Modulation depth threshold in the Compensation Comparison approach

Luuk Franssen
Netherlands Institute for Neuroscience, Ophthalmic Research Institute, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands

Joris E. Coppens
Netherlands Institute for Neuroscience, Ophthalmic Research Institute, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands

Thomas J. T. P. van den Berg
Netherlands Institute for Neuroscience, Ophthalmic Research Institute, Royal Netherlands Academy of Arts and Sciences, Amsterdam, The Netherlands

Recently, the “Compensation Comparison” method was introduced for measuring retinal straylight. In this article, basic aspects are described, in particular a generalization of the approach using the concept of “precompensation,” and including flicker threshold as parameter in the psychophysical model. The model was experimentally verified in lab measurements with and without artificially increased straylight and was tested on the data from the multi-center GLARE study. The resulting flicker threshold estimates were analyzed to better understand their origin. An effect of flicker adaptation over distance was found. The new approach proved suitable to describe Compensation Comparison measurements including precompensation, and also for subjects with poor psychometric behavior.

Keywords: straylight, glare, psychophysics, flicker, eye media, flicker threshold, Compensation Comparison


Introduction

Recently, the Compensation Comparison method was introduced as a new technique for measuring straylight on the retina of the human eye (Franssen, Coppens, & van den Berg, 2006). Retinal straylight is a clinically and practically important phenomenon, degrading visual function. It is caused by intraocular light scatter. This is the phenomenon that part of the light reaching the retina does not partake in normal image formation (van den Berg, 1986). Most rays originating from a certain point in space are converged by the refracting elements of the eye to the focal spot on the retina. However, some of the rays are dispersed to other areas by optical imperfections of the eye. This already occurs in the healthy eye (Vos, 1984), but to a much larger extent in pathological states, such as cataract (de Waard, IJspeert, van den Berg, & de Jong, 1992) and corneal dystrophy (van den Berg, Hwan, & Delleman, 1993). These dispersed rays are distributed all over the retina, but with decreasing densities at distances further away from the original focal spot.

It is important for assessment of functional integrity of a patient’s eye to develop a method to measure straylight in an accurate way. Due to straylight, the retinal light distribution in any visual environment is composed of two parts: the image of the external world based on the more or less properly focused rays, superimposed upon a background caused by the dispersed rays. As a result, contrast is lost in the retinal image. The severity of the contrast loss depends on the luminance ratio between background and image. This ratio is a function of the optical clarity of the eye and can be quantified and expressed in the physically well-defined retinal straylight parameter $s$ (van den Berg, 1986, 1995). The extreme situation of contrast loss due to intraocular light scatter is represented by the classical (disability) glare condition (Vos, 1984): strong light somewhere in the visual field while a weakly lit object has to be observed. In such a situation, the contrast of the retinal image may drop below the contrast threshold and can lead to complete blinding. The typical situation is blinding by oncoming traffic at night.

The Compensation Comparison method to measure the straylight parameter is based upon the previously used Direct Compensation method (van den Berg, 1986) but...
implemented in a two alternative forced choice (2AFC) paradigm (Franssen et al., 2006). In short, the Direct Compensation method works as follows (Figure 1): An annulus at a certain angular distance \( \theta \) from a (dark) test field is presented flickering. Due to intraocular scatter, part of the light from the bright straylight source will be projected on the retina at the location of the test field, inducing a (weak) flicker in the test field. To determine the exact amount of straylight, we presented variable counterphase compensation light in the test field. By adjustment of the amount of compensation light, the flicker perception in the test field can be extinguished. In this way, the straylight modulation caused by light scattered from the glare source is directly compensated.

In essence, the Compensation Comparison method presents the same stimuli to the subject as the Direct Compensation method. Note that in the Direct Compensation method, the amount of compensation light is varied until the straylight flicker has disappeared. In other words, in the Direct Compensation method, the subject compares different stimuli sequentially. Contrarily, in the Compensation Comparison method, two stimuli of the Direct Compensation method are presented to and compared by the subject simultaneously. This is achieved by splitting the test field in two halves (Figure 1). Compensation light is presented in one of the two half fields, whereas no compensation light is present in the other half field. As a result, two flickers are perceived, which differ in modulation depth: one results from straylight only, the other is a combination of straylight and compensation light, flickering in counterphase with this straylight. The subject’s task is to decide which test field half flickers stronger. In this way, the Direct Compensation method is implemented as a 2AFC approach.

The Compensation Comparison method has some important advantages with respect to the Direct Compensation method: (1) Subject-dependent bias as well as the ability to deliberately influence the test outcome has been eliminated (Franssen et al., 2006). (2) A measurement reliability parameter (called ESD for “expected standard deviation”) could be developed to assess the quality of individual measurements (Coppens, Franssen, van Rijn, & van den Berg, 2006). With the Direct Compensation technique, repeatability information had been limited to only population-based repeated measures standard deviations. Given these advantages, retinal straylight measurement was made possible on a large scale and in the clinical routine, as demonstrated in the European GLARE study (Franssen et al., 2006) (www.glare.eu). In this study, which aimed at assessing the prevalence of vision impairments in the driving population, several visual tests, including straylight measurements, were performed among a population of drivers in five European countries, spread over five age categories. The measured population
consisted of a wide range of subjects, including ages from 20 to 85, visual acuities below 0.5 (logMAR 0.3) to more than 1.0 (logMAR 0.0), visual field defects, and other ocular pathologies such as glaucoma and cataract. The straylight data from the GLARE study, which were already used to evaluate the Compensation Comparison method in clinical practice (Franssen et al., 2006), will be utilized in the present paper to evaluate a more complete psychophysical model for the comparison task.

