Brightness perception and the effect of synthetic glare in virtual lighting applications

Roskam, S.J.

Award date:
2015

Brightness Perception and the Effect of Synthetic Glare in Virtual Lighting Applications

S.J. Roskam
0654469

Master Thesis
February 2015

Supervisors:

Prof. dr. I.E.J. Heynderickx	Department of Industrial Engineering & Innovation Sciences, TU/e
Ir. M.G.M. Stokkermans	Department of Industrial Engineering & Innovation Sciences, TU/e
Dr. M.J. Murdoch	Philips Research
Dr. M.T.M. Lambooij	Philips Research

Eindhoven University of Technology
Department of Industrial Engineering and Innovation Sciences
Master's Program Human Technology Interaction

1 Contents

2 Summary

To improve virtual prototyping of lighting as a research and marketing tool, the visualizations should convey the same visual impression as in the real-world situation. Recent research concluded that computer generated lighting visualizations are still in the process of converging towards their maximal predictive value (Engelke et al., 2013). Brightness is a perceptual attribute considered in virtual prototyping for which an accurate match with the physical corresponding space has not yet been achieved consistently (Murdoch et al., 2013).

Considering the findings that synthetic glare can increase brightness in visualizations, it is interesting for the field of virtual prototyping to investigate the impact of synthetic glare on brightness perception further. The main research question of this study is therefore: "*How does synthetic glare affect brightness perception of a rendering of a luminaire?*" We hypothesize that adding synthetic glare to renderings increases the perceived brightness. Secondly, it is hypothesized that synthetic glare can improve the perceived brightness match between the rendering and reality.

In the current study, synthetic glare is applied to renderings of luminaires. By means of a psychophysical study it is investigated how the addition of synthetic glare influences brightness of the visualizations. Since brightness is not absolutely measurable, two relative scales (i.e., an indirect and a direct one) are used to quantify brightness. The indirect scale relates to a brightness-matching task, conducted in Experiment 1. The direct scale refers to evaluative questions that were answered on a 7-point Likert scale, as used in Experiment 2. Hence, this study quantifies the impact of synthetic glare on perceived brightness in visualizations by direct and indirect subjective measurements.

The main hypothesis that synthetic glare can increase the perceived brightness, was confirmed by both direct and indirect brightness assessments. We could not confirm the second hypothesis that synthetic glare can improve the brightness match with reality, but we found, on the contrary, that synthetic glare decreased realism. Furthermore, our results suggest that a luminous spot on the wall is a stronger virtual brightness cue than the (glary) reflectors of a luminaire, and the former seems to overrule the latter. All in all we can conclude that synthetic glare is an effective way to increase the brightness of a visualization, but is also likely to decrease its realism.

3 Introduction

In recent years computer-aided design (CAD) and computer-aided engineering (CAE) have made enormous progress in terms of realism. Virtual models and simulations have become able to serve as prototypes for many products and designs. This so-called *virtual prototyping* is a promising tool for design, research and business. Design processes can be accelerated because virtual prototyping can shorten the feedback loop. Photorealistic computer visualizations of a design can be viewed from any angle, which can be used for evaluation just like a physical prototype. Then, the feedback can be processed directly onto the virtual model, thereby shortening the feedback loop and design process. Marketers can use it to virtually show products to consumers that consider buying them. As a research tool, virtual prototyping can be used in experimental designs to increase time- and cost effectiveness (Engelke, Stokkermans, & Murdoch, 2013). In the current paper the application of virtual prototyping in lighting is examined, with the aim to create visualizations of light that elicit the same feeling as the actual physical space that is simulated.

Light is of vital importance to humans, since our functioning is largely dependent on vision. Our cognitive performance, mood, and biological clock are influenced by the lighting in our environment. Perception research contributes to a better understanding of also the user experience in various lighting situations. Virtual prototyping provides tools to accurately simulate these various lighting situations, and as such allows assessing these user experiences via display-based perception experiments. Virtual prototyping of lighting can also function as a business tool. For example, web-based selling of luminaires could greatly benefit from this technique if it would allow customers to examine visualizations of the luminaires' light appearance. It is important that these visualizations are realistic. The validation of the visualizations by end-users is therefore required. The goal is to create visualizations of light in spatial contexts that convey the same visual impression as in the real situation. Recent research concluded that computer generated lighting visualizations are still in the process of converging towards their maximal predictive value (Engelke et al., 2013). Although for many perceptual attributes that are considered in virtual prototyping (e.g. texture quality, atmosphere, uniformity, and contrast) an accurate match with the physical corresponding space has been achieved, for perceived brightness this is not yet the case (Murdoch, Engelke & Stokkermans, 2013).

Perceived brightness is the perceptual attribute of the amount of radiation a light source emits or a surface reflects, i.e., the perceived amount of luminance. The relationship between luminance and brightness however, is very complex, and in addition, is viewer- and situation dependent. A situational context may have an effect on the state of adaptation in the human visual system, and the adaptation mechanism is largely responsible for an individual's perception of brightness. In real life we could encounter scenes containing light sources mutually differing several orders of magnitude in luminance. Since conventional displays are limited in peak luminance and dynamic range, it is not possible to match such real world scenes in terms of luminance. The limited display luminance range, combined with the complex relation between luminance and brightness make accurate reproduction of brightness difficult. This is reflected in the literature, where virtual brightness assessments of images on a display consistently have a high inter-observer variation (Murdoch et al, 2013). The latter indicates that further research in brightness perception of lighting visualizations on displays is needed.

Although luminance in a virtual visualization is constrained by the peak luminance of the display, brightness is a perceptual attribute and is also susceptible to cognitive top-down processing. By reproducing the lighting in virtual scenes in the right way, a close match between virtual brightness and real-world brightness might be achievable. Virtual brightness is influenced by some factors that we are able to control. These are mainly factors in the process of image creation (e.g., the renderer and tone mapping operator (further explained in section 4.2.2)) and image reproduction (e.g., the display monitor), which all influence the appearance of the final picture. Another method that may influence virtual brightness is the introduction of synthetic glare. Glare is the perceptual effect of scattering in the eye caused by light from bright sources. Synthetic glare approximately simulates this effect by reducing the contrast in the image, especially around strong light sources. Synthetic glare can increase the perceived dynamic range of an image and has shown to be able to increase the apparent virtual brightness (Spencer, Sherley, Zimmerman & Greenberg, 1995) and might for example make luminaires in an online catalogue appear just as bright as in reality. However, the exact influence of synthetic glare on display-related brightness perception is still largely unclear, and therefore, needs further investigating.

The main research question of this study is therefore: "*How does synthetic glare affect brightness perception of a rendering of a luminaire?*" We hypothesize that adding synthetic glare to renderings increases the perceived brightness. Secondly, it is hypothesized that synthetic glare can improve the perceived brightness match between the rendering and reality. In the next chapter, the relevant literature to this research question is discussed. This literature overview is then followed by the description of the methodology used for the creation of the visualizations and for the perception experiment and data analysis. Finally, the last chapters describe and discuss the results.

4 Literary Overview

Several topics relevant to the research question are discussed in this chapter. It therefore starts with an explanation of light characteristics and visual perception. Two core parts can be distinguished in light perception: the physical characteristics of light and the perceptual characteristics of light. The former refers to the physical properties of light as well as the viewing circumstances, whereas the latter refers to the optics of the eye and the sensory and neurological mechanisms in the human brain. Both parts need to be explained in order to understand the complexity of visual perception.

Then, the process of virtual prototyping in lighting is discussed. This section debates rendering, HDR imaging, tone mapping, earlier findings in virtual brightness research and viewing methods. Ultimately, literature on glare is discussed.

4.1 Light Perception

4.1.1 Physics of Light

Physically, light can be defined as electromagnetic radiation that is visible to the human eye. Electromagnetic waves with wavelengths within the range of 400nm to 700nm generate a visual response in the human eye, allowing us to see light sources and objects in colour, depending on the spectral composition of wavelengths that are reflected into our eye (Fairchild, 2005). Electromagnetic energy can be measured in several ways. The radiant power of a source can be measured absolutely: this so-called radiometric approach is not an accurate method to measure the light humans perceive, since it measures the radiant power over the entire spectrum, while humans are only sensitive to a small range of wavelengths. Photometry, on the other hand, is a measurement method developed to measure visible light. This is done by measuring the power at each wavelength and weighing it with the sensitivity of the eye at that wavelength (Barati, 2012). Photometry uses several units and quantities to describe different aspects of light (see Figure 1). The most important are:

- Luminous Flux (lm): Measure of power of the light source in lumen.
- Luminous intensity (cd; lm / solid angle): Measure of luminous flux in a given direction, expressed in candela.
- Luminance (cd / m^2): Measure of luminous intensity per area of a surface.
- Illuminance (lx; lm / m^2): Total luminous flux on a surface per area, expressed in lux.

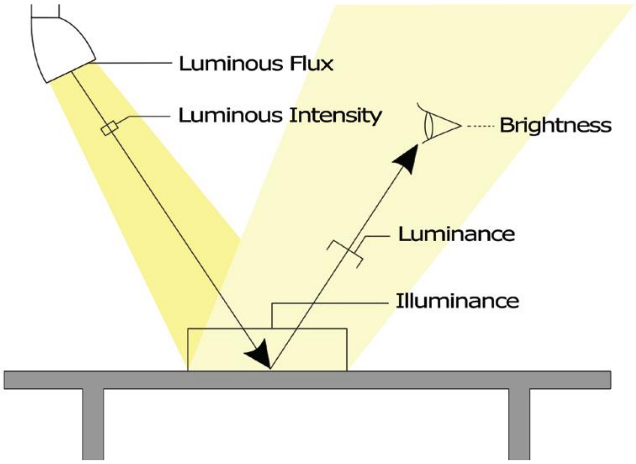


Figure 1: Illustration of luminance, illuminance, brightness, luminous intensity and luminous flux (Barati, 2012)

Visual perception is initiated when light enters the eye through the pupil and falls on the retina, where the photoreceptors are located. The cornea and lens work like the focussing of a camera lens; they are able to converge the light from the focused object to fall on the fovea. The fovea is the centre of the retina and contains the highest density of photoreceptors. Light falling on the fovea therefore generates the sharpest visual image. Still, even when light intensity is measured photometrically, the sensation of light intensity is a different quantity. Humans are very poor at judging the absolute radiant power of a light source. The reason for this is that the human visual system works like a change detector: instead of detecting absolute radiant power, it is sensitive in detecting spatial and temporal changes in radiance (Tumblin & Rushmeier, 1991). Based on these changes the visual system somehow reconstructs the sensation of brightness

4.1.2 Lightness

Lightness is an attribute that is coherent with brightness, but should not be mixed. Brightness describes the perceived luminance, while lightness describes the perceived reflectance. The role of lightness is probably best described by the definitions of Fairchild (2005):

Brightness: "Attribute of a visual sensation according to which an area appears to emit more or less light."
Lightness: "The brightness of an area judged relative to the brightness of a similarly illuminated area that appears to be white or highly transmitting."

From these definitions it can be concluded that lightness is a relative brightness, normalized for changes in the illumination and viewing conditions (Fairchild, 2005). The ability to perceive relative brightness is called lightness constancy and can be explained by the checker-block in Figure 2. The box appears to be painted with two shades of grey and illuminated from an oblique angle. By perceiving it this way, patch *p* and patch *q* seem to have the same reflectance, but different luminance. Patches *p* and *r* seem to have both different reflectance and different luminance. In other words, we perceive *p* and *r* to differ in brightness and lightness (Adelson, 2000). Because the box seems to be illuminated our visual system discounts the illumination, thereby making *p* and *q* appear to be painted in the same shade, and *p* and *r* in different shades, while *p* and *r* are actually the same shade. Although the underlying mechanism is not yet fully understood, lightness constancy suggests that brightness perception relies to some extent on top-down processing of contrast and contextual information (Kingdom, 2011).

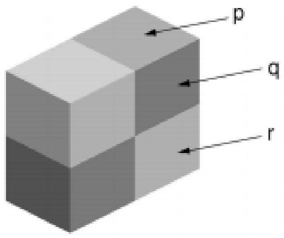


Figure 2: The "checker-block" (Adelson, 2000)

4.1.3 Measuring Brightness

Figure 1 visualizes the interrelations between the photometric quantities and brightness, which cannot be measured with an instrument. The relation between luminance and perceived brightness is complex, but despite the complexity there is some consistency in brightness perception (Stevens, 1961, cited by Boyce, 2003). When subjects were asked to assign a magnitude number to the perceived brightness in such a way that if the stimulus was twice as bright as the first one, they had to assign a magnitude number twice as high, the resulting measurements provided evidence that for simple visual fields the relation between luminance and brightness could be described by a power law with exponent. The exponent consistently differed between individuals, but Stevens found that the value of this exponent depended on the size, luminance of the surrounding field and colour of the target stimulus as well as on the observer's state of adaptation.

Room interiors are, however, made up out of more complex visual fields, which make it problematic to apply Steven's power law due to a large amount of mediating spatial factors (Marsden, 1969, cited by Tiller & Veitch, 1995). Besides the scene complexity making the judgements more difficult, it also introduces the problem of specifying visual adaptation. Visual adaptation, as will be explained in section 4.1.5, is a large mediating factor and can strongly influence the results from light experience research.

The psychological nature of brightness means no instrument can measure it. It can only be measured by using judgements from human observers. This is called psychometrics; this approach systematically measures a perceptual aspect, by combining objective measurements (e.g. luminance) with subjective measurements (e.g. perceived brightness) and potentially subjective preferences (e.g. perceived pleasantness) and by linking these measurements by means of various models and algorithms. An example of this is the Image Quality Circle as shown in Figure 3. This method is generally accepted in the social science community, because measuring perceptual experience with instruments ignores the ultimate end-user, the human customer (Engeldrum, 2000). Hence the current study will use psychometric methods.

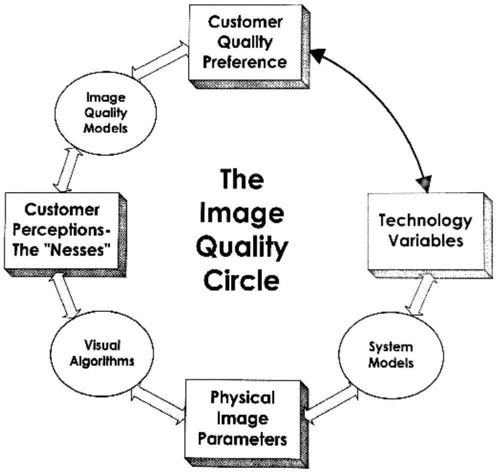


Figure 3: The complete Image Quality Circle showing all the connecting links (Engeldrum, 2000)

4.1.4 Mediating Factors of Brightness

Stevens (1961, cited by Boyce, 2003) discovered that his power law linking brightness and luminance was influenced by the size of the stimulus, the surrounding luminance, and the colour of the light. For a simple disk-shaped light stimulus, increasing background or surround luminance lowers the brightness perception of the stimulus (Boyce, 2003). This is mostly due to the decreased contrast between stimulus and background; however the perception of contrast also depends on visual adaptation (Shapley & Enroth-Cugell, 1984). Adaptation is further discussed in section 4.1.5.

The above mentioned factors are only a few that mediate perceived brightness of a luminaire. Literature indicates plenty of other factors that also affect brightness perception of light. Akashi and others (2000) found that the visibility of self-luminous elements can increase the overall brightness impression of a scene. Also the spectrum of the light can have an effect on the perception of brightness. Our eyes vary in spectral sensitivity, which means that some colours of light appear brighter. In addition, the light spectrum interacts with objects in the field of view. When objects are illuminated by light with a spectrum that makes their colours more saturated, they typically look brighter (Boyce, 2003). Also a higher Correlated Colour Temperature (CCT) is perceived brighter than a lower CCT. The Colour Rendering Index (CRI) is a measure of the ability of a light source to reveal the colours or the surfaces it illuminates in comparison with an ideal light source. It was found that, in the presence of coloured objects, scenes lit by lights with a higher CRI are perceived to be brighter (Boyce, 2003).

Brightness is also affected by the spatial context. The way light is distributed in a room influences its brightness perception, but also the lights' location in the field of view is of importance (Loe et al., 1994, cited by Boyce, 2003). Also illuminating the walls often increases brightness of a room as compared to situations with lights of the same luminous intensity that barely illuminate the walls. The use of the term illumination here suggests that surface reflectivity also affects brightness, which is indeed the case (Boyce, 2003).

4.1.5 Visual Adaptation

Besides the previously discussed factors, the observer's state of visual adaptation is also very influential on the perceived brightness. Natural light sources, from bright sunlight to dim starlight, produce an illumination range of more than 8 orders of magnitude. The human visual system is capable of perceiving this large range of illumination, but to achieve this it makes use of adaptation. The visual system is always bounded by a certain sensitivity window, a smaller interval of the total illumination range that may be perceived. This smaller sensitivity window determines the maximum perceivable contrast. The exact width of the sensitivity window, or the maximally perceivable instantaneous contrast range is not yet entirely clear; some claim the range to be between 1.5-2 orders of magnitude (Spillman et al., 1990, cited by Swinkels, Murdoch & Heynderickx, 2009) while others go up to a range of 5 orders of magnitude (Seetzen et al., 2004). The absolute level of the sensitivity window changes with visual adaptation. The effect of light adaptation in practice can be explained by an example. On a clear night one can easily see stars, while on a clear day they are not visible, although they are still there. This is caused by a decrease in contrast between the overall luminance levels of the daylight sky compared to the night sky. The adapted luminance level that made the stars perceivable at night is inadequate to make them perceivable during the day (Fairchild, 2005). In the night situation, the sensitivity window is near maximum sensitivity. If the clear sky would be viewed in this same sensitive state of adaptation, the photoreceptors would be overloaded and the light would be perceived blindingly bright.

Visual adaptation has the important functionality of normalizing the effect of illumination, thereby keeping the retinal response to contrast invariant. When the retinal responses to the contrast of an object are constant despite luminance variation, brightness constancy is achieved (Shapley & Enroth-Cugell, 1984). This means that a white piece of paper will keep appearing white, even if the luminance varies by several orders of magnitude, because the relative brightness remains the same. Visual adaptation therefore helps us to recognize objects despite changing luminance conditions by making use of regularities in the environment such as surface reflectances. Brightness constancy holds well in the middle range of illumination, but the limits of adaptation become apparent at very high or very low levels of illumination, i.e. we perceive that it gets dark at night (Fotios, 2006).

We distinguish light adaptation, dark adaptation and chromatic adaptation, since they are driven by different mechanisms (Fairchild, 2005). Light adaptation occurs when going from a dark to a bright environment; it reacts quickly, such that the eyes are sufficiently adapted after ten seconds. The other way around, dark adaptation from a bright to a dark environment, takes more time. Imagine for example walking into a dark cinema. Often you won’t be able to discriminate anything at first, but slowly objects in the room become visible. The third type of adaptation, chromatic adaptation, is the process of increased or decreased sensitivity for a certain colour, with the purpose to maintain colour constancy and relative colour contrast high in different lighting situations. Like we perceive brightness constancy for objects such as a piece of paper, we also perceive colour constancy: whether the paper is illuminated by daylight (predominantly blue light) or by incandescent light (predominantly yellow light), it still appears white (Fairchild, 2005).

4.2 Virtual Prototyping

The current study is concerned with brightness in virtual scenes. Modelling and simulating these scenes using a computational virtual model is called virtual prototyping. This section discusses the principals and considerations of virtual prototyping for lighting applications. Modelling in this context is the creation of a virtual 3D model, whereas simulation in the current context refers to the calculation of light transport in the scene using various algorithms. After light simulation the resulting image of the scene can be rendered by calculating the right value for each pixel. Compared to real-world scene creation, virtual models offer more flexibility in scene creation while maintaining accuracy. Because flexibility and accuracy are especially important for experience research, virtual models are suggested to complement or even replace traditional evaluation methods in real spaces. This section discusses the choices that need to be made during the creation of the final image. The creation process is referred to as the imaging pipeline and can be generalized by breaking it down into four separate blocks: Scene, Rendering, Reproduction and Viewing, as shown in Figure 4.

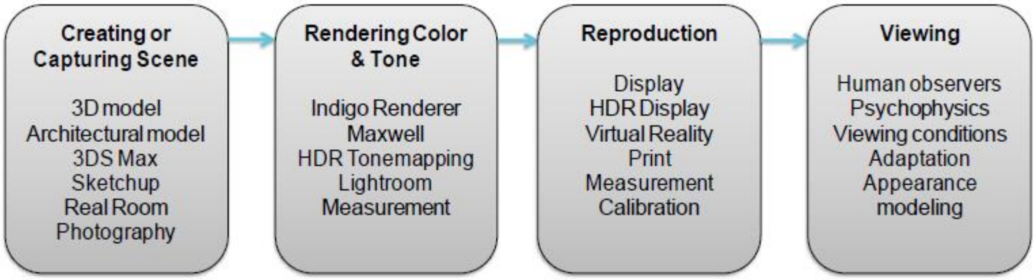


Figure 4: Imaging Pipeline (Murdoch et al, 2013)

4.2.1 Scene Creation

While a real-world scene is created by physical placement of the geometry and luminaires, the virtual creation of the lighting situation is done by using models of the geometry, materials and luminaires. The reusability of the models makes virtual scene creation much more flexible than physical decoration. IES files contain the photometric information (e.g. the beam profile) of a specific luminaire. After importing the IES profile to the luminaire model, the luminaire illuminates the virtual scene just as in reality. The current study is concerned with further improving computer visualizations. Modelling with 3DS Max will be used as scene creation method.

4.2.2 Rendering Colour and Tone

The virtual scene is captured by a virtual camera. The V-Ray renderer can mimic a physical camera. There are several options to control the camera's sensitivity to light, i.e. the diaphragm, exposure time and ISO value, where the latter can shift the sensitivity window to higher or lower levels of light intensity much like human visual adaptation. The camera should be set such that the scene appears as if it would appear in reality with fully adapted vision.

After modelling and configuring the virtual camera the scene is ready to be rendered. *Rendering* has two meanings:

1) In computer graphics, it refers to the computation of pixel values based on light transport simulation of the scene geometry, materials, and light sources.
2) In photography, rendering means the translation of "scene intensity" to "image intensity" – also known as tone mapping.

In this study we use both meanings: the illumination values of the pixels are calculated and then scene intensity is translated to image intensity, that is, to a range that is displayable for standard displays. For research purposes, it is desirable to create renderings that are as perceptually-predictive as possible. The current study focuses on rendering of light, so in this context "perceptually-predictive" refers to visualizations of light in a space that yield a visual impression of the scene that is similar to the real situation.

When an image is being graphically rendered, it means that for each pixel the colour value is calculated by combining the spatial, textural and lighting information of the entire scene (Lambooij, 2013). Different graphical renderers use different algorithms, and so have different methods of calculating the pixel values. Therefore, the quality of the rendering depends heavily on the renderer as well as the rendering settings. A distinction can be made between biased and unbiased renderers. The unbiased renderers use rendering algorithms that do not introduce systematic errors in the approximation. The biased renderers take calculation shortcuts, which introduce systematic errors to some degree (Damkat, 2013). Unbiased renderers are not flawless however, and given enough samples, biased renderers can create images that deviate insignificantly from unbiased renderers. Biased renderers have the large advantage that they are much faster, and some of them, such as V-Ray, even outperform other methods given the right setup of parameters. V-Ray is able to replace unbiased renderers for scenes that are relatively simple in terms of light rendering, and so for these cases, V-Ray's results are without significant perceptual differences in light characteristics (Lambooij, 2013).

The Tone Reproduction Problem

Capturing a scene is done by rendering the image through the view of the virtually set up camera. The resulting image then has pixel values corresponding to the camera settings as if the picture was taken by a camera in a real world scene, although the process is very different (luminance values are calculated instead of simply captured). Like in a real world scene, the light levels in a rendering might span eight or more orders of magnitude; more contrast than the human eye can handle at once. To prevent loss of information, the scene can be captured as a High Dynamic Range Image (HDRI). HDR imaging is a technique that can accurately capture scenes containing dynamic ranges of more than seven decades of magnitude by combining multiple exposures (Hoefflinger, 2007).

Displaying HDR images, however, is problematic, since regular display devices can reproduce only ranges of light that span a maximum of two to three orders of absolute dynamic range (Reinhard, Stark, Shirley & Ferwerda, 2002; Teoh, 2014). The contrast that can be achieved in these Low Dynamic Range (LDR) displays is limited, so the luminance values of the HDR image need to be compressed and mapped somehow to display the image in a proper way. Tone-mapping operators (TMO) are algorithms that convert HDR images to LDR images as authentically as possible (see Figure 5). They reduce the contrast of the original scene to a displayable range while preserving the image details and colour appearance.

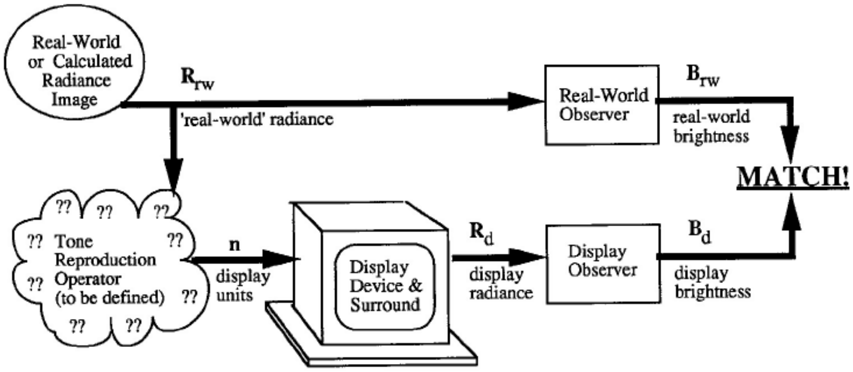


Figure 5: The Tone Reproduction Problem (Tumblin & Rushmeier, 1991)

The appearance of the resulting image, including its perceived brightness, strongly depends on the TMO-algorithm and the chosen parameters, but there are still some structural problems. Firstly, visual artefacts like visible clamping or ringing are introduced. The second problem is that matching contrast and/or brightness are not the only factors that influence visual appearance. Scene content, image medium, and viewing conditions should be considered as well (Fairchild, 1998, cited by Reinhard et al., 2002).

There is a wide range of TMO algorithms to choose from. The most important quality of a good TMO in our application is perceptual realism, however this cannot be determined purely objectively, and therefore should also be judged by preference. Villa and Labayrade (2012) performed psycho-visual tests to find the most preferred TMO out of six candidates. In the first part, participants observed six printed images - each corresponding to a different TMO - and had to rate how well each image matched the real world scene. In the second part, images were displayed in virtual reality, without real-world reference. Drago's and Reinhard's 2002 algorithms were found to obtain LDR images with the best representation of reality. Based on the consistent performance in realism and preference in the literature, Reinhard '02 can be considered as a good TMO (Čadík, Wimmer, Neumann & Artusi, 2008; Yoshida, Blanz, Myszkowski & Seidel, 2005; Ledda, Chalmers, Troscianko & Seetzen, 2005).

The Key Value

The Reinhard algorithm is based on photographic reproduction. In photography, the key of a scene gives an indication of its subjective lightness: "a white-painted room would be high-key, and a dim stable would be low-key" (Reinhard et al., 2002, p. 2). Low-key scenes need a relatively lower exposure value to look right, while high-key scenes need a relatively higher exposure value, compared to a typical "auto exposure". The Reinhard '02 TMO is based on this relation between key and relative exposure value. The algorithm has a parameter called the *key value*; the TMO allows different exposures across the image dependent on the key value. There is both a local and global version of the algorithm, and the global algorithm is well-suited when subjectively satisfactory and seemingly artifact-free images are the desired goal (Reinhard et al., 2002).

