Exploring daylight features in relation to a realistic daylight impression

Steven Bos Allart Hoekstra Martijn Rentmeester Frank van Wijk
Delft University of Technology
Faculty of EEMCS
Department Mediamatics

Abstract—Engineering artificial light sources capable of producing a realistic daylight impression has proven to be notoriously difficult. Light sources with a similar spectral power distribution as the sun, such as incandescent light, are unable to trick a human subject on both the psychological and physiological level. Breakthroughs in engineering artificial light sources capable of this feat will revolutionize architectural design decisions, enable more realistic virtual environments and might be beneficial for health treatments for example. In this paper we tried to find out what features contribute the most to a realistic daylight impression. We experimented with three features involved with a realistic daylight impression: directionality of light, colour temperature and window visibility. Our main research question was to discover which feature combination of three features resulted in the strongest realistic daylight impression. To answer this question we set up a paired comparison experiment, presenting two images with varying combinations of the three features. The participant selected one of the two images which perceptually gave the strongest realistic daylight impression. This process was repeated until all combinations were judged. Using a specialized analysis algorithm involving the GLM extended Thurstone model, statistically significant results were found. From these results we can conclude that directional light and cold light give a significantly stronger realistic daylight impression than respectively diffuse light and warm light. No statistical significance was found for window visibility.

Index Terms—Daylight impression, Directionality, Colour temperature, Window visibility, Paired comparison, GLM, Thurstone model.

1 Introduction

In this experiment we wanted to research which feature combination, given a set of three features of daylight, gives the strongest realistic daylight impression. Researching this question is important, since daylight can have an impact on human performance and workspace productivity, but also on human health [1]. Boyce mentions that absence of daylight effects the performance of both visual and non-visual tasks and might lead to long-term health problems. The work of Leather [2] describes the relation between mood changes and daylight.

The availability of artificial light sources capable of producing a realistic daylight impression in offices and homes for example, would thus be very helpful. Currently, it is hard to engineer these light sources since it is unknown what daylight exactly is and why humans recognize daylight as daylight and not as electrical light. When we know which features of daylight are important for a realistic daylight impression, it can be a breakthrough in engineering artificial light sources capable of producing a realistic daylight impression.

After doing a literature study about daylight in general and about which features can be important for a realistic daylight impression we came up with the following list:

- Brightness, an attribute of a visual sensation according to which an area appears to emit more or less light. The importance of this feature is proved in [3].
- Daylight area, an area filled with reflected daylight on a surface.
- Colour temperature, the temperature in Kelvin (K) of the reflected daylight on white objects within the image. We made

a difference in warm images (7600K) and cold images (9500K). The importance of this feature is proved in [3].
- Colour rendering, effect of an illuminant on the colour appearance of objects by conscious or subconscious comparison with their colour appearance under a reference illuminant [4].
- Dynamics, the dynamic of daylight in all aspects, such as the dynamic in intensity or colour temperature.
- Directionality, percepted by the sharpness of the edges of the shadows and the diffuseness of the (reflected) light. This feature is an important one in the daylight design of buildings [5] and therefore it will probably contribute to the impression of realistic daylight.
- Window visibility, is about if there is a window visible. It is widely known that windows are an important feature in the daylight design of buildings [5]. Because most visible daylight in buildings comes through the presence of windows, we thought that the visibility of a window should probably contribute to a realistic daylight impression.
- Airflow, a tactile sensation which can support a realistic daylight impression. When one feels a breeze, one might think he is outside and thus his environment should be illuminated by daylight.
- Temperature, a tactile sensation which can support a realistic daylight impression. An increase in the sensed temperature caused by the presence of light can influence a realistic daylight impression, because a higher temperature can feel more like you are illuminated by sunlight.

Since this experiment is part of a pilot study, we focused on only three features. After some discussion we chose the three most intuitive features: window visibility, directionality and colour temperature. These features have proven to have an effect on the well-being of a person [6] and have enough visible presence that we can use them in our experiment.

2 Protocol

We set up a paired comparison perception experiment to measure the effect of three features on a realistic daylight impression using images. In this chapter we will explain all the details regarding this experiment.

2.1 Stimuli

During the experiment, we have tested three features. These features are discussed below. Each possible combination of these features was represented in two images, both in a home setting. For each feature two (opposite) values of that feature has been tested. The type and style of the interior was constant in all images.

