HUMAN PERCEPTIONS OF DAYLIGHT COMPOSITION IN ARCHITECTURE: A PRELIMINARY STUDY TO COMPARE QUANTITATIVE CONTRAST MEASURES WITH SUBJECTIVE USER ASSESSMENTS IN HDR RENDERINGS

Siobhan Rockcastle[1] & Marilyne Andersen[1]
[1]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

Humans perceive daylit architecture as a rich and dynamic luminous composition and yet existing performance metrics most often evaluate natural light for its ability to provide adequate illumination to a two-dimensional task surface while avoiding glare-based visual discomfort. This rather limited task-driven approach places a disproportionate emphasis on surface illumination and glare-based discomfort and ignores the likelihood that contrast can provide a positive visual impact on our impression of space. Existing studies on perceptual daylight performance have linked subjective ratings of contrast and brightness to simple global measures in digital images, yet the conclusions are varied and lack robust consensus. These 'global' contrast measures do not account for the composition of luminance values within a scene and while more robust methods have been developed in computational graphics, vision research, and psychology, they have not been applied to studies in qualitative lighting research. As daylight-driven visual effects are not only dependent on composition, but are heavily influenced by dynamic sky conditions, this paper will introduce an experimental method for comparing subjective ratings of daylight composition in architecture against existing global and local contrast metrics under a range of annual moments. This preliminary study will test the effects of sun position and spatial composition on subjective ratings for contrast, uniformity, complexity, variation, stimulation, and excitement. It will then identify which quantitative measures (local and/or global contrast metrics) correlate most strongly to these ratings.

INTRODUCTION

In daylit spaces, our visual perception of architecture is largely influenced by the dynamic state of our surrounding environment. Conditions such as climate, sky type, time-of-day, and time-of-year are in constant flux and the resulting lighting conditions can create diverse compositions of light and shadow. In an essay on Architect Steven Holl's work, theoretician Stanford Kwinter writes, "For Holl, architecture is the science of experience…Light is not itself the plenum of matter, but rather what reveals and conveys it (like water in the paper into which pigment is placed" (Holl, 2011). While we understand how light can reveal spatial depth and material texture, the ephemerality of these effects under natural lighting conditions is far less intuitive and can produce un-anticipated and even surprising results. In his seminal book titled, The Eyes of the Skin: Architecture and the Senses, Juhani Pallasmaa states that "In great architectural spaces, there is a constant, deep breathing of shadow and light; shadow inhales and illumination exhales light" (Pallasmaa, 2005). Mary Ann Steane discusses the key differences between daylight and artificial sources in her book titled, *The Architecture of Light*. While artificial light sources can be carefully combined to match a desired composition, they produces a static effect and can never match the "nuance of mood created by the time of day and the wonder of the seasons" (Steane, 2011).

While the visual effects of daylight, such as shadow, depth, contrast, and texture are strongly valued by architects, they are most often defined as qualitative factors and research into measuring their effects on human perception has been limited. Aside from the obvious challenges associated with measuring qualities that are inherently subjective, our current discourse on daylight performance has been dominated by energy-related concerns, brought on by the energy crisis of the 1970s and strengthened by the shift toward sustainable building practices. In an effort to reduce energy consumption, daylighting research has gravitated toward the widespread development of task-based illumination metrics to assess general illumination thresholds while minimizing electric energy use (Reinhart & Mardaljevic, 2006). Visual comfort metrics, especially those pertaining to glare, have also gained predominance within the last two decades, as the emphasis on daylight integration has led to an increase in glazed facades and complex shading systems that can trigger occupant discomfort in places where visual tasks are performed (Konis, et al., 2011). While research in these areas is undeniably important, perceptual performance indicators such as daylight composition, contrast, and dynamics have taken a back seat and only gained momentum in the last decade due to concerns that existing illumination-based metrics are not addressing light perceived from an occupant's field-of-view (Cuttle, 2015). In this paper, the authors will apply existing contrast measures to high dynamic range (HDR) renderings for a series of nine existing architectural spaces which vary in daylight composition. These measures will then be compared to subjective ratings for contrast, uniformity, complexity, variation, excitement, and stimulation. Ratings have been gathered through a preliminary survey with a limited population size as a proof-of-concept for forthcoming online and lab-based experiments.

BACKGROUND

Those existing studies that have assessed the impact of contrast in daylit space have relied primarily on subjective surveys to explore the relationship between simple photometric measurements and perceived impressions of interior space (Flynn, et al., 1979) (Wymelenberg & Inanici, 2009) (Newsham, et al., 2005) (Newsham, et al., 2010). Two factors that are widely accepted to impact our perception of daylit architecture are average luminance and luminance variation (Veitch & Newsham, 2000). The former has been directly associated with perceived lightness and the latter with visual interest (Loe, et al., 1994). Some studies have found that both mean luminance and luminance variation within an office environment contribute to occupant impressions of preference (Cetegen, et al., 2008) (Newsham, et al., 2005) (Newsham, et al., 2010), whereas others have discovered that the distribution of luminance values across an occupant's field-of-view (Boubekri, et al., 1991) (Tiller & Veitch, 1995) as well as the strength of variation are factors of preference (Wymelenberg & Inanici, 2009) (Parpairi, et al., 2002).