According to feedback we got from the operators in the clinics that participated in the GLARE study, and who also had earlier experience with the Direct Compensation method, the Compensation Comparison test is easier, more intuitive, easier to explain, and needs less explanation from the operator. The measurement time is fixed and limited, making the test more pleasant for both patient and operator. We did however not collect systematic statistical data on these subjective assessments.

As a result of the proven suitability of the Compensation Comparison method for large-scale and clinical use, the method has been implemented in a commercially available measurement device (called C-Quant) by the Germany based firm Oculus.

Apart from the improvements mentioned above, the accuracy of the Compensation Comparison method appeared to be somewhat better with respect to the Direct Compensation method in field studies. The Direct Compensation method was evaluated in a study that compared several devices for measuring straylight and glare (van Rijn et al., 2005), involving 112 subjects drawn from the patients and visitors of the outpatient departments of three clinics. The standard deviations of differences between repeated measurements found in such a field study were 0.15 and 0.18 log units for two different implementations of the Direct Compensation method. The repeated measures standard deviation for the Compensation Comparison technique in the GLARE study was found to be 0.1 log units or lower, depending on the filter criterion for excluding low quality measurements (Coppens et al., 2006). Although for many applications a repeated measures standard deviation of 0.1 log units may be adequate, a need was felt for improvement, for example, when a more precise cutoff value is involved (e.g., for driver testing or clinical treatment decisions). In the present situation, the reliability criterion ESD is used for this purpose. ESD gives the expected standard deviation for individual measurements. Using this criterion, substandard measurements can be detected and redone in order to get a better measurement. In this way, the repeated measures standard deviation could be improved to values around 0.06 log units, depending on the ESD criterion used (Coppens et al., 2006).

A better way would be to improve the measurement precision directly. In the beginning, expectations were high in this respect, and for good reasons. Realize that the task in the Direct Compensation method is in essence the same as in a flicker threshold experiment. In the Direct Compensation method, counterphase flicker is adjusted until it is precisely equal to the straylight flicker, thus silencing the flicker percept. The precision to perform this task should be comparable to flicker threshold. Hence, the accuracy of the Direct Compensation method was originally expected to be of the order of the flicker threshold corresponding with the test field characteristics used. Flicker thresholds were measured by de Lange (1958) and many others after him and were found to be in the order of 1% (0.004 log units) for a flicker frequency of 8 Hz and average test field luminance of around 1–4 cd/m², values used in the Direct Compensation as well as the Compensation Comparison test. It was therefore somewhat disappointing to find the much higher values mentioned above for these straylight tests. Causes for these differences have not been systematically investigated so far. Speculatively, for the Direct Compensation method, it was thought that the continuous presence of a strong flicker in the periphery lowered sensitivity in the center. This was one other reason for a change to the Compensation Comparison method, in which short duration stimulus presentations are used. Some form of retinal lateral adaptation, induced in the test field by the flickering straylight ring, might be involved.

However, the Compensation Comparison method may have also introduced a disadvantage with respect to the sensitivity to be expected. It differs from the Direct Compensation method in that it is not similar to a flicker threshold test anymore: the task is now to compare two suprathreshold flicker signals. Such a task may be considerably less precise, depending on the absolute values of the flicker levels. Assuming a kind of Weber-like behavior, precision will suffer if the absolute flicker levels are higher. Presently, the comparison task is performed with the full straylight flicker in one half field or, in other words, at maximum flicker level. If this could be lowered to, say, threshold level, the straylight value could be determined with maximum accuracy. It is indeed possible to generate stimuli closer to threshold levels by adding counterphase light to both test field halves (Figure 2). This will be further explained in the Methods section. Such stimuli are not only closer to flicker threshold levels, but also closer to the “silent point” of the Direct Compensation method. This means that the (simultaneous) comparison task of the Compensation Comparison method comes closer to the (sequential) comparison task of the Direct Compensation method. In other words, stimuli that are in between the two extremes of full straylight flicker (current Compensation Comparison method) and threshold flicker (Direct Compensation method) can be chosen. In this way, the Compensation Comparison method can be optimized for maximum accuracy.

