Detection of the stroboscopic effect by young adults varying in sensitivity

JA Veitch PhD[a] and **C Martinsons** PhD[b]
[a]NRC Construction, National Research Council of Canada, Ottawa, Canada
[b]Division of Lighting and Electromagnetic Fields, Centre Scientifique et Technique du Bâtiment, Saint Martin d'Hères, France

The advent of LED lighting has renewed concern about the possible visual, neurobiological, and performance and cognition effects of cyclic variations in lighting system luminous flux (temporal light modulation). The stroboscopic visibility measure (SVM) characterises the temporal light modulation signal to predict the visibility of the stroboscopic effect, one of the visual perception effects of temporal light modulation. A SVM of 1 means that the average person would detect the phenomenon 50% of the time. There is little published data describing the population sensitivity to the stroboscopic effect in relation to the SVM, and none focusing on people subject to visual stress. This experiment, conducted in parallel in Canada and France, examined stroboscopic detection for horizontal and vertical moving targets when viewed under commercially available lamps varying in SVM conditions (SVM: ~0; ~0.4; ~0.9; ~1.4; ~3.0). As expected, stroboscopic detection scores increased with increasing SVM. For the horizontal task, average scores were lower than the expected 4/8 at ~0.90, but increased non-linearly with higher SVMs. Stroboscopic detection scores did not differ between people low and high in pattern glare sensitivity, but people in the high-pattern glare sensitivity group reported greater annoyance in the SVM ~1.4 and ~3.0 conditions.

1. Introduction

The introduction of solid-state lighting to the marketplace has brought renewed concern about possible adverse consequences of exposure to temporal cyclic or transient variations in lighting system luminous flux, known as temporal light modulation (TLM) (or more commonly referred to as 'flicker'). TLM may have visual, neurobiological, performance and cognitive effects on viewers.[1–3] Visual perception effects such as the stroboscopic effect occur very quickly, with very short exposures.[1–3] TLM can also cause ill effects on a longer time scale, such as disruptions to eye movements,[4] visual and task performance[5,6] and headaches and eyestrain[7] (see also the French expert report on health effects issued in May 2019[8] or the English summary[9]). There is not yet expert consensus about all of the possible health and behavioural effects of TLM, and no single metric to predict their occurrence. This remains an active area for research and standardisation.[10]

The visual perception effects are collectively known as temporal light artefacts (TLA), comprising flicker, the stroboscopic effect and the phantom array effect. The stroboscopic effect is said to occur when an observer who views a moving object under a

Address for correspondence: JA Veitch, NRC Construction, Building M-24, National Research Council of Canada, 1200 Montreal Road, Ottawa, ON K1A 0R6, Canada.
Email: jennifer.veitch@nrc-cnrc.gc.ca

lighting system that exhibits TLM experiences '*a change in motion perception*'.[11] Usually this change in motion perception is of the form that the motion appears interrupted or jerky, rather than smooth.

TLM can be characterised in several ways, all derived from measurements of the light output waveform. Among the parameters available to describe TLM are the dominant frequency, the modulation depth, the IES flicker index,[12] a predictor of visible flicker from the International Electrotechnical Commission (IEC), called P_{ST}^{LM},[13] the stroboscopic visibility measure (SVM)[11] and the ASSIST stroboscopic acceptability measure.[14] Others may be developed as research progresses, and to support the development of new metrics, the International Commission on Illumination (CIE) Technical Committee TC 2–89 is tasked with developing a technical note to describe a rigorous method for measuring and reporting TLM waveforms.

For the present investigation, which concerns the stroboscopic effect, we characterised the independent variable in terms of the SVM. The SVM is a visibility measure that is derived from measurements of the TLM of the light source or lighting system. This characterisation of the TLM relates to the visibility of the stroboscopic effect and is scaled such that, by definition, a value of 1 means that the average person would detect the phenomenon 50% of the time; thus, a light source having an SVM value of 1 would mean that the average person can detect the stroboscopic effect 50% of the time when that light source is the sole source of illumination.[11,15] As noted by the CIE and by the National Electrical Manufacturers' Association (NEMA), the visibility threshold (i.e. SVM = 1) is not a guarantee of acceptability of the visible phenomenon.[11,16]

Prior to the development of the SVM, Bullough *et al.*[17] examined the visibility and acceptability of flicker and of the stroboscopic effect across a range of TLM frequencies, modulation depths and duty cycles, and obtained data from 10 participants viewing each of nine conditions once, for less than 2 minutes. The stroboscopic effect was detectable at frequencies up to 300 Hz for a hand waving under the lamp, but acceptability of what had been seen was high for all frequencies at and above 120 Hz (controlling for modulation depth), and when modulation depth was 33% (but not 50% or 100%).