Existing Contrast Measures

The problem with studies that rely on simple photometric measures such as average luminance and luminance variation, is that they generally do not address the *spatial* diversity of luminance values within an occupant's field-of-view. In these studies, luminance variation or contrast is most commonly defined by a global measure, such as Michelson or root mean square (RMS) contrast. Where Michelson computes the ratio between two single points of extreme brightness, $Pmax$, and darkness, $Pmin$, (Michelson, 1927)

$$\text{MICHELSON} = \frac{P_{max} - P_{min}}{P_{max} + P_{min}} \quad (1)$$

RMS measures the root mean square of pixel intensities (Pavel, et al., 1987)

$$\text{RMS} = \sqrt{\frac{1}{WH}\sum_{i=0}^{W-1}\sum_{j=0}^{H-1}(P_{i,j} - \bar{P})^2} \quad (2)$$

Where P is the brightness of a given pixel and $\bar{P}$ is the average pixel intensity across an image where W is the image width, and H is the image height.

While these global contrast measures provide a single comprehensible value - which existing studies in daylight perception have utilized due to the ease of comparing this value to subjective rankings (Wymelenberg & Inanici, 2009) - they cannot effectively predict perceived contrast between two images that vary in composition (Simone, et al., 2012). 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 (Simone, et al., 2012). Local contrast measures were developed to overcome the limitations associated with global measures by quantifying the effect of composition on contrasting areas of brightness and darkness. Included within this group of measures are methods that measure spatial frequencies in the Fourier domain (Hess, et al., 1983), those that measure a weighted color contrast based on the distance between chroma regions (Tremeau, 2000), and those that calculate the difference between a single pixel and a surrounding region or neighborhood (Tadmor & Tolhurst, 2000) (Rizzi, et al., 2004) (Matekovic, et al., 2005) (Rockcastle & Andersen, 2014). The authors have focused on the latter group of neighborhood metrics for their ability to quantify the local contrast values between pixels within a neighborhood or sub-region and assign a singular measure which represents the strength of local variation across all pixels. Spatial Contrast (SC), a metric adapted from (Rizzi, et al., 2004) and simplified by the authors for ease of computing (Rockcastle & Andersen, 2014), measures the sum of local pixel variations across a single image resolution.

$$\text{SC} = \frac{\sum_{i=1}^{W}\sum_{j=1}^{H}\overline{\Delta p}_{i,j}}{p_{max}} \times 100 \quad (3)$$

Where $\overline{\Delta p}_{i,j}$ is the local difference between each pixel and its surrounding four pixels:

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

Where p_{max} is a hypothetical maximum value which is computed as a black and white checkerboard of size $(W\ x\ H)$ where every pixel has an average local contrast of 255:

$$p_{max} = 255 \times W \times H \quad (5)$$

RAMMG, a contrast algorithm developed by (Rizzi, et al., 2004), applies a multi-level approach to compute mean local pixel variations across a subsampled pyramid structure, to account for perceived differences in brightness across multiple image resolutions.

$$\text{RAMMG} = \frac{1}{N} * \sum_{k=1}^{N}\bar{c}_k \quad (6)$$

Where N is the number of levels and $\bar{c}_k$ is the mean contrast in the level k *(the image resolution is halved in each subsequent level):*

$$\bar{c}_k = \sum_{i=1}^{W_k}\sum_{j=1}^{H_k} c_{i,j} \quad (7)$$

Where W_k and H_k are the width and height of the image at level k and $c_{i,j}$ is the contrast of each pixel, calculated as:

$$c_{i,j} = \sum_{n \epsilon N_8} \alpha_n |P_{i,j} - P_n| \quad (8)$$

where pixels P_n are 8 neighbouring pixels of $P_{i,j}$ and the weight applied to each surrounding pixel in n is:

$$\alpha_n = \frac{1}{4+2\sqrt{2}} \begin{bmatrix} \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \\ 1 & & 1 \\ \frac{\sqrt{2}}{2} & 1 & \frac{\sqrt{2}}{2} \end{bmatrix} \quad (9)$$

Each of these metrics will be applied to a selection of nine case-study spaces to compare their response to subjective ratings of daylight composition.