This article will be devoted to understanding the psychophysics of the processes involved. As measurement technique, the extended Compensation Comparison method will be used, referred to as “generalized Compensation Comparison method” or “CC*” method. A model
for the associated psychometric function will be developed and tested with a small number of laboratory subjects. The model serves not only to develop the new approach as described above, but also as an improvement to the model used for analysis of data obtained with the Compensation Comparison method (Franssen et al., 2006). This analysis will be done for the data of 2,422 subjects that participated in the GLARE study. As essential part, a model for threshold behavior will be incorporated, which will also be of importance to study the causes of the discrepancy between the accuracy of the Direct Compensation and Compensation Comparison methods on the one hand and the classical flicker thresholds on the other hand. For this part of the study, flicker threshold experiments were performed with stimulus layouts ranging from the full straylight case to the classical case used by de Lange and others, with intermediate steps in between. We should explicitly state here that the current study does not assess the (improvement in) reliability of the CC* method in a practical/clinical environment. This would have required an extensive population study involving all three variations of straylight compensation techniques (Direct Compensation, Compensation Comparison, and CC*), which would be outside of the scope of this study.

Methods

Seven subjects (ages ranging from 21 to 57 years, with a mean age of 30 years) participated in the experiments. They were lab students and coworkers, including the authors. All subjects were without ocular disease. Testing was done monocularly on the subject’s preferred eye. For all types of refraction, habitual glasses were allowed, but contact lenses were replaced by trial glasses. The actual refraction values ranged from $-7$ to emmetropic. It must be noted here that the test does not require refractive
correction to be precise. Corrections were chosen for comfortable viewing, resulting in a +2 near addition for the older subjects because the tests were performed at a distance of 32 cm from the stimulus screen. The study adhered to the guidelines of the Declaration of Helsinki for research in human subjects.

To test the CC* method also for conditions of increased scattering, five subjects were additionally measured with a light diffusing filter (Tiffen Black Pro Mist 2, in short BPM2) in front of the tested eye. This filter, among a collection of 23 commercially available light diffusing filters, was found to have the best light scattering characteristics for mimicking (early) cataract or aging effects in the human crystalline lens (de Wit, Franssen, Coppens, & van den Berg, 2006).

As mentioned before, the CC* method was evaluated with the straylight data from the European GLARE study, involving 2,422 subjects in total. In the course of the study, some improvements were made on the implementation of the straylight test: (1) A three-trial instruction phase was added prior to the real measurement, to familiarize the subject with the flicker comparison task. (2) The subject’s responses were displayed to the operator during the measurement, making it possible to interfere in case the response pattern was erratic. In such a case, a new measurement can be started after additional explanation. (3) The luminance in the test fields was increased by a factor of 2 in the first part of the test, making the measurement easier for older subjects. In total 1,073 subjects were measured with this final version (including these improvements).

For stimulus generation, a computer system with either a CRT monitor or combination of DLP projector and back-projection screen was used. The straylight source was a white light annulus with angle \( \theta \) extending from \( 7^\circ \) to \( 14^\circ \). Because of the approximate \( 1/\theta^2 \) dependence of retinal straylight, this corresponds to an average scattering angle of \( 10^\circ \) (van den Berg, 1995).

Simplified, the measurement procedure runs as follows (for a full description, see Franssen et al., 2006): During the test, a series of limited duration stimuli is presented that differ in the amount of compensation light in one test field half. In the other test field half, no compensation light is presented (Figure 1). Following a two alternative forced choice (2AFC) paradigm, the task for the subject is to decide for each stimulus which test field half flickers stronger. The subject’s responses are recorded by means of two push buttons, representing the left and right test halves. A choice for the test field half with the compensation light is recorded as a 1 score, a choice for the test field half without the compensation light is recorded as a 0 score (Figure 2). Using the psychophysical model for this flicker comparison task, which will be described in detail below, a psychometric curve is fitted to the subject’s responses by means of a maximum likelihood procedure. From this fitted curve both the straylight parameter and a measure for the quality of the measurement (ESD) can be deduced. This procedure is explained in more detail in a separate publication (Coppens et al., 2006).

### Psychometric function including threshold behavior

As a basis to describe the psychometric function we started out from the well-known logistic function (Strasburger, 2001). Comparing two flickering test fields \( a \) and \( b \) with different modulation depths, the chance \( P \) of choosing one of the test fields as having the stronger flicker was written as

\[
P = \frac{1}{1 + e^{-\frac{MDCc}{s}}} \quad (1)
\]

\( MDC_c \) is the parameter in the equation, giving a critical value for the contrast between the two flickers. \( MDC \) stands for modulation depth contrast. \( MDC \) is the independent variable in the equation, giving the actual contrast between the two presented flickers, defined as

\[
MDC = \frac{MDb - MDa}{MDb + MDa}, \quad (2)
\]

where \( MDa \) and \( MDb \) represent the retinal modulation depths, or flicker levels, in both test fields.