A subsequent experiment,[18] also with 10 participants but changing the task to a wand waved under the light source, found that 100 Hz TLM was detected 80% or more of the time for modulation depths equal to or greater than 25%. The conditions were rated as just acceptable, on average, when the modulation depth was 25%, but unacceptable at higher values of modulation depth. The small sample sizes in these two studies (N = 10 each, with more males than females in each one) limit their generalisability to the general population, and the fact that individuals controlled the wand movement risks inconsistency in the visual task.

The most relevant paper to the question of acceptability is that of Perz and her colleagues[19] who reported a series of experiments as one data set with a combined range of SVM conditions from 0 to 4.9. An unknown number of participants performed various office-work-type tasks for variable amounts of time under varying sets of these conditions, and then rated the acceptability of the room conditions. Based on these data, Perz and his colleagues constructed a logistic function to predict the percentage of annoyed people in relation to the SVM, concluding that 20% of the population would be annoyed at an SVM of 1.5, and ~12% at an SVM of 1. It is problematic that there is no information available about the sample size or composition, because it is impossible to assess how generalisable the results might be. Moreover, by combining the results from

studies in this way, the function combines within-subject and between-groups variation, which adds to the difficulty of interpreting the results.

To the best of our knowledge, the published literature does not address the population rates for stroboscopic motion detection under varying levels of the SVM. The metric's development has included several experiments, each with samples ranging from 20 to 35 individuals ranging in age from a low of 17 years to a high of 47 years. The data from these experiments show considerable individual variability in responses to TLM.[15,20] Perz and her colleagues developed the SVM based on data for the visibility of the stroboscopic effect when viewing a white dot on a black disc rotating at 4 m/s, viewed under various TLM conditions. The waveform for the stimulus was captured and analysed in the frequency domain using a Fourier transform to identify the frequency components in the signal. These are then normalised using an average visibility threshold function derived from empirical data, and summed to provide the definition that at SVM = 1, the average individual should detect the stroboscopic effect 50% of the time (the visibility threshold).[11,15] Although this is a valid choice to characterise the population average, it does not provide information on the full range of perceptions across the population.

Data have not been published showing the effect of varying levels of the SVM on stroboscopic visibility in the population at large: That is, what proportion of the population do, in fact, detect the stroboscopic effect when SVM = 1? Put another way, at what value of the SVM does the stroboscopic effect become almost undetectable? Furthermore, are there differences in stroboscopic visibility between people who might be at risk for visual stress[21] and those who are not?

We report here the final findings of an experiment designed to answer these questions. (A preliminary paper from this experiment, based on a partial data set, was presented at the CIE 29th Quadrennial Session[22].) The objectives of the project were as follows:

- Test the visibility of the stroboscopic effect for five levels of the SVM (targeting 0, 0.4–0.6; ~1; ~1.6; and >2) using an experimental method as similar as feasible to the published work from which the metric was developed.[15,20]
- Examine the population frequency of pattern glare sensitivity [PGS],[23] which is known to predict sensitivity to headache and disrupted eye movements in response to TLM.
- If possible, establish preliminary information about the visibility of the stroboscopic effect by individuals high in PGS; and
- Collect preliminary information about how people judge the acceptability of the conditions.

2. Research design and hypotheses

This is a repeated-measures experiment with one independent variable, light source; it had five levels. The five light sources were commercially available products chosen because they are known to exhibit the chosen levels of the metric SVM.[15,24] They were chosen based on the following criteria: about 800 lm output; ~2600–3000 K correlated colour temperature; $P_{st}^{LM} \ll 1$, and having close to the target SVM values (0, 0.4–0.6; 1; 1.6; and >2). To the extent possible, chromaticity coordinates were matched, although this was dependent on the existence of alternatives at any given SVM value.

