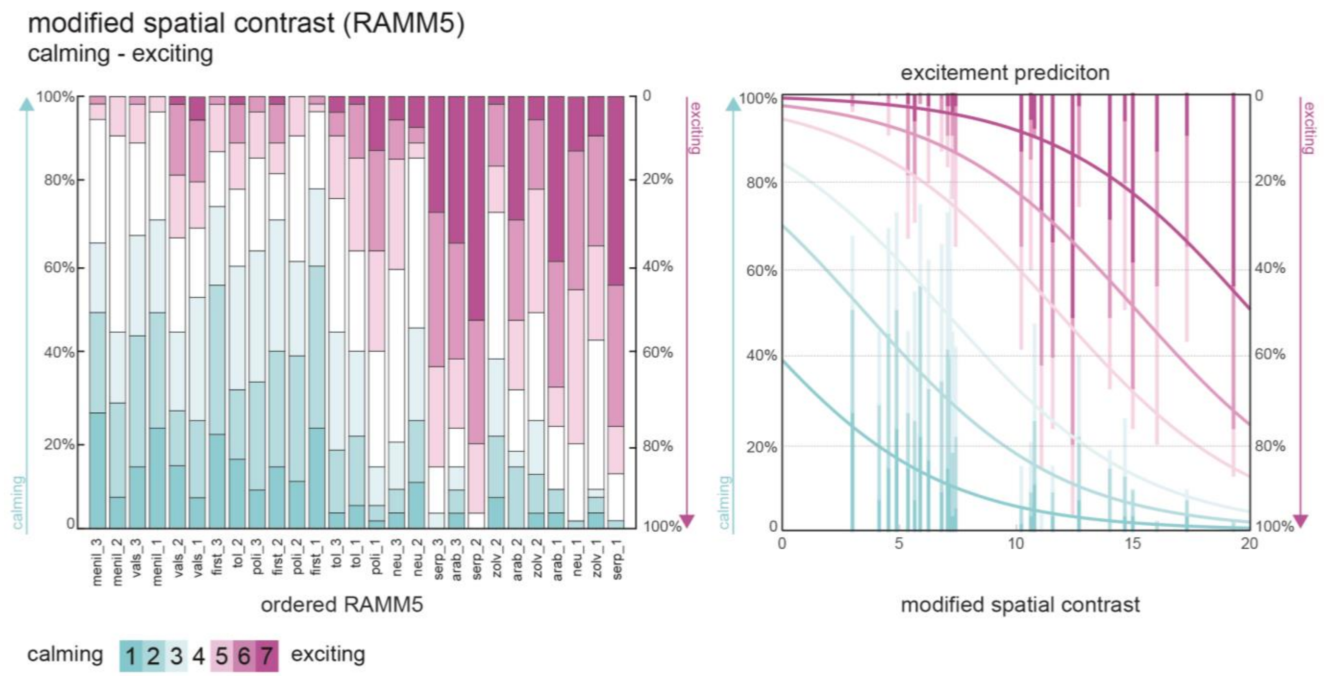

Figure 3 Ordered logistic regression through RAMM5 and ratings of calming – exciting.

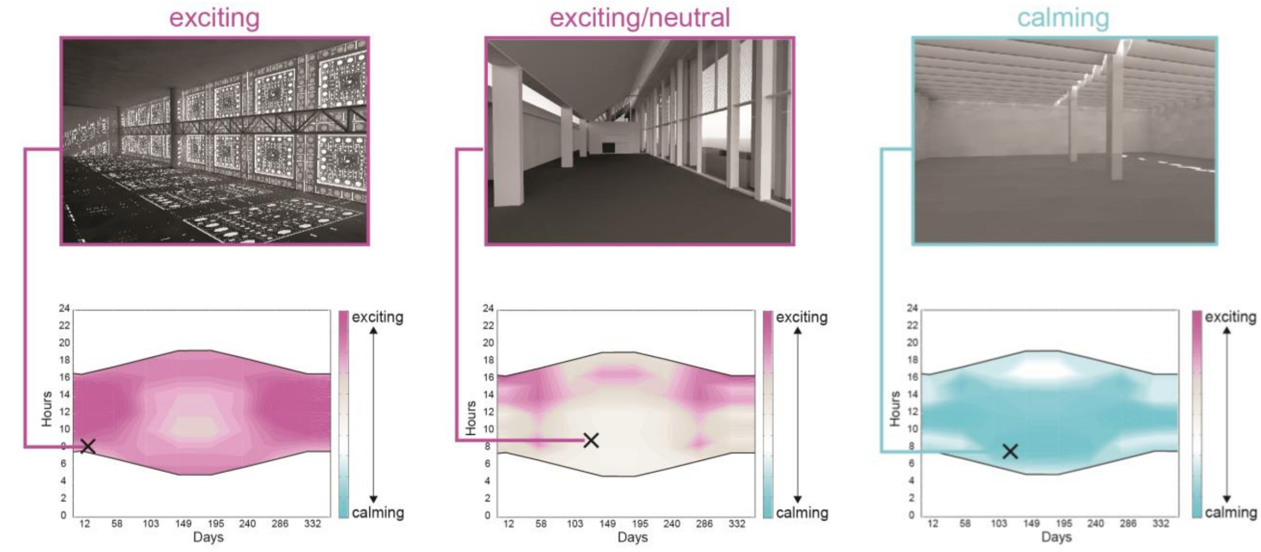

Figure 4 Application of the *Modified Spatial Contrast* measure to renderings of 56 symmetrical annual moments to predict ratings of calming (shown in cyan) or excitement (shown in magenta) in *arab*, *neug*, and *menil* (from left to right).