The key value essentially determines the overall brightness of the tone-mapped image. For the purpose of the current study, it is interesting to examine the role of the key value in brightness perception of the renderers. The key value has already been used in brightness matching experiments between real and virtual lighting settings (Murdoch et al., 2013; Tang, Chen, Chen & Xu, 2014). Participants were instructed to tune the rendering until it best matches the real scene, or adjust the lighting in the real room until it looks most like the rendering. The difficulty of this method is that brightness is a perceptual attribute that has no absolute metric to tune with. Key value-tuning offers a way to change the brightness of an image, while maintaining the right contrast, colours, and details. Key value tuning seems like a suitable method for measuring virtual brightness and the effect of synthetic glare (further discussed in section 4.3) on virtual brightness.

4.2.3 Reproduction

Image reproduction can be done on several media such as displays, prints or virtual reality goggles. The current study focuses on brightness reproduction on displays, as this is a generally accepted medium. Typical displays have a low dynamic range (LDR). Since the maximum achievable contrast on a LDR display is much smaller than the achievable contrast in reality, the display constitutes a limiting factor in virtual brightness reproduction. This is why tone mapping of HDR images is needed before they can be displayed on a typical display (See 4.2.2).

Increasing the peak luminance of a display may help reproduce higher brightness but at the expense of higher black levels because of limited dynamic range. Such displays are called High Brightness (HB) displays, but these displays are not very common. Since the luminance of HB displays is larger, the regular tone mapping operators do not give the right result, but linear mapping of HDR images generated more accurate results. However, in practice it was found that the dynamic range of the HB display was still too limited to display all images artifact-free (Murdoch et al., 2013), meaning better tone mapping is one improvement that can be made to achieve more accurate reproductions. Moreover, Murdoch and colleagues (2013) recognize that HDR display technology has the potential to improve visualizations of lighting, but conclude from their experiments that the current HB (LDR) technology not yet adds anything to the visual conveyance of lighting. Considering the overall poor performance of the HB displays (Murdoch et al., 2013), the normal LDR display currently seems to be the best option.

Furthermore it was found that a TV-sized display achieved the best results in terms of immersiveness and image quality, as compared to a laptop or beamer-sized display (Murdoch et al., 2013).

4.2.4 Brightness in Virtual Scenes

Previous research has shown that especially the reproduction of brightness of rendered light is hard (Murdoch et al., 2013). The brightness attribute is affected by a lot of parts in the imaging pipeline: the (physical) surround luminance, the choice of TMO, the TMO settings, the choice of renderer, the rendering settings, the display luminance, the field-of-view, the ambient illumination in the display lab, and the list goes on. Murdoch et al. (2013) investigated the comparison of renderings with real rooms on different attributes, and found that especially brightness was assessed differently between the renderings and the real scenes. For a lot of the dimmed scenes, the brightness was highly overestimated. This brightness overestimation of dim renderings seems to happen structurally, and regardless of the renderer (Lambooij, 2013). Perhaps viewers discount the medium, interpreting a dim image as an underexposed picture of a normal room rather than a properly-exposed picture of a dim room. Or, since the effect also happens with photographs (Lambooij, 2013), it could be an environmental effect: for the real room the surround luminance changes accordingly, but not for the image.

Key Value Tuning

Murdoch et al. (2013) performed brightness matching tasks using key value tuning. During a tuning task the observer chooses an image that best matches with a reference. In essence, the observer tunes along the key value dimension through a set of tone mapped images that are based on a single HDR image. It was found that matching the real room to the rendering had different results than matching vice versa. Furthermore, they discovered that the key parameter was not visually linear: steps in low values of the key are perceived as bigger effects than equally sized steps in high values of the key. Also, the preferred key value seemed to correlate with the ratio of total perceived brightness of the image and display over the surround luminance. Participants tended to keep this ratio constant (Murdoch et al., 2013). These findings also stressed the importance of the display luminance.

These issues with key value tuning can be largely overcome. Most of these effects can be controlled for by keeping values, such as the display brightness/surround luminance ratio or the natural key of the stimulus, constant throughout the experiments. Furthermore, recent research gained insight into the relation between key value and perceived brightness. Tang and colleagues (2014) performed brightness matching experiments with key value tuning to find this relation. Based on their results, they constructed a predictive model for perceived brightness as a function of key value, plotted in Figure 6. By using their formula, a range of key values can be calculated that describes a perceptually linear trend in overall brightness, which would make it a suitable scale for measuring brightness. The current study conducts key value tuning and will make use of Tang et al.'s (2014) model to map a range of key values to a scale that is perceptually linear.

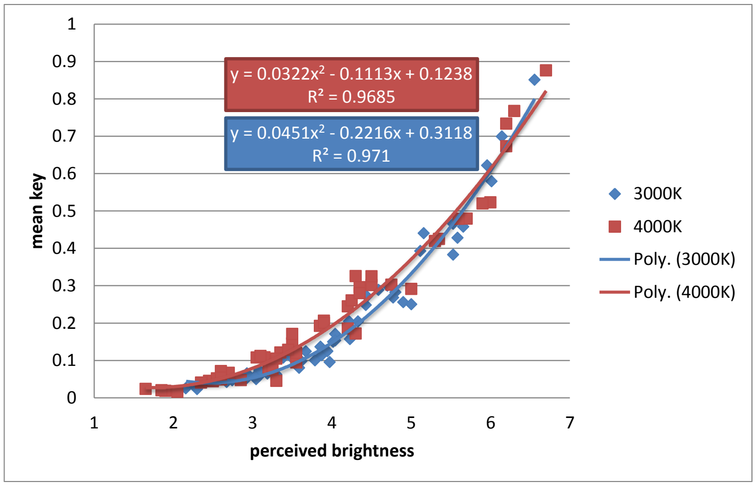


Figure 6: Relation between Reinhard's key parameter and perceived brightness of an image (Tang, Chen, Chen & Xu, 2014)

4.2.5 Viewing

The final block in the imaging pipeline concerns the human observer and the way the image is viewed. Answering the research question involves brightness comparisons between real-world luminaires and rendered luminaires. Since viewing and evaluation are so closely coupled, they are both discussed in this section.

Evaluation

Evaluating two stimuli can be done by either a joint evaluation or a separate evaluation (Hsee, Loewenstein, Blount & Bazerman, 1999). In a separate evaluation, each stimulus is evaluated individually on an absolute measurement scale. Joint evaluations can be sub-divided into temporal and spatial collocations. Joint evaluations are relative judgments; the stimulus can be compared to a reference stimulus, presented in a spatially adjacent field, or sequentially at the same spatial location. Joint evaluations are suitable for brightness comparisons, and so, are used in the current study (real-world vs virtual rendering). Responses can then be measured by forced choice discrimination (e.g. "Is the rendering brighter?") or adjustment (e.g. "Tune the rendering until its brightness matches the reference stimulus").

Joint side-by-side evaluations have been used in previous brightness studies (Takahashi, Yaguchi & Shioiri, 1999; Fotios & Cheal, 2007; Vidovszky-Nemeth & Schanda, 2012), but in different variations. In 1996, Braun, Fairchild and Alessi compared different side-by-side viewing techniques for the comparison between printed images and CRT images. Five viewing techniques were examined: memory, successive-binocular, simultaneous binocular, simultaneous-haploscopic, and successive-Ganzfeld-haploscopic viewing. Results of their study showed that the different techniques yielded different results, which obviously illustrates the importance of their work. So the viewing method should be carefully selected after considering all pros and cons of the different possibilities.

Binocular viewing

When both the original and reproduction are viewed with both eyes, the viewing technique is binocular. Binocular viewing can be done either simultaneously or successively. In *simultaneous binocular viewing* the stimuli are presented side-by-side and no memorization of the original is required. A downside of this method is that it does not control for visual adaptation. When the viewing conditions differ, the observer adapts to either one of the conditions and this makes accurate comparisons impossible.

This visual adaptation can be controlled for by successive viewing of the two conditions, allowing stimulus-specific adaptation and some adaptation time in-between. The disadvantage of successive binocular viewing is that it requires memorization. Braun et al. (1996) define two variants of this viewing type. In *memory viewing*, the observer first views the original after a fixed adaptation period and then views the reproduction after another adaptation period. In the other type, *successive binocular viewing*, the only difference is that observers are allowed to look back preceded by the adaptation period. Compared to successive binocular viewing, memory viewing was preferred by the observers and had slightly more sensitive results (Braun et al., 1996).

Haploscopic viewing

Haploscopic viewing methods present each eye with a different stimulus. If we assume that the eyes adapt separately, *simultaneous haploscopic viewing* would allow side-by-side viewing in a fully adapted state for both conditions. Unfortunately, studies have shown that in practice it becomes difficult to make brightness judgments in haploscopic viewing with large luminance differences between both eyes (Takahashi et al., 1999). Also the assumption that each eye independently adapts may not be entirely true which means the actual state of adaptation is unknown (Braun et al., 1996). In *successive-Ganzfeld-haploscopic* viewing, the stimuli are not presented to both eyes simultaneously, but one eye views one condition, while a neutral diffuse filter covers the other eye, and vice versa. The purpose of this method is to avoid the confusion of simultaneous viewing, without losing the state of adaptation in one of the eyes. Nonetheless also for successive-haploscopic viewing the state of adaptation is not fully controllable.

Conclusion Viewing Method

Table 1 shows a summary of the advantages of each viewing technique. It clearly demonstrates that there is no generally ideal viewing method. It visualizes the tradeoff that has to be made between full-adaptation and no memorization. Comparisons between real and virtual scenes are easier when going back and forth between them without the need to memorize the reference; however, this doesn’t guarantee adaptation. When two stimuli are presented simultaneously, the participant can be in only one state of adaptation at a time. This means that for the slightest difference in luminance, at least one stimulus will be judged in brightness in a mismatching state of adaptation, which introduces a bias in the judgments.

	Natural	Quick	Adapted	No memorization
Memory	✓		✓	
Successive binocular	✓		✓	
Simultaneous binocular	✓	✓		✓
Simultaneous haploscopic		✓	?	✓
Successive haploscopic		✓	?	✓

Table 1: Overview of advantages per viewing method (Braun et al., 1996)

Since adaptation is one of the most important mediating factors of brightness perception, it should have priority to control for it. Simultaneous binocular viewing and haploscopic viewing do therefore not qualify as appropriate viewing methods for the current study. This leaves memory viewing and successive binocular viewing as useful methods to choose from. Though memorization will introduce some noise in the results, the final use of visualizations will often be isolated from the reference scene, and will therefore be viewed in an independently adapted state. Thus, memory viewing and successive binocular viewing are both appropriate methods in the context of the current study.

4.3 Glare

Bright sources of light often appear to be surrounded by a halo of light, sometimes accompanied by radial streaks of light and a veiling luminance over the retinal image. This perceptual effect is called glare and is caused by intra-ocular scattering of light. It is most visible when there is a significant ratio in luminance between the glare source and the area around it, typically lowering detail visibility in that area. For this reason, glare occurs a lot at night, when the environment is much darker than the luminaires. In some situations glare causes stress and fatigue and makes people want to look away from the light source. The International Commission on Illumination (CIE) distinguishes two types of glare (CIE, 1987):

- Disability Glare: “glare that impairs the vision of objects without necessarily causing discomfort”
- Discomfort Glare: “glare that causes discomfort without necessarily impairing the vision of objects”

In this research we are mainly interested in the application of glare as a visual effect in visualizations; hence we will use the definition of Disability Glare. Disability Glare is caused by relatively small but intense light sources. The light is scattered in the eye, which results in a perceived veil of light around the source. Objects near the source become more difficult to perceive because of a lowered contrast. Discomfort Glare is of a different nature, namely annoyance (Vos, 2003). Applying virtual glare is not likely to significantly increase the actual luminance of the visualization because that depends on the display; therefore it can be assumed to not cause any discomfort.

The effect of glare on visual performance depends on several parameters: illuminance at the eye, angle of the glare source, luminance, size, and spectral power distribution of the glare source and the duration of exposure. Glare also increases with the observer’s age and visual health problems (Mace, Garvey, Porter, Schwab & Adrian, 2001).

4.3.1 Synthetic glare

Due to the limited peak luminance and contrast range, images including light sources displayed on standard display devices cannot produce the same magnitude of glare in the eye as the real light sources would do. However, glare effects can be mimicked in images by changing the luminance levels around a light source. The light source will then seem to have a glow around it, which resembles the effect of intraocular scattering/diffraction that occurs when looking directly into a bright light. This is called synthetic glare. Artists have long been using similar techniques of painting halos around objects to make them appear luminous on paper (Zavagno & Caputo, 2001). This suggests brightness perception is divergent from the physical reality and easily influenced by illusions, which is supported by the earlier discussed checker-box illusion (Figure 2). During the discussion of lightness constancy in section 4.1.2, the checker-box illusion demonstrated that top-down processes discounted a non-existent illuminant, thereby interpreting light grey as illuminated dark grey. It is hypothesized that brightness perception may also be cognitively influenced when synthetic glare is perceived.

The hypothesis that synthetic glare increases brightness perception in visualizations can also be substantiated from a bottom-up perspective. Keil (2007) further explains the neurological theory behind the perceptual effect of luminance gradients, which form an essential part of synthetic glare's shape (see Figure 7). It is theorized that luminance gradients are a perceptual feature of luminosity, so an object that has no luminance gradients around the edges will not be perceived as self-luminous. Adding luminous gradients to the image, for example in the form of a point spread function, might thus be able to simulate luminosity and enhance virtual brightness perception in renderings. Zavagno and Caputo (2005) found that the left stimulus in Figure 7 was perceived as self-luminous. Zavagno and Caputo (2005) and Keil (2007) support the existence of separate perceptual pathways for self-luminosity perception and for surface-colour perception. From a neurological point of view, an object with sharp edges usually produces a surface representation and is therefore perceived on the lightness scale, and not in terms of brightness.

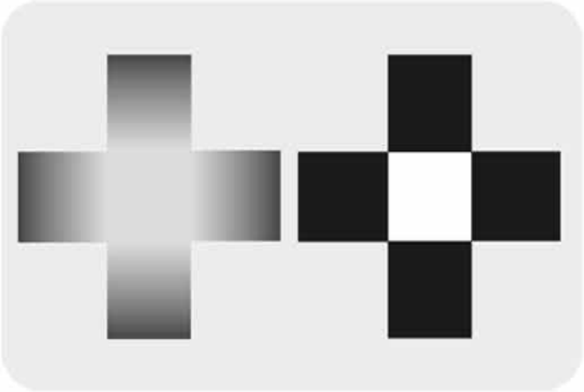


Figure 7: The squares on the left have luminance grading towards the center, the squares on the right have no

These findings suggest that halos in images can indeed simulate glare and increase the perceived dynamic range in a rendering. The presence of luminance gradients seems to trigger the glare illusion.

The shape of the synthetic glare has been investigated as well. Yoshida et al. (2008) employed two different glare models (i.e., a simple Gaussian convolution model and Spencer et al.'s (1995) physically based model). They found that glare correlated positively with perceived luminance: synthetic glare was able to increase perceived brightness by 20-35%. Comparison between the two glare functions suggested that the differences found were due to stimulus size rather than to differences in the functions; hence functions that attempt to accurately simulate the optics in the human eye are not more effective in boosting perceived luminance than simple Gaussian convolution approaches (Yoshida et al, 2008). Thus for the application of synthetic glare, the current study will use a point spread function that is based on the CIE Glare Formula of Vos and Van den Berg (1999). CIE has studied Disability Glare and developed this formula to predict glare based on intensity of the light source, glare angle, age, and eye pigmentation (Vos & Van den Berg, 1999).

4.4 Summary

The human visual system is not accurate in estimating absolute levels of luminance, but detects spatial and temporal changes relatively accurately. The window of sensitivity in luminance of the human visual system is limited but adjustable to brighter or darker settings. Absolute brightness perception is normalized to a great extent by these adjustments, and so light/dark adaptation can strongly influence perceived brightness.

Besides observer-dependent factors such as state of adaptation or age, perceived brightness also depends on the physical characteristics of the light sources and their environment. Light characteristics that influence brightness are the light source's type, luminance, spatial distribution, spectral distribution/colour, CRI, and CCT. Environmental characteristics that influence brightness can be background or surround luminance, location of the light in the visual field, illumination of the walls and reflectivity of the scene materials.

Although tone mapping operators are able to compress HDR images to a displayable range without losing details, the perceived brightness is still an attribute that is hard to replicate virtually and is often inconsistently judged by observers. Firstly, brightness of visualizations is constrained by the low dynamic range of displays and their limited peak luminance. Furthermore, real-world surround luminance does often not correspond with the virtual scene. The mismatch of surround luminance can also cause a mismatch of adaptation with the virtual scene.

Adaptation is a valuable ability of the human visual system as it allows us to perceive relative brightness under different lighting situations. The maximum perceivable dynamic range is estimated to be between 2 and 5 orders of magnitude: not much larger than the dynamic range of a standard LDR display. This theoretical match gives reason to believe that a more accurate reproduction of brightness is achievable.

Conclusively it was discussed that Disability Glare can be mimicked in visualizations by changing the luminance levels around a light source, which is known as synthetic glare. Synthetic glare has been shown to be able to increase brightness visualizations (Yoshida et al., 2008). How synthetic glare affects brightness visualizations is not yet fully understood, but there are arguments for top-down processing effects as well as for bottom-up processing effects. Furthermore, the impact of synthetic glare on brightness in visualizations has not yet been accurately quantified in a robust way. Further research on this topic is valuable for the domain of virtual prototyping.

5 Aim of this report

To improve virtual prototyping of lighting as a research and marketing tool, the visualizations should convey the same visual impression as in the real-world situation. Recent research concluded that computer generated lighting visualizations are still in the process of converging towards their maximal predictive value (Engelke et al., 2013). Brightness is a perceptual attribute considered in virtual prototyping for which an accurate match with the physical corresponding space has not yet been achieved consistently (Murdoch et al., 2013).

Considering the findings that synthetic glare can increase brightness in visualizations, it is interesting for the field of virtual prototyping to investigate the impact of synthetic glare on brightness perception further. The main research question of this study is therefore: "*How does synthetic glare affect brightness perception of a rendering of a luminaire?*" We hypothesize that adding synthetic glare to renderings increases the perceived brightness. Secondly, it is hypothesized that synthetic glare can improve the perceived brightness match between the rendering and reality.

In the current study, synthetic glare is applied to renderings of luminaires. By means of a psychophysical study it is investigated how the addition of synthetic glare influences brightness of the visualizations. Since brightness is not absolutely measurable, two relative scales (i.e., an indirect and a direct one) are used to quantify brightness. The indirect scale relates to a brightness-matching task, conducted in Experiment 1. The scale consists of a series of renderings that only vary along the key value dimension, which in turn relates to brightness (see section 4.2.2: The key value). The brightness-matching is conducted by means of a tuning task within a controlled lab experiment with side-by-side comparisons between real-world luminaires and renderings of these luminaires. The participant is asked to tune the key value of the renderings until the brightness of the virtual luminaires matches the brightness of the real-world luminaires. The direct scale refers to evaluative questions that are answered on a 7-point Likert scale, as used in Experiment 2. Here participants are asked to directly rate aspects related to the brightness of the luminaires and of the scene, not only in terms of absolute appearance, but also in terms of the match between reality and the renderings.

In short, this study quantifies the impact of synthetic glare on perceived brightness in visualizations by direct and indirect subjective measurements. Taking the previous findings of Murdoch et al. (2013) as a foundation, the current study builds upon it by investigating whether synthetic glare can improve the accuracy of brightness virtualization.

6 Experiment 1

6.1 Introduction

The purpose of Experiment 1 (E1) is to gain insight on the effect that synthetic glare has on the brightness of a rendering with respect to a real-world reference. In E1 this effect is quantified by tuning the key value of the rendering, since the key value determines the overall brightness of the rendering. Therefore, a brightness-matching task with key-value tuning, in which real-world luminaires are compared to renderings of luminaires in terms of brightness, is conducted.

In addition, such a tuning experiment may contribute to our understanding of the relation between Reinhard '02's key value and the brightness of visualizations. In this chapter the design and procedure of the experiment are described.

6.2 Method

6.2.1 Design

E1 was a brightness-matching experiment in which the participants' task was to tune the renderings of the luminaires until their brightness matched the brightness of the real-world luminaires. The design of the experiment is a full-factorial within-subjects design with luminaire intensity (3 levels), glare (3 levels), and direction (3 levels) as independent variables, resulting in 27 conditions. An overview of all conditions can be found in Appendix 11.2.1. Image intensity was the dependent variable, and was measured with the tuned key value of the Reinhard '02 TMO. The levels of the independent variables are clarified in more detail in section 0.

In summary, the procedure was as follows. Observers viewed the real-world luminaires for a fixed time interval before they viewed the virtual stimuli. During the latter they tuned the key value with the task to match the brightness from memory. Each condition represented a unique tuning task, so there were 27 unique tuning tasks in total. All participants performed all 27 tunings in a randomized presentation order. The experimental procedure is discussed in full detail in section 6.2.6.

6.2.2 Experimental Setup

The brightness-matching task was done with a side-by-side presentation of the stimuli, with the real-world scene on the left, and the virtual scene on the right, shown on a display. The scene consisted of two luminaires aimed in different directions, positioned in front of a wall. Both scenes had the same content, with the virtual scene being a visualization of the real-world scene as exactly as possible.

Care was taken that the observers viewed both scenes under the same perspective. Two same-sized windows were placed in front of the scenes (i.e., the real scene and the display). They were made by cutting 105.5 x 66 cm rectangles out of two 140 x 100 cm foam boards. One window exactly enclosed the size of the display. At the real scene, the other window frame simplified the scene, i.e. the visual stimulus, because it hided irrelevant details from sight. A chin rest was used to fix the observers' field of view and viewing distance. The distance between the chin rest and each window was 190 cm. The window frames were angled towards the observer on both sides, such that the observer had a perpendicular view on the left window as well as the right window from the chinrest, as indicated by the dashed lines in Figure 8. The center of each windows of the stimulus was visible under a horizontal gaze angle below 31 degrees, which can be considered as comfortable, because it limits eye exertion to less than 20% (Menozzi, v. Buol, Krueger & Miege, 1994). For the visualization, the virtual camera in the 3D model was placed at the location where the chin rest was in reality, such that the perspective in the visualization matched the perspective of the real-world scene as viewed from the chin rest.

The setup was built in the Display Lab at Philips Research in Eindhoven, The Netherlands. The lab had white walls with black carpet and no windows. At one side of the lab, two tables were positioned next to each other, on which the stimuli were presented. Two luminaires were positioned on the left table and a 49 inch display (Philips 49PUS7809 4K/Ultra HD LED, driven at a resolution of 1920x1080) was positioned on the right table. To control the surround luminance it was chosen to build a "viewing box". It was constructed around the observer rather than around the stimulus, because this kept the appearance of the scenes more natural, which according to previous research was found to be very important in order to obtain reliable judgments from the participants (Murdoch et al., 2013). The viewing box was made out of black foam board of 140 x 100 cm, just like the window frames. The two window frames were merged (i.e. taped together) into a 280 x 100 cm board and placed at an angle of 142 degrees (see Figure 8). Side panels were taped to the outer edges of the window frame boards and two foam boards were placed on top of the structure as roof panels. The center of Figure 8 shows a "separator panel", existing of another foam board vertically attached in between the window frames. The observer sat at a table with chinrest and keyboard, directly in front of the separator panel, which separated the real scene from the visualization. The resulting setup is shown in Figure 9 to Figure 13.

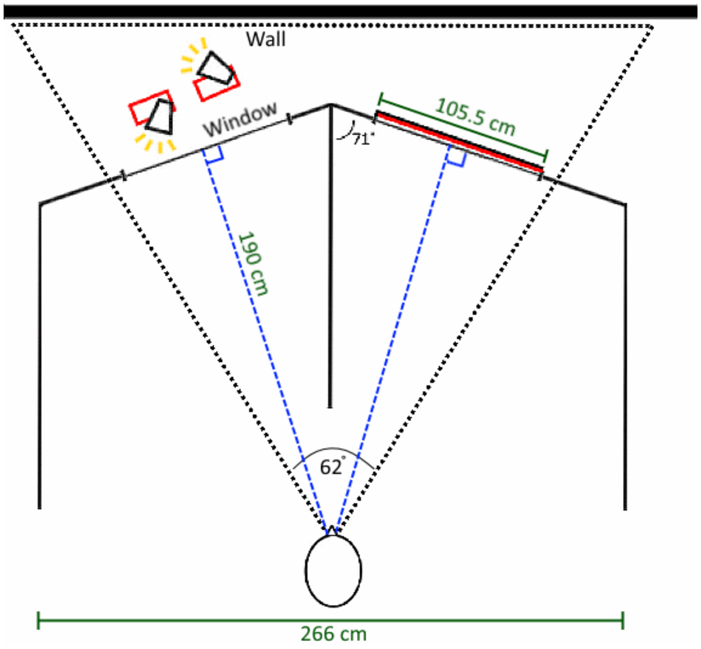


Figure 8: Schematic top view of the experimental setup. The dotted lines show that, during resting position of the eyes, both windows are within the field of near-peripheral vision, where colour perception is adequate (Abramov, Gordon, & Chan, 1991). The dashed lines show the perpendicular angle with the windows.

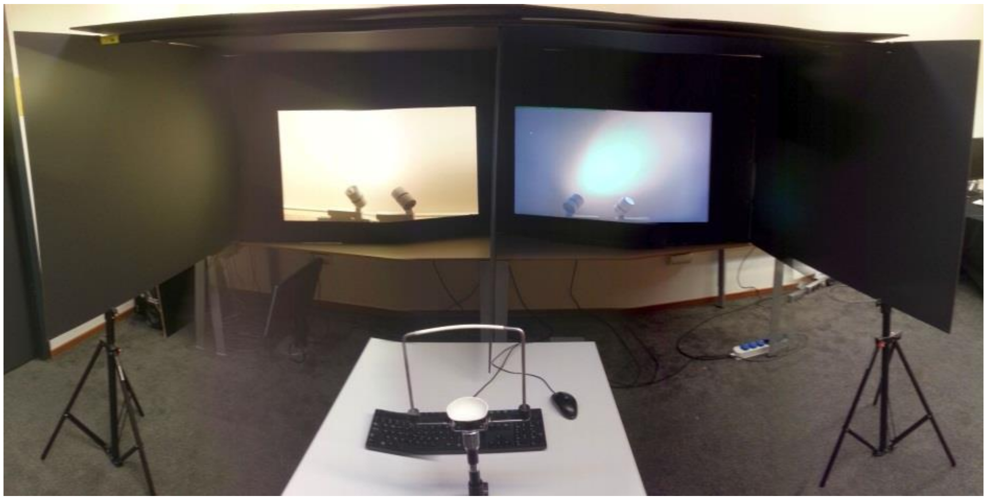


Figure 9: Panorama photograph of the viewing box; on the left the real-world scene; on the right the display showing the virtual scene

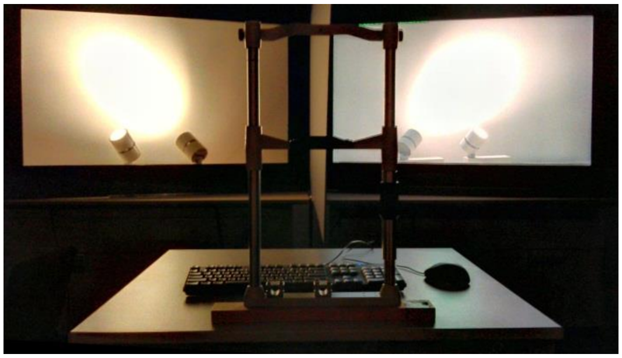


Figure 10: Position of the chin rest in front of the stimuli

Figure 11: Photograph of a participant observing the luminaires

Figure 12: Photograph of a participant observing the real-world luminaires

Figure 13: Photograph of a participant observing the reproduction of the luminaires on the display

6.2.3 Stimuli

The Real-World Scene

The scene contained two luminaires, each aimed in a different direction (direction of the luminaire being one of the independent variables, explained later in this section), as shown in Figure 14. Using two luminaires instead of one allowed us to change the direction of the light beam without having to change the aiming angle of the luminaire. Instead of having to manually adjust the luminaire's direction, each luminaire could be switched on individually to change the direction of the light beam.