The hypotheses tested were:

H1: Participants will detect the stroboscopic effect on 50% of trials for SVM = 1.

H2: Participants will detect the strobo-scopic effect on an increasing percentage of trials with increasing SVM.

H3: Comfort and pleasantness drop with increasing SVM, and annoyingness increases with increasing SVM.

Data were collected in parallel in Canada and in France, using the same protocol. In Canada, this research protocol was reviewed and approved by the National Research Council of Canada Research Ethics Board (Protocol 2018-139) and by the Carleton University Research Ethics Board-B (CUREB-B Clearance # 109982). This research complies with the EU and French General Data Protection Regulation (GDPR).

2.1 Participants

This paper is based on data from 85 people, 27 tested in France and 58 tested in Canada. Table 1 summarises their characteristics. All participants were university students, and none was older than 32 years of age. The limited age range was chosen because there is evidence that younger people may be more sensitive to TLM.[25] All had self-reported normal or corrected-to-normal vision, and normal hearing.

The Wilkins and Evans Pattern Glare Sensitivity test asks the observer to look, in sequence, at three patterns of horizontal stripes of three spatial frequencies, pattern 1 having the lowest spatial frequency and pattern 3 the highest.[23,26] Sensation scores on Pattern 2 were used as the indicator of a higher risk of visual stress, as suggested in Wilkins *et al*.[26] There was no difference between NRC and CSTB on this variable (Mann–Whitney U test, $Z = -0.50$, $p = 0.62$). Another suggested indicator of pattern glare sensitivity is the difference between responses to pattern 3 and pattern 2.[23] We have not reported this value because we have learned that scores on pattern 3 may be unduly

Table 1 The number of participants in each demographic group shown by location.

	Sex		Age (years)		Pattern 2 high score (High PGS) Group
	Male	Female	18 to 29	30 to 39	
CSTB	16	11	26	1	7
NRC	18	40	56	2	19

affected by visual capabilities, making it less reliable than is desirable.[26]

The relationship between pattern 2 sensation scores and discomfort ratings for pattern 2 was high for CSTB ($r = 0.53$, $p = 0.005$, $N = 27$), whereas it was moderate for NRC ($r = 0.26$, $p = 0.05$, $N = 58$). We considered this adequate evidence for the validity of the pattern 2 sensation scores as the indicator of sensitivity and of risk of visual stress. Therefore, we created a grouping variable (PGS) for this purpose by selecting the top 30% of the full sample, which were the individuals having scores of 4 or greater on this variable ("high-PGS"), versus the lower 70% of the sample ("low-PGS"). We chose this value based on the evidence that for the highest scores on that pattern, approximately 30% of the population report adverse effects (discomfort).[26]

2.2 Lighting conditions, apparatus and setting

At both sites, testing occurred in a dedicated, small, windowless room. The room was minimally furnished with a desk and chairs. During the session trials, all the illumination came from the custom desktop luminaire described below. During the instructions and demographic questionnaires, a desk lamp with the low-SVM condition was used on a separate desk from the one holding the apparatus (see Figures 1 and 2)).