The retinal light levels can be expressed in (equivalent) straylight parameter units, referred to as \( s \) units in this article (explained in more detail in Franssen et al., 2006):

\[
MDa = \frac{Sa^{off} - Sa^{on}}{Sa^{off} + Sa^{on}} \quad \text{and} \quad MDb = \frac{Sb^{off} - Sb^{on}}{Sb^{off} + Sb^{on}}, \quad (3)
\]

with

\[
Sb^{on} = s \quad \text{Sb}^{off} = S_{\text{comp}}, \quad (4)
\]

\[
Sa^{on} = s + 0.5 \cdot S_{\text{comp}} \quad \text{Sa}^{off} = 0.5 \cdot S_{\text{comp}}(+S_{\text{prec}}), \quad (5)
\]

if field \( b \) is defined as the half field with compensation light and field \( a \) as the half field without compensation light. \( Sa^{off} \) and \( Sb^{off} \) represent the light in the off-phase of the straylight ring, whereas \( Sa^{on} \) and \( Sb^{on} \) represent the light in the on-phase of the straylight ring. The on-phase light in the test fields is the straylight \( s \) originating from the flickering ring, summed in field \( a \) with half of the compensation light in field \( b \) to equalize the average luminance in both half fields (luminance equalizing light, explained in more detail in Franssen et al., 2006). The off-phase light is the compensation light \( S_{\text{comp}} \) in field \( b \). Half of this amount is again added as offset to field \( a \), serving as luminance equalizing light. Plotting \( P \) against \( S_{\text{comp}} \) or
log(S_{\text{comp}}) results in psychometric curves as they are measured in the practice of the Compensation Comparison method.

The right part of Equation 5 also considers the more general case that counterphase light can be added to both half fields (Figure 2). A fixed amount of light, called “precompensation,” is added in field a in the off-phase, and the term \( S_{\text{prec}} \) is added to \( S_{\text{off}} \) in Equation 5. In that case, the luminance equalizing term \( 0.5S_{\text{comp}} \) changes to \( 0.5(S_{\text{comp}} - S_{\text{prec}}) \) in Equation 5 when \( S_{\text{comp}} > S_{\text{prec}} \), or to \( 0.5(S_{\text{prec}} - S_{\text{comp}}) \) in Equation 5 when \( S_{\text{comp}} < S_{\text{prec}} \). With precompensation, very small modulation depths in both test fields are possible. Therefore, the model needs to be further refined by considering near and below threshold behavior. A formulation was chosen that gives the above Equation 1 as limit case for large suprathreshold flicker, and a threshold function (see below) as limit case for small flicker:

\[
P = \left( \frac{1}{1 + e^{-\text{MDT}} \varepsilon} \right)^{tr} \left( \frac{1}{1 + e^{-\text{MDT} \beta}} \right)^{1-tr}.
\]  

The exponent \( tr \) controls the transition between the two domains, the suprathreshold domain (left part of Equation 6), and the threshold domain (right part of Equation 6). MDT stands for modulation depth threshold. The threshold function was also based on psychometric functions often used in literature (Strasburger, 2001). When MDA (or MDB) is equal to zero, this function reduces to a form that compares well to logistic or Weibull distributions. It is easily checked that its key values are as they should be: the function value is 0.5 when also MDB (or MDA) is equal to zero (corresponding to guessing chance if both half fields are identical (or equal zero)); it is 0 for large MDA and 1 for large MDB; and it is precisely halfway if the other field is at threshold level: 0.75 when MDA = 0 and MDB = MDT, and 0.25 when MDB = 0 and MDA = MDT. The parameter \( \beta \) determines the slope of the threshold function. A value of \( \beta = 10/3\ln3 \approx 3 \) gives the same slope as a logistic function with \( \beta = 5 \) or a Weibull function with \( \beta = 3.5 \), values often found in literature (Strasburger, 2001). For the analysis of the data presented here, \( \beta \) was set to 3 (see below).

The transition parameter \( tr \) was further defined as

\[
tr = \frac{1}{1 + \left( \frac{\text{MDT}}{\text{MDA}} \right)^{\varepsilon}},
\]

with MDA = \( \sqrt{\text{MDA} \cdot \text{MDB}} \), the geometrical average of the modulation depths, and \( \varepsilon \) a parameter (set to 3 also, see below) that determines the speed of the transition between the two domains. The definition (Equation 7) of the transition parameter ensures that the transition is precisely halfway (\( tr = 0.5 \)) if the average modulation depth equals the threshold value (MDA = MDT).

The model parameters \( (s, \text{MDC}_c, \text{MDT}) \) were fitted by means of a maximum likelihood procedure (see, for example, Harvey, Jr., 1986) to the seven-subject laboratory data. After some preliminary experiments, the transition speed parameter \( \varepsilon \) was fixed at a value of 3. To better include the threshold domain, we performed measurements with different fixed values of compensation light in test field a. These precompensation values \( S_{\text{prec}} \) in Equation 5 used were chosen depending on the straylight parameter \( s \) values of the individual subjects. For example, for the oldest subject \( (s = 3.9) \), precompensation values up to 12.6 were used. For the youngest subject \( (s = 3.9) \), precompensation values up to 3.2 were used. As mentioned before, measurements were repeated with artificially increased straylight values for five subjects, achieved by holding a light scattering filter, found to represent early cataract (BPM2 filter, see de Wit et al., 2006), in front of the tested eye.