The type of luminaire used for the experiment was the Philips StyliD Compact 930[1], narrow beam, containing 25 LEDs (See Figure 15), which is typically used in retail applications. The luminaire had a CCT of 3000K, being a common colour temperature for indoor lighting. The lamp had a Colour Rendering Index (CRI)[2] of 90, which is considered high (EFI, 2014). The luminaires were individually dimmable by means of a DALI control connected via Ethernet to the control computer.

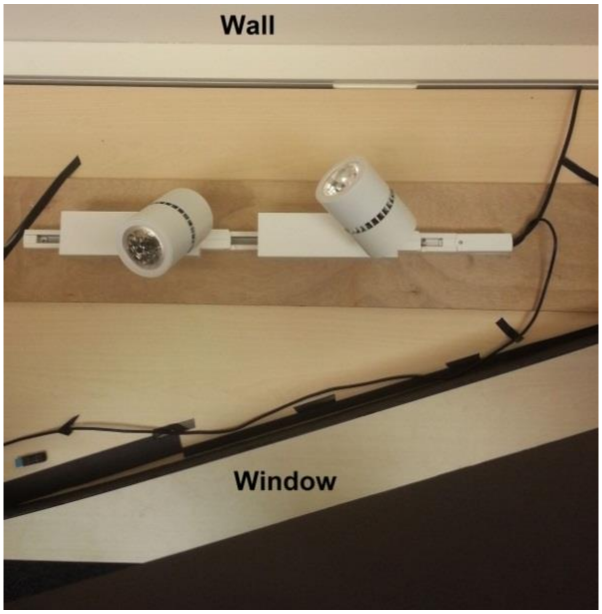


Figure 14: Top view of the real-world scene

Figure 15: Photograph of the real-world luminaire

---

[1] http://www.ecat.lighting.philips.com/l/new-products/new-luminaires/stylid-compact-track-and-surface-mounted/910500454485_eu/

[2] The CRI describes the ability of a light source to accurately reveal the colourfulness of reflective object colours. CRI closer to 100 are closer to displaying colours fully authentic (Lighting Research Center, 2014), therefore colours viewed under a high CRI are perceived as being more natural.

The Virtual Scene

The virtual scene was modeled in 3DS Max, using a preexisting model of the Light Lab that already contained the right models of the geometry, including measurements, reflectance properties, materials and bump maps. Besides, two luminaire models were imported, and foam board models were manually created. Subsequently, the real-world scene as previously discussed was virtually replicated, but mirrored along the vertical axis, such that the stimuli fell at the same location on the retina (albeit in the other eye). This is important since location in the visual field can be a mediating factor of brightness (Loe et al., 1994, cited by Boyce, 2003). A top view of the resulting geometry is shown in Figure 16. Figure 17 depicts the scene seen from the V-Ray Physical Camera's point of view. The camera was placed at the same height (vertically) and distances (190 cm) from the window as the participants' eyes were from the display. The V-Ray camera was set to a field-of-view of 31 degrees; the angle of view for the TV at a distance of 190 cm. Furthermore the camera was set to an f-number of 8.0 and a shutter speed of 50^{-1} s.

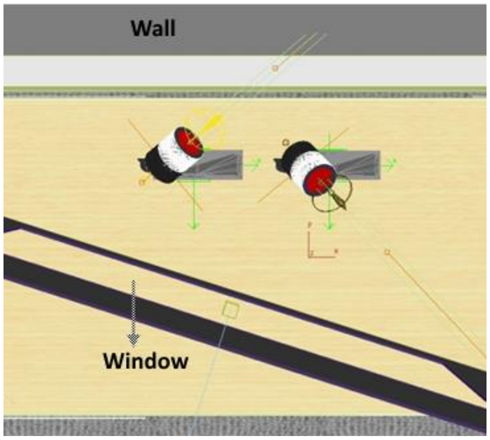


Figure 16: Top view of the virtual scene

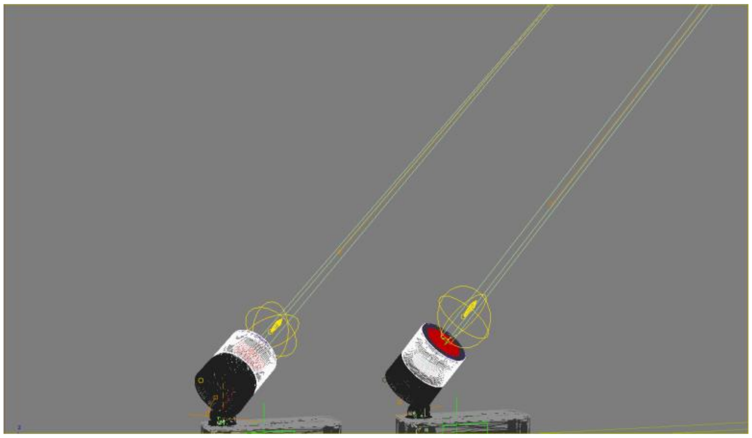


Figure 17: Wireframe model as seen from the camera's viewpoint

The virtual luminaire was a StyliD Compact 930, accurately modeled in 3DS Max (See Figure 18). To simulate the beam profile, the IES file[3] for the StyliD Compact 930 (narrow beam, 25 LEDs) was imported into the virtual scene via a so-called IES light. To accurately replicate the light effect of the luminaire, a modeling structure as depicted in Figure 19 was used. The so-called IES light was placed at location (4), because the IES profile already accounted for the luminaire's housing and optics. Hence, it was placed just outside the luminaire's housing, thereby preventing disruptions of the housing in the IES profile (e.g., shadows that are not supposed to appear). Consequentially, the inside of the luminaire model was still unilluminated. Therefore, to give the impression that the luminaire was the light source, a dummy light source was placed at location (3) in Figure 19; i.e., at the location of the LEDs in the actual physical model. To prevent the dummy light from entering the scene and disrupting the IES, a one-way glass was placed at (2) such that the light could not come outside of the luminaire, while at the same time it maintained its visibility to the camera. The CCT of both the IES and the dummy light was set to 3000K, just as for the real luminaire.

---

[3] The Illuminating Engineering Society of North America (IES) has defined IES files: files that contain formatted photometric data that can be used in lighting design software such as Dialux or 3DS Max.

In order to create the visualizations (i.e. the virtual stimuli) from the virtual model, several processing steps were needed such as rendering an HDR image, intensity calibration, tone mapping and synthetic glare application. V-Ray was used to render the scene as captured by the camera. It was set to Adaptive DMC image sampling with the maximum subdivision set to 16. An area-antialiasing filter and sub-pixel mapping was used. The scene was rendered in a resolution of 1920x1080 pixels and saved in HDR format. In other words, the output contained pixel values representing the full dynamic range of the virtual scene.

Figure 18: Rendering of the virtual luminaire

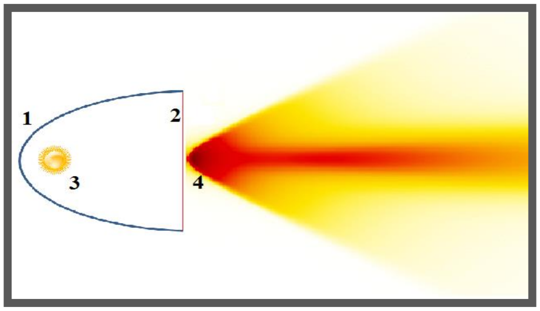


Figure 19: Schematic side view of the virtual luminaire. 1 = The reflector part of the luminaire, 2 = The one-way glass, 3 = The dummy light, 4 = The IES light

Direction

The scene, both virtual and real-world, consisted of two juxtaposed spot luminaires, each aimed in a different, fixed direction. The luminaire closest to the center of the visual field was directed away from the participant, while its beam aimed towards the wall, angled towards the far upper corner of the scene, thereby leaving a luminous spot on the wall in the center of the scene. The other luminaire aimed away from the wall, also angled towards the far upper corner of the scene. It was directed such that the observer did not look directly at the LEDs, but saw the reflectors of the luminaire.

The independent variable *direction* was hence manipulated as follows (for three levels):

A. Only the luminaire aiming to the wall was switched on (see Figure 20, left).
B. Only the luminaire aiming away from the wall was switched on (see Figure 20, center).
C. Both spots were switched on (see Figure 20, right).

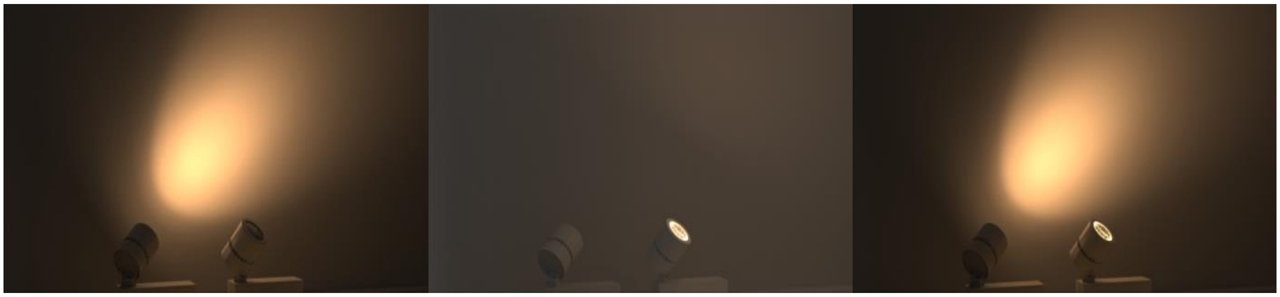


Figure 20: Different conditions for the direction of the light, from left to right: A (only the "wall spot" is on), B (only the "ceiling spot" is on), C (both lamps are on)

Luminaire Intensity

Luminaire intensity related to both the real-world and virtual stimuli, but in a different way. Firstly, for the real-world scene, luminaire intensity was varied to different luminance levels. In determining the luminance level of a setting (i.e., *low*, *medium* or *high*) we made sure that the different conditions were visibly different. A Photo Research SpectraScan® spectroradiometer measured the luminance of a spot on the wall[4] at maximum intensity of the luminaire to be 4175 cd/m^2. This value was then chosen as the luminance level for the high intensity condition. For the low intensity condition the luminance had to depart from 0 in order to prevent floor effects in the tuning task (i.e., to prevent that all participants tuned the image intensity all the way down). Since the peak luminance of the display was measured to be 377 cd/m^2, we chose to match the real-world luminance, again measured on the spot on the wall[2], to this value for the low intensity condition. For the luminance level of the medium intensity condition, we used the average between the high and low intensity value, but then in terms of CIELAB Lightness L*. The high intensity condition with a luminance of 4175 cd/m^2 had a L* of 100. The L* of the low intensity condition was calculated to be 36. Translating the average L* value of 68 back to linear lighting levels resulted in a luminance of 1572 cd/m^2. A pilot test confirmed that all three intensity conditions were visibly different from each other.

In the virtual model, luminaire intensity was not a variable; instead *image intensity* was tuned by the participants to match the apparent brightness of the real-world luminaires. Each possible step in image intensity was a tone mapping of the same HDR image with a different key value. The luminous intensity of the virtual luminaires (as set in 3DS Max) defined the HDR image, hence remained the same during tone mapping. The intensity of the virtual luminaires was thus not an experimental variable, but a constant. Still, the right proportion for the luminous intensities of the IES light and dummy light in the virtual model had to be deduced. The latter was done by means of a ratio calculation between the brightest part of the spot on the wall over the brightest part of the reflectors of the luminaire. The locations of these two points are marked in Figure 21 with an "x". The real-world wall/reflector intensity ratio was found to be 1:7.6. Subsequently, this ratio had to be the same for the pixel intensities located at the brightest part on the wall and at the brightest part on the reflectors in the HDR image. Setting the dummy light (representing the brightest part on the reflectors) to 1,500 lm created a wall/reflector pixel intensity ratio of 1:7.6, thereby matching the real world ratio (see Figure 21).

Having the right relative pixel intensities was not only important for a realistic visualization, but also for the point-spread function that was used to calculate the synthetic glare, as explained in the next section.

---

[4] Based on luminance measurements of brightest spot on the wall (see Figure 21 for exact location on wall)

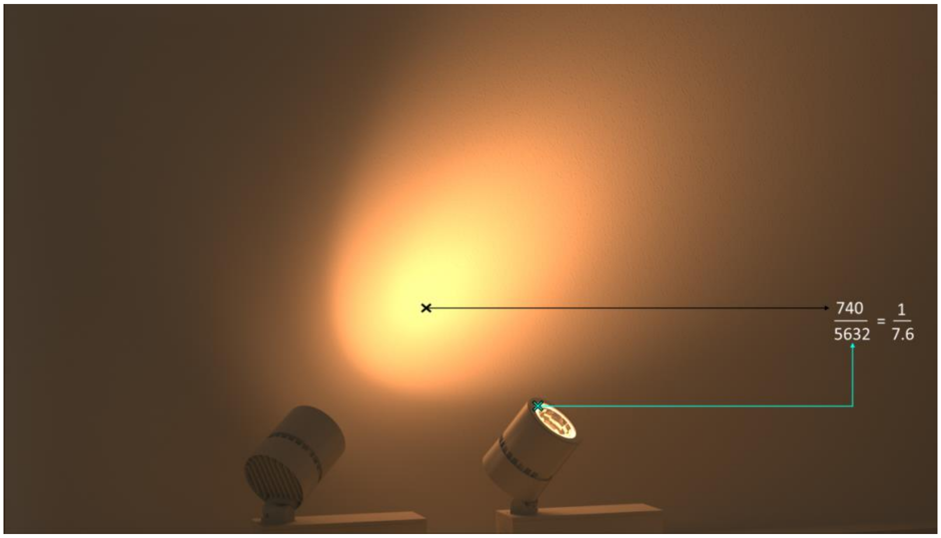


Figure 21: Pixel intensity ratio between brightest spot on the wall and brightest spot in the reflectors

Synthetic Glare

Synthetic glare (explained in section 4.3.1) was applied to the visualizations to vary the independent variable “glare”. It was applied by means of a Matlab script that read an image, applied the point spread function, and added it to the image. As a result, point sources in the image were spread out. In essence, the point spread function spread the high pixel values representing light sources in an image over the surrounding pixels, thereby lowering contrast around the luminous parts in the image, just like disability glare does. The point spread function had a parameter called the center/rest ratio that determined how strongly the point source was spread out.

The variable glare in the experiment had three levels: *no glare*, *glare* and *exaggerated glare*. In the *no glare* condition, no point spread function was applied to the visualizations (see Figure 22, A). In the *glare* condition, synthetic glare was applied to a moderate extent with a center/rest ratio of 0.9, such that a luminance gradient around the luminous ring of the luminaire became visible (see the blurry edges around the luminous part in Figure 22, B). In the third condition of “exaggerated glare”, synthetic glare was applied to a larger extent than it would occur in reality. A center/rest ratio of 0.7 was chosen which created a halo around the luminous ring of the luminaire, and also resulted in a haze over the image, resembling a veiling luminance (see Figure 22, C).

Figure 22: ***Glare*** **conditions used in the experiment, from left to right: A (no glare), B (glare), C (exaggerated glare)**

The Key Value Parameter

The key value is a parameter in rendering visualizations. For this experiment, we started from three HDR images; i.e., one for each of the three conditions in the light direction (as shown in Figure 20). These HDR images were then tone mapped multiple times using different key values, as such generating images of the same condition at different image intensity. In practice, each HDR image was tone mapped 50 times using 50 different key values. The range in key values was not chosen to be linear, but rather perceptually linear. To retrieve a scale that is perceptually linear, the formula by Tang and colleagues (2014) for 3000K light was used:

$$y = 0.0451x^2 - 0.2216x + 0.31118$$

where y is the key value and x represents perceived brightness (see Figure 6 for the graph). Hence, the formula maps a linearly increasing range of x values to a set of key values that increases perceptually linear in brightness. The formula is only valid for $x > 2.3$, so to find the key values below this point, a simple linear function ($y = 0.014x+0.005$) was extrapolated from $x = 0$ to $x = 2.3$. Accordingly, to obtain a range of 50 key values that increases perceptually linearly in brightness, the formula was given an input of x values ranging from 0.01 to 9, with steps of 0.18. The white limit, another parameter of the Reinhard ’02 TMO that influences the apparent brightness of an image, was kept constant to isolate the effect of the key value.

The resulting image intensity scale consisted of 50 images that approximately linearly increased in brightness. The scale with its corresponding parameters is displayed in Table 2. The selected image intensity in the tuning was the dependent variable in our first experiment, and hence corresponds to a key value, which in turn relates to perceived brightness. At this stage, we should, however, be cautious with the interpretation of perceived brightness in terms of key value, because a best match in tuning does not necessarily mean a close match with reality. Yet within-subject differences in key values are usable for detecting effects in perceived brightness.

Image Intensity	Key $f(x)$	Brightness x	Image Intensity	Key $f(x)$	Brightness x	Image Intensity	Key $f(x)$	Brightness x	Image Intensity	Key $f(x)$	Brightness x	Image Inten-sity	Key $f(x)$	Brightness x
0	0.0051	0.01	10	0.0303	1.81	20	0.0996	3.61	30	0.4329	5.41	40	1.0585	7.21
1	0.0077	0.19	11	0.0329	1.99	21	0.1198	3.79	31	0.4823	5.59	41	1.1372	7.39
2	0.0102	0.37	12	0.0354	2.17	22	0.1429	3.97	32	0.5347	5.77	42	1.2187	7.57
3	0.0127	0.55	13	0.0401	2.35	23	0.1689	4.15	33	0.5899	5.95	43	1.3032	7.75
4	0.0152	0.73	14	0.0398	2.53	24	0.1978	4.33	34	0.6481	6.13	44	1.3906	7.93
5	0.0177	0.91	15	0.0425	2.71	25	0.2297	4.51	35	0.7092	6.31	45	1.4809	8.11
6	0.0203	1.09	16	0.0481	2.89	26	0.2645	4.69	36	0.7732	6.49	46	1.5742	8.29
7	0.0228	1.27	17	0.0566	3.07	27	0.3022	4.87	37	0.8402	6.67	47	1.6704	8.47
8	0.0253	1.45	18	0.0680	3.25	28	0.3429	5.05	38	0.9100	6.85	48	2.3732	8.65
9	0.0278	1.63	19	0.0823	3.43	29	0.3864	5.23	39	0.9828	7.03	49	1.8715	8.83

Table 2: Mapping of key value on the image intensity scale and their relationship with brightness as defined by Tang et al. (2014)

6.2.4 Viewing Method

As discussed in section 4.2.5, the ideal viewing method ensures brightness adaptation without a need to memorize the perceived intensity of the scene. As this is impossible, having participants to memorize the brightness of the scene could not be avoided. This might introduce noise to the tuning results, but without proper visual brightness adaptation the tuning results would even be far less useful. Besides, in the scope of virtual prototyping, understanding brightness from memory matching is most useful, since in practice the end user views virtual brightness typically without a real-world reference, and so, has to rely on memory matching. As such, the two methods identified in section 4.2.5 as most appropriate are memory viewing and successive binocular viewing. Compared to successive binocular viewing, memory viewing was preferred by observers and had slightly more sensitive results (Braun et al., 1996). To decrease the bias of memorization, however, we chose for an intermediate approach between memory viewing and successive binocular viewing: i.e., memory viewing with one memory refreshment. The real-word luminaires were viewed first (while the display was black). Then, after a short dark pause, the image appeared on the display with an initial key value that was randomized to counterbalance the conservative bias (Fotios, 2007). The participant matched the image as well as possible with his/her memory of the real scene. After this first matching the participant got a second and final look at the real-world scene as a way to refresh his/her memory. After a short pause the image appeared as they matched it the first time. They now had the opportunity to adjust their match.

It should be noted that in this way adaptation could only be controlled for to a certain extent. In general more prolonged periods of adaptation are needed for full brightness adaptation to occur. The duration of the experiment, however, had to be kept within a decent timespan. Therefore, the real-world stimulus was temporally separated from the image by only two seconds of darkness. This period was not made any longer because too much dark adaptation was not beneficial, especially when considering the fact that both the real scene and the virtual scene were supposed to be close in brightness. It was also made sure that each participant viewed the real-world stimulus for the same amount of time: the luminaires went on for ten seconds, offering some time to memorize the brightness. The fixed viewing time guaranteed *equal adaptation times* between observers. However, since individual differences may cause observers to have different rates of adaptation, we could not guarantee that the states of adaptation were exactly equal between observers.

6.2.5 Participants

Twenty-six participants (i.e., 8 female, 18 male, mean age = 29 years, standard deviation = 5.3 years, range = 21-45 years) volunteered for the experiment. All participants were students and employees at Philips Research with limited knowledge on virtual brightness perception. All participants performed a Landolt C visual acuity test to make sure they had a normal or corrected to normal acuity.

6.2.6 Procedure

Upon entering the experiment room, the participant was asked to have a seat and was given a moment to read and sign the informed consent. We then tested the visual acuity before starting the experiment.

An oral introduction was given about the scope of the project (see Appendix 11.2.2 for the full text), after which any questions the participant had were answered. Then the experiment started with two training tunings. During the training the experiment leader explained the brightness matching task, and how to perform the tuning with the keyboard. For clarification, brightness was defined as the amount of light produced by the luminaire(s). After the training the actual experiment started. For each condition, the participant followed four steps:

1. The real-world luminaires in the left window went on for 10 seconds. During this step the participants tried to memorize how bright the luminaires appeared.
2. When the luminaires in the left window turned off, there was a 2 seconds pause of darkness before the image of the corresponding condition appeared in the right window (on the display). During this step, the participants performed the first tuning on the image.
3. After confirming the first tuning, the real-world scene for the same condition was viewed for another 10 seconds to refresh the subject's memory
4. Again there was a 2 seconds pause of darkness, after which the participants had the opportunity to adjust the image tuning if they wanted.

After step 4 the experiment continued with the next condition at step 1, and so on for all 27 conditions.

6.2.7 Statistical analysis

All data were analyzed with SPSS Statistics version 17.0. Before analyzing the effect of the independent variables *direction*, *luminaire intensity* and *glare* on the tuned image intensity, the results of the first and second tuning were mutually compared via violin plots and with a paired-samples t-test. Subsequently, the assumption of normality, implying that the sampled distribution of variables should be normal when included in significance tests, was evaluated. The assumption doesn't mean that the overall distribution of the dependent variable should be normal, but it means that it should be normally distributed at each unique level of the predictor variables (Field, 2013). But, even when normality is rejected at the individual levels of the predictor variables, which may be plausible since other predictor variables may interact with the tuning distribution, it makes more sense to test for normality per unique combination of predictor variables' levels, i.e., per condition. Hence, Shapiro-Wilk (S-W) tests were performed for each of the six test conditions. In addition, sphericity, being roughly defined as similarity of the relationship between pairs of experimental conditions, was tested with the Mauchly's Test of Sphericity. We made sure that, when the assumption of sphericity was rejected, the appropriate correction was used to adjust the degrees of freedom.

After checking the assumptions, we performed a General Linear Model Repeated-Measures ANOVA (hereafter abbreviated to rANOVA) that included the independent variables and the dependent variable *image intensity*. The main effect of the independent variable glare on the tuned image intensity answered the research question. To further investigate our hypotheses, pairwise comparisons or post-hoc contrast analyses were done, i.e., (interaction) effects of the independent variables on the dependent variable (i.e., tuned image intensity) were evaluated. Where pairwise comparisons were required, we used Bonferroni adjustments and where post-hoc analysis was required, we used the Tukey HSD algorithm.

6.3 Results

6.3.1 Introduction

The experiment yielded a tuned image intensity value for each of the 27 conditions. These conditions consisted of combinations of the three independent variables (i.e., direction, glare, and luminaire intensity) with each three levels. In the remainder of this chapter they are denoted by a combination of three numbers: the first number refers to the luminaire intensity level (1 = low, 2 = medium, 3 = high), the second number to the glare level (1 = no glare, 2 = glare, 3 = exaggerated glare), and the third number to the direction (1 = wall, 2 = ceiling, 3 = both) (see Appendix 11.1. for the definition of the variables, and Appendix 11.2.1 for the list of all conditions).

6.3.1 Data Exploration

Before performing the rANOVA, standard ANOVA assumptions should be met (except for independence of samples) to prevent inflation of a type 1 error (Field, 2013). In this section the assumptions of normality and homogeneity are discussed.

The first and second tuning

During the experiment, the dependent variable image intensity, was measured twice: a first tuning (T1) and an adjustment of the first tuning (T2). A paired-samples t-test compared the image intensity scores of T1 and T2. The result showed that there was a significant difference in the scores ($M_{T1} = 16.99$, $SD_{T1} = 9.36$ versus $M_{T2} = 17.40$, $SD_{T2} = 9.60$; $t(701) = -3.00$, $p = .003$). This suggests that participants made use of the second tuning opportunity to adjust their choice of T1. Figure 23 visualizes all tuning distributions per condition with violin plots. The left sides of the violins are the distributions from T1; the right sides are the distributions from the final tunings T2. Inside each distribution is a box plot. Visual inspection of the box plots gives the impression that the image intensity scores for T2 are less spread than those for T1. Comparison of the distributions per condition indeed shows that T2 has a smaller standard deviation for 23 out of all 27 conditions (see Appendix 11.2.4 for the full table). Since the image intensity scores of T2 are less spread, we assume that these scores are more reliable.

We also compared the results of T1 and T2 based on their relation with the independent variables, by comparing two separate rANOVAs: one with the image intensities of T1 as dependent variable and one with the image intensities of T2 as dependent variable. The results showed that the rANOVA with the scores from T2 contained more significant effects with larger effect sizes. Thus, since the tuning results from T2 seem more reliable and show stronger relations with the independent variables, only the image intensity scores from T2 are used in further analyses.