With these parameters combined, the total number of stimuli was $2^3 \times 2 = 16$. All three features were present in each image, where each feature took one of the values in table 1. In total there were 8 (2^3) unique stimuli. These eight stimuli can be found in table 1.

TABLE 1
Overview of the 8 different stimuli

Window visibility	Colour temperature	Directionality of light
Window	Warm	Directional
Window	Warm	Diffuse
Window	Cold	Directional
Window	Cold	Diffuse
No Window	Warm	Directional
No Window	Warm	Diffuse
No Window	Cold	Directional
No Window	Cold	Directional

The features were varied within this way:
- **Window visibility**
  Do people automatically think that a room is illuminated by daylight when they see a window, because then it might be obvious that the light that illuminated the room has come through this window and thus from outside and thus from the sun? To test this, the set of stimuli contained eight images with a window visible and eight without.
- **Colour temperature**
  Because sunlight has a range of colour temperatures over a day, the colour temperature of a room lit by the sun can give

the participant a realistic daylight impression.
To test this hypothesis, the set contained eight images where white objects in the interior had a colour temperature of 7600K (warm) and eight images where these objects had a colour temperature of 9500K (cold). These temperatures were the temperatures of the images when they were projected on the TV screen that was used and then measured with a colorimeter under a two degree angle at 1 meter distance.
- **Directionality**
  How sharp are the shadows seen in the image? Are the shadows placed in one direction? Does the light spread via (light) walls or is there many direct light?
  To see if the feature directionality has influence on a realistic daylight impression, the set of stimuli contained eight images which were illuminated with diffuse light and eight images which were illuminated with direct light.

To create colour temperature differences, we used Adobe Photoshop LBA photo filters, defined hue/saturation and did colour balance adjustments. Examples of two images we have used in the experiment are figure 1 and 2.

Fig. 1. Image with directional and warm light, and no window visible

2.2 Equipment

In this experiment we have used a PHILIPS Cineos 42PFL9732D/10 LCD TV [7] and set the resolution at 1280 x 800 pixels, with a refresh rate of 60 Hz. We measured a whitepoint of

Fig. 2. Image with diffuse and cold light, and a window visible

9100K. This was done with a colorimeter under a two degree angle at 1 meter distance.

This TV is located in the Experience Lab at the TU Delft [8] where this experiment took place. In this lab the light level can be adjusted by completely darkening the windows and by adjusting the light level of the TL lamps. These TL lamps provide an uniform light distribution and can be adjusted in steps in intensity from 0% to 100%.

During the experiment we made the room almost fully dark. We only had some light on the wall behind the testing display with a light level of 80 lux, measured at a distance of 2 meters perpendicular to the wall.

The software we have used for the paired comparison is a java-written program for doing paired comparisons and ranking images, custom build by Philips. The analysis software was an implementation of the paper by [9] in Matlab.

2.3 Participants

In this experiment a total of 25 participants participated. But only the results of 20 participants could be used, since the results of 2 participants were lost and 3 participants did not pass the colour blindness test. The resulting participants all met our predefined characteristics:

People (both male and female) in the age between 18 and 60 years, not suffering from colour blindness.

This upper boundary in age is chosen since from the age of 60, people can suffer from

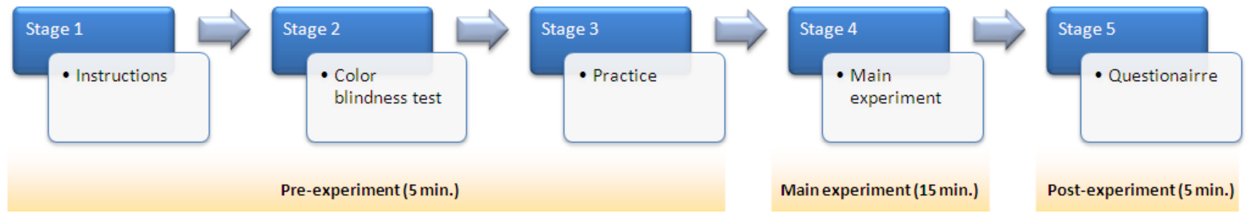

Fig. 3. Five stages of the experiment

cataract, diabetes and other eye diseases causing vision loss which could influence our results [10]. In chapter 5 there is a bias discussion about the people who actually participated in our experiment.

2.4 Procedure

The complete experiment consisted of three phases. The pre-experiment phase, the main experiment and the post-experiment phase. All three phases will be explained below.