Existing Experimental Studies

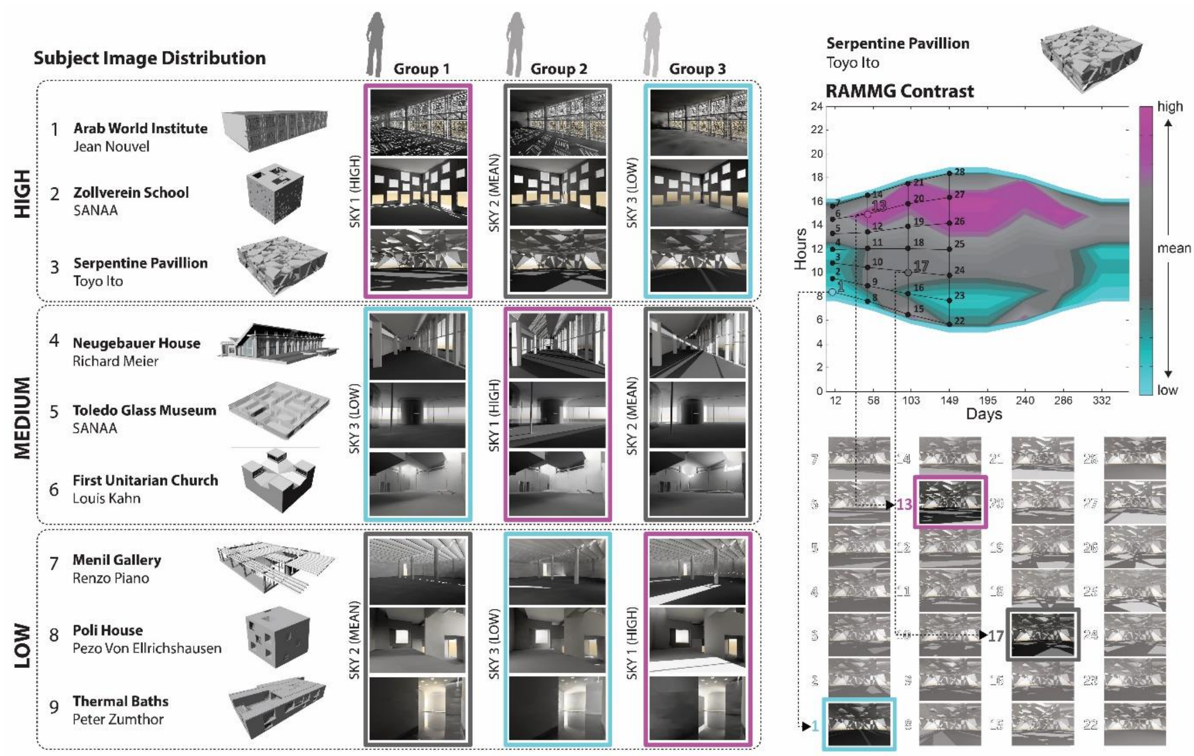
To conduct qualitative lighting research using digital images, existing studies have applied subjective rating methods to measure impressions of lighting composition in HDR photographs (Newsham, et al., 2002) (Newsham, et al., 2010) (Cauwerts & Bodart, 2011) (Cauwerts, 2013) as well as rendered images of a simulated office environment (Newsham, et al., 2005). These experiments have asked participants to view a series of images and then respond to semantic differential ratings (Flynn, et al., 1979) for pleasantness, contrast, brightness, spaciousness, and/or distribution which are then compared to photometric measurements taken from the digital images.

When using images to collect subjective impressions of daylight related to perceptual factors such as brightness and contrast, it is essential that light levels are accurately captured or generated (in the case of renderings) and displayed with a broad a range of luminance levels using proper tone-mapping algorithms calibrated for the specific display. 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 on the market which can display luminance values up to 4,000 cd/m2 (Whitehead, et al., 2005), a study by Cauwerts in 2013 found no significant difference in subjective assessments of contrast between images displayed on HDR displays and those that were tone-mapped to conventional LDR displays. While there were some differences in ratings of pleasantness, distribution, and spaciousness between real world scenes and image displays, it was concluded 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 (Cauwerts, 2013). In 2012, Villa & Labayrade developed an online protocol to limit the impacts of uncontrolled experimental conditions (i.e. screen resolution, brightness, background, etc) on the assessment of digital images for lighting quality research. Their study found that significant effects could be identified with 40 subjects despite systematic error due to uncontrolled conditions and the variance in p-value was constant with ≥100 subjects (0.002%) (Villa & Labayrade, 2012). Where controlled experiments are time-consuming, an online protocol allows for more widely distributed experiments, which the authors will take advantage of in the next phase of planned research.