Figure 4 shows the application of modified spatial contrast (RAMM5) to a selection of three spaces: *arab*, *neug*, and *menil*. This measure was applied to 56 renderings for each space, representing a symmetrical distribution of hourly and daily instances and plotted temporally to show an annual prediction of excitement. Values in magenta show point-in-time predictions of excitement while cyan shows predications of calming.

5 CONCLUSION & OUTLOOK

In conclusion, the experiment presented in this paper resulted in the following findings: 1) both space and sky condition have a significant effect on subject ratings of contrast, direction, complexity, excitement, and stimulation and 2) local neighborhood contrast measures such as RAMMG and specific levels within than metric (RAMM5, i.e. Modified Spatial Contrast) were found to be good predictors of contrast-based visual effects, especially ratings of diffuse – direct, calming – exciting and subdued – stimulating. Using a cumulative logistic model, this paper introduces a novel probabilistic model for predicting subject responses to excitement in simulated daylight renderings using an objective contrast measure.

While a single point-in-time quantitative analysis may be less useful to designers who can evaluate this performance qualitatively, modified spatial contrast is useful in its ability to predict *dynamic effects* which may be unanticipated. By predicting how visually engaging a space may be (and how this changes over time), this research offers a new dimension in daylight performance assessment. Rather than be satisfied with the knowledge that a space achieves *enough* or *too much* daylight, this model evaluates human arousal to daylight composition.

To further validate this approach, the authors will conduct a series of upcoming experiments with an expanded set of architectural spaces and view parameters. To limit potential error due to screen size, brightness, and tone-mapping, these forthcoming experiments will be conducted under controlled laboratory conditions using screen technologies with an extended view and luminance range. Future experimental parameters will include the assessment of daylight composition using immersive viewing techniques achieved through a virtual reality headset. This virtual method is currently being tested as a surrogate for extracting qualitative lighting assessments in live space and initial findings suggest a positive result. While a single view is convenient for the application of digital image measurements, architecture is rarely composed of a single space or view position and requires more immersive evaluation techniques.

ACKNOWLEDGMENTS

Funding for this research is provided by the EPFL and a grant awarded by the Velux Stiftung Foundation (project 936).

APPENDIX A

A.1 Global Measures

$$\text{Michelson} = \frac{P_{max} - P_{min}}{P_{max} + P_{min}},$$

where P_{max} and P_{min} represent the highest and lowest pixel intensity.

$$\text{RMS} = \sqrt{\frac{1}{WH}\sum_{i=1}^{W}\sum_{j=1}^{H}(p_{i,j} - \bar{p})^2}$$

where $p_{i,j}$ are the pixels intensities at position (i, j) in an image of size W by H and $\bar{p}$ is the average pixel intensity.

A.2 Local Measures

DOG calculates local differences between two bi-dimensional Gaussian filters with a center component $R_c(x, y)$ and a surround component $R_s(x, y)$:

$$DOG(x,y) = \frac{R_c(x,y) - R_s(x,y)}{R_c(x,y) + R_s(x,y)}$$

Center and surround components $R_c(x,y)$ and $R_s(x,y)$ can be found in [18]. The authors have chosen to compute the average $DOG(x,y)$ across all pixels in a given image with width W and height H:

$$\overline{\mathrm{DOG}} = \sum_{i=1}^{W} \sum_{j=1}^{H} \mathrm{DOG}(x_i, y_j)$$

RSC combines the pyramid subsampling method used in RAMMG (eq.3) with the DOG measure:

$$\mathrm{RSC} = \frac{1}{N} * \sum_{l=1}^{N} \overline{DOG_l},$$

where N is the number of levels and $\overline{DOG_l}$ is the mean contrast in level l [19].