2.4.1 Pre-experiment

One or two of us were present in the experiment room. The participant was welcomed and guided to the test environment. One of us gave the participant a sheet with instructions and an explanation of the experiment (see the appendix). The verbal explanation included:

- **Global instructions**
  We told the participant that the experiment was about measuring the impression of realistic daylight. We also have explained all the phases of the experiment.
- **Mental preparation**
  We also have prepared the participant mentally so he or she knew what was coming and what he or she was expected to do. Some important notes were that the participant knew that there were no good or wrong answers, that there was no time limit, that the choice between images had to be taken intuitively and that if the participant could not make a choice, he or she had to select one image at random.

During the instructions the eyes of the participant are accustomed to the environment. The instructions are called stage 1 (see figure 3).

After the instructions a mandatory Ishihara colour blindness test was performed. Note that the participant was not allowed to skip this. The colour blindness test is called stage 2. The participant passed the colour blindness test if (s)he scores 100%.

When completed, one of us guided the participant to the television. The participant had to take a seat in a chair in front of the screen with a distance of about 1 meter from the head to the screen. One of us opened the experiment application and instructed the participant what he/she had to do during the main experiment.

Next, two example pairs, which were not used in the main experiment, were presented to get the participant accustomed to the experimental setup. Every participant got the same two test pairs. These pairs were just like the pairs presented in the main experiment, so only images in a home setting were presented.

When the participant finished the two comparisons, one of us asked if everything was clear. If the participant confirmed this, one of us mentioned again that the main experiment would be about the impression of realistic daylight. One of us also told that he will be present somewhere in the back of the room during the whole experiment to answer any questions if there are any. The practice phase is called stage 3.

The total time for the pre-experiment was approximately 5 minutes. Only a single participant was participating in the experiment at a time.

2.4.2 Main experiment

The fourth stage was the main experiment. The experiment consisted of comparing two images, the so called *paired comparison method*. Every participant was presented a randomly ordered sequence of pairs. All possible combinations of images were made (see 2.1). Between every question a 2 second pause was given (50% grey screen).

The participant was asked which of the two displayed images had the most realistic daylight impression. All selections were saved in a big database, to be used in the analysis. With k items, there are

$$\frac{k(k-1)}{2}$$

pairs in total [11]. In our case $k = 2^3 \times 2 = 16$, so we got

$$\frac{16(16-1)}{2} = 120$$

pairs per participant with a total of 20 participants. The total time for this phase was on average 15 minutes.

2.4.3 Post-experiment

After the experiment the participant received a questionnaire (see the appendix), allowing the participant to provide feedback on the experiment and procedure. This phase is called stage 5. After the participant handed in the questionnaire (s)he was acknowledged. The total time for the post-experiment was on average 5 minutes.

3 Analysis method

In the previous chapter, we discussed how the raw data was obtained. In this chapter we will discuss how we transformed the raw data into meaningful results. This process encompasses four phases:

1) Data acquisition
2) Data conversion
3) Data visualization
4) Interpretation of visualization results

Each phase is discussed in some detail, parts are discussed more elaborately elsewhere in this paper and/or can be found in [9].

3.1 Data acquisition

The method of data acquisition is discussed in chapter 2. Table 2 displays raw data outputted by the paired comparison application. An example (first 8 rows) of the raw data of participant #15 is presented.

TABLE 2
Raw data: first 8 rows of results of participant #15

#	Pair presented[1]	Chosen value[2]	Reaction time
0	0	1	6.766
1	1	1	12.266
2	2	1	3.328
3	3	1	10.703
4	71	0	4.078
5	52	1	3.656
6	29	1	3.813
7	35	0	4.594

[1] random, except for first 4 test images
[2] left is 0, right is 1

3.2 Data conversion

To get from raw data to data that can be compared, we used the method described in [9] and is based on the extended Thurstone model. First the basic components of the method are briefly described, next are the steps of the method itself. Because the paper is very scarce in presenting the underlying mathematical assumptions, we found it necessary to include an overview of the assumptions.

Thurstone model

Thurstone in 1927 [12], defined that discriminal processes taking place in the human brain and needed for paired comparison, have a normal distribution although he admits there is no scientific basis. With this definition he created "*the law of comparative judgement*" involving 5 practical cases in which under more assumptions and approximations differ in degree of simplification. An example of raw data of three features fitted on a case 5 Thurstone model is presented in figure 4. In recent literature the simplest, the case 5 model, is mostly used and often without explicit mentioning. Mosteller [13] strikingly captured the essence of describing the case 5 model:

> "It is assumed that we have a set of stimuli (i.e. features) which, when presented to a subject, produce sensations. These sensations are assumed to

be normally distributed, perhaps with different means. However the standard deviations of each distribution are assumed to be the same, and the correlations between pairs of stimuli sensations are assumed equal. [..]"