In summary, there are a number of methods for creating and displaying interior images to assess subjective qualities of daylight, each of which has its own set of advantages and limitations. While real spaces obviously produce the most accurate impression of light for subjective assessment, experimental conditions are limited by the physical sky conditions available on site and it is difficult to compare a range of spatial configurations or temporal conditions in an efficient manner. HDR renderings using Radiance (Ward, 1994) allow for a broad range of spatial and temporal lighting conditions, but are limited by the luminance output of the display device and must therefore apply tone-mapping algorithms to achieve an acceptable range. As the first step towards a more extensive online survey, the authors have conducted this preliminary study with 9 subjects, which will then be distributed to a larger pool of subjects in the next phase of research. While the preliminary survey discussed in this paper used paper medium, it is intended as a proof-of-concept for a more extensive online method. In that phase, images will be shown on conventional computer displays to a broad pool of test subjects. These will be further validated in a controlled laboratory experiment (also forthcoming).

METHOD

The aim of this paper is three-fold: 1) To measure the impact of sun position and space on subjective impressions of daylight composition through a Latin Square experimental design, 2) To compare the semantic differential pairs used in the pilot study to find redundancies, and 3) To investigate the relationship between subjective assessments of contrast-related characteristics in daylight composition and existing quantitative measurements for contrast. To compare existing contrast metrics to subjective ratings, the authors modelled nine contemporary architectural spaces which display a range of contrast-based visual effects and were selected based on a prominent interior view.



a. b.

Figure 1 Shows nine Architectural Spaces selected for the study (a), each of which is rendered across three distinct sun positions and shown to three distinct subject groups. The three sun positions chosen for each space were selected from 28 symmetrical semi-annual instances under sunny sky conditions and represent high, mean, and low levels of RAMMG contrast (b). Thresholds for high and low were relative to each space. Each group therefore sees all nine architectural spaces under one of the three sun positions for each space.

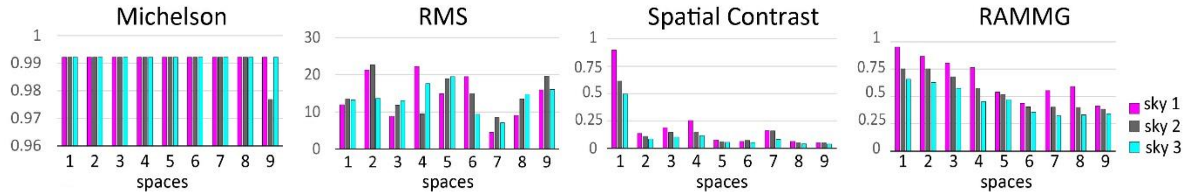


Figure 2 Results for global (Michelson & RMS) and local contrast metrics (Spatial Contrast and RAMMG) across the 9 spaces and 3 sky conditions selected for this study.

In this paper, the authors have chosen a repetitive 3 x 3 Latin-Square design of experiments (Montgomery, 2012), which allows for the comparison of three variables – spaces, subjects, and sky - while limiting experimental fatigue by showing each subject 9 images, rather than the 27 which are required by a full factorial experiment. While the experiment is intended to study the effects of architectural design or 'space' and 'sky type' on a subject's impression of contrast, there is a risk that the architecture may bias subjective assessments of daylight composition when repeated under multiple sky conditions. To overcome this, a Latin-Square design of experiments uses three distinct subject groups. Each subject within a group is shown a single rendering for of each of the nine spaces, with one of the three sun positions in each rendering. With three distinct groups, we can then compare the effect of architectural composition within groups and the effect of sun position between groups. While this proposed experimental method will be applied to a larger population through an online survey, this initial pilot study was composed of 5 female and 4 male researchers in building performance, with varying competencies in architectural design, computational and civil engineering. The authors acknowledge that this is a limited, and potentially biased (with existing knowledge of daylight analysis) sample and that a larger and more diverse sample is necessary to draw any concrete conclusions.

Each of the selected spaces was modelled in Rhinoceros version 5 sr6 based on available building plan and section drawings, and exported to Radiance using the Diva 2.0 toolbar to produce HDR daylight renderings. These were tone-mapped to a range of 0.5 to 255 using the PCOND mapping algorithm (Ward, et al., 1997), which produces a range of pixel values acceptable for printed medium. 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 in each of the 28 moments and analyzed in MATLAB R2012b using the RAMMG contrast metric (Rizzi, et al., 2004), which was selected to represent a larger group of neighborhood metrics not included in this initial study. 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 (Figure 1a). Based on the mean RAMMG contrast for each architectural space, the 9 spaces were then ordered and divided into three sub-groups: high, medium, and low (see Figure 1b).

Figure 2 shows the results for both local (Michelson & RMS) and global contrast metrics (spatial contrast & RAMMG) when applied to the renderings selected by the RAMMG metric. It is apparent through these quantitative results that global contrast measures such as Michelson produce less differentiated predictions of contrast between spaces and sky conditions, despite the highly varied composition of pixels in each image.