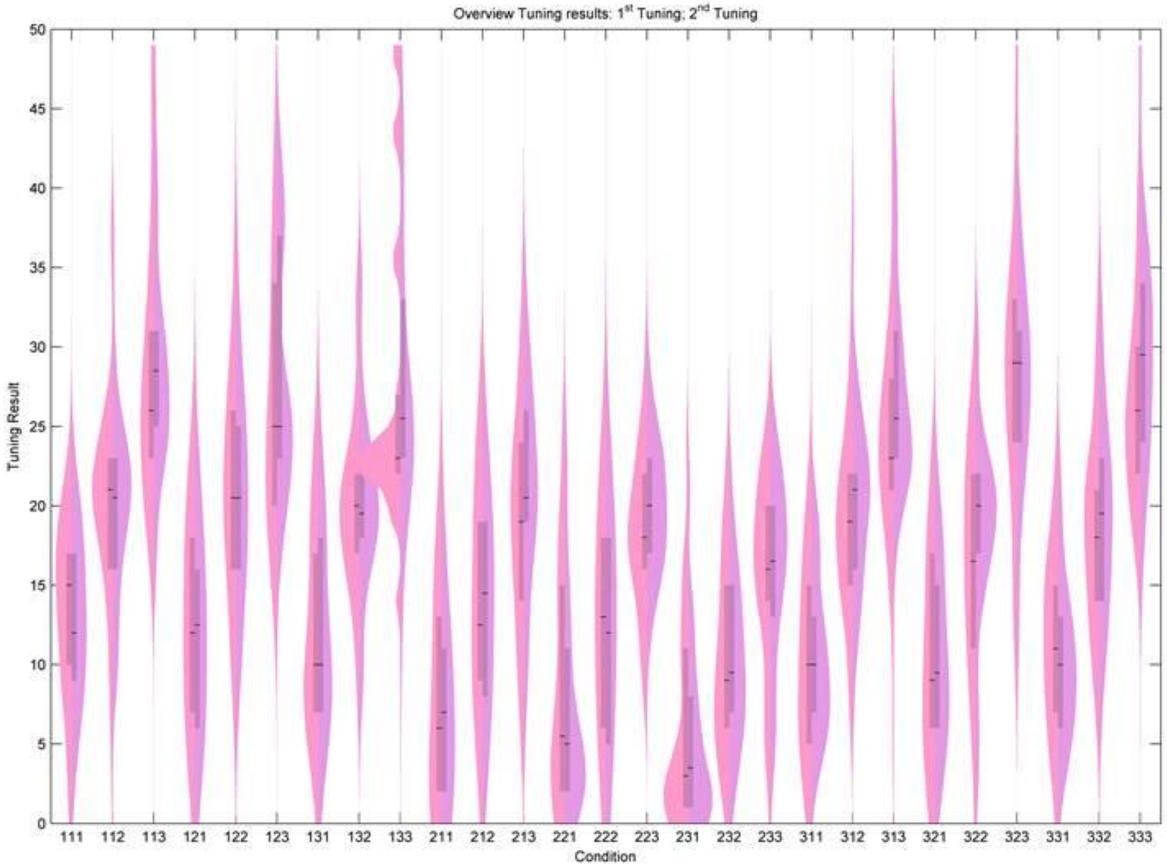


Figure 23: Violin plots displaying the tuning results of the first tuning (left side of the violin) and second tuning (right side of the violin) per condition.

Assumptions

All data were evaluated on the assumptions of normality and homogeneity. The results of the S-W tests on normality per condition are presented in Table 3, and indicate normality for all but six conditions. Of these six, the conditions 221, 231, 311 are distributed near the bottom part of the tuning scale, as can be seen in Figure 23. Considering that these conditions are all low-intensity conditions, there is reason to believe that the skewed distribution in these conditions is caused by floor effects. Condition 123 has a distribution at the center of the tuning scale, but a small group of participants is distributed higher on the scale. Normality was rejected because there seems to be a bimodal distribution. By means of a hierarchical cluster analysis, indeed, two groups of participants were identified; a smaller group of people tuned the image intensity significantly higher, especially for conditions in which the wall luminaire was on (see Appendix 11.2.4 for the dendogram and graph). The distribution in condition 313 is skewed towards the higher end of the scale, which can also be caused by this bimodality. Lastly, condition 322 has a distribution that just deviates from normality. The reason for this is not clear, but the long downwards tail may indicate a bimodal distribution that is slightly overlapping around an image intensity value of 10.

Condition	Skewness (SE = .456)	Kurtosis (SE = .887)	S-W Statistic	Significance
111	-.087	-.892	.968	.567
112	.375	.986	.963	.455
113	.851	.565	.928	.068
121	.258	-.989	.947	.196
122	.236	-.369	.978	.819
123	.601	-.947	.903	.018*
131	.386	-.781	.950	.233
132	.407	.994	.955	.295
133	.678	.265	.935	.103
211	.542	-.538	.929	.073
212	.000	-.869	.975	.761
213	-.132	-.254	.965	.490
221	1.278	.953	.855	.002*
222	-.097	-1.252	.930	.076
223	.081	-.632	.962	.428
231	1.373	1.707	.842	.001*
232	.351	-.561	.971	.640
233	-.643	-.376	.931	.084
311	.537	-.846	.913	.032*
312	.706	2.103	.941	.145
313	.997	.354	.908	.023*
321	.307	-1.068	.939	.125
322	-.725	1.401	.921	.047*
323	.338	1.654	.957	.329
331	.784	.340	.942	.152
332	.248	-.236	.980	.867
333	.543	1.245	.965	.503

Table 3: Shapiro-Wilk tests per condition. For conditions marked with a " * " normality was rejected.

6.3.1 Image Intensity

A rANOVA was conducted to compare the effect of all independent variables on image intensity. The independent variable most relevant to our research question, i.e., glare, is discussed first. Glare had a significant effect on image intensity, but with a small effect size (*F*(2, 50) = 4.66, *p* = .014, $\eta^2_{partial} = .157$). The estimated marginal means of no glare (*M* = 18.10), glare (*M* = 17.43), and exaggerated glare (*M* = 16.67) showed that glare decreased the chosen image intensity. Hence, the brightness of the real-world stimulus was matched with a less bright image when a higher level of glare was present. In other words, the brightness-match was achieved with a lower key value when glare was present, which is in line with our main hypothesis that adding synthetic glare to renderings increases the perceived brightness. The Tukey HSD post hoc test found two homogeneous subsets. One group contained no glare and glare, the other group contained glare and exaggerated glare. Hence, only the means of no glare and exaggerated glare were statistically significantly different from each other (*p* = .001).

The rANOVA also tested if glare interacted with the other independent variables. Glare showed no significant interaction with luminaire intensity, but it did with direction (*F*(4, 100) = 5.51, *p* < .001, $\eta^2_{partial}$ = .180). The estimated marginal means, as displayed in Table 4, provided insight into this effect. Whereas the means in image intensity were relatively unaffected by glare for the light directed to the wall and for the light directed to both the wall and the ceiling, relatively larger differences in mean image intensity between different glare levels were found for the light only directed to the ceiling. This indicates that the main effect of glare on perceived brightness is primarily caused by its strong effect on the light directed to the ceiling. This was, however, expected. In the ceiling direction, the reflectors of the luminaire were the only source of light in the visualization. This point-source-like type of light source in front of an unilluminated background made the application of the glare function more visible in comparison to the other directions. The effect of glare within the ceiling direction was analyzed further by pairwise comparisons of the mean image intensity of the glare levels within the ceiling direction. No glare (*M* = 13.99) and glare (*M* = 13.05) did not significantly differ (*p* = .731). The mean of exaggerated glare (*M* = 10.47) was statistically significantly different from both the no glare condition (*p* < .001) and the glare condition (*p* < .001).

Direction	**Glare**	**Mean**	**Std. Error**	**95% Confidence Interval**	
				Lower Bound	Upper Bound
Wall	No glare	20.859	.925	18.953	22.765
	Glare	20.192	1.167	17.789	22.596
	Exaggerated glare	19.872	1.226	17.348	22.396
Ceiling	No glare	13.987	.980	11.968	16.006
	Glare	13.051	.831	11.341	14.762
	Exaggerated glare	10.474	.822	8.781	12.168
Both	No glare	19.423	.880	17.610	21.236
	Glare	19.051	.966	17.062	21.041
	Exaggerated glare	19.654	.979	17.638	21.670

Table 4: Estimated marginal means of image intensity for "direction x glare"

The rANOVA also showed that direction had a significant effect on image intensity (*F*(1.58, 39.44) = 64.58, *p* <.001, $\eta^2_{partial}$ = .721). In addition, also luminaire intensity had a significant effect on image intensity (*F*(1.69, 42.14) = 64.58, *p* <.001, $\eta^2_{partial}$ = .919). And, the same was true for the interaction between direction and luminaire intensity (*F*(4, 100) = 5.40, *p* =.01, $\eta^2_{partial}$ = .178). Figure 24 shows the averaged image intensity values and their 95% confidence interval per direction, clustered by luminaire intensity. These values were also averaged over the glare conditions. Obviously, the tuned image intensity increased with the luminaire intensity, each level of the latter being mutually significantly different. The significant effect of direction indicated that the average image intensity was lower when the light was directed to the ceiling, while the tuned image intensity was nearly equal when the light was directed to the wall or to both the ceiling and the wall. The significant interaction between luminaire intensity and direction seems to be primarily caused by a less strong effect of luminaire intensity in the ceiling condition. The means of luminaire intensity low, medium, and high for the ceiling direction are 6.35, 12.17, and 19.00, respectively. Similarly, the means of luminaire intensity for the wall condition (12.22, 20.42, and 28.28, respectively) and the both condition (10.42, 19.47, and 28.23, respectively) have larger between-level differences than for the ceiling condition.

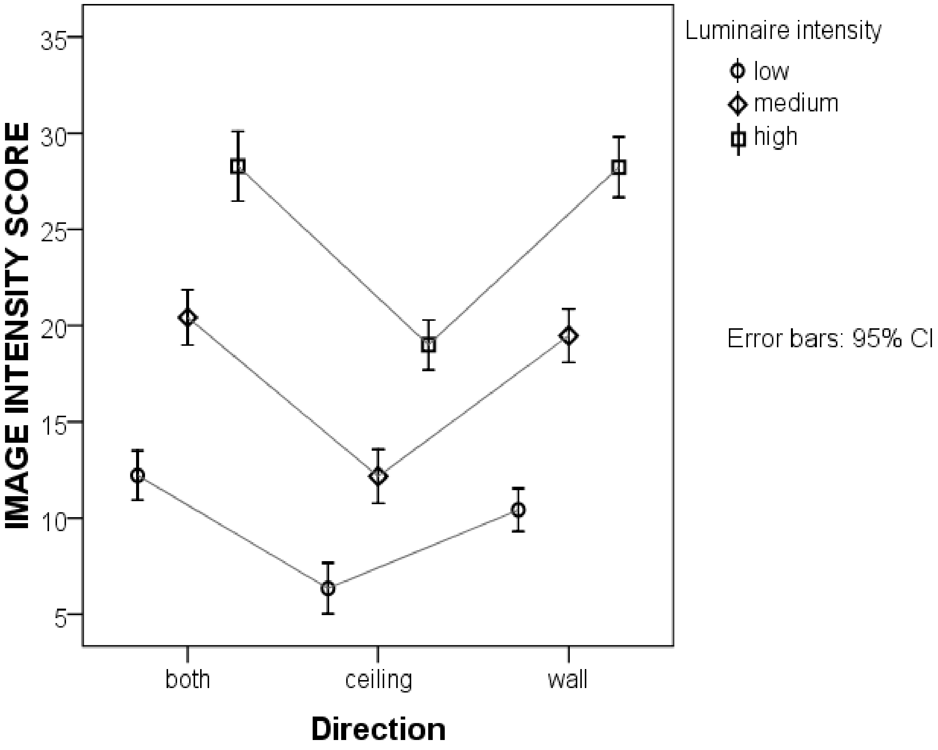


Figure 24: The effect of direction (x-axis) on image intensity (y-axis) clustered by luminaire intensity

6.4 Discussion

6.4.1 Introduction

An overview of the results from E1 is given in the graphs of Figure 25, including all variables. Each graph represents a direction, with the luminaire intensity on the x-axis and the tuned image intensity on the y-axis. Each line represents a different glare level.

Firstly, it shows that participants tuned to a higher image intensity when the intensity of the real-world stimulus increased. The increase of image intensity, and thus also of key value, with luminaire intensity demonstrates that the key value was a significant predictor for perceived brightness. Secondly, Figure 25 visualizes that the image intensity scores were clearly lower and had a less steep slope with respect to luminaire intensity in the ceiling condition. Lastly, but most importantly, Figure 25 illustrates that the effect of glare was visible in several conditions. Glare showed no significant effect in the conditions wall direction and both directions, so when the luminaire aiming towards the wall was on. From the high similarity in image intensity scores between the conditions both directions and wall direction we may derive that the luminous spot on the wall functioned as a strong brightness indicator for the participants. In the ceiling condition, however, glare had a statistically significant negative effect on image intensity. This means that the presence of glare caused participants to choose a less bright visualization, which suggests that glare increased the perceived brightness of the visualization. This supports our hypothesis.

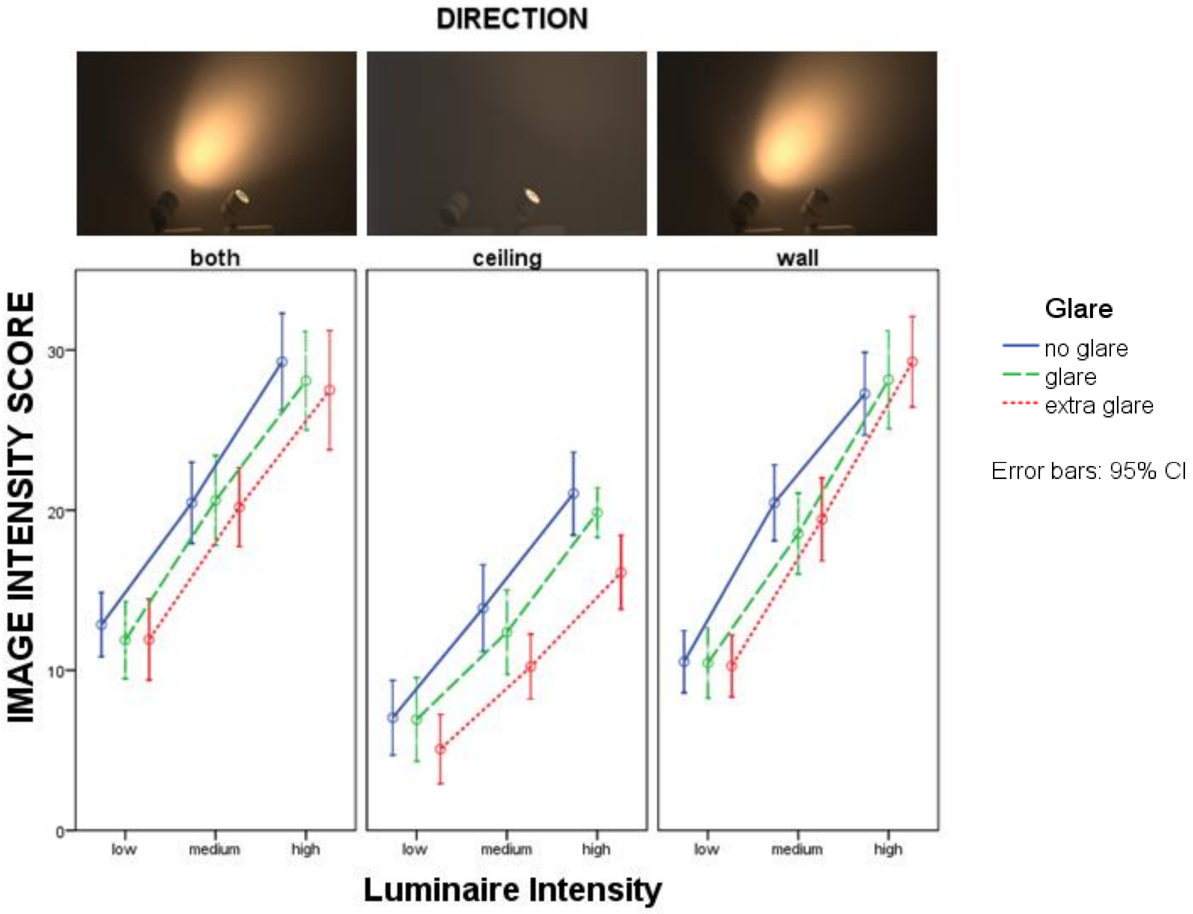


Figure 25: Error bar graphs of brightness versus intensity, per direction

6.4.2 Repeated Tuning

The conclusion that the repeated tunings improved the reliability was based on the comparison between the first and second tuning results (T1 and T2), which revealed that the T2 results had a smaller standard deviation for 23 out of all 27 conditions (see Appendix 11.2.1 for the full table). Apparently, participants needed to see the stimuli twice in order to become more consistent in their assessment.

A second difference between the T1 and T2 results is a change towards more extreme image intensity values. The mean image intensity decreased in the low intensity conditions (M_{T1} = 10.3, M_{T2} = 9.7) and increased in the high intensity conditions (M_{T1} = 23.9, M_{T2} = 25.2) from the T1 to the T2 results. No well-founded conclusion can be drawn from this, but there are two plausible factors to consider, namely memory and adaptation. The difference in mean image intensity between T1 and T2 may be caused by the effect of a memory "refresher" of T2 with respect to T1, but such a memory effect would not explain the interaction with luminaire intensity. Adaptation, on the other hand, *could* explain the interaction with luminaire intensity. In the medium and high luminaire intensity conditions, the real-world stimulus had a much higher luminance than the visualization. Whenever a participant adapted too strongly to the luminance level of the real-world stimulus, the display luminance of the visualization would be experienced as relatively dim, which would cause them to overcompensate that by tuning to a higher image intensity value. This was indeed the case for the high luminaire intensity condition and also for the medium luminaire intensity condition (M_{T1} = 16.83, M_{T2} = 17.35). Although the viewing method with repeated tuning seemed to have improved the accuracy of the tuning results, we should be cautious since it cannot be excluded that the negative effect of adaptation was stronger than the positive effect of a memory "refresher".

6.4.3 Synthetic Glare

We learned from our analysis that the effect of glare exclusively occurred when the light was directly to the ceiling. Glare was expected to have the strongest effect in this condition, because it had the largest ratio in luminance between the glare source and the area around it, which is the main cause of disability glare. Since the synthetic glare formula replicates glare in a realistic way, glare was indeed the most apparent in the ceiling condition due to the high luminance contrast with the background. The extent of the luminance ratio of the glare source over the surrounding area can also explain the trend that the effect of glare increases with higher intensity (as is visible in Figure 25).

The absence of a significant effect of glare when both luminaires are switched on can also be related to the luminance ratio. When both luminaires are on, the luminance of the surrounding area is higher because of the additional light source, even though the luminance of the glare source remains just as high. So, the luminance ratio becomes smaller, which could have made the synthetic glare appear less bright. On the other hand, the latter does not explain why the tuned image intensity was nearly equal when light was directed only to the wall or to both the wall and the ceiling. Even with the decrease in luminance ratio, it would still be expected that the condition with light to both ceiling and wall would be experienced as brighter than when light was only directed to the wall. The reason why the latter is not the case may possibly be found in the way the matching task was executed. The condition with both luminaires on was the only one in which the image intensity of two light sources had to be matched, which might have been a difficult task for the participants. Participants instead might have based their matching task on one luminaire more than on the other. More specifically, maybe the spot on the wall was a stronger brightness cue that drew away the attention from the “ceiling” luminaire completely. We can, however, not ascertain how participants made these – perhaps unconscious – judgments.

The increase in image intensity as a consequence of adding glare in the conditions with light directed to the ceiling can give an estimate of how much glare increased perceived brightness. Without synthetic glare, the mean image intensity was 14, while with glare the mean image intensity was 13. Apparently, the point spread function with a center/rest ratio of .9 created synthetic glare that was too subtle to significantly influence brightness perception. With exaggerated glare, however, the mean image intensity was 10. Assuming that image intensity was approximately linearly related to brightness, we can conclude that exaggerated glare increased perceived brightness by 29% (because the chosen image intensity decreased by 29%). This finding is in agreement with the findings from Yoshida et al. (2008), who reported an increase of 20-35% in “perceived luminance” by applying synthetic glare.

6.4.4 Intensity and Direction

Luminaire intensity had a significant effect on the tuned image intensity. Thus, as expected, the key value gives a representation of the perceived brightness.

Moreover, the analysis pointed out a difference in tuned image intensity between different directions of the light sources. As is visible in Figure 25, mainly when the light is directed to the ceiling the tuned image intensity was significantly lower. There are several possible explanations for this finding. It could be the case that the real-world luminaire towards the ceiling appeared less bright than the one towards the wall. This reasoning is supported by some of the brightness-mediating factors discussed in the literature, i.e., a larger size of a luminous spot and illumination on the wall can increase brightness (Boyce, 2003). The effect of glare in the condition that both the luminaire towards the wall and towards the ceiling were switched on also suggest that the luminous spot on the wall was a stronger brightness cue, as discussed in the previous subsection. This is interesting, because it suggests that luminous spots on surfaces are a very important way to convey a certain brightness experience in a visualization. Looking at a spot on the wall is probably a more natural way to assess brightness than looking directly at the reflectors of a luminaire. However, if the luminaire directed to the ceiling really appeared less bright, then the *visualization* of this condition should also appear less bright. This would cancel out a tuning difference between the different directions of the luminaires. It is thus more plausible that the lower perceived brightness for the luminaire directed to the ceiling is primarily caused by some factors of the visualization. Visual inspection of the stimuli gives reason to believe that the behavior of the TMO may have caused the bias. The Reinhard ’02 algorithm seems to underestimate the luminance difference between the luminaire’s reflectors and the background. The consequence is that increasing the key value has a lot of influence on the background as well, even though it is barely illuminated. Due to this unusual behavior of the TMO, the visualizations with image intensities of level 20 or higher for the condition with the light directed to the ceiling look so light that they start to cause a mismatch with the appearance of the real-world reference. That would explain why almost all tuned image intensities in the ceiling conditions were kept within the lower half of the scale.

7 Experiment 2

7.1 Introduction

The purpose of Experiment 2 (E2) is to answer some additional questions that can be asked based on the results from E1. Firstly, we tested whether the image intensity values chosen in E1 are also the best brightness match when assessed by different people, i.e., are the means of the image intensities chosen in E1 better matches than image intensities that slightly deviate from the mean? Secondly, the match scores will give us an indication of the absolute match, i.e. how good was the best match? Furthermore, we investigate whether images with glare are perceived to be brighter than image without glare. E1 has shown that participants have chosen lower image intensities for images with (exaggerated) glare, which suggests they indeed have been perceived as brighter, but we want to add a direct question to this indirect measure. Finally, we are also interested in the effect that glare has on realism, and what the effect is of the presence of a real-world reference.

7.2 Method

7.2.1 Design

E2 is an experiment, comparing brightness of image stimuli with their real-world reference. The task is to observe the brightness of the real-world reference, and subsequently to assess an image stimulus by answering evaluative questions. The experimental design is within-subjects with luminaire intensity (2 levels), glare (3 levels), direction (2 levels), image intensity (3 levels) and reference (2 levels; with and without reference) as independent variables. Since glare only had an effect on the perceived brightness for the “ceiling” direction and not for the "wall" direction in E1, E2 primarily uses the “ceiling” direction for the evaluations, and only add some of the conditions for the “wall” direction as control conditions (see Figure 26). Consequently, the design is full-factorial for the "ceiling" direction, and only includes the variation of luminaire intensity and image intensity for the "wall" direction. Accordingly, there are (3 (glare) x 3 (image intensity) x 2 (luminaire intensity) x 2 (reference)) + (3 (image intensity) x 2 (luminaire intensity) x 2 (reference)) = 48 conditions. An overview of all conditions can be found in Appendix 11.2.4.

In short, E2 consists of two parts (i.e., E2A and E2B). In E2A, the participants view the real-world scene for a fixed time interval before viewing the corresponding image of the luminaires on the display. During the latter they are asked to answer three evaluative questions, which measure the dependent variables “match”, “brightness”, and “realism”. E2A is directly followed by E2B, which consists of observations of the images, without the real-world scene as a reference. Per image the participants are asked to answer two evaluative questions, which measure the dependent variables “brightness without reference” and “realism without reference”. All participants observed all 24 conditions of E2A, in a randomized presentation order, followed by all 24 conditions of E2B in a randomized presentation order.

7.2.2 Experimental Setup

The same set-up as described for E1 in section 6.2.2 is used.

7.2.3 Stimuli

The same real-world scene and images as described for E1 in section 0 are used. In E2, however, the direction has only two levels: i.e., ceiling and wall (see Figure 26). Finally, glare has again 3 levels: no glare, glare, and exaggerated glare (see Figure 22). Also the intensity of the real-world luminaires has only two levels: low (377 cd / m^2)[5] and high (4175 cd / m^2). *Luminaire intensity* also relates to image intensity. Image intensity will be high or low in correspondence with the level of luminaire intensity. The exact image intensity values are explained in the subsection below.

Figure 26: Illustration of the *direction* conditions: on the left: the "ceiling" condition and on the right: the "wall" condition

Image Intensity

Image intensity (related to the key value) is an additional independent variable. It has three categorical levels which are chosen based on the results of E1 to be mean, above mean, and below mean. "Mean" here refers to an image intensity that equals the mean image intensity chosen during the tuning task in E1 for that condition. To be more specific, it refers to the chosen image intensity for a given direction and luminaire intensity, averaged over the three glare levels and over all participants from E1. "Above mean" refers to a condition in which the image intensity is five steps higher than the image intensity that was chosen on average in E1. Similarly, "below mean" refers to a condition in which the image intensity is five steps lower than the brightness chosen on average during the tuning. Table 5 shows the mean of the results from E1, as well as the new values for the image brightness per experimental condition. It was visually determined that image intensity differences of 5 steps were notably different from each other in brightness.

Table 5: Overview of the image intensity values for the image stimuli, and the means from the tuning data of E1. The values of the "mean" column are computed from the "tuning mean" column, by taking the average per direction and intensity. For the "below mean" and "above mean" columns, 5 is subtracted from and added to the mean, respectively.

Condition			Data E1	Image Intensity		
Direction	Luminaire Intensity	Glare	*Tuning Mean*	Below Mean	Mean	Above Mean
Ceiling	Low	No	*7*	1	6	11
Ceiling	Low	Yes	*7*	1	6	11
Ceiling	Low	Extra	*5*	1	6	11
Ceiling	High	No	*21*	14	19	24
Ceiling	High	Yes	*20*	14	19	24
Ceiling	High	Extra	*16*	14	19	24
Wall	Low	No	*11*	6	11	16
Wall	High	No	*27*	22	27	32

---

[5] Based on luminance measurements of brightest spot on the wall (see Figure 21 for exact location on wall)

7.2.4 Participants

Twenty-five participants (i.e., 7 female, 18 male, mean age = 29 years, standard deviation = 6.88 years, range = 21-45 years) volunteered for the experiment. All participants were students and employees at Philips Research with limited knowledge on virtual brightness perception. All participants performed a Landolt C visual acuity test to make sure they had normal or corrected to normal acuity.

7.2.5 Viewing

E2A involved the comparison of the image stimulus with the real-world reference. For this comparison we used the two-windows set-up (as illustrated in Figure 8, Figure 9, and Figure 10). This set-up allowed side-by-side viewing with the reference on the left side and the image on the right side of the participants' field of view. As in E1, memory viewing was chosen as the viewing method. Because only the first question related to the reference, it was not necessary to have the participants view the same reference multiple times. Hence, the reference was viewed for ten seconds, followed by a two-seconds pause of darkness, before the image appeared on the display on the other side. E2B did not involve the real-world scene, and so, participants viewed only the images as long as they needed to answer both questions. The next image/condition then appeared after two seconds of darkness.

7.2.6 Procedure

Although the experiment consisted of two parts, it was possible to do both parts in a single session of 20 minutes.

When entering the experiment room, the participant was asked to have a seat and was given a moment to read and sign the informed consent. We then performed the visual acuity test and started the experiment. An oral introduction on the scope of the experiment (see Appendix 11.2.4 for the full text) was given, after which all questions the participant had were answered. The actual experiment started with two training conditions, during which the experiment leader explained the questions, and how to answer them with the keyboard. The questions sequentially appeared at the top of the image. After the training E2A started, which consisted of two steps:

1. The real-world luminaires were turned on for 10 seconds.
2. When the luminaires went off there was a 2-seconds pause. Then the image for the specific condition appeared, with an assessment question at the top of the screen. An answer could be selected by means of the arrow and enter keys. The following questions were asked:
    - “How good do you think the MATCH is between this picture and the real luminaires in terms of brightness?”
    - “How bright does the luminaire IN THIS PICTURE appear to you?”
    - “How REALISTIC does this picture appear to you?”