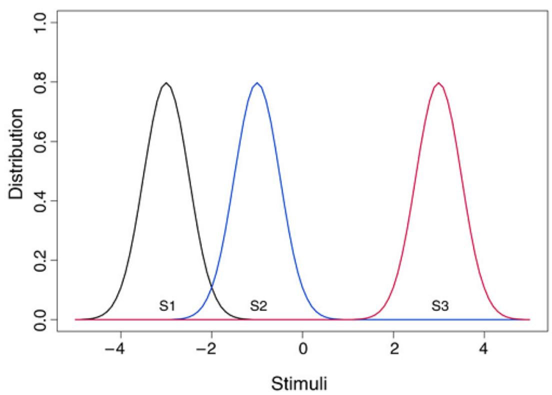

Fig. 4. Example of case 5 Thurstone model of three features, source: [14].

Generalized Linear Models

Case 5, the simplest case of Thurstone's models, is actually a linear case and belongs to the group of Generalized Linear Models (GLM). The case 5 Thurstone model is called in statistics literature the "probit model". Viewing the Thurstone model within the GLM methodology allows calculation of more statistics and enables testing the z-score differences on a confidence interval.

Z-score

The z-score represents the amount of standard deviation from the mean. Through z-scores, different features can be compared. This is possible since the z-score is calculated from a normalized distribution such that standard deviation has the same meaning (see figure 5). In our case the z-score represents the distance of the mean of a feature from the mean of the benchmark feature. This distance is expressed in the amount of standard deviation from the benchmark mean. The benchmark is the lowest ranked feature. The mean of the lowest ranked feature is shifted to zero, adding the difference to all other z-scores.

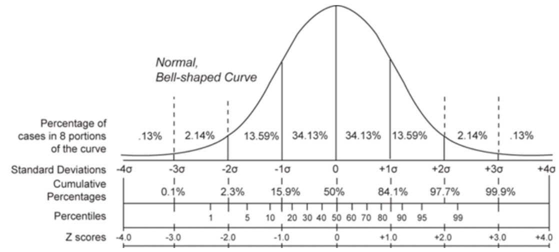

Fig. 5. The normal distribution with amongst other the z-score measurement scale, source: [15]

Conversion process

The method by Rajae-Joordens and Engel [9] consist of six steps, using the components briefly discussed above.

1) Either a raw paired comparison (binary) dataset or ranked data per participant is provided as input (in this experiment a binary dataset is used).
2) An intermediate regression matrix is calculated based on the raw data.
3) The intermediate regression matrix is checked and possibly corrected if extreme fractions exist (0 or 1 as probability), since these fractions cannot be fitted on the Gaussian distribution ($-\infty$, respectively $+\infty$). Rajea-Joordens and Engel propose two rules for correction in the paper.
4) The GLM enhanced Thurstone model is fitted on the (corrected) regression matrix. The output is a 2D matrix of z-score pairs. Now by simply adding over a single feature, the total z-score for that feature can be calculated and be placed on a Thurstone interval scale.
5) The lowest z-score is mapped to zero, all other z-scores are shifted with the same amount. After the mapping all z-scores are positive.
6) All possible pairs of z-scores are tested on a 95% confidence interval and are judged to be either significant or not. The output are z-score pairs being either significant or not.

Assumptions

- Thurstone case 5 (probit) model assumes zero, near zero or spherical (equal) correlation between features. Since our set

of stimuli are required to be realistic, the stimuli always contain more than one feature. Hence uncertainty that the other features are also weighted in scoring the image cannot be avoided. Philips experts selected a subset from the set of stimuli based on the dominant presence of the three wanted features, such that the correlation is expected to be near zero.
- Thurstone case 5 model assumes equal variance for all features.
- The ”law of comparative judgement” by Thurstone assumes a Gaussian distribution with unknown μ. Other distributions have been used in paired comparison experiments.
- GLM assumes a distribution from the exponential family, which in this case is the standard normal distribution.
- Assuming a complete dataset per participant, not discarding any results. If we do discard results (for example by long reaction time or due to large variation), some results could be statistically insignificant because of the small test size used.