The model was also evaluated using the data of the GLARE study, but with two considerations: (1) The wide variation in ocular conditions found in this population can be expected to reflect itself in different psychophysical behaviors, and therefore in psychometric functions that differ between these 1,073 individuals. However, the limited number of trials (around 25) in clinical cases does not allow estimation of the full model parameters on an individual basis. (2) Precompensation was set to zero early in our studies on clinical or practical use of the method, such as the GLARE study. This makes the suprathreshold part of Equation 6 the dominating factor in most cases. Therefore, it was not possible to accurately estimate the modulation depth threshold MDT independently from these measurements.

To solve these issues, the straylight parameter value \( s \) of individual subjects was estimated by shifting a fixed shape psychometric function (with \( S_{\text{prec}} = 0 \)) to fit the data set of that individual. Fitting was again done by means of the maximum likelihood procedure. The straylight value was determined by the horizontal position of the minimum of the curve, where \( \text{MDB} = 0 \) and \( S_{\text{comp}} = s \). Once the \( s \) values for the individual measurements were obtained, the different data sets of the GLARE study could be summed to sufficient numbers to test the full model described above. All GLARE study measurements were performed twice and divided in nine groups of equal size, sorted on the differences between the two repeated measurements. In each group, Equation 6 was fitted to all data of that group together, after normalizing each individual curve for the individual straylight value. In this way, the data of the GLARE study could be used to study the psychometric behavior of a clinical or practical population, according to the model developed above.

Threshold experiments

The new parameter MDT found with the above approach was researched with some independent experiments to try to better understand the discrepancy between
repeatability values and flicker sensitivity, as mentioned in
the Introduction section. Flicker thresholds were measured
for five test screens with the same geometrical layout as
the Compensation Comparison test (7–14° radius ring and
2° radius test field, see Figures 3 and 4). In these five
experiments, luminance values in the test screen were
altered in order to approach the screen layout for a
classical flicker threshold experiment in a stepwise manner
(Table 1). Hence, with the different screen layouts the
effects on flicker threshold of different aspects of the
layout in the Compensation Comparison experiment could
be estimated. In the same way as with the Compensation

Figure 3. Field geometry for the flicker threshold experiments.
During a test, the flickering stimulus was presented randomly in
either Area I or II, whereas the luminance values of the other
fields were constant. Between different tests, some of these
luminance values were different (see Figure 4).

![Figure 3](image)

Table 1. Luminance values in the different field areas as
percentages of the maximum luminance for the five threshold
experiments. Field areas refer to Figure 3. The resulting field
geometrics are depicted in Figure 4.

<table>
<thead>
<tr>
<th>Experiment</th>
<th>Field area</th>
<th>IV</th>
<th>III + V</th>
<th>VI</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>IV (before</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>stimulus)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>IV (during</td>
<td>50%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>stimulus)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>IV (during</td>
<td>100%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>stimulus)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4 + 5</td>
<td>IV (during</td>
<td>0%</td>
<td>0%</td>
<td>0%</td>
</tr>
<tr>
<td></td>
<td>stimulus)</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 4. Field layouts for the different flicker threshold experiments. The luminance of Area IV (the straylight ring) in between stimuli was 100% in Experiments 1 and 3, 50% in Experiment 2, and 0% in Experiments 4 and 5.

![Figure 4](image)
Comparison test, the threshold experiments were performed with short stimuli, following a 2AFC procedure.

The general field geometry for the experiments is depicted in Figure 3. During a test, the flickering stimulus was presented randomly in either field Area I or II, whereas the luminance values of the other field areas were constant. Between different experiments, some of these luminance values were different. Table 1 gives the

Figure 5. Measured psychometric curves and corresponding model curves (Equation 6) for 7 subjects. Data points are averages over 8 (TK, DT), 10 (TB, JC, LF), or 12 (LR, GS) responses. Results are given for measurements with various levels of precompensation $S_{\text{prec}}$ (depending on the individual straylight parameter value). For each subject, Equation 6 was fitted to all data points with $s$ and $\text{MDC}_{c}$ (=MDT) as parameters (see Table 3).
luminance values in the different field areas as percentages of the maximum luminance (100 cd/m²) for all five experiments. The separation ring (Area VI in Figure 3) luminance was 100% in all experiments. The only difference between Experiments 4 and 5 is the width of this separation ring (1 pixel (about 3 arcmin), as opposed to 10 pixels (30 arcmin) in the other experiments). The resulting field geometries are depicted in Figure 4. Note that in Experiments 1 and 3 there was a step in the ring luminance (Area IV) at the beginning of each trial (as also present in the Compensation Comparison test), whereas no step occurred in the other experiments.

In all experiments, one of the two test field halves had a constant luminance, corresponding to the straylight value of the subject ($s = 14$ or $L = 1.8$ cd/m²). The other (flickering) test field half had the same average luminance. Both test field halves were dark (0%) between trials.

The data were fitted to a threshold psychometric function of the following shape (logistic function Strasburger, 2001):

$$P = \frac{0.5}{1 + 10^{-\beta \left(\log M_{Da} - \log M_{DT}\right)}}$$

In this function, the threshold value is situated at the 75% point of the curve ($P = 0.75$ when $M_{Da} = M_{DT}$).