The high contrast group includes the Arab World Institute by Jean Nouvel, the Zollverein School by SANAA, and the Serpentine Pavilion by Toyo Ito. The medium contrast group contains the Neugebauer House by Richard Meier, the Toledo Glass Museum by SANAA, and the First Unitarian Church by Louis Kahn, while the low contrast group holds the Menil Gallery by Renzo Piano, the Poli House by Pezo Von Ellrichshausen, and the Thermal Baths at Vals by Peter Zumthor. To test the strength of our latin-square design, each subject group was shown all nine spaces, under a mix of sun positions: sky 1 (high), sky 2 (mean), or sky 3 (low) RAMMG contrast. For example, Group 1 (column one in Figure 1a) 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. If there is an effect of sun position on the perception of space, then there will be a significant effect between groups.

For each space, the subjects were shown an image and asked to rate the daylight composition using the following seven point semantic differential scales: low contrast – high contrast, uniform – non-uniform, unvaried – varied, simple – complex, calming - exciting, sedating - stimulating (Figure 3). 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 (Flynn, et al., 1979). Numerous studies thereafter have employed the use of these scales to conduct daylight quality research in real spaces and simulated or photographed views (Newsham, et al., 2002) (Newsham, et al., 2005) (Vogels, 2008) (Demers, 2007). For the proposed study, the authors have focused on semantic differential scales associated with complexity and spatial modifiers as well as visual interest.

RESULTS

For each of the rating scales, subject responses were processed to measure the effect of sky and space factors. ANOVA analyses in table 1 show that Sky type has a significant effect on ratings of contrast, variation, complexity, excitement, and stimulation while Space (the selected architecture and view) had a significant effect on ratings of complexity, excitement, and stimulation.

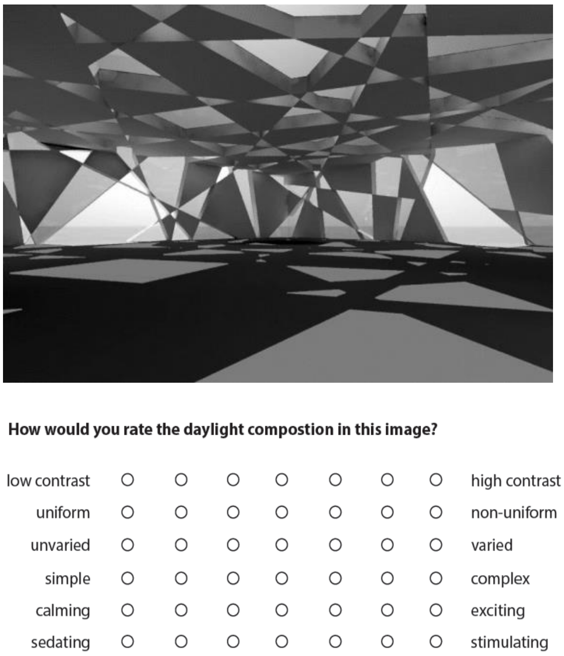


Figure 3 A sample page from the survey, showing the six semantic rating scales selected for this study.

None of the three factors were shown to have a significant effect on ratings of uniformity, and were weaker for variation than the other four rating scales. Based on the ANOVA results, we can see that the null hypothesis is always true for uniformity ratings under these experimental conditions, but can be rejected for some factors in the remaining five ratings, which show significant effects from at least one of the three factors. The effect of subject is not shown to be significant for any of the semantic rating scales, indicating that the experimental design used in this study is working, despite the limited population size.

To assess potential redundancies in the semantic rating scales selected for this study, Pearson's linear correlation coefficients (PCC) was conducted between responses for each rating pair. 'sedating - stimulating' and 'calming exciting' ratings were found to be strongly correlated, with PCC>80%. While 'uniform – non-uniform' and unvaried – varied' had a PCC>50%, no other rating pairs showed significant correlation.

To understand the relationship between semantic ratings and quantitative measurements, a correlation analysis was conducted between mean subject responses per space and sky condition for each of the 9 subjects. In Figure 4, each of the four quantitative metrics were fit against each of the six semantic differential ratings to find which, if any, can serve as a prediction model for subjective assessments. Data points represent a mean rating for each of the 3 subjects who rated the same space and sky condition with error bars showing the standard deviation between subject ratings. It is important to note that PCC values are shown for the linear regression fit through *mean subject responses*. Most significant in this analysis are the PCC values ≥ 60% for RAMMG – 'calming - exciting' and RAMMG – 'sedating - stimulating'. For all of the semantic rating scales, including 'low contrast – high contrast', global metrics such as RMS and Michelson show no significant trend. What is perhaps most interesting about these initial results it that ratings of 'calming-exciting' and 'sedating - stimulating' are more strongly correlated to pyramidal subsampled metrics such as RAMMG than subjective ratings of 'contrast' for which the metrics were designed. This finding raises some interesting questions about subjective interpretations of the word 'contrast' and whether subjects are responding to micro or macro compositional effects within the images. The authors believe that further research must be conducted to understand which compositional characteristics subjects are responding to when asked to rate the contrast of a daylight composition. Given the lack of consensus on contrast measurements, it is not surprising that each subject may define it differently.