Limitations of the current method

- Unable to cope with incomplete data. Various methods exist to cope with incomplete data such as [16] and [13]. Mosteller’ method shows that full ranking of all features remains possible even when each participant is provided with few features of a total set.
- Other fitting method. For fitting the data a maximum likelihood method is used, however other techniques such as Bayesian could also be tried.
- Not using aggregated data. Instead of using aggregated frequency data, the raw data with scores per participant could be used [14].
- Not reusable. The end result (rank of three features) of this experiment cannot be reused when more features are considered using this method. The method of [17] allows features not yet considered, to be added to the acquired data. This is also more in the philosophy of an ”Analytic Hierarchy Process”, which this visual perception experiment is part of.

3.3 Data visualization

The raw z-scores from step 4 of the conversion algorithm and the significant z-score pairs from step 6 can be visualized in many ways. For paired comparison experiments often Thurstone interval scale visualizations (see figure 6) are used, sometimes extended with a significance bar. When two significance bars of two features overlap they are considered not significant. Another visualization option is to visualize the significant z-score pairs using a 1D histogram plot or a 2D matrix showing all possible combinations. The values of the matrix can be coloured green and red representing significant respectively not significant. When all values are green, a ranking can be done without a-cyclic relations.

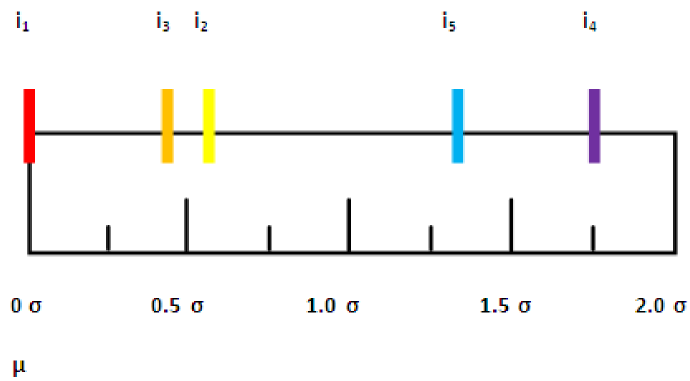

Fig. 6. Thurstone interval scale (insignificant)

3.4 Interpretation of visualization results

For this paired comparison experiment two general types of results exist, significant results and not significant, each having another interpretation. Depending on the visualization used, extra interpretations can be made and possibly more complex conclusions can be drawn.

- **Insignificant (based on z-score alone)** Ranking of all features is possible, however it is not checked whether these results (the inter-relationships between features) are significant. This could lead to a completely different ranking when more data is available. With large datasets significance is often assumed.

- **Significant**
When only considering significant results (relationships), ranking of all features is not always possible since interrelationships of features could be statistically weak. It is possible that these relationships switch positions (i.e. $a > b$ becomes $b > a$). This is likely when considering small datasets. Often few significant results could prove to be valuable itself or indicative to further explore when experiments have tight time and budget control.

4 RESULTS

In this chapter we will discuss and interpret the raw results gained from the analysis phase, which was done by using a Matlab implementation of the analysis method of Rajae-Joordens and Engel [9]. In chapter 5 the results are put in a broader daylight: the results are discussed as well as guesses to why they appear.

One of the first and most important results is the z-score ranking of the 16 stimuli plotted on a Thurstone interval (see figure 7. The numbers in this figure are explained in table 3).

TABLE 3
Ranking of images (each item represents a combination of three features)

#	Image features
1	dif nwv w 2
2	dif wv w 2
3	dif nwv w 1
4	dif wv w 1
5	dif nwv c 2
6	dir nwv w 2
7	dir nwv w 1
8	dif wv c 1
9	dif wv c 2
10	dir wv w 1
11	dir wv w 2
12	dif nwv c 1
13	dir wv c 1
14	dir nwv c 1
15	dir nwv c 2
16	dir wv c 2

dir: directional light
dif: diffuse light
wv: window visible in image
nwv: no window visible in image
w: whitepoint with lower colour temperature (warm image)
c: whitepoint with higher colour temperature (cold image)

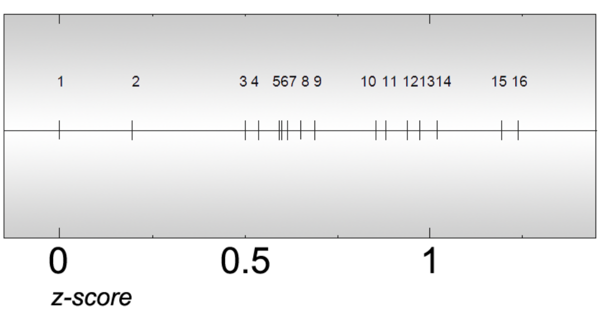

Fig. 7. Thurstone interval scale showing raw z-score ranking of feature combinations. Significance has not been tested!