### Results

Results of the laboratory experiments are presented in Figure 5 (without BPM2 filter) and Figure 6 (with BPM2 filter). Similar to Figure 5, only five subjects with BPM2 filter in front of their eye. Data points are averages over 4 (TB, JC, LF) or 8 (TK, DT) responses. Model parameter values are given in Table 3.
The results of each subject are plotted in a separate graph, and model fit curves are drawn for each precompensation condition. Values for the independently fitted parameters \( s \), \( \text{MDC}_c \), and \( \text{MDT} \) are presented in Table 2. From the results in this table, it seems that \( \text{MDC}_c \) and \( \text{MDT} \) are more or less equal. Moreover, because in the fit process both parameters counteract each other, we tested to set them equal. A renewed analysis was carried out, of which results are given in Table 3. The \( \log(s) \) values here are almost identical to those in the original fit (Table 2), and also the (average) \( \text{MDT}/\text{MDC}_c \) values did not change much. Moreover, the choice \( \text{MDT} = \text{MDC}_c \) virtually did not influence the precision of the fit. Therefore, this simplification was adopted. The curves in Figures 5 and 6 were fitted with this condition.

Tables 2 and 3 and Figures 5 and 6 show that in this small laboratory population the model fitted all data equally well, with small differences in the model parameters, apart from the parameter \( \log(s) \). The \( \log(s) \) parameter not only differed because of inter-individual differences (such as age), but also because of the addition of the BPM2 filter. Note that the inter-individual differences found are smaller for the BPM2 data, which is as expected, because the straylight from the filter dominates over the straylight differences between the subjects.

The model was further validated by applying it to the field measurements of 1,073 subjects, performed in the GLARE study, as described in the previous section. For these measurements, it would not have been possible to accurately estimate the modulation depth threshold \( \text{MDT} \) independently from these measurements because the precompensation was set at zero, as mentioned in the previous section. Hence, in this case only the condition \( \text{MDT} = \text{MDC}_c \) was fitted. Data were sorted from the best to the worst observers and split in nine groups, in the same way as reported earlier (Franssen et al., 2006). Results for the first group (the best observers) are shown in the top row of Figure 7; results for the last group (the worst observers) are shown in the bottom row. The areas of the circles in this figure indicate the amount of trial responses that were averaged for the corresponding data points. The largest circles represent around 500 trial responses. Also note the differences between the slopes of the fitted psychometric functions of the different subgroups, accounted for in the model by different \( \text{MDT} = \text{MDC}_c \) values.

Results of the flicker threshold experiments under the different field layout conditions, as well as the psychometric function fits and their corresponding \( \text{MDT} \) values are given in Figure 8 and Table 4. The parameter \( \beta \) (Equation 8) was simultaneously fitted and found to be 4.88, well in correspondence with the value of five commonly found in literature (Strasburger, 2001).

<table>
<thead>
<tr>
<th>Subject (age)</th>
<th>log(s)</th>
<th>log(( \text{MDC}_c ))</th>
<th>log(( \text{MDT} ))</th>
<th>0.5 log(( \text{MDC}_c ) \cdot ( \text{MDT} ))</th>
</tr>
</thead>
<tbody>
<tr>
<td>TK (24)</td>
<td>0.76</td>
<td>-0.99</td>
<td>-1.27</td>
<td>-1.13</td>
</tr>
<tr>
<td>DT (23)</td>
<td>0.83</td>
<td>-0.89</td>
<td>-1.10</td>
<td>-0.99</td>
</tr>
<tr>
<td>TB (57)</td>
<td>1.15</td>
<td>-0.94</td>
<td>-1.08</td>
<td>-1.01</td>
</tr>
<tr>
<td>JC (35)</td>
<td>1.05</td>
<td>-1.09</td>
<td>-0.73</td>
<td>-0.91</td>
</tr>
<tr>
<td>LF (29)</td>
<td>0.73</td>
<td>-0.93</td>
<td>-0.68</td>
<td>-0.80</td>
</tr>
<tr>
<td>LR (21)</td>
<td>0.57</td>
<td>-0.70</td>
<td>-1.05</td>
<td>-0.87</td>
</tr>
<tr>
<td>GS (22)</td>
<td>0.65</td>
<td>-0.79</td>
<td>-0.97</td>
<td>-0.88</td>
</tr>
<tr>
<td>TK BPM2</td>
<td>1.29</td>
<td>-1.09</td>
<td>-1.08</td>
<td>-1.09</td>
</tr>
<tr>
<td>DT BPM2</td>
<td>1.26</td>
<td>-1.01</td>
<td>-0.92</td>
<td>-0.96</td>
</tr>
<tr>
<td>TB BPM2</td>
<td>1.39</td>
<td>-1.33</td>
<td>-0.95</td>
<td>-1.14</td>
</tr>
<tr>
<td>JC BPM2</td>
<td>1.39</td>
<td>-0.93</td>
<td>-0.90</td>
<td>-0.91</td>
</tr>
<tr>
<td>LF BPM2</td>
<td>1.26</td>
<td>-1.14</td>
<td>-0.99</td>
<td>-1.06</td>
</tr>
<tr>
<td>Average</td>
<td>-0.98</td>
<td>-0.97</td>
<td>-0.97</td>
<td>-0.98</td>
</tr>
<tr>
<td>SD</td>
<td>0.17</td>
<td>0.16</td>
<td>0.12</td>
<td></td>
</tr>
</tbody>
</table>