    They were answered on 7-point scales, ranging from “Not at all good” to “Very good”, “Not at all bright” to “Very bright”, or “Not at all realistic” to “Very realistic” (see Figure 27). For clarification, brightness was defined as the amount of light produced by the luminaire(s), and realism was defined as the realism of the visualization in general. After answering the first question, the second question appeared right away, etcetera.

Figure 27: Illustration of the presentation method of the assessment questions with respect to an exemplary image stimulus. The "O" could be moved along the seven points of the scale to select an answer.

After the third question step 1 started for the next condition, and so on, until all 24 conditions with a reference were assessed. Thereafter E2B started, containing the conditions without a reference. In this case, the image for the first condition appeared, with an assessment question at the top of the screen. An answer could be selected by means of the arrow and enter keys. In this part of the experiment the following questions were asked:

- "How bright does the luminaire IN THIS PICTURE appear to you?"
- "How REALISTIC does this picture appear to you?"

Both questions were answered on a 7-point scale, ranging from "Not at all bright/realistic" to "Very bright/realistic". After answering the first question, the second question appeared right away. Answering the second question was followed by a 2-seconds pause of darkness (i.e., the display was black for 2 seconds). Then the next image appeared for the assessment of the two questions, until all 24 conditions were assessed.

7.2.7 Statistical Analysis

The data is analyzed with SPSS Statistics version 17.0. Before analyzing the effect of the independent variables *direction, luminaire intensity, image intensity, glare,* and *reference* on the dependent variables *match, brightness,* and *realism*, the assumption of normality, implying that the sampled distribution of variables should be normal when included in significance tests, was evaluated by testing if the dependent variables showed normal distributions for each level of the independent variables, i.e., per condition. The dependent variables were measured on a 7-point Likert scale, which means the scores are distributed over only seven values. Consequently, the samples were too chunky to describe a Gaussian function. Therefore, normality was checked not by relying on the test statistics, but by inspection of normal probability plots. In addition, homogeneity, which is the assumption that the error term is approximately the same across all values of the independent variables, was investigated. Levene's test for homogeneity of variance could not be used to check this assumption, because Levene's test assumes independent samples. However, since we have groups of equal sample size, our ANOVA is relatively robust, so even if homogeneity was not met, it would not be fatal to our analysis (Garson, 2012).

After checking for normality, three General Linear Model Univariate within-subject ANOVAs (hereafter abbreviated to ANOVA) were conducted: one for the analysis of the effect on each of the dependent variables (i.e., match, brightness, and realism). A Univariate ANOVA with *participant* as random factor generates the same results as a Repeated Measures ANOVA and has the benefit that it supports incomplete factorial designs, such as ours: [3 (glare) x 3 (image intensity) x 2 (luminaire intensity) x 2 (reference)] + [3 (image intensity) x 2 (luminaire intensity) x 2 (reference)]. The Univariate ANOVA is able to include the unbalanced independent variable *direction*, thereby incorporating the parts of the design before and after the plus sign.

The first ANOVA included all independent variables (i.e., *direction*, *luminaire intensity*, *image intensity*, *glare*, and *reference*) as fixed factors and *participant* as random factor, to analyze the effect on the dependent variable *brightness*. The model included all main effects and 2-way interactions of the independent and random factors. The main effect of glare pointed out if a direct brightness assessment conveyed the same results as the indirect brightness assessment of E1. The second ANOVA included all independent variables (i.e., *direction*, *luminaire intensity*, *image intensity*, *glare*, and *reference*) as fixed factors and *participant* as random factor to analyze the effect on the dependent variable *realism*, thereby revealing factors that have influenced the realism assessments of the visualization. The model included the main effects and all 2-way interactions of the fixed factors glare, direction, luminaire intensity, image intensity, and reference, and the random factor participant. Lastly, to analyze the effect on the dependent variable *match*, we performed an ANOVA that included the fixed factors *direction*, *luminaire intensity*, *image intensity*, and *glare,* and the random factor *participant*. The match was only measured when a reference was shown, so the condition 25-48, where no reference was shown, were not included in the model. Since there were only four fixed factors (glare, direction, luminaire intensity, and image intensity), the model not only included the main effects and all 2-way interactions but also all 3-way interactions of the fixed and random factors. The effect of glare on match is relevant to our second hypothesis. Pairwise comparisons with Bonferroni adjustments were done to analyze several significant effects in more detail. Where post-hoc analysis was required, we used the Tukey HSD algorithm.

7.3 Results

7.3.1 Introduction

The experiment yielded scores for brightness and realism for each of the 48 conditions, and scores for the goodness of match for 24 of the conditions, which were analyzed with the ANOVA models as specified in section 7.2.7. The conditions consisted of combinations of the independent variables direction, reference, glare image intensity, and luminaire intensity. See Appendix 11.1 for the variable definitions and Appendix 11.3.1 for the full list of conditions.

Before performing the ANOVAs, the assumption of normality was checked by inspecting the normal probability plots of the dependent variables for each condition. These yielded satisfactory results: the distributions of the observed values did not deviate from normality with values greater than 1 for all but two conditions. The two deviating conditions were condition 14 and 20, where the distribution in brightness scores was negatively skewed. However, no consistent outliers were identified.

7.3.2 Brightness Score

An ANOVA analyzed the effect on the brightness score, the dependent variable that is relevant to our main hypothesis.

Firstly, the analysis showed that the wall direction had significantly higher brightness scores ($F(1, 24) = 64.145$, $p < .001$, $\eta^2_{partial} = .728$), which is in line with the results from E1.

The ANOVA also showed that glare had a significant effect on the brightness score ($F(2, 48) = 9.257$, $p < .001$, $\eta^2_{partial} = .278$). Because glare only varied in the ceiling direction, we examined the estimated marginal means of glare x direction. As shown in Table 6, for the ceiling direction, the estimated marginal means increased with the level of glare, which indicates that there was a positive correlation between level of glare and brightness score. Furthermore, pairwise comparisons of the means showed that exaggerated glare is significantly higher than no glare ($p < .001$) and glare ($p < .001$), but glare and no glare did not differ significantly ($p = .198$).

Direction	Glare	Mean	Std. Error	95% Confidence Interval	
				Lower Bound	Upper Bound
Ceiling	No glare	2.553	0.049	2.457	2.650
	Glare	2.643	0.049	2.547	2.740
	Exaggerated glare	3.017	0.049	2.920	3.113
Wall	No glare	4.100	0.049	4.003	4.197

Table 6: Estimated marginal means of brightness for "direction x glare"

Luminaire intensity had a very strong effect on the brightness score ($F(1, 39.3) = 354.6$, $p < .001$, $\eta^2_{partial} = .900$). The marginal means in the low intensity condition ($M = 2.1$) are significantly lower ($p < .001$) than the high intensity condition ($M = 4.0$). This is not surprising, since the high intensity condition contained images of a higher intensity and thus a higher key value.

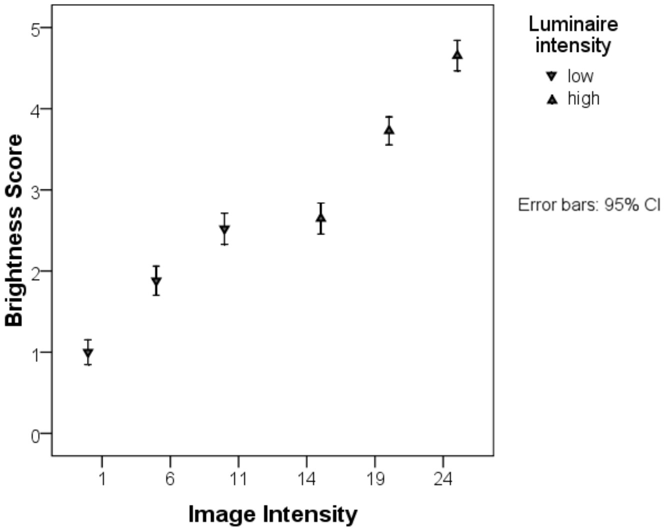
We further investigated the effect of image intensity with the dedicated variable, i.e., "image intensity". The ANOVA showed that image intensity had a significant effect on brightness score ($F(2, 145.7) = 144.4$, $p < .001$, $\eta^2_{partial} = .665$). The levels "below mean" ($M = 2.23$), "mean" ($M = 3.16$), and "above mean" (*M* = 3.84) all had significantly different brightness scores, indicating a positive correlation between key value and perceived brightness.

Image intensity also interacted with other independent variables. Firstly, there was an interaction effect with luminaire intensity with a small effect size ($F(2, 981) = 3.95$, $p = .019$, $\eta^2_{partial} = .008$). The effect can be explained with the graph in Figure 28, which displays the brightness score for each image intensity that was used in the ceiling direction. Each level of image intensity can be split in "high" and "low", as can be seen in Table 7. The interaction effect seems to be caused by a trend difference between level 2 and 3 of image intensity, where the mean difference in the "low" cluster becomes smaller as compared to the mean difference in the "high" cluster.

Table 7: The image intensities in the ceiling condition

Image Intensity		Luminaire Intensity *Low*	Luminaire Intensity *High*
Level 1	*Below Mean*	1	14
Level 2	*Mean*	6	19
Level 3	*Above Mean*	11	24

Figure 28: Error bar graph of mean brightness score versus image intensity, clustered by luminaire intensity



There is also an interaction effect of image intensity x direction ($F(2, 983) = 3.018$, $p = .049$, $\eta^2_{partial} = .006$). There seems to be a difference between the trend for wall and the trend for ceiling: the mean differences are larger for the ceiling condition. We should take into account that the wall direction had no conditions with synthetic glare applied to the visualization, which could explain the difference in brightness scores.

Lastly, the analysis showed that reference significantly affected the brightness score ($F(1, 19.3) = 10.11$, $p = .003$, $\eta^2_{partial} = .205$). Brightness scores were significantly higher ($p < .001$) without a reference (*M* = 3.24) than with a reference (*M* = 2.92). A possible reason for this effect is visual adaptation, which is supported by the interaction effect reference x luminaire intensity ($F(1, 983) = 35.23$, $p < .001$, $\eta^2_{partial} = .035$): in the low intensity conditions, the brightness score with reference (*M* = 2.12) does not differ significantly from the brightness score without reference (*M* = 2.14), while in the high intensity conditions, the scores were significantly higher ($p < .001$) when there was no reference (*M* = 4.33) than when there was a reference (*M* = 3.72).

7.3.3 Realism Score

The second ANOVA analyzed the effect on the realism score. Most of the variance in the model was explained by luminaire intensity ($F(1, 34.69) = 25.77$, $p < .001$, $\eta^2_{partial} = .426$). The visualizations in the low luminaire intensity condition were assessed as more realistic ($M = 3.40$) than the visualizations in the high luminaire intensity conditions ($M = 2.85$). Luminaire intensity also had an interaction effect with direction on realism ($F(1, 975) = 6.775$, $p = .009$, $\eta^2_{partial} = .007$). On average the realism scores were higher for the wall condition. Visual inspection of the interaction effect also revealed that the difference in realism between the luminaire intensity conditions high and low is larger for the wall condition.

Glare had the second largest effect on realism ($F(2, 48.1) = 15.53$, $p < .001$, $\eta^2_{partial} = .392$). The estimated marginal means of the individual levels “no glare”, “glare”, and “exaggerated glare” showed a decreasing trend ($M = 3.36$, $M = 3.14$, and $M = 2.64$, respectively). A post hoc Tukey HSD test confirmed that each mean was statistically significantly different from the other means, i.e., “no glare” was higher than “glare” ($p = .002$) and higher than “exaggerated glare” ($p < .001$), and “glare” was in turn higher than “exaggerated glare” ($p < .001$). Synthetic glare thus decreased the realism of the visualizations.

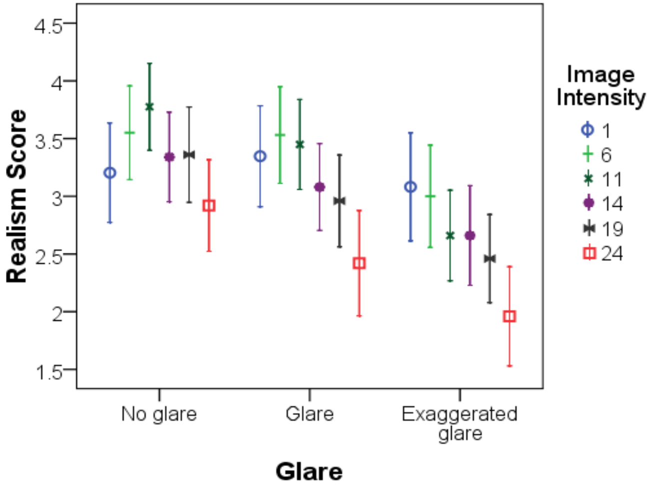
Also the effect of the independent variable image intensity was significant ($F(2, 92.81) = 4.895$, $p = .010$, $\eta^2_{partial} = .095$). Interestingly, the “mean” image intensity did not differ significantly from the “below mean” image intensity, but the “above mean” image intensity had a significantly lower realism score than both “mean” image intensity ($p = .004$) and “below mean” image intensity ($p = .002$). Realism hence decreased mainly when the image intensity was above the mean. The ANOVA also revealed that image intensity is involved in two two-way interactions: image intensity x glare ($F(4, 975) = 2.704$, $p = .029$, $\eta^2_{partial} = .011$) and image intensity x reference ($F(2, 975) = 4.685$, $p = .009$, $\eta^2_{partial} = .010$).

The estimated marginal means of image intensity x glare are shown in Table 8. We can recognize an increasingly decreasing trend of the means with increasing levels of both independent variables. Our presumption seems to be correct: the low realism scores of the “above mean” image intensity could have been caused by a reinforced effect of glare, since synthetic glare was applied *after* tone mapping of the image. To investigate this effect further, the different image intensities used in the ceiling condition

Table 8: Estimated marginal means of realism for “glare x image intensity”

Glare	Image intensity	Mean	Std. Error
No glare	Below mean	3.382	.070
	Mean	3.398	.070
	Above Mean	3.312	.070
Glare	Below mean	3.204^{a}	.100
	Mean	3.253^{a}	.100
	Above Mean	2.952^{a}	.100
Exagge-rated glare	Below mean	2.866^{a}	.100
	Mean	2.743^{a}	.100
	Above Mean	2.310^{a}	.099

Figure 29: Error bar graph of the realism scores versus glare level, clustered by image intensity



are plotted in an error bar graph, grouped by glare level (see Figure 29). The interaction effect follows a consistent pattern. For the image intensities in the high luminaire intensity condition (i.e., 14, 19, and 24 (see Table 7)) realism decreased with image intensity, and this effect is reinforced by glare. For image intensities in the low luminaire intensity condition (i.e., 1, 6, and 11, (see Table 7)) the direction of the effect of image intensity on realism is inverted by glare. In the no glare condition realism increases with image intensity, in the glare condition this trend flattens, and in the exaggerated glare condition this trend has been inverted: the realism score decreases with image intensity. Both glare and image intensity seem to decrease realism, and since synthetic glare was better visible for higher image intensities, the variables reinforce each other's negative effect on realism.

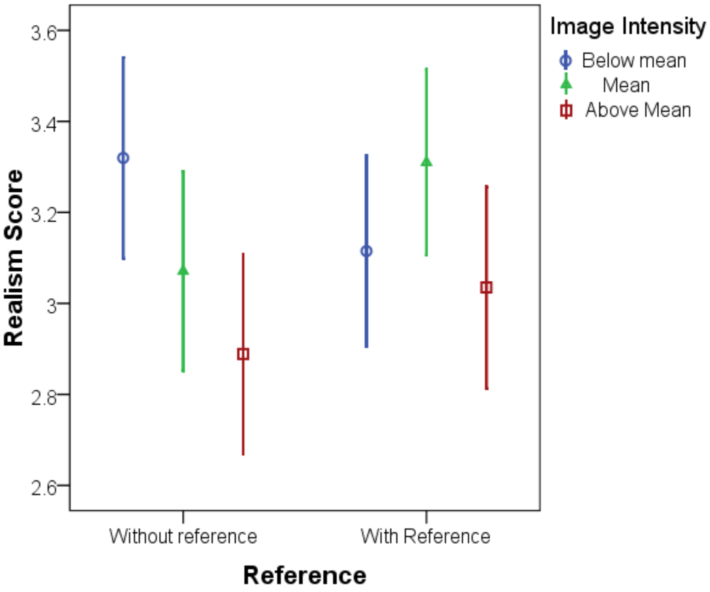
The estimated marginal means of the other interaction effect with image intensity, image intensity x reference, are shown in Table 9. Plotting the means in an error bar graph (see Figure 30) helps to distinguish two different trends. When a reference was viewed first, participants rated the realism the highest at mean image intensity, but when no reference was viewed, the visualization was rated the most realistic when the image intensity was below the mean.

Table 9: Estimated marginal means of realism for "image intensity x reference"

Reference	Image intensity	Mean	Std. Error
Without reference	Below mean	3.302[a]	.071
	Mean	3.086[a]	.071
	Above Mean	2.908[a]	.071
With reference	Below mean	3.115[a]	.070
	Mean	3.310[a]	.070
	Above Mean	3.035[a]	.070

[a] Based on modified population marginal mean

Figure 30: Error bar graph of the effect of reference on realism score, clustered by image intensity



7.3.4 Goodness of Match

We also analyzed the effect on the goodness of match (in terms of brightness) with the real-world reference. First of all, the match was significantly affected by luminaire intensity (*F*(1, 11.54) = 9.051, *p* = .011, $\eta^2_{partial}$ = .440). A pairwise comparison pointed out the low intensity level had a significantly higher score (*p* < .001) than the high intensity level (*M* = 3.38 and *M* = 2.80, respectively). Even more significant, although of a smaller effect size, was the interaction effect luminaire intensity x direction (*F*(1, 24) = 9.609, *p* = .005, $\eta^2_{partial}$ = .286). Table 10 contains the estimated marginal means of direction x intensity x glare. If we, for now, ignore the variance between levels of glare, we see that the effect of luminaire intensity only occurred when the light was directed to the ceiling. While in the low luminaire intensity condition the match score was relatively high when the light was directed to the ceiling, the match score is significantly lower ($p < .001$) in the high luminaire intensity condition. When the light was directed towards the wall, on the other hand, luminaire intensity has no effect.

Glare did not affect the match score ($F(2, 48) = .365$, $p = .696$, $\eta^2_{partial} = .015$), which means we have to reject our hypothesis that synthetic glare would improve the brightness match between the visualization and reference. A small effect of glare is visible in the interaction effect direction x luminaire intensity x glare. In Table 10 we already saw that high luminaire intensity decreased the match in the ceiling condition, but glare seems to counteract this effect. A pairwise comparison showed that the condition “ceiling direction, high luminaire intensity, glare” achieved a better match ($p = .048$) than the condition “ceiling, high luminaire intensity, no glare”.

Direction	**Luminaire Intensity**	**Glare**	**Mean**	**Std. Error**	**95% Confidence Interval**	
					Lower Bound	Upper Bound
Ceiling	Low	No glare	3.453	.146	3.164	3.743
		Glare	3.320	.146	3.031	3.609
		Exaggerated glare	3.373	.146	3.084	3.663
	High	No glare	2.400	.146	2.111	2.689
		Glare	2.813	.146	2.524	3.103
		Exaggerated glare	2.667	.146	2.377	2.956
Wall	Low	No glare	3.360	.146	3.071	3.649
	High	No glare	3.307	.146	3.017	3.596

Table 10: Estimated marginal means of match for “direction x luminaire intensity x glare”

Finally, we found that image intensity had no significant main effect on match ($F(2, 23.9) = 2.884$, $p = .075$, $\eta^2_{partial} = .194$). However, we recoded the variable based on the image intensity’s congruence with the E1 tunings, by redefining the levels as follows:

- “mean” image intensity → “congruent” image intensity
- “above mean” image intensity and “below mean” image intensity → “incongruent” image intensity

When we, in the ANOVA, replaced *image intensity* with *image intensity congruence*, we found a significant main effect ($F(1, 8.38) = 10.25$, $p = .012$, $\eta^2_{partial} = .550$). If image intensity was congruent with the E1 mean tuning, the brightness match was significantly better.

Lastly, image intensity x direction also showed some interesting, though not significant, effects, which are discussed in Appendix 11.3.4.

7.4 Discussion

The main conclusion is that glare had a positive effect on brightness, a negative effect on realism, and no effect on the match. Furthermore, it was found that the independent variables luminaire intensity and image intensity accounted for most of the variance in the dependent. The relations of the independent variables glare and image intensity with the dependent variables (in the ceiling direction) are summarized in Figure 31. In general, higher luminaire intensity (and thus also higher image intensity) led to higher brightness scores, but worse realism scores and a worse brightness match with the reference. Interestingly, the interpolation lines of the different graphs have resembling shapes with respect to each dependent variable. The best match was achieved by the mean image intensity, so we can conclude that the image intensity that was chosen in E1, was assessed to be the best match with the reference in terms of brightness. In absolute numbers, however, the “best” match was still not very good. Glare showed no effect on match. Both image intensity and glare positively correlate with brightness, but describe a slightly different curve. The effect of image intensity on brightness shows a steeper trend, than the effect of glare on brightness. Moreover, glare and image intensity both show a negative correlation with realism, which seems to be optimal around or just below the mean image intensity and with complete absence of synthetic glare. Furthermore, we found that the presence of the reference had a negative effect on the brightness scores. Reference also interacted with image intensity on the perceived realism. When the reference was absent, the realism ratings decreased, especially for the mean and above mean image intensities. Lastly, it was found that the wall condition was perceived to be significantly brighter.

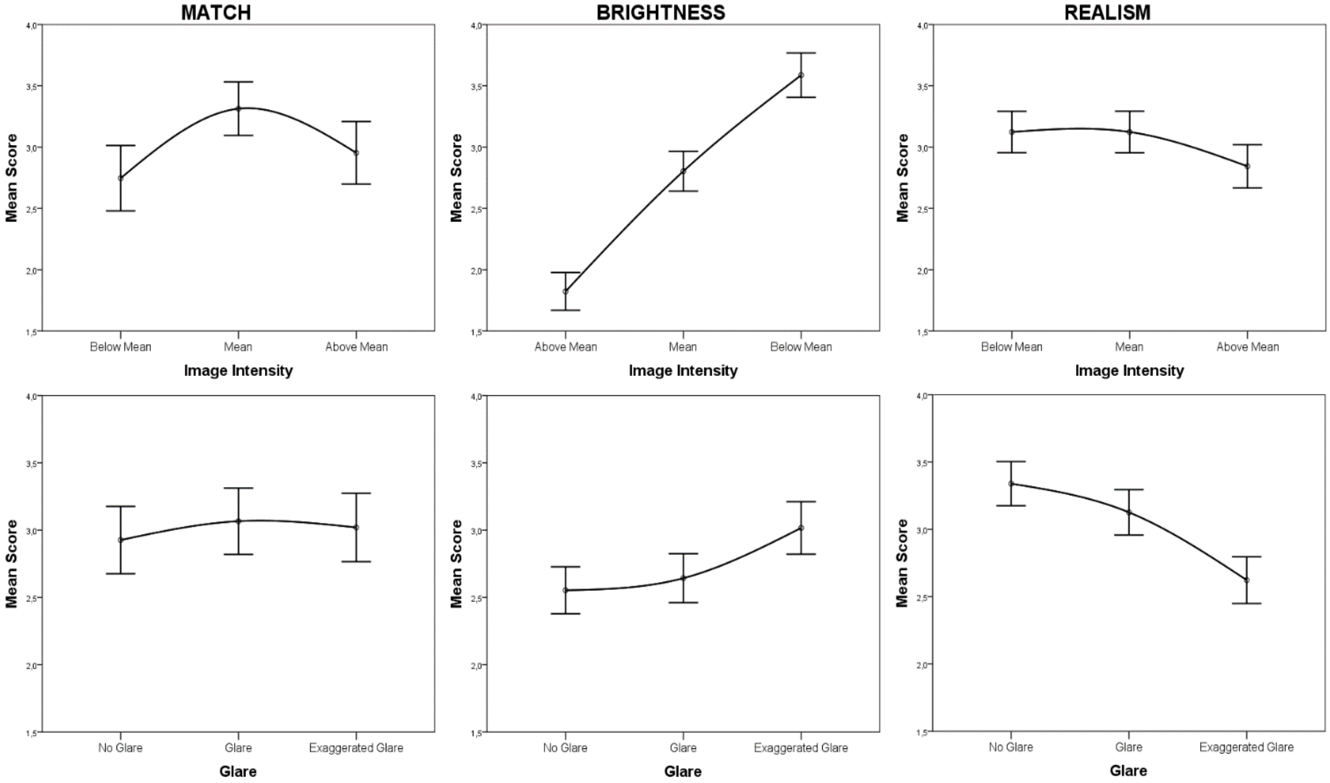


Figure 31: Error bar plots from the E2 data for the ceiling direction. First row from left to right, the effect of: image intensity on match, image intensity on brightness, and image intensity on realism. Second row from left to right, the effect of: glare on match, glare on brightness, and glare on realism. The error bars indicate the 95% confidence intervals.

Firstly, we see that the results are in line with E1. Image intensity can be translated to key value, which again shows that key value is a significant predictor of perceived brightness. Besides, glare was again found to increase perceived brightness, so our main hypothesis can be confirmed. Based on the visible trends in Figure 31, we observe that, compared to glare, image intensity caused a stronger increase in brightness and a less strong decrease in realism. Taking into account our finding that image intensity and glare reinforce each other's effects, we identify an inevitable tradeoff between brightness and realism (in the context of the used imaging pipeline).

Also, the results of match are noteworthy, because the effect of image intensity shows that the tuning results of E1 (represented by "mean image intensity") was the best match. Agreement between two different samples, each with a different task, over the image intensity that best matches the reference indicates that an accurate representation of the perceived brightness was achieved. For glare, however, no large effects on match were found. We cannot confirm our hypothesis that glare would improve the match with reality, especially when we take the negative effect on realism into account as well. A small effect was found, however, in the high intensity condition: synthetic glare had a positive, just significant effect on the match score. Hence, this effect was in correspondence with our hypothesis. Maybe the low realism also decreased the perceived match in brightness, which would imply that realism of synthetic glare is a contributing factor to its effect on brightness. This implication would contradict our assumption in section 4.3.1 that realism does not contribute substantially to the glare illusion.

Moreover, the presence of the reference negatively affected the perceived brightness. This suggests that the real-world luminaires introduced some adaptation effects. The analysis supports this suggestion by showing that the effect of the reference mainly occurred after viewing the luminaires on high intensity. The by-effects of potentially increased adaptation seem to only have decreased the size of the effect, but not the direction, and thus adaption did not change the interpretability of our results, i.e., the effects that were found with reference are even stronger without reference. Besides, the results without reference are more valuable for the domain of virtual prototyping, since in practice the end user views virtual brightness typically without a real-world reference. The finding that brightness scores were higher without reference is beneficial for the field, because it shows higher quantities of brightness were experienced in the absence of a real-world reference, despite the limited peak luminance of the display.

8 Discussion and Conclusion

In the introduction of this report we explained that, in the domain of virtual prototyping of lighting, brightness is an attribute that is especially hard to replicate in visualizations of lighting situations, mainly due to the limited display luminance range and the complex relation between luminance and brightness. The main objective of this report was to further investigate brightness in visualizations and explore the applicability of synthetic glare to influence brightness in these visualizations. The main hypothesis was that synthetic glare could increase perceived brightness. Secondly it was hypothesized that synthetic glare might improve the brightness match between the visualization and reality.