The Thurstone interval visualizes a ranking of z-scores between the features. This ranking is however not tested for significance. Still, by visual inspection two interesting points can be deduced:

1) **The second set of the 8 combinations matches the first set nearly everywhere.** As explained in the protocol with 3 binary scaled features only 8 combinations are possible and of each combination two images were chosen resulting in 16 stimuli. The overlap of the first and second image of an identical feature combination indicate that the participants rated the images consistently, therefore strengthening the results. More interestingly is that the ranking appears to be done subconsciencely since many participants expressed in the questionnaire afterwards that they changed their judgement criteria. If this were to be true, few overlap should appear, which is not the case.
2) **The list already shows a realistic daylight impression involving at least directional light (6 results in the top 8 (actually the bottom 8 in table 3)) and cold light (6 results in the top 8 (actually**

the bottom 8 in table 3)). Although still nothing can be said if these results are statistically significant at least by visual inspection of the distribution on the Thurstone scale, they appear to be very distinct from their counterpart (diffuse light, respectively warm light).

The analysis software then tests all possible z-score differences (120 pairs) of the feature combinations on a 95% confidence interval. Since in our experiment we involved only 20 participants, we expected the list to contain more insignificant results than significant. Beforehand we also noted that most inter-results (one of the 120 results such as dir_wv_c_1 vs. dir_wv_c_2) will be not of any interest since the goal of this experiment is to find the combination that leads to the most realistic daylight impression. We are also interested in the aggregated results where features are isolated and ranked against their counterpart such as directional light vs. diffuse light.

1) Colour temperature (figure 8)
   Z-scores:
   cold: 0.365586
   warm: 0.000000
   Comparisons:
   cold-warm: 0.3656+-0.0656
   Significant

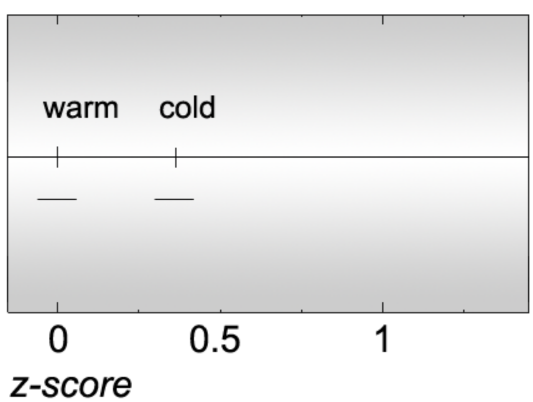

Fig. 8. Colour temperature with significance bar

2) Window visibility(figure 9)
   Z-scores:
   no window: 0.000000
   window: 0.061342
   Comparisons:
   window-no window: 0.0613+-0.0641
   Not Significant

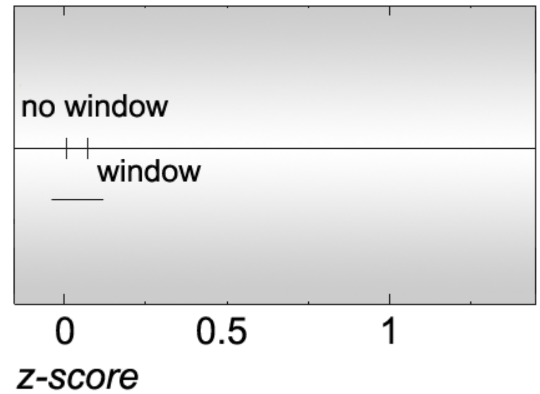

Fig. 9. Window visibility with significance bar

3) Directionality (figure 10)
   Z-scores:
   diffuse: 0.000000
   directional: 0.385687
   Comparisons:
   directional-diffuse: 0.3857+-0.0658
   Significant