Table 2. Results from maximum likelihood fits of Equation 6 to the Compensation Comparison measurements of 7 subjects and 5 subjects with BPM2 filter (artificial straylight increase). The last column gives the logarithm of the geometrical average of \( \text{MDC}_c \) and \( \text{MDT} \).
figure and table show a general decline from around 4% (0.02 log units) to around 2% (0.01 log units) in modulation threshold (MDT) with field layouts closer to the classical case. The step function in the ring (Area IV) at the start of each trial seems to have little effect on the threshold value (Experiments 1 and 2), and the same seems to hold for the width of the separation ring (Area VI, Experiments 4 and 5). The largest effect seems to be caused by the area surrounding the test field halves (Area III). It must be noted here that each field layout causes a different amount of straylight in the test fields (Areas I and II). To evaluate the importance of straylight in this experiment, we calculated equivalent luminance values in the test fields from each stimulus area and presented them in Table 4, along with the fitted modulation threshold values for each stimulus condition. As outlined elsewhere (van den Berg, 1995), calculation of equivalent luminance involves the luminance of the straylight source, the straylight parameter of the subject, and the ratio between the outer and inner radius of the straylight source.

Note that the threshold value for the field layout most resembling the Compensation Comparison field layout (Experiment 1) is more than twice as low as the average threshold value found in the Compensation Comparison laboratory experiments (Table 2). The only difference between these conditions is the flickering aspect of the luminance in Area IV (the straylight ring in the Compensation Comparison test). This points to a significant sensitivity-lowering effect of flicker in Area IV per se. Hence, a kind of flicker adaptation effect over distance seems to play a role here.

Discussion

In this paper, a more general model for the psychophysics of the Compensation Comparison method for measuring retinal straylight was introduced. The generalization comprised the addition of counterphase flickering light, called precompensation, to the test field half originally without compensation light. To take into account the low modulation depths that can occur in the test fields as a result of this precompensation, we extended the psychophysical model for the flicker comparison task with a component describing the behavior near flicker threshold. Actual flicker thresholds were further investigated with measurements under various screen geometry conditions. A second importance of the extension is to incorporate psychometric behavior of subjects with low flicker sensitivity.

Figure 7 shows that the extended model is capable to describe the psychophysical behavior of a population that varies widely with respect to physical condition of the eye. The model fits very well to the data, even for the subgroup with the largest repeated measurement differ-
ences (bottom right graph). Yet it is clear that for this subgroup there are real differences between model and reality. However, it must be noted that for some cases in this subgroup response behavior was so erratic that reliably fitting a psychometric curve, and therefore reliably estimating the \( \log(s) \) value, is not possible. For the best subgroup, the model fits to the data equally well as the simple model without threshold (Equation 1); but for the worst subgroups (bottom row Figure 7), the extended model (Equation 6) performs clearly better. As already mentioned in the Methods section, the supra-threshold part of Equation 6 was dominant for most measurements of the GLARE study. However, this might be less so for the worst measurements, which would explain the better performance of the extended model for these cases.

The generalized approach with precompensation was evaluated with laboratory experiments on a small number of different field layout conditions. Each point in the graphs is the average of 20 trials. The data were fitted to a logistic psychometric function (Equation 8) with MDT and \( \beta \) as parameters. The MDT values are given in the corresponding graphs, and also indicated by a vertical line at the 75% point in each graph. The parameter \( \beta \) was found to be 4.88.
of subjects. The new model describes the measured data well (Figures 5 and 6) for a wide range of straylight values (Tables 2 and 3). The log(s) values without BPM2 filter all fall within the normal population range, which in the past was shown to increase with age (van den Berg, 1995). The log(s) values with BPM2 filter show less variation, as explained in the previous section. The straylight values found with or without the model restriction MDT = MDCc (Tables 2 and 3) are virtually the same, which is another indication that this model restriction is justified.

In Figure 5 as well as in Figure 6, the minimum of the psychometric curves drifts away from 0 with increasing precompensation value. In this minimum, the straylight flicker is completely extinguished by the compensation flicker in one test half (MDb = 0 and Scomp = s). In the other test half, the retinal flicker is determined by the straylight and the precompensation value. In other words, these minima correspond precisely to a threshold experiment (one half with no flicker at all). With increasing precompensation value, the flicker in this test half varies from above threshold to near zero, causing the minimum in the curve to vary from near 0 to near 0.5. In case the precompensation would precisely equal the straylight, the curve would have a minimum at \( P = 0.5 \).

The choice for the logistic function as the basis for our psychometric model was mainly made for simplicity reasons. Because the data could be fitted to a high level of precision with this model, we saw no reason to switch to a different function, such as the Weibull function. It cannot be excluded, though, that a different function might have worked equally well. Several authors (e.g., Strasburger, 2001) have pointed out that the different mathematical descriptions for the psychometric function effectively work out in a very similar way.