table 1 - ANOVA RESULTS

low contrast – high contrast

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	24.765	8	3.0957	1.23	0.2995
sky	**37.062**	**2**	**18.5309**	**7.34**	**0.0014****
space	29.877	8	3.7346	1.48	0.1835
error	156.617	62	2.5261		
total	248.321	80			

uniform – non-uniform

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	20.444	8	2.55556	0.79	0.6156
sky	7.63	2	3.81481	1.18	0.3155
space	46.667	8	5.83333	1.8	0.0947
error	201.259	62	3.24612		
total	276	80			

unvaried - varied

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	23.556	8	2.9444	0.99	0.4517
sky	**24.963**	**2**	**12.4815**	**4.2**	**0.0194***
space	55.556	8	6.9444	2.34	0.029*
error	184.148	62	2.9701		
total	288.222	80			

simple - complex

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	15.778	8	1.9722	0.96	0.4746
sky	**34.296**	**2**	**17.1481**	**8.35**	**0.0006****
space	**122.222**	**8**	**15.2778**	**7.44**	**<0.000****
error	127.259	62	2.0526		
total	299.556	80			

calming - exciting

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	27.432	8	3.429	1.41	0.2102
sky	**20.469**	**2**	**10.2346**	**4.21**	**0.0193***
space	**98.099**	**8**	**12.2623**	**5.04**	**0.0001****
error	150.765	62	2.4317		
total	296.765	80			

sedating - stimulating

Source	*Sum Sq.*	*d.f.*	*Mean Sq.*	*F*	*Prob>F*
subject	23.778	8	2.9722	1.66	0.1257
sky	**21.407**	**2**	**10.7037**	**5.99**	**0.0042****
space	**66**	**8**	**8.25**	**4.62**	**0.0002****
error	110.815	62	1.7873		
total	222	80			

*****Prob>F is less than 5% ****** Prob>F is less than 1%

CONCLUSION

In conclusion, the results suggest that both sky and space factors have significant effects on three out of six proposed semantic rating scales. Considering the time-savings of a Latin-Square over a Full-Factorial experimental approach, the authors will be able to distribute the forthcoming online survey to a broader audience and increase response rates. When semantic ratings for contrast, complexity, excitement, and stimulation were compared to quantitative contrast measures, we can see that local measures such as SC and RAMMG show higher correlation than global