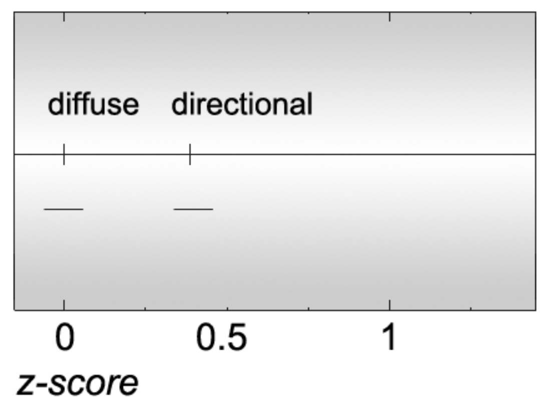

Fig. 10. Directionality with significance bar

The isolated features with their counterparts are numerically and visually presented. Four interesting points can be deduced:

1) Directional light gives a significantly stronger realistic daylight impression than diffuse light, since the significance bar do not overlap.
2) Cold light gives a significantly stronger realistic daylight impression than warm

light, since the significance bar do not overlap.
3) No significant effect of window visibility was found on a realistic daylight impression, since the bars overlap.
4) Due to the last result no combination of three features are significantly strong enough to be classified as the strongest realistic daylight impression triplet. However, it is likely that the triplet contains directional and cold light since they were both significant from their counterpart.

5 Discussion

This experiment was performed with a total of 25 participants. The results of 20 participants were useful, since 2 results were lost and 3 participants were measured colour blind. The 20 participants that were left had an average age of 26,8 years. Five of these participants were female and fifteen were male. Biases occur in gender, age and educational background.

The dataset is thus heavily biased. Despite this bias, we did not found a significant relation between age and concentration difficulties for example, according to the results from our questionnaire. There is also no indication that age or gender is related to be more or less sensitive for distracting factors according to the questionnaire. From this, one might think that the results are not influenced by the bias in age and gender. Nevertheless, we can not conclude that the bias in age and gender has totally no influence on our test results (we had too few participants to measure this), because other unmeasured perceptual differences might have influenced the results.

Colour temperature

From the results of the experiment we can conclude that cold light gives a significantly stronger realistic daylight impression than warm light, where cold light and warm light are defined in section 2.1. This might be called a little surprise, since we expected that warm light would give a more realistic daylight impression, based on our own intuition. One possible answer might be that sunlight is colder during normal days and that sunlight is warmer in the evening and on very hot days. Those days are very scarce in The Netherlands. It would be interesting to see if the results are very different when the same experiment is performed in a country with a different climate.

Window visibility

Another conclusion that can be drawn from the results is that the visibility of a window has no significant influence on a realistic daylight impression. Like the previous conclusion, this is different from what we expected. We expected that the visibility of a window would be of influence, since people might reason that the image is illuminated by daylight when a window is visible. Our results show this is not the case. It is hard to come up with a good explanation for this result. We guess that it has to do with experience. People spent much time inside and outside, in environments with and without windows. Therefore they can judge images for a realistic daylight impression if the scene is plausible (such as a domestic or office environment). Thus, the visibility of a window does not add extra value to the realistic daylight impression of an image. We can not ground this guess with facts. It is just a guess. Another explanation for the found result can be that the other two features (colour temperature and directionality) have a significantly stronger influence on a realistic daylight impression than the window visibility feature, through which the window visibility feature has no significant influence anymore. In a experiment with other features, window visibility may still have a significant influence on a realistic daylight impression.

Directionality

The only result that we expected is the significant influence of directional light. When directional light is visible instead of diffuse light, it influences the impression of realistic daylight in a positive way. The reason might be that people experience directional light (coming from the sun) continuously during the day and use that as a reference. Inside, directional light is visible when it falls through a window (and thus might come from the sun). In the case that an image is illuminated by diffuse light, it mostly does not contain parallel shadows and lines. People immediately recognize this difference since they know how shadows look like when objects are illuminated by sunlight. Again this is just a guess.

Most realistic daylight impression

The answer to our main research question could not be completely given with our results since window visibility gave an insignificant result. However we can significantly say that images which contain a combination of directional and cold light give a stronger realistic daylight impression than other combinations.

6 Conclusion

Based on the results and discussion, there are some conclusions we can draw from this experiment. The four most interesting conclusions are:

- Visibility of a window does not significantly influence the realistic daylight impression
- Directional light gives a stronger realistic daylight impression than diffuse light
- Cold light (9500K) gives a stronger realistic daylight impression than warm light (7600K)
- Images with directional and cold light give the strongest realistic daylight impression

It is important to mention that further experiments are needed to verify if these results hold for other populations than the one used in this experiment, since the participants in this experiment are heavily biased in age, gender, educational background and nationality. The participants were mostly Dutch males with an average age of 26,8 years and an academic background.