As with most models, however, there are some limitations. On close inspection, some of the curves in Figures 5 and 6 show a discontinuity in the minimum of the curve (where MDb = 0 and Scomp = s). This is caused by the transition between the two domains in Equation 6. We could have remedied this by much more complicated modeling. Because the discontinuities are small, we adhered to the present model.

One of the questions for the study was to understand the high thresholds suggested by the relatively low level of accuracy in the Direct Compensation method. The dependence of MDT on screen layout was summarized in Table 4. To understand this, the fact must be considered that steady light sources in the surroundings (Areas III to VI in Figure 3) cause a straylight offset in the test fields (Areas I and II), thereby decreasing the retinal modulation depth. Assuming Weber law behavior, the threshold level is proportional to mean light level. Table 4 illustrates how much light in total reaches the fovea, apart from the intended light in a straylight experiment (\( L = 1.85 \text{ cd/m}^2 \) for this subject). From column 6 (total), it seems clear that light scattered from the other areas plays a significant role in the reduction of sensitivity found in the fovea. However, although the MDT values found in stimulus layout situations 4 and 5 (0.021 and 0.025) are close, they still seem a bit high as compared to classical threshold values (order of 0.01 or 1%, as mentioned in the Introduction section). It must be noted though that in these classical experiments, the luminance of the foveal test fields was kept at the value of the average luminance of the test field, which appeared to yield the lowest possible modulation threshold (de Lange, 1958). The test field layout in two half fields may also play a role here.

To summarize the discrepancy between the classical flicker threshold (around 1%) and the flicker threshold in a straylight measurement (around 10%), three factors seem to play about equal roles: (i) the test field itself, (ii) the straylight from the surroundings; and (iii) the flicker adaptation effects over distance from the straylight source (Area IV). Further study is needed to evaluate this adaptation over distance. The strength of this effect might vary from subject to subject, and it would be of interest to establish whether subjects with poorer psychometric

---

### Table 4

<table>
<thead>
<tr>
<th>Stimulus condition</th>
<th>Stimulus luminance</th>
<th>Area III</th>
<th>Area IV</th>
<th>Area VI</th>
<th>Total luminance</th>
<th>MDT</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1.85</td>
<td>2.94</td>
<td>0.93</td>
<td>0.60</td>
<td>6.32</td>
<td>0.041</td>
</tr>
<tr>
<td>2</td>
<td>1.85</td>
<td>2.94</td>
<td>0.93</td>
<td>0.60</td>
<td>6.32</td>
<td>0.043</td>
</tr>
<tr>
<td>3</td>
<td>1.85</td>
<td>2.94</td>
<td>0.00</td>
<td>0.60</td>
<td>5.39</td>
<td>0.029</td>
</tr>
<tr>
<td>4</td>
<td>1.85</td>
<td>0.00</td>
<td>0.00</td>
<td>0.60</td>
<td>2.45</td>
<td>0.021</td>
</tr>
<tr>
<td>5</td>
<td>1.85</td>
<td>0.00</td>
<td>0.00</td>
<td>0.07</td>
<td>1.92</td>
<td>0.025</td>
</tr>
</tbody>
</table>

Table 4. Effects in the central test field (Areas I and II in Figure 3) for the different experimental stimulus conditions. Column 2: mean luminance really presented in Areas I and II. Columns 3–5: Equivalent luminance in the two test fields resulting from light presented in the other stimulus areas. These values were calculated (van den Berg, 1995) for the maximum test screen luminance of 100 cd/m\(^2\) and for the subject’s straylight parameter s = 14. Column 6: total luminance in the test fields (sum of columns 2 to 5). Column 7: fitted modulation depth threshold (MDT) (from Figure 8).
behavior exhibit higher foveal increases in flicker thresholds in response to peripheral flicker.

The generalized psychophysical model for the Compensation Comparison method, as presented in this paper, fits to laboratory data as well as field data. If no precompensation is used, the model performs equally well as the previous model that did not take into account flicker threshold. However, the new model does perform better with subject groups showing less reliable psychometric behavior.

It is to be expected that implementation of precompensation in the Compensation Comparison method will result in better efficiency and more accurate straylight measurements in clinical as well as research applications. To assess this issue, large-scale population studies are necessary. How much precompensation should be used is another central question. The amount of precompensation may depend on several factors but will at any rate be related to the straylight value of the subject. Adaptive procedures as trial strategies are currently under investigation, “adaptive” meaning that the amount of precompensation depends on the previous responses of the subject (Treutwein, 1995). Preliminary results show promise for significant improvement of efficiency and accuracy.

Acknowledgments

This research was conducted in the Ocular Signal Transduction group at the Netherlands Ophthalmic Research Institute in order to develop a better scientific basis for the C-Quant straylight meter (Oculus GmbH).

Commercial relationships: none.
Corresponding author: Luuk Franssen.
Email: l.franssen@nin.knaw.nl.
Address: Netherlands Institute for Neuroscience, Ophthalmic Research Institute, Royal Netherlands Academy of Arts and Sciences, Meibergdreef 47, 1105 BA, Amsterdam, The Netherlands.

References