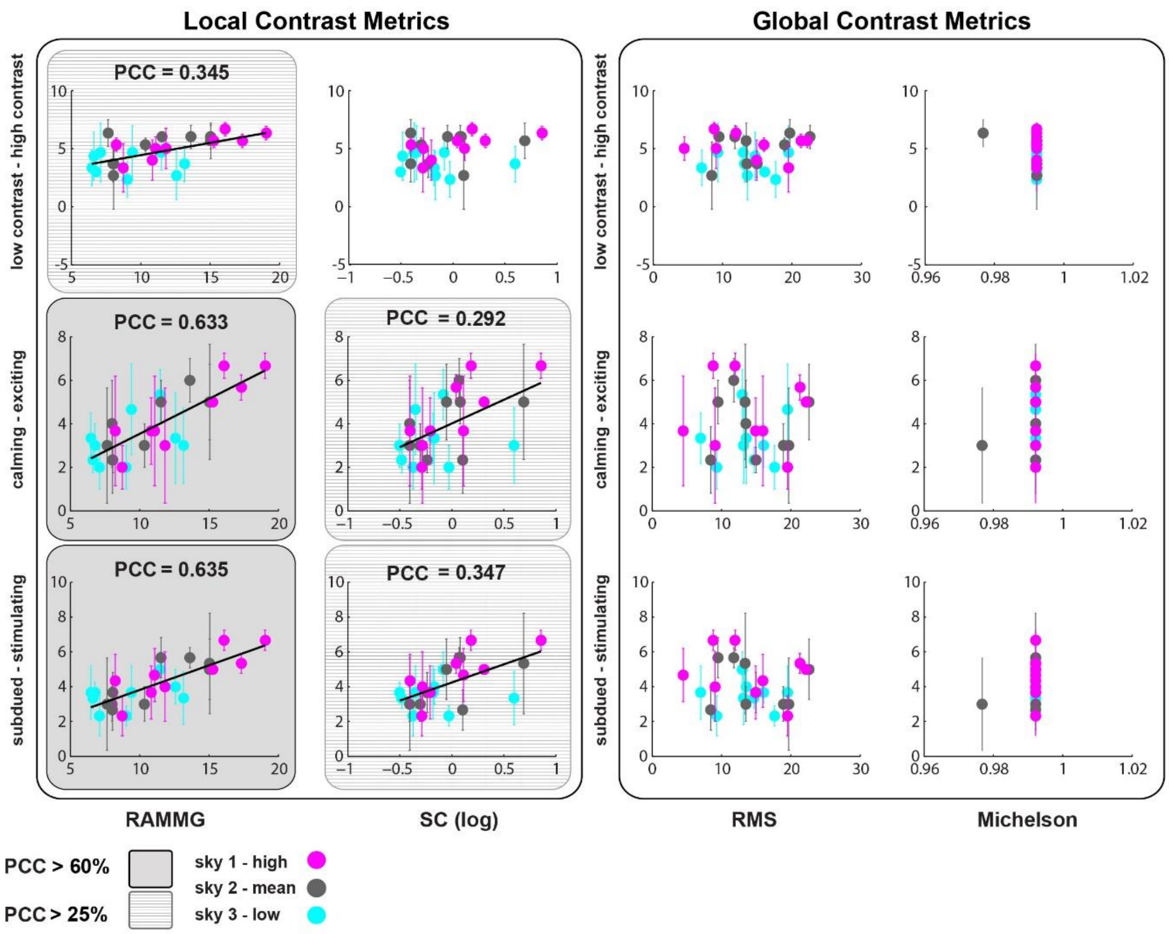


Figure 4 Correlation cluster study on the ***mean subject response per space and sky condition*** for each of the 3 subjects in each group. Error bars show the standard deviation between all three subjects for each space and sky condition. The three sky conditions are shown in magenta (sky 1), grey (sky 2), and cyan (sky 3). Pearson's linear correlation coefficients (PCC) show the linear regression through *mean subject responses* for each metric and rating pair.

measures like Michelson and RMS. Local contrast measures therefore appear to predict contrast and contrast-based effects better than global measures, as hypothesized by the authors. While the linear regression between mean subject responses for excitement - stimulation ratings and RAMMG shows particular promise, a larger sample size, provided by the forthcoming online experiment, is necessary to perform a more sophisticated analysis. The current sample is too small to extract conclusive trends, whether linear (shown in Figure 4), non-linear, or threshold driven.

As RAMMG was the most highly correlated measurement to subjective responses used in this study, the forthcoming experiment will consider an expanded group of neighbourhood metrics which include variations on weighting methods for multi-level subsampling (Matekovic, et al., 2005) as well as Gaussian filters (Tadmor & Tolhurst, 2000). Furthermore, the authors hypothesize that a combination of existing contrast metrics or a refined weighting of resolution channels within a single neighbourhood metric could improve the link between quantitative and subjective measures of contrast and its resulting visual effects. A broader sample size and range of quantitative metrics will allow us to explore the data further and develop a more robust measure. Using simulation, this validated measure can assess dynamic contrast-based visual effects such as excitement and stimulation alongside existing illumination, glare-based discomfort, and health-based metrics. This unified framework is essential to creating more holistic evaluation tools for measuring human responses to daylight in architecture (Andersen, 2015).

ACKNOWLEDGEMENTS

Funding for this research is provided by the Ecole Polytechnique Fédérale de Lausanne and through a grant awarded by the Velux Stiftung Foundation.

REFERENCES

Andersen, M., 2015. Unweaving the human response in daylighting design. *Building and Environment,* Volume 91, pp. 101 - 117.

Boubekri, M., Hull, R. & Boyce, L., 1991. Impact of Window Size and Sunlight Penetration on Office Worker's Mood & Satisfaction: A Novel Way of Assessing Sunlight. *Environment & Behavior,* 23(4), pp. 474-493.

Cauwerts, C., 2013. *Influence of presentation modes on visual perceptions of daylit space.* Louvain-la-Neuve: Universite Catholique de Louvain.

Cauwerts, C. & Bodart, M., 2011. *Investigation of 3D projection for qualitative evalutation of daylit spaces.* Louvain-la-Neuve, s.n., pp. 819-824.

Cetegen, D., Veitch, J. & Newsham, G., 2008. *View Size and Office Illuminance Effects on Employee Satisfaction.* Ljubljana, Slovenia, s.n., pp. 243-252.

Cuttle, C., 2015. *Lighting Design: a percetion-based approach.* New York: Routledge.

Demers, C., 2007. *A Classification of Daylighting Qualities Based on Contrast and Brightness Analysis.* Cleveland, Ohio, s.n., pp. 243-252.

Flynn, J., Hendrick, C., Spencer, T. & Martyniuk, O., 1979. A guide to methodology procedures for measuring subjective impressions in lighting. *Journal of IES,* pp. 95-110.