Two experiments (E1 and E2) were conducted to investigate our hypotheses. In E1 an indirect brightness assessment task was carried out by means of image intensity tuning. In E2 brightness, as well as realism and goodness of match, were assessed directly by means of evaluative questions. The results confirmed our main hypothesis: glare increased perceived brightness, whether it was assessed directly or indirectly. Our second hypothesis, however, was not confirmed by our results. Synthetic glare had no effect on the brightness match between visualizations and reality in general, and even had a negative effect on realism. Synthetic glare did have a small effect on the brightness match with reality only for the high intensity levels of the luminaire. Visualizations with a moderate amount of synthetic glare were assessed as a better brightness match with reality than the visualization without glare. Perhaps, synthetic glare would be able to improve the match with reality in general, if it would also achieve higher realism scores.

The results of E1 also revealed that glare only affected perceived brightness for the light directed to the ceiling, where actually the only source of light was aimed towards the observer. This conclusion suggests that glare only affects brightness when it is clearly visible, e.g., when coming from a point light source in front of a dark background. The effect of glare on brightness was consistent in both experiments: in both cases the “exaggerated glare” condition was assessed to be significantly brighter than the “glare” and “no glare” conditions. Application of (exaggerated) glare increased the perceived brightness by 29%, which is in line with the findings of Yoshida et al. (2008), who reported an increase of 20-35% in “perceived luminance” when applying synthetic glare. In E2, exaggerated glare also significantly increased the brightness score, but by no more than half a point on the 7-point scale.

Apparently, the chosen levels of glare did not increase brightness linearly. This is not unexpected, since the values selected for the synthetic glare (controlled via the center-rest ratio of the glare equation) were chosen visually. Further research could investigate the relation between center-rest ratio of synthetic glare and brightness, and potentially generalize this relation in a mathematical formula, as was done with the key value parameter.

Also, we have strong suspicion that our viewing methods did not fully control for light adaptation. In both experiments we detected by-effects suggesting incomplete adaptation. In E1 this was related to the repeated tuning: in the second tuning the mean image intensity increased in the high and medium condition of the luminaire intensity. In other words, an additional observation of the luminaires (at a luminous intensity that was much higher than the peak luminance of the display) increased the image intensity that was chosen to match the reference. A likely explanation is that the participants were not fully adapted during their first exposure with the high-intensity stimulus, and so, further adapted during the second exposure. Due to the increased adaptation during the second tuning, the display luminance of the visualization would be experienced as relatively dim, which would cause them to overcompensate that by tuning to a higher image intensity value. We cannot conclude with certainty that the increased memorization time of the repeated tuning led to better results than a single tuning, but the fact that the distributions of the results became tighter is an indication that a memory “refresher” improved the *reliability* of the tuning results. On the other hand, the increased light adaptation decreased their *validity* since under more common circumstances people don't have a high-intensity reference sample when judging a visualization. We know, however, that the lowered validity results in an underestimation of the image intensity, which can be taken into account when designing visualizations. Also in E2, where the reference was viewed only once, adaptation was suspected to influence the brightness scores. The results showed that the image intensity was underestimated when the real-world luminaires were viewed at high intensity. In this case, the brightness scores evaluated in the absence of a reference have higher validity, as they were assessed in a more complete state of adaptation. This is a beneficial effect, because the results without real-world reference are also more relevant for the domain of virtual prototyping.

A suggestion for future research would be to work towards a general formula for estimating the brightness of visualizations. Psychophysical formulas, such as the one we used to map the key value to a perceptually linear scale, are useful in the process of optimizing perceptual accuracy of visualizations. Besides the key parameter, the Reinhard ’02 TMO has another parameter that influences brightness, namely the white limit, which determines the cut-off point at which pixels of certain intensity get clipped to white. There is yet much to learn about the relation between the white limit and perceived brightness. Also, as was mentioned above, the relation between the center-rest ratio of synthetic glare and brightness could be investigated further. For the field of virtual prototyping of lighting, it would be most useful to work towards an algorithm that translates luminance to brightness by incorporating the effect of parameters, such as key value, white limit, and center-rest ratio.

8.1 Conclusions

This paper investigated the perceived brightness of visualizations of luminaires which varied in key value and synthetic glare. The main hypothesis that synthetic glare can increase the perceived brightness, was confirmed by both direct and indirect brightness assessments. We could not confirm the second hypothesis that synthetic glare can improve the brightness match with reality, but we found, on the contrary, that synthetic glare decreased realism. The realism of our visualizations, however, also decreased with high key values. Furthermore, our results suggest that a luminous spot on the wall is a stronger virtual brightness cue than the (glary) reflectors of a luminaire, and the former seems to overrule the latter. All in all we can conclude that the key value and synthetic glare are effective ways to increase the brightness of a visualization, but are also likely to decrease its realism.

9 Acknowledgements

This research was supported by the Visual Experience department of Philips Research, Eindhoven. Many thanks to my supervisors - Ingrid Heynderickx (TU/e), Michael Murdoch (Philips), Marc Lambooij (Philips), and Mariska Stokkermans (TU/e) - for their guidance, feedback, and pleasant collaboration. Furthermore, I want to thank all my colleagues at Philips Research for giving me a great time during my internship, and special thanks to Guy Dubois for helping me to build my experimental setup. Lastly, I want to express my gratitude to all the participants for volunteering in my experiments.

10 References

Abramov, I., Gordon, J., & Chan, H. (1991). Color appearance in the peripheral retina: effects of stimulus size. J. Opt. Soc. Am. A 8(2), 404-414.

Adelson, E.H. (2001). Lightness Perception and Lightness Illusions. In Gazzaniga, M (Ed.), *The New Cognitive Neurosciences* (2nd ed, pp. 339-351), Cambridge, MA: MIT Press.

Barati, B. (2012). *Lighting Probe for Apparel Retail Application.* Delft: Delft University of Technology.

Boyce, P.R. (2003). Lighting and the perception of spaces and objects. In Boyce, P.R. (Ed.) *Human Factors in Lighting* (pp 192-121). London: Taylor & Francis.

Braun, K. M., Fairchild, M. D., & Alessi, P. J. (1996). Viewing techniques for cross-media image comparisons. Color Research & Application, 21(1), 6-17.

Čadík, M., Wimmer, M., Neumann, L. & Artusi, A. (2008). Evaluation of HDR tone mapping methods using essential perceptual attributes. *Computers & Graphics, 32*(6), 716-719.

Commission Internationale de l‟Éclairage (CIE), 1987. International Lighting Vocabulary, *Bureau Central de la Commission Electrotechnique International,* CIE publication 17(4).

Damkat, C. (2013). Technical comparison of 3D renderers in terms of speed and quality. Eindhoven: Philips Research.

Engeldrum, P. (2000). *Psychometric Scaling: A Toolkit for Imaging Systems Development*, Winchester, MA: Imcotek Press, 1st edition.

Engelke, U., Stokkermans, M.G.M., Murdoch, M. (2013). Visualizing lighting with images: converging between the predictive value of renderings and photographs, *Proc. SPIE 8651, Human Vision and Electronic Imaging XVIII,* 86510L.

Fairchild, M. D. (2013). *Color appearance models*. John Wiley & Sons.

Fotios, S. (2006). Chromatic adaptation and the relationship between lamp spectrum and brightness, Lighting Res. Technol., 38(1), 3-17.

Fotios S.A., Cheal C. (2007). Lighting for subsidiary streets: Investigation of lamps of different SPD. Part 2 – Brightness. *Lighting Research and Technology,* 39(3), 233–252.

Field, A.P. (2013). *Discovering Statistics using SPSS: and sex and drugs and rock 'n' roll* (4th Edition). London: Sage.

Garson, G. D. (2012). *Univariate GLM, ANOVA, & ANCOVA*. Asheboro, NC: Statistical Associates Publishers.

Hoefflinger, B. (Ed.). (2007). *High-Dynamic-Range (HDR) Vision*. Springer Berlin Heidelberg.

Hsee, C., Loewenstein, G., Blount, S., Bazerman, M. (1999). Preference Reversals Between Joint and Separate Evaluations of Options: A Review and Theoretical Analysis. *Psychological Bulletin,* 125(5), 576-590. American Psychological Association, Inc.

Keil, M.S. (2007). Gradient representations and the perception of luminosity. *Vision Research*, 47(27), 3360‑3372.

Kingdom, F. (2011). Lightness, brightness and transparency: a quarter century of new ideas, captivating demonstrations and unrelenting controversy. *Vision Research,* 51(7), 652–673.

Lambooij, M. (2013). Perceived virtual lighting in relation to rendering quality and speed. Eindhoven: Philips Research.

Ledda, P., Chalmers, A., Troscianko, T. & Seetzen, H. Evaluation of tone mapping operators using high dynamic range display. *ACM Transactions on Graphics*, 24(3), 640-648. New York: ACM.

Lighting Research Center. (2004). *What is colour rendering index?*, retrieved on 14-08-2014 http://www.lrc.rpi.edu/programs/nlpip/lightinganswers/lightsources/whatisColorRenderingIndex.asp)

Lund Research Ltd (2013), Laerd Statistics, *Repeated Measures ANOVA – Understanding Repeated Measures ANOVA.* Retrieved 05/01/2015 from https://statistics.laerd.com/statistical-guides/repeated-measures-anova-statistical-guide.php.

Mace, D., Garvey, P., Porter, RJ., Schwab, R., Adrian, W. (2001). *Countermeasures for reducing the effects of headlight glare*. Washington, D.C.: The AAA Foundation for Traffic Safety.

McCann, J.J. (2007). Art, science, and appearance in HDR. *Journal of the Society for Information Display,* 15(9), 709–719.

Menozzi, M., v. Buol, A., Krueger, H., & Miege, Ch. (1967). Direction of gaze and comfort: discovering the relation for the ergonomic optimization of visual tasks. *Opthalmic and physiological optics*. Vol 14(4), Bartleson.

Murdoch, M.J., Engelke, U., & Stokkermans, M.G.M. (2013). *Virtual Prototyping 2012: Visualization of Lighting Systems in Context.* Eindhoven: Philips Research.

Reinhard, E., Stark, M., Shirley, P., & Ferwerda, J. (2002). Photographic tone reproduction for digital images. *ACM Trans. Graph.* 21(3), 267-276.

Seetzen, H., Heidrich, W., Stuerzlinger, W., Ward, G., Whitehead, L., Trentacoste, M., Ghosh, A., Vorozcovs, A. (2004). High Dynamic Range Display Systems. In *Proceedings SIGGRAPH '04,* 23(3), 760-768. New York: ACM.

Shapley, R., & Enroth-Cugell, C. (1984). Visual adaptation and retinal gain controls. In N. N. Osbourne & G. J. Chader (Eds*.), Progress in retinal research*, 263–346. Oxford: Pergamon Press.

Spencer, G., Shirley, P., Zimmerman, K., & Greenberg, D.P. (1995). Physically-based glare effects for digital images. In *Proceedings of ACM SIGGRAPH*, 325-334. New York: ACM.

Spillman, L. & Werner, J.S. (Eds.). (1989). *Visual perception: The neurophysiological foundations*. San Diego, California: Academic Press, Inc.

Swinkels, S.H., Murdoch, M.J., Heynderickx, I.E.J. (2009). Limits of Contrast Perception in LCD Displays. Eindhoven: Philips Research.

Takahashi, H., Yaguchi, H., Shioiri, S. (1999). Estimation of Brightness and Lightness in All Adaptation Levels. *Journal of Light & Visual environment*, 23(2).

Tang, X., Chen, Y., Chen, Y. & Xu, C. (2014). Matching the brightness of the rendering with the real room. Eindhoven: Philips Research.

Teoh, V. (2014, October 17). Panasonic TX42AS650B LED TV Review. Retrieved January 24, 2015, from http://www.hdtvtest.co.uk/news/tx42as650b-201410173927.htm

Tiller, D.K. & Veitch, J.A. (1995). Perceived room brightness: Pilot study on the effect of luminance distribution. *Lighting Research & Technology,* 27 (2), 93-101.

Tumblin, J., Rushmeier, H. (1991). Tone Reproduction for Realistic Computer Generated Images. Georgia Institute of Technology.

Vidovszky-Németh, A., & Schanda, J. (2012). White light brightness–luminance relationship. Lighting Research and Technology, 44(1), 55-68.

Villa, C., & Labayrade, R. (2010). Psychovisual assessment of tone-mapping operators for global appearance and colour reproduction. *Proc. of Colour in Graphics Imaging and Vision*, *2*.

Vos, J., Van den Berg, T., (1999). Report on Disability Glare. CIE Collection 1999, Vision and Colour; Physical Measurement of Light and Radiation.

Vos, J. (2003). On the Cause of Disability Glare and its Dependence on Glare Angle, Age and Ocular Pigmentation. *Clinical and Experimental Optometry,* 86(6), 363-370.

Yoshida, A., Blanz, V., Myszkowski, K., & Seidel, H. (2005). Perceptual evaluation of tone mapping operators with real-world scenes. In *Stereoscopic Displays and Virtual Reality Systems XII, Proc. SPIE*, 5666, 192-203. (2005).

Yoshida, A., Ihrke, M., Mantiuk, R., & Seidel, H. P. (2008, August). Brightness of the glare illusion. In Proceedings of the 5th symposium on Applied perception in graphics and visualization (pp. 83-90). ACM.

Zavagno, D., & Caputo, G. (2001). The glare effect and the perception of luminosity. *Perception,* 30(2), 209–222.

11 Appendix

11.1 Variable Definitions

EXPERIMENT 1	
Independent variables	
Glare	Level of synthetic glare applied to the visualization
Direction	Direction of the light, relating to both the real-world scenes and the visualizations
Luminaire Intensity	Intensity of the real-world luminaires
Dependent variable	
Image Intensity	Intensity of the visualization, translatable to key (and related to perceived brightness)

EXPERIMENT 2	
Independent variables	
Glare	Level of synthetic glare applied to the visualization
Direction	Direction of the light, relating to both the real-world scenes and the visualizations
Luminaire Intensity	Intensity of the real-world luminaires, translatable to luminance
Image Intensity	Intensity of the visualization, translatable to key (and related to perceived brightness)
Reference	Presence of the real-world reference
Dependent variable	
Match Score	Assessment of how good the visualization matched the real-world reference in terms of brightness
Brightness Score	Assessment of the brightness of the luminaire in the visualization
Realism Score	Assessment of the realism of the visualization

11.2 Experiment 1

11.2.1 Conditions

Condition	Intensity	Glare	Direction
111	Low	No	Wall
112	Low	No	Ceiling
113	Low	No	Both
121	Low	Yes	Wall
122	Low	Yes	Ceiling
123	Low	Yes	Both
131	Low	Exaggerated	Wall
132	Low	Exaggerated	Ceiling
133	Low	Exaggerated	Both
211	Medium	No	Wall
212	Medium	No	Ceiling
213	Medium	No	Both
221	Medium	Yes	Wall
222	Medium	Yes	Ceiling
223	Medium	Yes	Both
231	Medium	Exaggerated	Wall
232	Medium	Exaggerated	Ceiling
233	Medium	Exaggerated	Both
311	High	No	Wall
312	High	No	Ceiling
313	High	No	Both
321	High	Yes	Wall
322	High	Yes	Ceiling
323	High	Yes	Both
331	High	Exaggerated	Wall
332	High	Exaggerated	Ceiling
333	High	Exaggerated	Both

11.2.2 Briefing

The text below is the verbal explanation given to the participants prior to Experiment 1. The explanation is given in combination with step by step demonstrations:

Welcome and thank you for participating in my experiment. I am investigating brightness in virtual lighting applications. In this experiment I want you to compare the brightness of real-world luminaires to images of the luminaires, which will appear on the display. In the context of this experiment, brightness is defined as the amount of light you perceive to be produced by the luminaire, if that luminaire is on. But first I will explain the procedure to you. I would like to ask you to put your head on the chin rest. Firstly, the luminaires will go on for 10 seconds. During this period, try to memorize the brightness you perceive as well as possible. Then the luminaires go off again and an image of the luminaires will appear on the display. This image is adjustable; using the arrow keys you can tune the image's brightness. The goal of the task is to match the brightness of the luminaires in the image to the previously seen real luminaires. With the 'left' and 'right' arrow keys small steps are taken, with the 'up' and 'down' arrows big steps in brightness can be taken. When you have tuned to the right brightness, you can confirm by pressing enter. Then, you will get to see the lights one more time, for 10 seconds, such that you can refresh your memory. Finally, the image appears again with the previously chosen brightness. It can now be adjusted if deemed necessary. On pressing enter, you will proceed to the next scene.

11.2.3 Informed Consent

Perception of Visualizations

Dear sir/madam,

You are kindly invited to participate in a perceptual experiment. Participation in this experiment is strictly voluntary and you may choose to stop participating and withdraw at any time, with or without providing reason. Please indicate the researcher if you wish to stop.

The goal of this study is to understand lighting perception in a virtual scene and how this compares to a real scene. Such information can be used for technology development, perceptual experiments, concept validation, and marketing communication.

In this experiment the brightness of a real luminaire seen has to be matched with a virtual luminaire which is viewed thereafter. Since the scope of the study is visual perception, preceding the experiment you are asked to participate in a short visual acuity test. The task is to match 27 images to the real scene in terms of brightness followed by four evaluative questions.

All vision screening results and test responses will be handled anonymously and confidentially. Your name will not be included or in any other way associated with the data collected in this study.

Consent:

- I have been informed about the objective of the investigation and my role in it and all my questions about the experiment have been answered by the responsible researcher.
- I had sufficient time to consider my participation in this investigation and I am aware that it is completely voluntarily.
- The potential risks associated with my participation in this investigation and the anticipated benefits have been discussed with me.
- I realize that I may decide to refuse participation or stop participation at any time, with or without providing reason.
- I understand and agree that data about me will be collected and processed, either manually or by computer, by the responsible researcher and other researchers in the project.
- I understand that my directly identifying personal data (e.g., name, address, etc.) will be separated from the research data and replaced by an assigned number/code
- I understand that I am entitled to access the personal information collected about me and to have inaccuracies corrected.

Participant name: ______________________

Participant signature: ______________________

Date: ____________

11.2.4 Statistical tables

First and Second Tuning Comparison

Comparisons of mean and standard deviation of the first and second tuning per condition

Condition #		Tuning 1	Tuning 2
1	Mean	13.31	12.85
	Std. Deviation	5.417	4.953
2	Mean	19.69	20.46
	Std. Deviation	6.329	6.275
3	Mean	28.38	29.27
	Std. Deviation	8.434	7.481
4	Mean	12.73	11.88
	Std. Deviation	6.116	5.949
5	Mean	21.04	20.62
	Std. Deviation	7.074	6.934
6	Mean	26.73	28.08
	Std. Deviation	8.283	7.61
7	Mean	12.27	11.92
	Std. Deviation	6.521	6.254
8	Mean	19.88	20.19
	Std. Deviation	5.771	6.125
9	Mean	26.88	27.5
	Std. Deviation	9.395	9.201
10	Mean	7.62	7.04
	Std. Deviation	6.5	5.758
11	Mean	14.12	13.88
	Std. Deviation	6.784	6.689
12	Mean	18.96	21.04
	Std. Deviation	6.576	6.422
13	Mean	8.65	6.92
	Std. Deviation	7.456	6.468
14	Mean	12.77	12.38
	Std. Deviation	6.766	6.518
15	Mean	18.31	19.85
	Std. Deviation	5.129	3.823
16	Mean	6.15	5.08
	Std. Deviation	6.577	5.344
17	Mean	10.35	10.23
	Std. Deviation	5.872	5.006
18	Mean	15.58	16.12
	Std. Deviation	5.825	5.722
19	Mean	10.12	10.54
	Std. Deviation	5.48	4.81
20	Mean	18.88	20.46
	Std. Deviation	6.029	5.874
21	Mean	24.38	27.27
	Std. Deviation	7.762	6.385
22	Mean	10.46	10.46
	Std. Deviation	6.307	5.413
23	Mean	16.42	18.54
	Std. Deviation	6.469	6.243
24	Mean	28.04	28.15
	Std. Deviation	8.563	7.561
25	Mean	11.38	10.27
	Std. Deviation	5.419	4.805
26	Mean	18.31	19.42
	Std. Deviation	6.071	6.401
27	Mean	27.38	29.27
	Std. Deviation	7.392	6.983
Total	Mean	16.99	17.4
	N	702 (27 * 26)	702 (27 * 26)
	Std. Deviation	9.355	9.594

Cluster Analysis

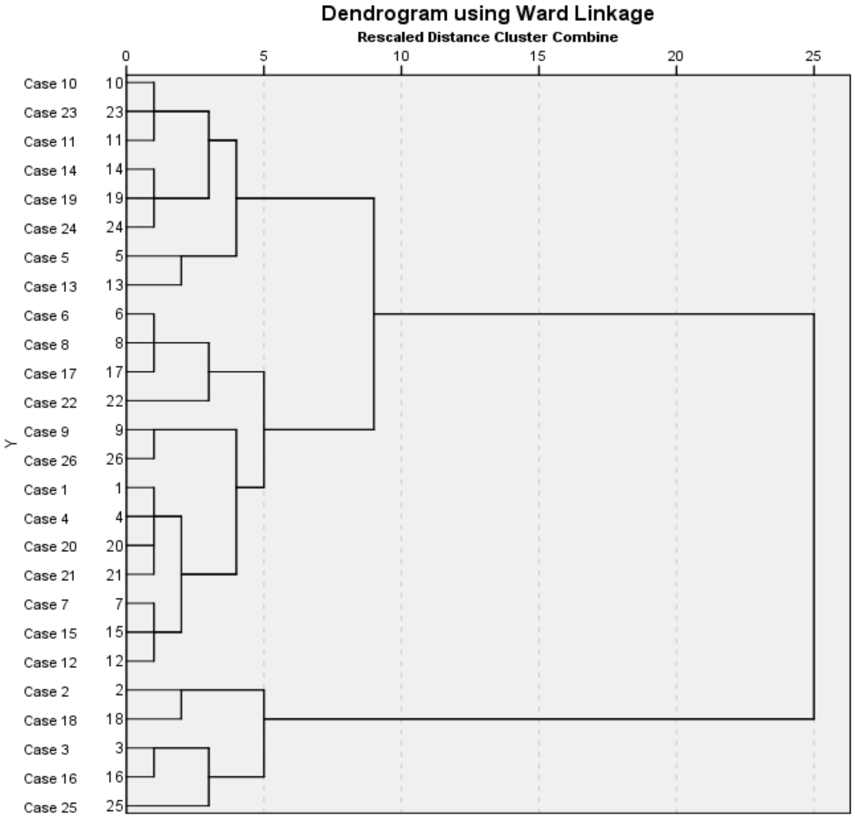


Figure 32: Hierarchical cluster analysis

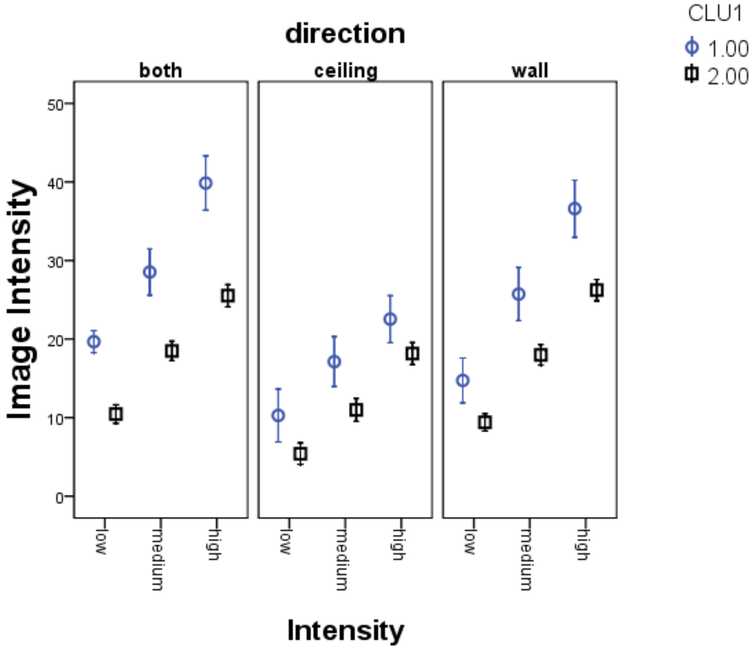
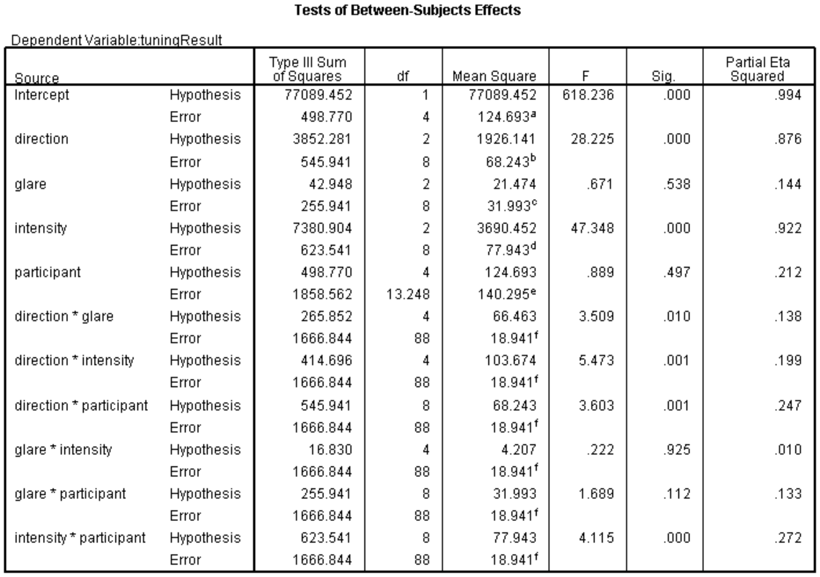


Figure 33: Error bar plots (Image Intensity (y-axis) versus Luminaire Intensity (x-axis)) for each direction, clustered based on the dendogram in Figure 32

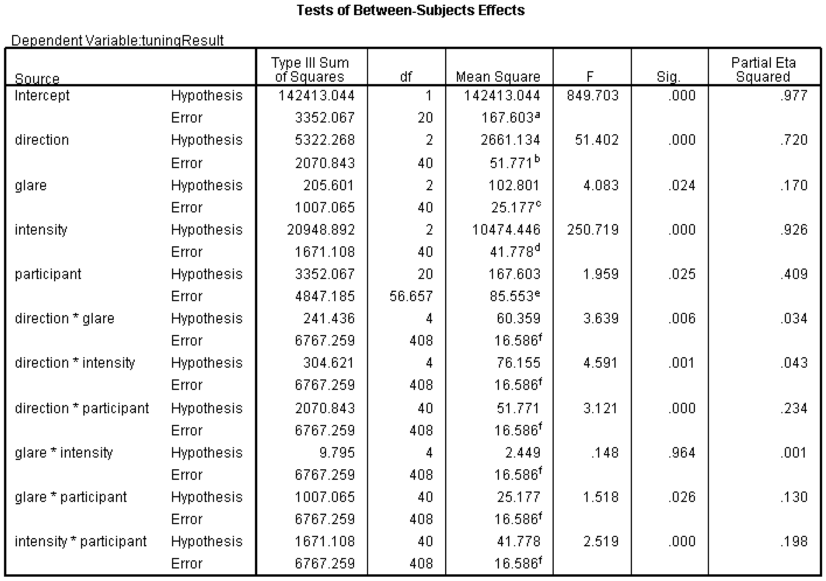
The most obvious difference between the clusters is the absolute difference in chosen image intensities. Cluster 1 consistently tuned much higher. Since both clusters have effects in the same direction, there is no reason to remove any cases from the data. To make sure they did not behave differently, an ANOVA on Image Intensity was done with both clusters. Besides stronger effects than cluster 2, cluster 1 also had an additional main effect: glare (ignoring the main effect of participant).