Each site designed its own custom luminaire consisting of an aluminium frame supporting a light box, although the details differed between sites. The central light box
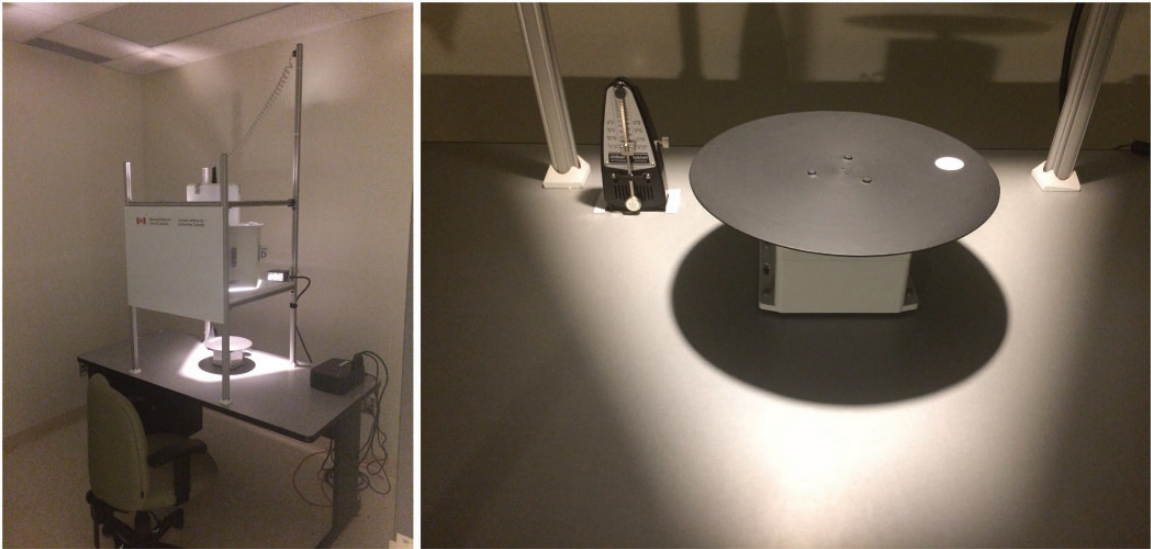
Figure 1 These are images of the NRC apparatus installed in the test room. For the image on the left, the hallway outside provided ambient light for the photograph, but the door was closed during testing. The image on the right shows the desk surface as seen by the participant.

Figure 2 This shows the CSTB apparatus installed in the test room.

had six chambers, in each of which was a standard E27 socket. Five locations were used, each with one of the five lamps (described below). All lamps were turned on, warmed for an hour before the first session and remained on until the end of the testing

session to maintain constant temperature and light output.

At NRC, the light box drum could be rotated such that one chamber was located over an aperture that allowed light to fall onto the desk surface, while the other lamps were blocked by the plywood base of the light box. The chambers in the light box were painted black (NRC) or covered with black adhesive velvet sheets, but some were modified with white reflective plastic to increase the illuminance on the desk below when that chamber was in use. There was no diffuser over the lamps, but the participant was shielded from any view of the aperture by a cover on the frame. At CSTB, each chamber of the light box had an individual shutter. The selection of the chamber was also done by rotating the device.

The height of the luminaire was adjusted once so that all the lamps delivered ~330 lx on the surface of the principal task, a rotating disk (see below). There were no sides to the

Table 2 Test lamp characteristics claimed CCT and luminous flux, and light characteristics measured horizontally at the location of the rotating disk under the light box for each of the five lighting conditions.

Condition	CCT [K]	Luminous. Flux [lm]	Illuminance [lx]	CCT [K]	R_a	Duv	Dominant Frequency [Hz]	Modulation [%]	Flicker Index [%]	P_{st}^{LM}	SVM
NRC-1	2700	800	341	2872	83	−0.0008	120	4.7	0.43	0.05	0.04
NRC-2	2700	800	319	3018	84	−0.0016	120	14.0	3.79	0.07	0.42
NRC-3	2700	800	354	2717	83	−0.0001	120	32.0	8.47	0.08	0.91
NRC-4	3000	800	334	3094	83	−0.0023	120	55.6	13.25	0.06	1.38
NRC-5	3000	800	335	3027	83	−0.0003	120	91.5	29.99	0.33	2.80
CSTB-1	2700	806	344	2756	83	0.00001	100	2.1	0.6	0.39	0.00
CSTB-2	2700	810	330	2810	82	0.0009	100	11.8	3.7	0.05	0.43
CSTB-3	2600	720	318	2559	90	0.0022	100	27.8	7.9	0.08	0.96
CSTB-4	2700	810	312	2641	81	0.0016	100	40.2	12.3	0.26	1.47
CSTB-5	2700	600	324	2799	80	0.0022	100	79.4	26.9	0.38	3.09

Note: For CCT and luminous flux, we show the manufacturer's package information for the lamps in columns 2 and 3. All of the other characteristics were measured at NRC and CSTB, respectively.
CCT: correlated colour temperature.

frame, so that viewers had the full field of view available to them and light from the luminaire could provide ambient illumination for the rest of the room beyond the desk on which it sat. An uninterruptible power supply was used by NRC to ensure clean power for the luminaire during testing. A laboratory specification AC power supply was used by CSTB for the same purpose.