Hess, R., Bradley, A. & Piotrowski, L., 1983. *Contrast-coding in amblyopia. I. Differences in the neural basis of human amblyopia.* London, s.n., pp. 309-330.

Holl, S., 2011. *Color Light Time.* Zurich: Lars Muller Publishers.

Konis, K., Lee, E. & Clear, R., 2011. 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,* 7(3).

Loe, D., Mansfield, K. & Rowlands, E., 1994. Appearance of Lit Environment and its Relevance in Lighting Design: Experimental Study. *Lighting Research and Technology Vol 26,* pp. 119-133.

Matekovic, K. et al., 2005. *Global Contrast Factor: a New Approach to Image Contrast.* Aire-la-ville, Computational Aesthetics in Graphics, Visualizationg, and Imaging.

Michelson, A., 1927. *Studies in Optics.* Chicagp: University of Chicago Press.

Montgomery, D., 2012. *Design and Analysis of Experiments.* 8th edition éd. s.l.:Wiley.

Newsham, G., Cetegen, D., Veitch, J. & Whitehead, L., 2010. Comparing Lighting Quality Evaluations of Real Scenes with those from High Dynamic Range and Conventional Images. *Transactions on Applied Perception.*

Newsham, G., Marchand, R. & Veitch, J., 2002. *Preferred Surface Luminance in Offices: A Piolet Study.* Salt Lake City, IESNA Annual Conference.

Newsham, G., Richardson, C., Blanchet, C. & Veitch, J., 2005. Lighting quality research using rendered images of offices. *Lighting Res. Technol.,* pp. 93-115.

Pallasmaa, J., 2005. *The Eyes of the Skin: Architecture and the Senses.* s.l.:Academy Press.

Parpairi, K., Baker, N., Steemers, K. & Compagnon, R., 2002. The Luminance Differences index: a new indicator of user preferences in daylit spaces. *Lighting Research and Technology Vol 34 (1),* pp. 53-68.

Pavel, M., Sperling, G., Riedl, T. & Vanderbeek, A., 1987. Limits of visual communication: the effect of signal-to-noise ratio on the intelligibility of American Sign Language. *Opt. Soc. Am. A,* 4(12), pp. 2355 - 2365.

Reinhart, C. & Mardaljevic, J. R. Z., 2006. Dynamic Daylight Performance Metrics for Sustainable Building Design. *Leukos Vol 3 (1),* pp. 1-25.

Rizzi, A., Algeri, G., Medeghini, D. & Marini, A., 2004. *A proposal for contrast measure in digital images.* Aachen, s.n.

Rockcastle, S. & Andersen, M., 2014. Measuring the dynamics of contrast & daylight variability in architecture: A proof-of-concept methodology. *Building and Environment,* Volume 81, pp. 320-333.

Simone, G., Pedersen, M. & Hardeberg, J. Y., 2012. Measuring perceptual contrast in digital images. *J. Vis. Commun. Image R.,* pp. 491-506.

Steane, M. A., 2011. *The Architecture of Light: Recent Approaches to Designing with Natural Light.* New York: Routledge.

Tadmor, Y. & Tolhurst, D., 2000. Calculating the contrasts that retinal ganglion cells and LGN neurones encounter in natural scenes. *Vision Research 40,* pp. 3145-3157.

Tiller, D. & Veitch, J., 1995. Perceived Room Brightness: Pilot Study on the Effect of Luminance Distribution. *Lighting Research and Technology Vol 27 (2),* pp. 93-101.

Tremeau, A., 2000. *Color contrast parameters for analysing image differences.* Derby, s.n., pp. 11-23.

Veitch, J. & Newsham, G., 2000. Preferred Luminous Conditions in Open Plan Offices: Research amd Practice Recommendations. *Lighting Research and Technology Vol 32,* pp. 199-212.

Villa, C. & Labayrade, R., 2012. Validation of an Online Protocol for Assessing the Luminous Environment. *Lighting Research & Technology,* 2012(0), pp. 1-20.

Vogels, I., 2008. Atmosphere metrics: Development of a tool to quantify experienced atmosphere. Dans: *Probing Experience.* s.l.:Springer, pp. 25-41.

Ward, G., 1994. *The RADIANCE Lighting Simulation and Rendering System.* s.l., s.n., pp. 459-472.

Ward, G., Rushmeier, H. & Piatko, C., 1997. *A Visibility Matching Tone Reproduction Operator for High Dynamic Range Scenes,* San Francisco : Lawrence Berkeley National Lab, University of California.

Whitehead, L., Ward, G., Stuerzlinger, W. & Seetzen, H., 2005. *High Dynamic Range Display Devices.* US, Brevet n° 6,891,672.

Wymelenberg, K. & Inanici, M., 2009. *A Study of Luminance Distribution Patterns and Occupant Preference in Daylit Offices.* Quebec City, PLEA, Annual Conference Proceedings.