Tests of Between-Subjects Effects

Dependent Variable:tuningResult

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	77089.452	1	77089.452	618.236	.000	.994
	Error	498.770	4	124.693[a]			
direction	Hypothesis	3852.281	2	1926.141	28.225	.000	.876
	Error	545.941	8	68.243[b]			
glare	Hypothesis	42.948	2	21.474	.671	.538	.144
	Error	255.941	8	31.993[c]			
intensity	Hypothesis	7380.904	2	3690.452	47.348	.000	.922
	Error	623.541	8	77.943[d]			
participant	Hypothesis	498.770	4	124.693	.889	.497	.212
	Error	1858.562	13.248	140.295[e]			
direction * glare	Hypothesis	265.852	4	66.463	3.509	.010	.138
	Error	1666.844	88	18.941[f]			
direction * intensity	Hypothesis	414.696	4	103.674	5.473	.001	.199
	Error	1666.844	88	18.941[f]			
direction * participant	Hypothesis	545.941	8	68.243	3.603	.001	.247
	Error	1666.844	88	18.941[f]			
glare * intensity	Hypothesis	16.830	4	4.207	.222	.925	.010
	Error	1666.844	88	18.941[f]			
glare * participant	Hypothesis	255.941	8	31.993	1.689	.112	.133
	Error	1666.844	88	18.941[f]			
intensity * participant	Hypothesis	623.541	8	77.943	4.115	.000	.272
	Error	1666.844	88	18.941[f]			



Tests of Between-Subjects Effects

Dependent Variable:tuningResult

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	142413.044	1	142413.044	849.703	.000	.977
	Error	3352.067	20	167.603[a]			
direction	Hypothesis	5322.268	2	2661.134	51.402	.000	.720
	Error	2070.843	40	51.771[b]			
glare	Hypothesis	205.601	2	102.801	4.083	.024	.170
	Error	1007.065	40	25.177[c]			
intensity	Hypothesis	20948.892	2	10474.446	250.719	.000	.926
	Error	1671.108	40	41.778[d]			
participant	Hypothesis	3352.067	20	167.603	1.959	.025	.409
	Error	4847.185	56.657	85.553[e]			
direction * glare	Hypothesis	241.436	4	60.359	3.639	.006	.034
	Error	6767.259	408	16.586[f]			
direction * intensity	Hypothesis	304.621	4	76.155	4.591	.001	.043
	Error	6767.259	408	16.586[f]			
direction * participant	Hypothesis	2070.843	40	51.771	3.121	.000	.234
	Error	6767.259	408	16.586[f]			
glare * intensity	Hypothesis	9.795	4	2.449	.148	.964	.001
	Error	6767.259	408	16.586[f]			
glare * participant	Hypothesis	1007.065	40	25.177	1.518	.026	.130
	Error	6767.259	408	16.586[f]			
intensity * participant	Hypothesis	1671.108	40	41.778	2.519	.000	.198
	Error	6767.259	408	16.586[f]			

The difference in the effect of glare is explored by examining Figure 34. Judging from the trends in the both direction, it seems that participants reacted differently to glare. This is possible, because the both direction has two brightness cues, i.e., the spot and the reflectors. It seems that the cluster 1 did not pay any attention to the reflectors, and even chose higher image intensities when glare was present, while cluster two did react to glare, by choosing lower image intensities. Because of the small size of cluster 1, it did not change our general results, but it is interesting that some people ignored the reflectors and only focused on the spot on the wall.

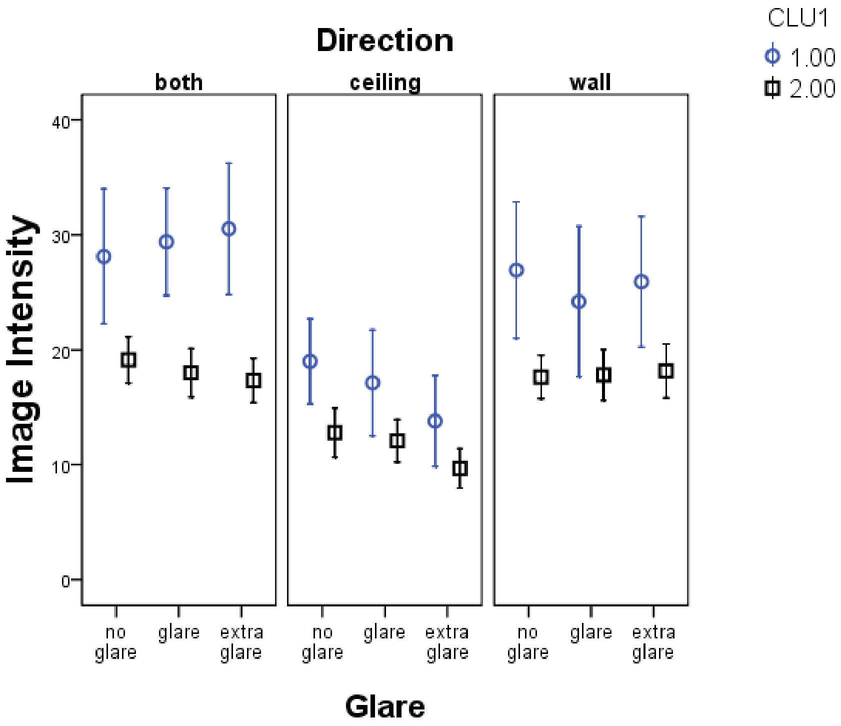
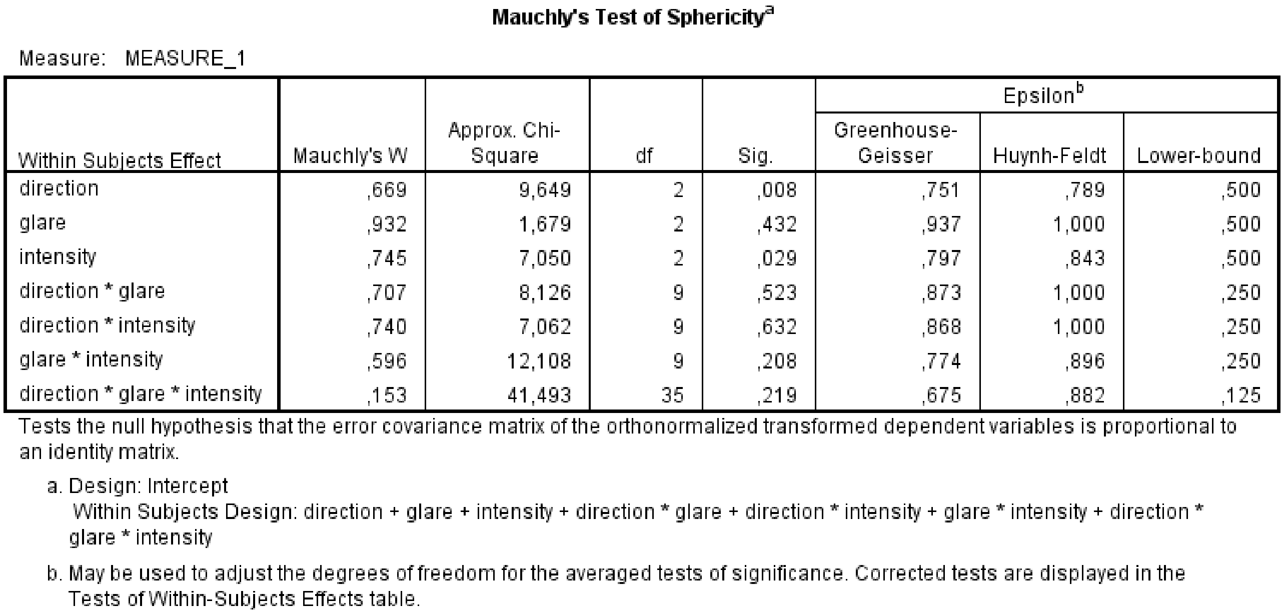


Figure 34: Error bar plots (Image Intensity (y-axis) versus Glare (x-axis)) for each direction, clustered based on the dendogram in Figure 32

Repeated Measures ANOVA: Complete Model



Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

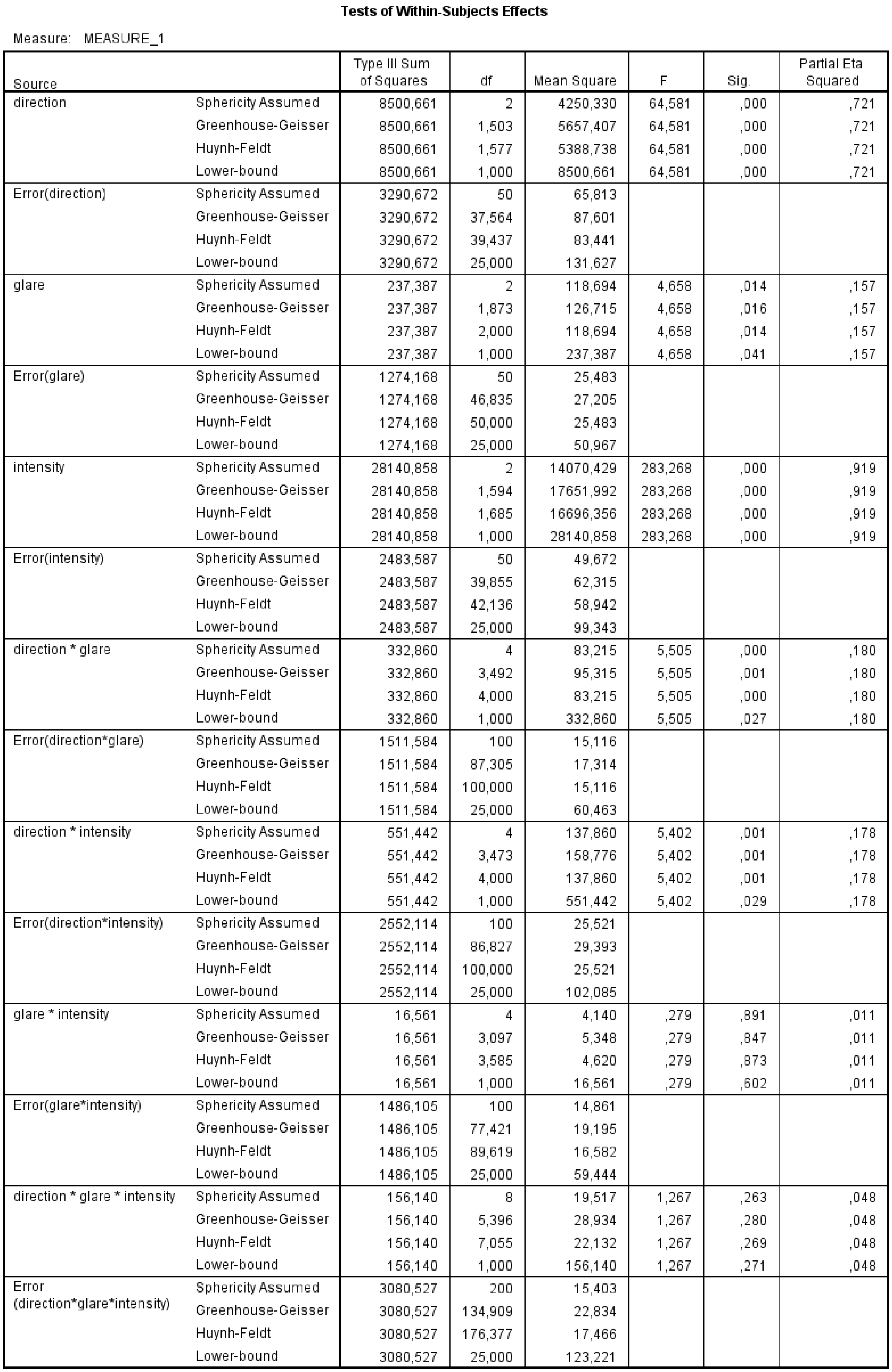
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
direction	,669	9,649	2	,008	,751	,789	,500
glare	,932	1,679	2	,432	,937	1,000	,500
intensity	,745	7,050	2	,029	,797	,843	,500
direction * glare	,707	8,126	9	,523	,873	1,000	,250
direction * intensity	,740	7,062	9	,632	,868	1,000	,250
glare * intensity	,596	12,108	9	,208	,774	,896	,250
direction * glare * intensity	,153	41,493	35	,219	,675	,882	,125

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: direction + glare + intensity + direction * glare + direction * intensity + glare * intensity + direction * glare * intensity

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Sphericity can be assumed for glare, direction x glare, direction x intensity, glare x intensity, and direction x glare x intensity, as Mauchly's Test of Sphericity is rejected. Mauchly's Test of Sphericity shows that the assumption of sphericity has not been met for direction and intensity. Therefore, to correct for the violation of sphericity, the degrees of freedom of direction were adjusted using Huynh-Feldt (ε = .79). The Huynh-Feldt correction was also used to adjust the degrees of freedom of intensity (ε = .84).

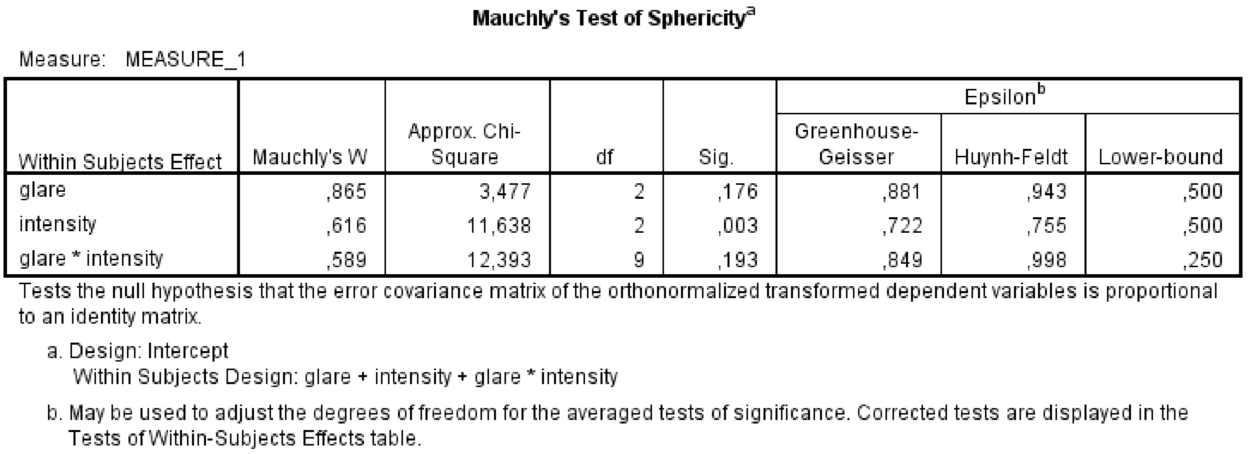


Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
direction	Sphericity Assumed	8500,661	2	4250,330	64,581	,000	,721
	Greenhouse-Geisser	8500,661	1,503	5657,407	64,581	,000	,721
	Huynh-Feldt	8500,661	1,577	5388,738	64,581	,000	,721
	Lower-bound	8500,661	1,000	8500,661	64,581	,000	,721
Error(direction)	Sphericity Assumed	3290,672	50	65,813			
	Greenhouse-Geisser	3290,672	37,564	87,601			
	Huynh-Feldt	3290,672	39,437	83,441			
	Lower-bound	3290,672	25,000	131,627			
glare	Sphericity Assumed	237,387	2	118,694	4,658	,014	,157
	Greenhouse-Geisser	237,387	1,873	126,715	4,658	,016	,157
	Huynh-Feldt	237,387	2,000	118,694	4,658	,014	,157
	Lower-bound	237,387	1,000	237,387	4,658	,041	,157
Error(glare)	Sphericity Assumed	1274,168	50	25,483			
	Greenhouse-Geisser	1274,168	46,835	27,205			
	Huynh-Feldt	1274,168	50,000	25,483			
	Lower-bound	1274,168	25,000	50,967			
intensity	Sphericity Assumed	28140,858	2	14070,429	283,268	,000	,919
	Greenhouse-Geisser	28140,858	1,594	17651,992	283,268	,000	,919
	Huynh-Feldt	28140,858	1,685	16696,356	283,268	,000	,919
	Lower-bound	28140,858	1,000	28140,858	283,268	,000	,919
Error(intensity)	Sphericity Assumed	2483,587	50	49,672			
	Greenhouse-Geisser	2483,587	39,855	62,315			
	Huynh-Feldt	2483,587	42,136	58,942			
	Lower-bound	2483,587	25,000	99,343			
direction * glare	Sphericity Assumed	332,860	4	83,215	5,505	,000	,180
	Greenhouse-Geisser	332,860	3,492	95,315	5,505	,001	,180
	Huynh-Feldt	332,860	4,000	83,215	5,505	,000	,180
	Lower-bound	332,860	1,000	332,860	5,505	,027	,180
Error(direction*glare)	Sphericity Assumed	1511,584	100	15,116			
	Greenhouse-Geisser	1511,584	87,305	17,314			
	Huynh-Feldt	1511,584	100,000	15,116			
	Lower-bound	1511,584	25,000	60,463			
direction * intensity	Sphericity Assumed	551,442	4	137,860	5,402	,001	,178
	Greenhouse-Geisser	551,442	3,473	158,776	5,402	,001	,178
	Huynh-Feldt	551,442	4,000	137,860	5,402	,001	,178
	Lower-bound	551,442	1,000	551,442	5,402	,029	,178
Error(direction*intensity)	Sphericity Assumed	2552,114	100	25,521			
	Greenhouse-Geisser	2552,114	86,827	29,393			
	Huynh-Feldt	2552,114	100,000	25,521			
	Lower-bound	2552,114	25,000	102,085			
glare * intensity	Sphericity Assumed	16,561	4	4,140	,279	,891	,011
	Greenhouse-Geisser	16,561	3,097	5,348	,279	,847	,011
	Huynh-Feldt	16,561	3,585	4,620	,279	,873	,011
	Lower-bound	16,561	1,000	16,561	,279	,602	,011
Error(glare*intensity)	Sphericity Assumed	1486,105	100	14,861			
	Greenhouse-Geisser	1486,105	77,421	19,195			
	Huynh-Feldt	1486,105	89,619	16,582			
	Lower-bound	1486,105	25,000	59,444			
direction * glare * intensity	Sphericity Assumed	156,140	8	19,517	1,267	,263	,048
	Greenhouse-Geisser	156,140	5,396	28,934	1,267	,280	,048
	Huynh-Feldt	156,140	7,055	22,132	1,267	,269	,048
	Lower-bound	156,140	1,000	156,140	1,267	,271	,048
Error (direction*glare*intensity)	Sphericity Assumed	3080,527	200	15,403			
	Greenhouse-Geisser	3080,527	134,909	22,834			
	Huynh-Feldt	3080,527	176,377	17,466			
	Lower-bound	3080,527	25,000	123,221			

Repeated Measures ANOVA: Ceiling Condition



Mauchly's Test of Sphericity[a]

Measure: MEASURE_1

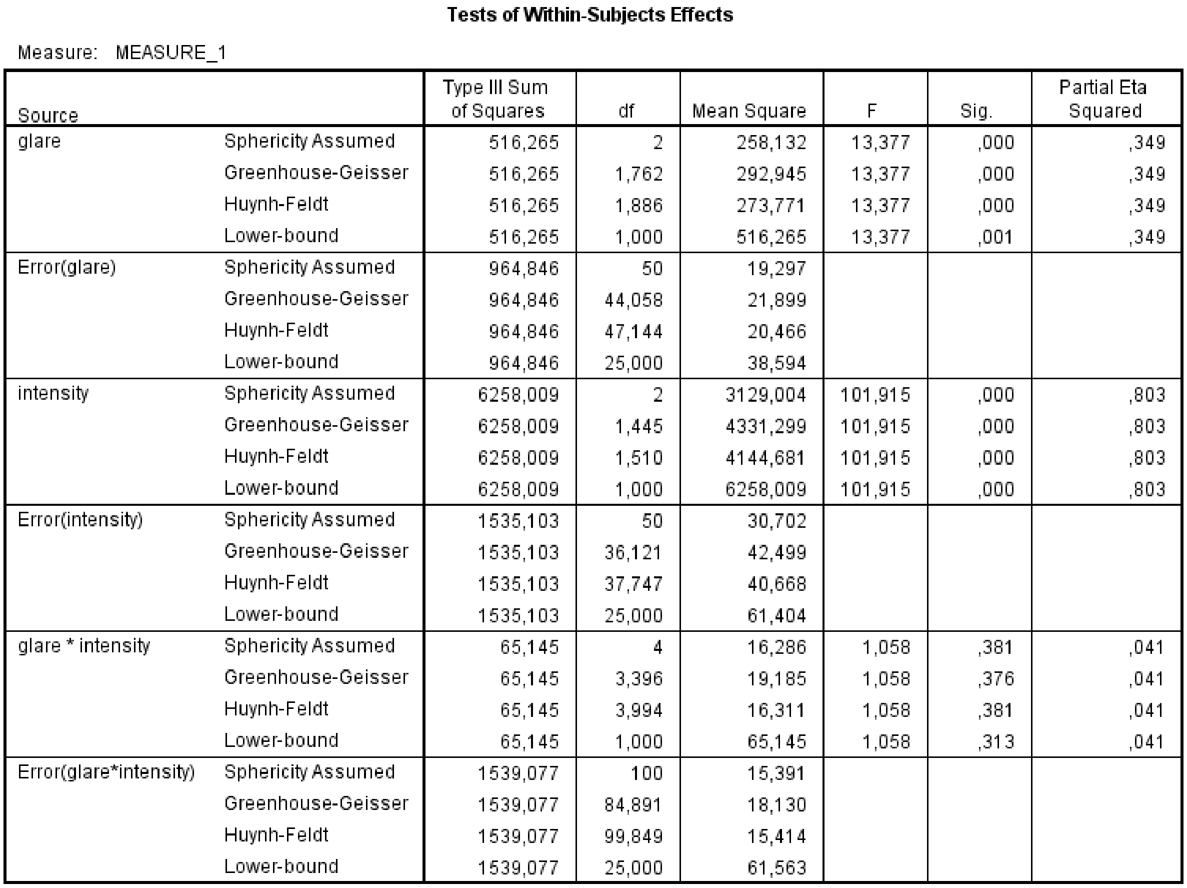
Within Subjects Effect	Mauchly's W	Approx. Chi-Square	df	Sig.	Epsilon[b]		
					Greenhouse-Geisser	Huynh-Feldt	Lower-bound
glare	,865	3,477	2	,176	,881	,943	,500
intensity	,616	11,638	2	,003	,722	,755	,500
glare * intensity	,589	12,393	9	,193	,849	,998	,250

Tests the null hypothesis that the error covariance matrix of the orthonormalized transformed dependent variables is proportional to an identity matrix.

a. Design: Intercept
Within Subjects Design: glare + intensity + glare * intensity

b. May be used to adjust the degrees of freedom for the averaged tests of significance. Corrected tests are displayed in the Tests of Within-Subjects Effects table.

Sphericity can be assumed for glare, as Mauchly's Test of Sphericity was rejected.



Tests of Within-Subjects Effects

Measure: MEASURE_1

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
glare	Sphericity Assumed	516,265	2	258,132	13,377	,000	,349
	Greenhouse-Geisser	516,265	1,762	292,945	13,377	,000	,349
	Huynh-Feldt	516,265	1,886	273,771	13,377	,000	,349
	Lower-bound	516,265	1,000	516,265	13,377	,001	,349
Error(glare)	Sphericity Assumed	964,846	50	19,297			
	Greenhouse-Geisser	964,846	44,058	21,899			
	Huynh-Feldt	964,846	47,144	20,466			
	Lower-bound	964,846	25,000	38,594			
intensity	Sphericity Assumed	6258,009	2	3129,004	101,915	,000	,803
	Greenhouse-Geisser	6258,009	1,445	4331,299	101,915	,000	,803
	Huynh-Feldt	6258,009	1,510	4144,681	101,915	,000	,803
	Lower-bound	6258,009	1,000	6258,009	101,915	,000	,803
Error(intensity)	Sphericity Assumed	1535,103	50	30,702			
	Greenhouse-Geisser	1535,103	36,121	42,499			
	Huynh-Feldt	1535,103	37,747	40,668			
	Lower-bound	1535,103	25,000	61,404			
glare * intensity	Sphericity Assumed	65,145	4	16,286	1,058	,381	,041
	Greenhouse-Geisser	65,145	3,396	19,185	1,058	,376	,041
	Huynh-Feldt	65,145	3,994	16,311	1,058	,381	,041
	Lower-bound	65,145	1,000	65,145	1,058	,313	,041
Error(glare*intensity)	Sphericity Assumed	1539,077	100	15,391			
	Greenhouse-Geisser	1539,077	84,891	18,130			
	Huynh-Feldt	1539,077	99,849	15,414			
	Lower-bound	1539,077	25,000	61,563			

11.3 Experiment 2

11.3.1 Conditions

Condition	Reference	Direction	Luminaire Intensity	Glare	Image Intensity
1	Yes	Ceiling	Low	No	Lower (1)
2	Yes	Ceiling	Low	No	Mean (6)
3	Yes	Ceiling	Low	No	Higher (11)
4	Yes	Ceiling	Low	Yes	Lower (1)
5	Yes	Ceiling	Low	Yes	Mean (6)
6	Yes	Ceiling	Low	Yes	Higher (11)
7	Yes	Ceiling	Low	Extra	Lower (1)
8	Yes	Ceiling	Low	Extra	Mean (6)
9	Yes	Ceiling	Low	Extra	Higher (11)
10	Yes	Ceiling	High	No	Lower (14)
11	Yes	Ceiling	High	No	Mean (19)
12	Yes	Ceiling	High	No	Higher (24)
13	Yes	Ceiling	High	Yes	Lower (14)
14	Yes	Ceiling	High	Yes	Mean (19)
15	Yes	Ceiling	High	Yes	Higher (24)
16	Yes	Ceiling	High	Extra	Lower (14)
17	Yes	Ceiling	High	Extra	Mean (19)
18	Yes	Ceiling	High	Extra	Higher (24)
19	Yes	Wall	Low	No	Lower (6)
20	Yes	Wall	Low	No	Mean (11)
21	Yes	Wall	Low	No	Higher (16)
22	Yes	Wall	High	No	Lower (22)
23	Yes	Wall	High	No	Mean (27)
24	Yes	Wall	High	No	Higher (32)
25	No	Ceiling	Low	No	Lower (1)
26	No	Ceiling	Low	No	Mean (6)
27	No	Ceiling	Low	No	Higher (11)
28	No	Ceiling	Low	Yes	Lower (1)
29	No	Ceiling	Low	Yes	Mean (6)
30	No	Ceiling	Low	Yes	Higher (11)
31	No	Ceiling	Low	Extra	Lower (1)
32	No	Ceiling	Low	Extra	Mean (6)
33	No	Ceiling	Low	Extra	Higher (11)
34	No	Ceiling	High	No	Lower (14)
35	No	Ceiling	High	No	Mean (19)
36	No	Ceiling	High	No	Higher (24)
37	No	Ceiling	High	Yes	Lower (14)
38	No	Ceiling	High	Yes	Mean (19)
39	No	Ceiling	High	Yes	Higher (24)
40	No	Ceiling	High	Extra	Lower (14)
41	No	Ceiling	High	Extra	Mean (19)
42	No	Ceiling	High	Extra	Higher (24)
43	No	Wall	Low	No	Lower (6)
44	No	Wall	Low	No	Mean (11)
45	No	Wall	Low	No	Higher (16)
46	No	Wall	High	No	Lower (22)
47	No	Wall	High	No	Mean (27)
48	No	Wall	High	No	Higher (32)

11.3.2 Briefing

The text below is the verbal explanation given to the participants prior to Experiment 2. The explanation is given in combination with step by step demonstrations:

Welcome and thank you for participating in my experiment. I am investigating brightness in virtual lighting applications. In this experiment I want you to compare the brightness of real-world luminaires to images of the luminaires, which will appear on the display. In the context of this experiment, brightness is defined as the amount of light you perceive to be produced by the luminaire, if that luminaire is on. Firstly I will explain you the procedure. I would like to ask you to put your head on the chin rest. Firstly, the luminaires will go on for 10 seconds. During this period, try to memorize the brightness you perceive as well as possible. Then the luminaires go off again and an image of the luminaires will appear on the display. At the top of this image is a question with a seven-point scale, which can be answered by using the arrow keys 'left' and 'right' and confirm with 'enter'. There are three questions per image. After answering the last one you will continue with the next condition and the real luminaires will go on again.