REFERENCES

1. C. Reinhart and J. R. Z. Mardaljevic, "Dynamic Daylight Performance Metrics for Sustainable Building Design," *Leukos Vol 3 (1),* pp. 1-25, 2006.
2. K. Konis, E. Lee and R. Clear, "Visual Comfort Analysis of Innovative Interior and Exterior Shading Systems for Commercial Buildings Using High Resolution Luminance Images," *LEUKOS, The Journal of the Illuminating Engineering Society of North America*, vol. 7, no. 3, 2011.
3. C. Cuttle, Lighting Design: a percetion-based approach, New York: Routledge, 2015.
4. J. Flynn, C. Hendrick, T. Spencer and O. Martyniuk, "A guide to methodology procedures for measuring subjective impressions in lighting," *Journal of IES,* pp. 95-110, 1979.
5. K. Wymelenberg and M. Inanici, "The Effect of Luminance Distribution Patterns on Occupant Preference in a Daylit Office Environment," *LEUKOS: The Journal of the Illuminating Engineering Society of North America,* vol. 7, no. 2, pp. 103 - 122, 2010.
6. G. Newsham, C. Richardson, C. Blanchet and J. Veitch, "Lighting quality research using rendered images of offices," *Lighting Research and Technology,* pp. 93-115, 2005.
7. G. Newsham, D. Cetegen, J. Veitch and L. Whitehead, "Comparing Lighting Quality Evaluations of Real Scenes with those from High Dynamic Range and Conventional Images," *Transactions on Applied Perception,* 2010.
8. J. Veitch and G. Newsham, "Preferred Luminous Conditions in Open Plan Offices: Research amd Practice Recommendations," *Lighting Research and Technology Vol 32,* pp. 199-212, 2000.
9. D. Loe, K. Mansfield and E. Rowlands, "Appearance of Lit Environment and its Relevance in Lighting Design: Experimental Study," *Lighting Research and Technology Vol 26,* pp. 119-133, 1994.
10. D. Cetegen, J. Veitch and G. Newsham, "View Size and Office Illuminance Effects on Employee Satisfaction," in *Proceedings of Balkan Light,* Ljubljana, Slovenia, 2008.
11. M. Boubekri, R. Hull and L. Boyce, "Impact of Window Size and Sunlight Penetration on Office Worker's Mood & Satisfaction: A Novel Way of Assessing Sunlight," *Environment & Behavior,* vol. 23, no. 4, pp. 474-493, 1991.
12. D. Tiller and J. Veitch, "Perceived Room Brightness: Pilot Study on the Effect of Luminance Distribution," *Lighting Research and Technology Vol 27 (2),* pp. 93-101, 1995.
13. K. Parpairi, N. Baker, K. Steemers and R. Compagnon, "The Luminance Differences index: a new indicator of user preferences in daylit spaces," *Lighting Research and Technology Vol 34 (1),* pp. 53-68, 2002.
14. A. Michelson, Studies in Optics, Chicago: University of Chicago Press, 1927.
15. M. Pavel, G. Sperling, T. Riedl and A. Vanderbeek, "Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language," *Opt. Soc. Am. A,* vol. 4, no. 12, pp. 2355 - 2365, 1987.
16. G. Simone, M. Pedersen and J. Y. Hardeberg, "Measuring perceptual contrast in digital images," *Journal of Visual Communication and Image Representation,* vol. 23, no. 3, pp. 491-506, 2012.
17. A. Rizzi, G. Algeri, D. Medeghini and A. Marini, "A proposal for contrast measure in digital images," in *Second European Conference on Color in Graphics, Imaging and Vision*, Aachen, 2004.
18. Y. Tadmor and D. Tolhurst, "Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes," *Vision Research 40,* pp. 3145-3157, 2000.
19. G. Simone, M. Pedersen, J. Y. Hardeberg and A. Rizzi, "Measuing Perceptual Contrast in a Multi-Level Framework," San Jose, 2009.
20. S. Rockcastle and M. Andersen, "Human Perceptions of Daylight Composition in Architecture: A Preliminary Study to Compare Quantitative Contrast Measures with Subjective User Assessments in HDR Renderings," in *14th International Conference of the International Building Performance Simulation Association*, Hyderabad, 2015.
21. G. Newsham, R. Marchand and J. Veitch, "Preferred Surface Luminance in Offices: A Piolet Study," Salt Lake City, 2002.
22. C. Cauwerts, M. Bodart and R. Labayrade, "Assessing Lighting Appearance Using Pictures: Influence of Tone-Mapping Parameters and Lighting Conditions in the Visualization Room," in *CISBAT*, Lausanne, 2013.
23. L. Whitehead, G. Ward, W. Stuerzlinger and H. Seetzen, "High Dynamic Range Display Devices". US Patent 6,891,672, 2005.
24. C. Villa and R. Labayrade, "Validation of an Online Protocol for Assessing the Luminous Environment," *Lighting Research & Technology,* vol. 2012, no. 0, pp. 1-20, 2012.
25. G. Ward, H. Rushmeier and C. Piatko, "A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes," Lawrence Berkeley National Lab, University of California, San Francisco , 1997.
26. A. Rizzi, G. Simone and R. Cordone, "A Modified Algorithm for Perceived Contrast Measure in Digital Images," Barcelona, 2008.
27. S. Rockcastle and M. Andersen, "Measuring the dynamics of contrast & daylight variability in architecture: A proof-of-concept methodology," *Building and Environment,* pp. v 81, 320-333, 2014.