Other notable points for further experiments are:

- **Use the Analytic Hierarchical Process (AHP) approach**
  With this approach, new features (and thus new combinations of pairs) can be added without the need of resetting the experiment and throwing away all the results. The AHP framework also allows to work with unbalanced or incomplete data.
- **Use a larger set of stimuli and more participants**
  With a larger set of stimuli and more participants, more significant results can appear.
- **Use a more diverse set of participants**
  With a more diverse and larger set of participants, stronger conclusions can be made about the influence of characteristics such as age, gender, study and nationality on the results.

Acknowledgements

We are thankful to our supervisors from Philips Eindhoven, Pieter Seuntiens and Marco van Boven, whose feedback helped us to create a better protocol for performing the experiment and helped us with performing the analysis of the raw data.

We also want to thank Prof. Dr. Ingrid Heynderickx, our supervisor from the TU Delft, for giving support and feedback to steer this project in the right direction.

References

[1] Boyce et al. (2003) *The Benefits of Daylight through Windows*
[2] Leather et al. (1998) *Windows in the workplace*, Environment and behaviour, 30(6), 739-762.
[3] Susanne Fleischer, Helmut Krueger, Christoph Schierz (2001) *Effect of brightness distribution and light colours on office staff*
[4] CIE 17.4 *International Lighting Vocabulary*
[5] Nick Baker, Koen Steemers (2002) *Daylight design of buildings*
[6] Fleischer et al. (2001) emphEffect of brightness distribution and light colours on office staff
[7] PHILIPS Cineos 42PFL9732D/10 LCD TV http://download.p4c.philips.com/files/4/42pfl9732d_10/42pfl9732d_10_pss_eng.pdf
[8] TU Delft Experience Lab, http://mmi.tudelft.nl/experiencelab/
[9] Rajae-Joordens et al. (2005), *Paired comparisons in visual perception studies using small sample sizes*, Displays 26, 1-7
[10] Harvey (2003), *Common Eye Diseases of Elderly People: Identifying and Treating- Causes of Vision Loss*
[11] Thomas C. Brown and George L. Peterson (2009) *An Enquiry Into the Method of Paired Comparison: Reliability, Scaling, and Thurstones Law of Comparative Judgment*
[12] Thurstone (1927), *A law of comparative judgment*, Psychological Review 34, 273-286

[13] Mosteller (1951), *Remarks on the method of paired comparisons: III. A test of significance for paired comparisons when equal standard deviations and equal correlations are assumed*, Psychometrika vol 16:2

[14] Lipovetsky et al. (2003), *Thurstone scaling via binary response regression*, Journal of Statistical Methodology, 93-104

[15] Wiki z-score (Dutch), http://nl.wikipedia.org/wiki/Z-score

[16] David (1987), *Ranking from unbalanced paired-comparison data*, Biometrika vol 74:2, 432-6

[17] Lipovetsky et al. (2003), *Priority estimations by pair comparisons: AHP, Thurstone scaling, Bradley-Terry-Luce, and Markov stochastic modeling*, Proceedings of the Joint Statistical Meeting August 3-7, 2003

Daylight Experiment

1. Instruction

Instruction about what you should do.

2. Color blindness test

Check if you are colour blind or not.

3. Training session

The experiment program will be started.

You will see a couple of consequent image pairs.

Do the following with each given pair:

CHOOSE THE IMAGE WHERE YOU'RE EXPERIENCING THE MOST REALISTIC DAYLIGHT IMPRESSION

4. Main experiment

The main experiment program will be started.

Do the same thing as in the Example questions until the test is ended.

We thank you very much for participating in this experiment!

Kind regards,

Steven Bos, Frank van Wijk, Martijn Rentmeester, Allart Hoekstra

Questionnaire Daylight Experiment

1. What is your age

....

2. What is your gender?

❒ Male
❒ Female

3. What is your highest fulfilled study? And if you have studied, what did you study?

....

4. Did you experience any concentration problems during the experiment? If true, do you have any explanation for these problems.

....
....
....

5. The total time per question was

❒ Too less
❒ Enough
❒ Too much

6. Were there any factors in the environment that extracted you from your task? If true, which?

....
....
....

7. Was it clear what was expected from you during the experiment? Did any questions pop-up during the experiment?

....
....
....

8. Any further comments

....
....
....

TU Delft