LED lamps were selected and purchased from the market to match the five SVM levels, and hence represent SSL products available in the market in North America and Europe. Although the specific lamps used by the laboratories were different in the two countries, the five experimental conditions were matched in terms of the SVM for methodological consistency.

Table 2 summarises the key information about the lamps used at each site, showing that they were similar enough to be expected to produce comparable visual perceptions. At both sites, the illuminance on the surface of the rotating disc was between 330 and 350 lux, the light source correlated colour temperatures (CCTs) were in the range of 2700–3000 K and the general colour rendering index, R_a, was between 80 and 85. P_{st}^{LM} is a metric that characterises TLM in the range of 0–80 Hz, which is the range in which viewers can report seeing the temporal variation in light output (flicker). The light sources were chosen to keep P_{st}^{LM} well below 1 to avoid experimental confounding. The choice of light sources in each country was made on the basis of matching the SVM values as closely as possible, but we have also reported other common TLM metrics to permit comparison with other investigations. Both the modulation depth and flicker index values for the light sources give the same rank order, and both correlate very highly with the SVM, $r = 0.97$ for modulation depth and $r = 0.98$ for flicker index.

2.3 Dependent variables

The same questions and tasks were used in both countries. They were originally written in English and translated to French at CSTB.

2.3.1 Stroboscopic effect

Within each trial, there were two probes for the stroboscopic effect. The first of these used a white dot on a rotating horizontal black disk, as was used by prior researchers[11,15,20,27] (see Figure 1). The dot on the disk rotates at a
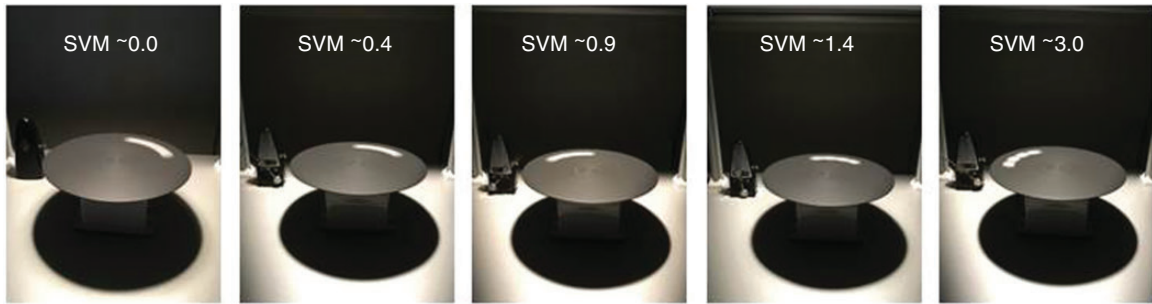
Figure 3 This series of images shows conceptually the effect of increasing the SVM on the detection of the stroboscopic effect on the rotating disc.

speed of 4 m/s, which the prior researchers considered to be the upper limit of the speed of hands moving in an office context. The participant was asked to look at the disk and to report whether or not they saw individual dots (stroboscopic effect) (see Figure 3). The rotating disks used in this experiment were designed, assembled and programmed at NRC using a programmable DC motor, two being shipped to CSTB with a suitable power cable for operation in France. The reflectance of the black surface was $\rho = 6.96\%$ and that of the white dot was $\rho = 90.85\%$, making the luminance ratio 13.05:1, as similar as possible to the original paper.[20] For each trial, the participant was asked to answer 'yes' or 'no' to the question 'Do you see white dots?' at NRC and 'oui' or 'non' to 'Voyez-vous des points blancs ?' at CSTB.

We also added a vertical task. Participants were asked to look at a black dot on the end of the arm of a mechanical metronome moving at 208 beats per minute (bpm) in the data reported here. (Veitch and Martinsons[22] reported preliminary data for 180 bpm in Canada and 150 bpm in France in the first round of data collection, but increased the metronome speed to 208 bpm for both locations for the remaining sessions.) For each trial, the participant was asked to answer 'yes' or 'no' to the question 'Do you see black dots?' at NRC and 'oui' or 'non' to 'Voyez-vous des points noirs ?' at CSTB (see Figure 1 for the metronome used in Canada and Figure 2 for France). The metronomes were identical in the two countries, but each team made and attached its own black dot. The metronome location differed slightly for the NRC and CSTB. At NRC, it was farther from the observer than the rotating disk, and slightly to the side, with the white wall ($\rho \sim 0.7$) of the room behind (see Figure 1, right image). At CSTB, the line of sight placed the very light grey desk surface behind the moving metronome arm (see Figure 2).