After completing the first parts, but before continuing to the second part of the experiment, a short explanation was given as well:

You have completed the first part. In the second part you are only going to observe the images, so the real luminaires will no longer go on. The procedure is the same, simply answer the questions, except there will be only two questions per image this time.

11.3.3 Informed Consent

Perception of Visualizations

Dear sir/madam,

You are kindly invited to participate in a perceptual experiment. Participation in this experiment is strictly voluntary and you may choose to stop participating and withdraw at any time, with or without providing reason. Please indicate the researcher if you wish to stop.

The goal of this study is to understand lighting perception in a virtual scene and how this compares to a real scene. Such information can be used for technology development, perceptual experiments, concept validation, and marketing communication.

In this experiment the brightness of a real luminaire is compared with a virtual luminaire which is viewed thereafter. Since the scope of the study is visual perception, preceding the experiment you are asked to participate in a short visual acuity test. The task is to answer questions about 48 images.

All vision screening results and test responses will be handled anonymously and confidentially. Your name will not be included or in any other way associated with the data collected in this study.

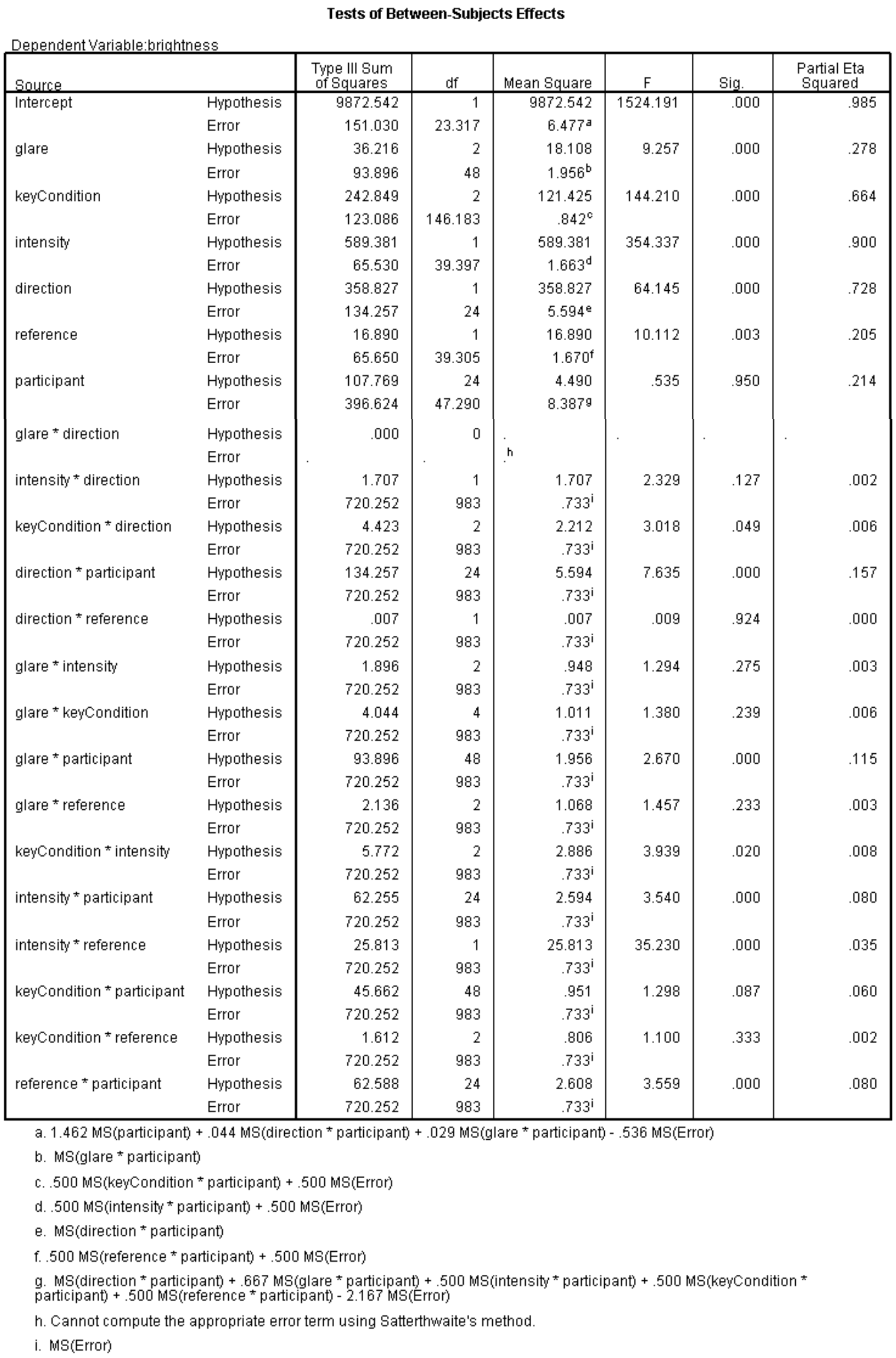
Consent:

- I have been informed about the objective of the investigation and my role in it and all my questions about the experiment have been answered by the responsible researcher.
- I had sufficient time to consider my participation in this investigation and I am aware that it is completely voluntarily.
- The potential risks associated with my participation in this investigation and the anticipated benefits have been discussed with me.
- I realize that I may decide to refuse participation or stop participation at any time, with or without providing reason.
- I understand and agree that data about me will be collected and processed, either manually or by computer, by the responsible researcher and other researchers in the project.
- I understand that my directly identifying personal data (e.g., name, address, etc.) will be separated from the research data and replaced by an assigned number/code
- I understand that I am entitled to access the personal information collected about me and to have inaccuracies corrected.

Participant name: Participant signature: Date:

11.3.4 Statistical tables

Univariate ANOVA on Brightness Score



Tests of Between-Subjects Effects

Dependent Variable:brightness

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	9872.542	1	9872.542	1524.191	.000	.985
	Error	151.030	23.317	6.477[a]			
glare	Hypothesis	36.216	2	18.108	9.257	.000	.278
	Error	93.896	48	1.956[b]			
keyCondition	Hypothesis	242.849	2	121.425	144.210	.000	.664
	Error	123.086	146.183	.842[c]			
intensity	Hypothesis	589.381	1	589.381	354.337	.000	.900
	Error	65.530	39.397	1.663[d]			
direction	Hypothesis	358.827	1	358.827	64.145	.000	.728
	Error	134.257	24	5.594[e]			
reference	Hypothesis	16.890	1	16.890	10.112	.003	.205
	Error	65.650	39.305	1.670[f]			
participant	Hypothesis	107.769	24	4.490	.535	.950	.214
	Error	396.624	47.290	8.387[g]			
glare * direction	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[h]			
intensity * direction	Hypothesis	1.707	1	1.707	2.329	.127	.002
	Error	720.252	983	.733[i]			
keyCondition * direction	Hypothesis	4.423	2	2.212	3.018	.049	.006
	Error	720.252	983	.733[i]			
direction * participant	Hypothesis	134.257	24	5.594	7.635	.000	.157
	Error	720.252	983	.733[i]			
direction * reference	Hypothesis	.007	1	.007	.009	.924	.000
	Error	720.252	983	.733[i]			
glare * intensity	Hypothesis	1.896	2	.948	1.294	.275	.003
	Error	720.252	983	.733[i]			
glare * keyCondition	Hypothesis	4.044	4	1.011	1.380	.239	.006
	Error	720.252	983	.733[i]			
glare * participant	Hypothesis	93.896	48	1.956	2.670	.000	.115
	Error	720.252	983	.733[i]			
glare * reference	Hypothesis	2.136	2	1.068	1.457	.233	.003
	Error	720.252	983	.733[i]			
keyCondition * intensity	Hypothesis	5.772	2	2.886	3.939	.020	.008
	Error	720.252	983	.733[i]			
intensity * participant	Hypothesis	62.255	24	2.594	3.540	.000	.080
	Error	720.252	983	.733[i]			
intensity * reference	Hypothesis	25.813	1	25.813	35.230	.000	.035
	Error	720.252	983	.733[i]			
keyCondition * participant	Hypothesis	45.662	48	.951	1.298	.087	.060
	Error	720.252	983	.733[i]			
keyCondition * reference	Hypothesis	1.612	2	.806	1.100	.333	.002
	Error	720.252	983	.733[i]			
reference * participant	Hypothesis	62.588	24	2.608	3.559	.000	.080
	Error	720.252	983	.733[i]			

a. 1.462 MS(participant) + .044 MS(direction * participant) + .029 MS(glare * participant) - .536 MS(Error)

b. MS(glare * participant)

c. .500 MS(keyCondition * participant) + .500 MS(Error)

d. .500 MS(intensity * participant) + .500 MS(Error)

e. MS(direction * participant)

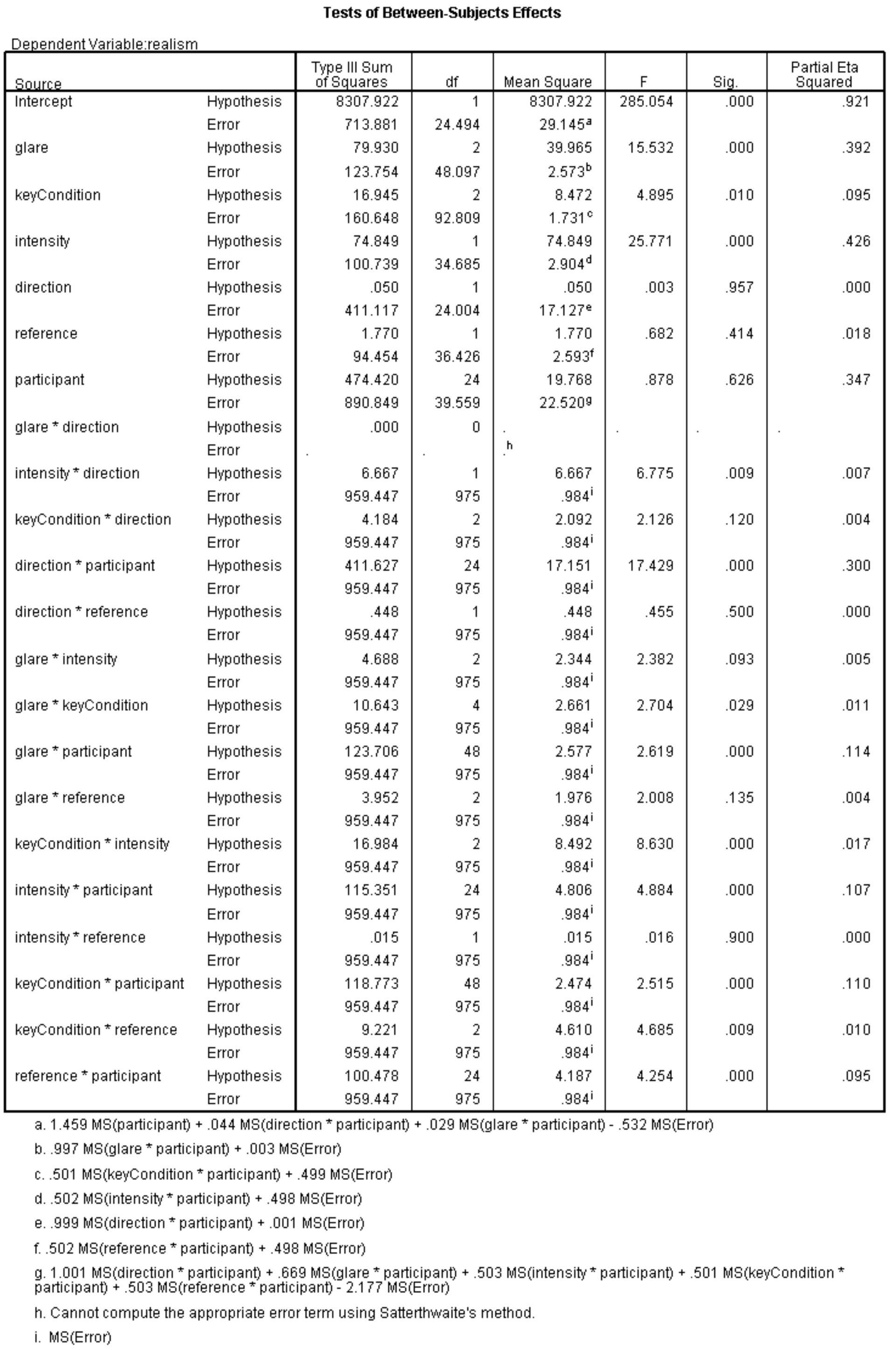
f. .500 MS(reference * participant) + .500 MS(Error)

g. MS(direction * participant) + .667 MS(glare * participant) + .500 MS(intensity * participant) + .500 MS(keyCondition * participant) + .500 MS(reference * participant) - 2.167 MS(Error)

h. Cannot compute the appropriate error term using Satterthwaite's method.

i. MS(Error)

Univariate ANOVA on Realism Score



Tests of Between-Subjects Effects

Dependent Variable:realism

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	8307.922	1	8307.922	285.054	.000	.921
	Error	713.881	24.494	29.145[a]			
glare	Hypothesis	79.930	2	39.965	15.532	.000	.392
	Error	123.754	48.097	2.573[b]			
keyCondition	Hypothesis	16.945	2	8.472	4.895	.010	.095
	Error	160.648	92.809	1.731[c]			
intensity	Hypothesis	74.849	1	74.849	25.771	.000	.426
	Error	100.739	34.685	2.904[d]			
direction	Hypothesis	.050	1	.050	.003	.957	.000
	Error	411.117	24.004	17.127[e]			
reference	Hypothesis	1.770	1	1.770	.682	.414	.018
	Error	94.454	36.426	2.593[f]			
participant	Hypothesis	474.420	24	19.768	.878	.626	.347
	Error	890.849	39.559	22.520[g]			
glare * direction	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[h]			
intensity * direction	Hypothesis	6.667	1	6.667	6.775	.009	.007
	Error	959.447	975	.984[i]			
keyCondition * direction	Hypothesis	4.184	2	2.092	2.126	.120	.004
	Error	959.447	975	.984[i]			
direction * participant	Hypothesis	411.627	24	17.151	17.429	.000	.300
	Error	959.447	975	.984[i]			
direction * reference	Hypothesis	.448	1	.448	.455	.500	.000
	Error	959.447	975	.984[i]			
glare * intensity	Hypothesis	4.688	2	2.344	2.382	.093	.005
	Error	959.447	975	.984[i]			
glare * keyCondition	Hypothesis	10.643	4	2.661	2.704	.029	.011
	Error	959.447	975	.984[i]			
glare * participant	Hypothesis	123.706	48	2.577	2.619	.000	.114
	Error	959.447	975	.984[i]			
glare * reference	Hypothesis	3.952	2	1.976	2.008	.135	.004
	Error	959.447	975	.984[i]			
keyCondition * intensity	Hypothesis	16.984	2	8.492	8.630	.000	.017
	Error	959.447	975	.984[i]			
intensity * participant	Hypothesis	115.351	24	4.806	4.884	.000	.107
	Error	959.447	975	.984[i]			
intensity * reference	Hypothesis	.015	1	.015	.016	.900	.000
	Error	959.447	975	.984[i]			
keyCondition * participant	Hypothesis	118.773	48	2.474	2.515	.000	.110
	Error	959.447	975	.984[i]			
keyCondition * reference	Hypothesis	9.221	2	4.610	4.685	.009	.010
	Error	959.447	975	.984[i]			
reference * participant	Hypothesis	100.478	24	4.187	4.254	.000	.095
	Error	959.447	975	.984[i]			

a. 1.459 MS(participant) + .044 MS(direction * participant) + .029 MS(glare * participant) - .532 MS(Error)

b. .997 MS(glare * participant) + .003 MS(Error)

c. .501 MS(keyCondition * participant) + .499 MS(Error)

d. .502 MS(intensity * participant) + .498 MS(Error)

e. .999 MS(direction * participant) + .001 MS(Error)

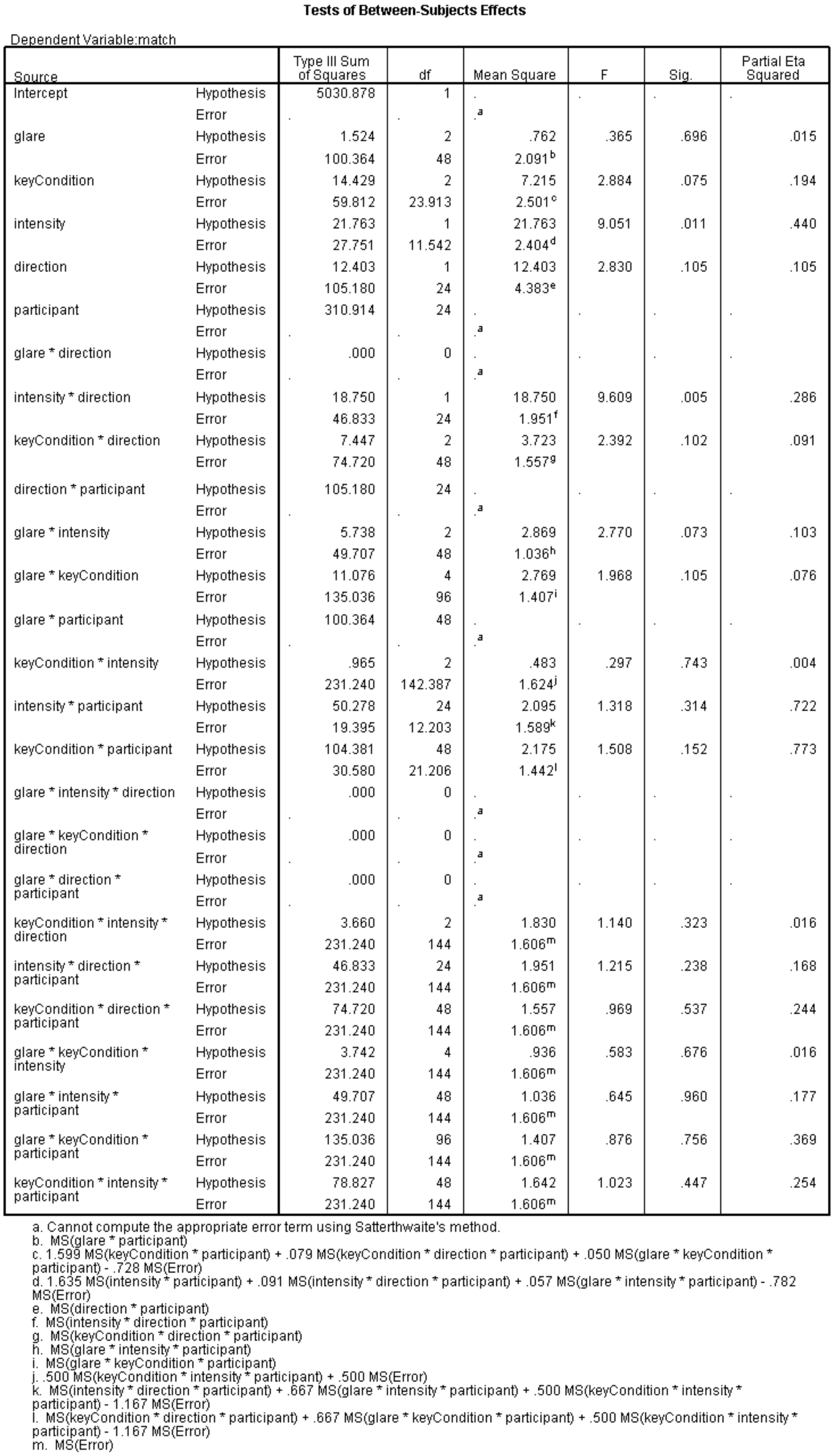
f. .502 MS(reference * participant) + .498 MS(Error)

g. 1.001 MS(direction * participant) + .669 MS(glare * participant) + .503 MS(intensity * participant) + .501 MS(keyCondition * participant) + .503 MS(reference * participant) - 2.177 MS(Error)

h. Cannot compute the appropriate error term using Satterthwaite's method.

i. MS(Error)

Univariate ANOVA on Match Score



Tests of Between-Subjects Effects

Dependent Variable:match

Source		Type III Sum of Squares	df	Mean Square	F	Sig.	Partial Eta Squared
Intercept	Hypothesis	5030.878	1	.	.	.	.
	Error	.	.	.[a]			
glare	Hypothesis	1.524	2	.762	.365	.696	.015
	Error	100.364	48	2.091[b]			
keyCondition	Hypothesis	14.429	2	7.215	2.884	.075	.194
	Error	59.812	23.913	2.501[c]			
intensity	Hypothesis	21.763	1	21.763	9.051	.011	.440
	Error	27.751	11.542	2.404[d]			
direction	Hypothesis	12.403	1	12.403	2.830	.105	.105
	Error	105.180	24	4.383[e]			
participant	Hypothesis	310.914	24	.	.	.	.
	Error	.	.	.[a]			
glare * direction	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[a]			
intensity * direction	Hypothesis	18.750	1	18.750	9.609	.005	.286
	Error	46.833	24	1.951[f]			
keyCondition * direction	Hypothesis	7.447	2	3.723	2.392	.102	.091
	Error	74.720	48	1.557[g]			
direction * participant	Hypothesis	105.180	24	.	.	.	.
	Error	.	.	.[a]			
glare * intensity	Hypothesis	5.738	2	2.869	2.770	.073	.103
	Error	49.707	48	1.036[h]			
glare * keyCondition	Hypothesis	11.076	4	2.769	1.968	.105	.076
	Error	135.036	96	1.407[i]			
glare * participant	Hypothesis	100.364	48	.	.	.	.
	Error	.	.	.[a]			
keyCondition * intensity	Hypothesis	.965	2	.483	.297	.743	.004
	Error	231.240	142.387	1.624[j]			
intensity * participant	Hypothesis	50.278	24	2.095	1.318	.314	.722
	Error	19.395	12.203	1.589[k]			
keyCondition * participant	Hypothesis	104.381	48	2.175	1.508	.152	.773
	Error	30.580	21.206	1.442[l]			
glare * intensity * direction	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[a]			
glare * keyCondition * direction	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[a]			
glare * direction * participant	Hypothesis	.000	0	.	.	.	.
	Error	.	.	.[a]			
keyCondition * intensity * direction	Hypothesis	3.660	2	1.830	1.140	.323	.016
	Error	231.240	144	1.606[m]			
intensity * direction * participant	Hypothesis	46.833	24	1.951	1.215	.238	.168
	Error	231.240	144	1.606[m]			
keyCondition * direction * participant	Hypothesis	74.720	48	1.557	.969	.537	.244
	Error	231.240	144	1.606[m]			
glare * keyCondition * intensity	Hypothesis	3.742	4	.936	.583	.676	.016
	Error	231.240	144	1.606[m]			
glare * intensity * participant	Hypothesis	49.707	48	1.036	.645	.960	.177
	Error	231.240	144	1.606[m]			
glare * keyCondition * participant	Hypothesis	135.036	96	1.407	.876	.756	.369
	Error	231.240	144	1.606[m]			
keyCondition * intensity * participant	Hypothesis	78.827	48	1.642	1.023	.447	.254
	Error	231.240	144	1.606[m]			

a. Cannot compute the appropriate error term using Satterthwaite's method.
b. MS(glare * participant)
c. 1.599 MS(keyCondition * participant) + .079 MS(keyCondition * direction * participant) + .050 MS(glare * keyCondition * participant) - .728 MS(Error)
d. 1.635 MS(intensity * participant) + .091 MS(intensity * direction * participant) + .057 MS(glare * intensity * participant) - .782 MS(Error)
e. MS(direction * participant)
f. MS(intensity * direction * participant)
g. MS(keyCondition * direction * participant)
h. MS(glare * intensity * participant)
i. MS(glare * keyCondition * participant)
j. .500 MS(keyCondition * intensity * participant) + .500 MS(Error)
k. MS(intensity * direction * participant) + .667 MS(glare * intensity * participant) + .500 MS(keyCondition * intensity * participant) - 1.167 MS(Error)
l. MS(keyCondition * direction * participant) + .667 MS(glare * keyCondition * participant) + .500 MS(keyCondition * intensity * participant) - 1.167 MS(Error)
m. MS(Error)

Direction x image intensity

As can be seen from Figure 35, the relation between match and image intensity behaved very differently in the ceiling direction. While the match remained relatively unaffected by image intensity in the wall condition, there were large between-level differences in the ceiling condition. The lowest image intensity (1) did not achieve a good brightness match. It is most likely that the image was simply too dark to create a brightness match with a real luminaire. The "mean" image intensity 6, but also the above mean image intensity 11, reached relatively good matches. For the higher image intensities, the match score dropped again: except for the "mean" image intensity 19, which almost reaches a mean match score of 3, the incongruent intensities 14 and 24 achieved very low scores. The low score of 24 is interesting, because it had the highest brightness score, but the worst match in brightness. If the participants managed to give the match scores solely based on perceived brightness, it would mean that the brightness of the image was too high. It could also be the case that the match score was low because the realism of the image was low.

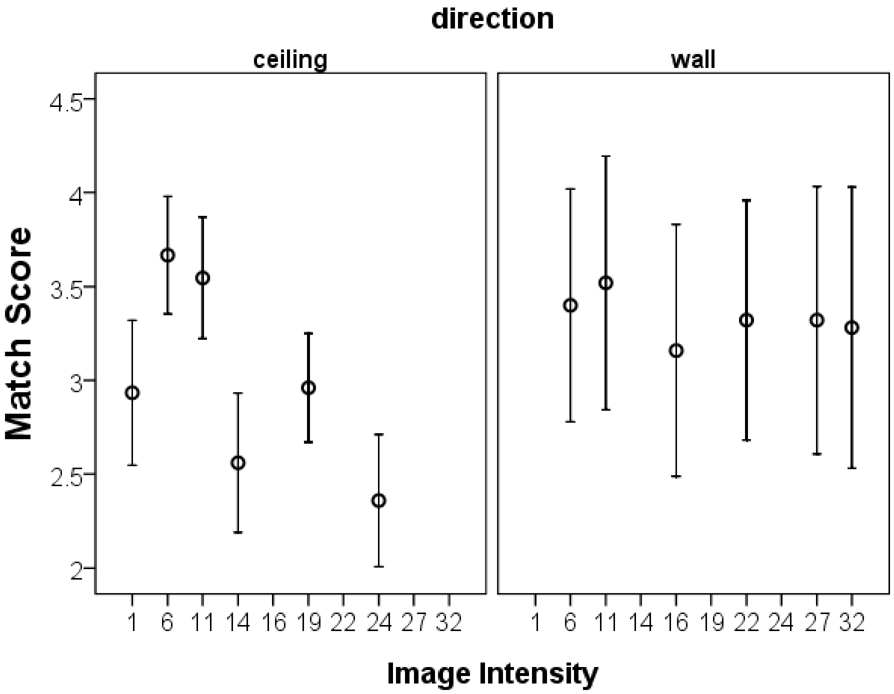


Figure 35: Error bar plot of mean match score versus Image intensity, per direction.