2.3.2 Judgements of light sources.

On the last trial for each lamp (i.e. after having repeated the visibility task 10 times), participants were asked to rate the comfort, pleasantness and annoyingness of that condition, each on a 5-point Likert scale.[20] See Table 3 for the exact wording used at NRC and CSTB. At NRC, the experimenter was prepared with definitions, if needed: 'Pleasantness refers to whether or not the condition gives a sense of satisfaction when one looks at it. Comfort refers to how one feels when one looks at the scene, a state that is one of physical ease and free from pain'.

Table 3 The light source judgements were asked in English at NRC and in French at CSTB.

NRC	Please rate the comfort of this condition		Please rate the pleasantness of this condition		Please rate the annoyingness of this condition
CSTB	Evaluez le confort de cet éclairage		Evaluez l'aspect agréable de cet éclairage		Evaluez la gêne que procure cet éclairage
	0	1	2	3	4
NRC CSTB	Not at all Pas du tout Confortable Agréable Gênant	A little Un peu Confortable Agréable Gênant	Moderately Modérément Confortable Agréable Gênant	Very much Très Confortable Agréable Gênant	Extremely Extremement Confortable Agréable Gênant

Note: The top part of the table shows the three questions in the two languages, and the bottom part of the table shows the response anchors in the two languages.

2.4 Procedure

When participants arrived, a lamp at the lowest SVM level provided the room light. Participants received information about the study and signed the consent form in this condition. They also completed a short paper-based questionnaire to record demographic information (age, sex, education, eye colour [28] and visual corrections).

For the visual perception trials, the participant was asked to rotate away from the desk that held the task (facing the opposite wall) and to close his or her eyes while the researcher set up each trial. Setting up involved moving the light box to reveal one or another light source. Light sources were presented in blocks of five with the conditions in random orders in each block. The random orders of presentation were listed on a pre-printed data sheet for that session. The experimenter asked the participant to turn around, and to look at first the rotating disk to answer the question 'Do you see white dots?' with an answer 'yes' or 'no'. Next, the same question was asked for the metronome. After this second question, the participant turned away and closed his or her eyes while the next trial was set up.

In the final block of five trials, the participant was asked to rate the appearance of the condition on the three scales described above after the metronome question.

At the conclusion of the session, the participant was provided the debriefing information sheet and asked not to share the information with other potential participants.

Participation took approximately 50 minutes. In Canada, participants either received an honorarium of $20 for their participation or were awarded 1% bonus credit for a Psychology undergraduate course. In France, participants received a 15 € gift card for their participation.

A total of 96 people were tested in the two countries. Five cases from France were excluded because one of the lamps presented was the wrong lamp and participants had not seen the condition with the SVM = 0.40. The assistants who collected the data reported that some participants had not paid attention to the instructions and appeared not to have attended to the visual task. Rather than relying only on these impressions, we identified six cases (two from Canada and four from France) for exclusion from analysis because they scored 4 or higher positive responses to the condition SVM = 0 (i.e. greater than chance reporting of the stroboscopic effect where none should exist). These cases were consistent with the assistants' notes

taken at the time, which suggested that these individuals might be responding randomly regardless of what they saw rather than giving a true response. Thus, the total sample on which results are based numbered 85.

2.5 Data analysis

We used non-parametric tests because the distributions for the stroboscopic effect detection scores did not meet the assumptions of normality required for parametric tests (e.g. analysis of variance), and applied the same statistical model to all analyses. In all analyses in this report, the alpha criterion for statistical significance was set to $p < 0.05$. Sample sizes vary between the analyses because of missing data, as explained below.

For every level of the SVM, we compared the data from the two locations (CSTB vs. NRC). We also combined the data from both sites, and then formed groups based on individuals' PGS pattern 2 scores (see Section 2.1). The comparisons between locations and the comparisons between PGS groups used the Mann–Whitney U test. The tests were repeated for each individual SVM level and for an overall score formed by averaging the detection scores for the five SVM levels.

We also tested for differences between the SVM levels in four successive planned comparisons (0 vs. 0.4; 0.4 vs. 0.9; 0.9 vs. 1.4; 1.4 vs. 3). For these repeated-measures, we used the Wilcoxon Signed Ranks test. We performed these repeated measures tests both for the combined sample and separately for the two locations and for the high and low pattern sensitivity glare (PGS) groups.

3. Results

3.1 Stroboscopic effect – Rotating disc

Data from the first two trials were treated as training runs, and not included in the analysis. For each participant, we averaged the responses to trials 3–10 and scaled them

out of a score of 8 (detection score, with theoretical minimum $= 0$ and maximum $= 8$). The SVM metric was originally developed from this task. It is intended that the average performance for an SVM $= 1$ light source should be a detection rate of 0.50, which would mean a score of 4 in this experiment.

There were no statistically significant differences between the two locations in rotating disc detection score: the outcomes for CSTB and NRC were the same for each SVM level (Table 4) and for the overall average (Mann–Whitney U test, $Z = -0.39$, $p = 0.69$). Therefore, we present the combined data for the full sample of 85 participants graphically in Figure 4. Looking to the Wilcoxon tests between the SVM levels (Table 4), we see that the distributions of rotating disc stroboscopic detection scores for SVM ~ 0 and SVM ~ 0.4 did not differ. The rotating disc detection scores did increase for each successive SVM level, with results that reached the criterion level of statistical significance. The pattern of Wilcoxon test results was similar when the sample was split into two groups by location.

Similarly, there were no statistically significant differences in the Mann–Whitney test results for comparisons between the low and high PGS groups for any SVM level, nor for the overall average (Mann–Whitney U test, $Z = -0.01$, $p = 1.00$). When the sample was split into these groups, the pattern of results for the Wilcoxon comparisons between the different SVM levels was the same for both groups (Table 4). PGS, an indicator of sensitivity to ill-effects of visual stress, did not predict performance on this stroboscopic visibility task in this experiment.

3.2 Stroboscopic effect – Metronome

The data reported here are for the detection of the stroboscopic effect on a metronome (vertical task) moving at 208 bpm, replicating Bullough and Marcus.[29] This rate was adopted partway through the study, so there are fewer participants. As for the

Table 4 Results of the stroboscopic detection scores for the rotating disc (N = 85).

Condition	Descriptive statistics					Wilcoxon signed ranks test			
SVM	1 ~0	2 ~0.4	3 ~0.9	4 ~1.4	5 ~3.0	1 vs. 2 Z (p)	2 vs. 3 Z (p)	3 vs. 4 Z (p)	4 vs. 5 Z (p)
Both sites (N = 85)									
Means	0.25	0.34	1.26	4.12	7.92	−1.23	**−4.72**	**−7.03**	**−7.09**
(StDev)	(0.60)	(0.88)	(2.11)	(3.05)	(0.54)	(0.22)	**(0.000)**	**(0.000)**	**(0.000)**
Minimum	0.00	0.00	0.00	0.00	4.00				
25th percentile	0.00	0.00	0.00	1.00	8.00				
50th percentile	0.00	0.00	0.00	4.00	8.00				
75th percentile	0.00	0.00	2.00	7.00	8.00				
Maximum	3.00	4.00	8.00	8.00	8.00				
CSTB (N = 27)									
Means	0.22	0.26	1.48	4.30	8.00	−0.45	**−3.08**	**−3.83**	**−4.04**
(StDev)	(0.51)	(0.71)	(2.36)	(3.20)	(0.00)	(0.66)	**(0.002)**	**(0.000)**	**(0.000)**
Minimum	0.00	0.00	0.00	0.00	8.00				
25th percentile	0.00	0.00	0.00	0.00	8.00				
50th percentile	0.00	0.00	0.00	5.00	8.00				
75th percentile	0.00	0.00	2.00	7.00	8.00				
Maximum	2.00	3.00	8.00	8.00	8.00				
NRC (N = 58)									
Means	0.26	0.38	1.16	4.03	7.88	−1.11	**−3.59**	**−5.92**	**−5.86**
(StDev)	(0.64)	(0.95)	(1.99)	(3.01)	(0.65)	(0.27)	**(0.000)**	**(0.000)**	**(0.000)**
Minimum	0.00	0.00	0.00	0.00	4.00				
25th percentile	0.00	0.00	0.00	1.00	8.00				
50th percentile	0.00	0.00	0.00	4.00	8.00				
75th percentile	0.00	0.00	2.00	7.00	8.00				
Maximum	3.00	4.00	8.00	8.00	8.00				
CSTB vs. NRC Z	−0.06	−0.48	−0.52	−0.27	−0.97				
p	0.95	0.63	0.60	0.79	0.33				
Low PGS (N = 59)									
Means	0.19	0.32	1.10	4.20	7.93	−1.57	**−3.91**	**−5.80**	**−5.80**
(StDev)	(.43)	(0.88)	(1.90)	(3.14)	(0.52)	(0.12)	**(0.000)**	**(0.000)**	**(0.000)**
Minimum	0.00	0.00	0.00	0.00	4.00				
25th percentile	0.00	0.00	0.00	1.00	8.00				
50th percentile	0.00	0.00	0.00	5.00	8.00				
75th percentile	0.00	0.00	2.00	8.00	8.00				
Maximum	2.00	4.00	8.00	8.00	8.00				
High PGS (N = 26)									
Means (StDev)	0.38	0.38	1.62	3.92	7.88	−0.72	**−2.70**	**−4.06**	**−4.12**
	(0.85)	(0.90)	(2.52)	(2.88)	(0.59)	(0.94)	**(0.000)**	**(0.000)**	**(0.000)**
Minimum	0.00	0.00	0.00	0.00	5.00				
25th percentile	0.00	0.00	0.00	2.00	8.00				
50th percentile	0.00	0.00	0.00	3.50	8.00				
75th percentile	0.00	0.00	2.25	7.00	8.00				
Maximum	3.00	3.00	8.00	8.00	8.00				
Low vs. high PGS Z	−0.51	−0.29	−0.68	−0.36	−0.58				
p	0.61	0.77	0.50	0.72	0.56				

Note: For all tests, the criterion for rejecting the null hypothesis of no effect (statistical significance, α) was $p < 0.05$. Statistically significant test results are shown in bold text. Splitting the sample by testing location, or by PGS, did not change the results.
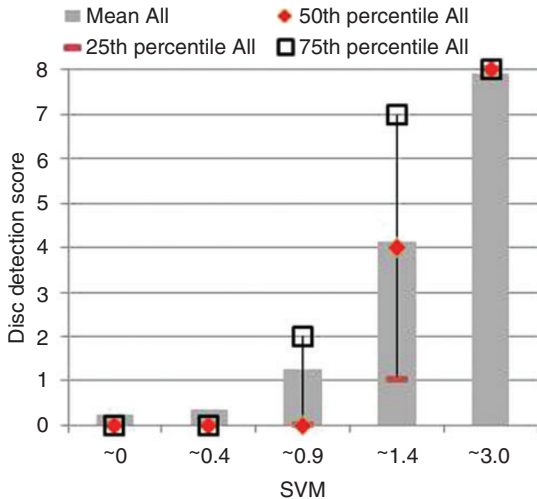PGS: pattern glare sensitivity.

Figure 4 This chart shows the detection score for the rotating disc for all participants at both sites at each SVM level, displaying the means in bars and the 25th, 50th, and 75th percentiles overlaid.

horizontal disc detection task, we used only the responses from the last eight trials, averaging the responses and scaling them to a maximum of eight. One person from CSTB missed several trials of the metronome question and was not included for this variable. The total sample size for this outcome was N = 50.

Table 5 shows the results for the full sample and for the split groups by country and PGS. The two sites showed no differences for the metronome detection scores at each level of the SVM, and for the overall average (Mann–Whitney U test, $Z = -0.13$, $p = 0.89$). There also were no differences between the high and low PGS groups on the individual SVM scores, nor on the overall average metronome detection score (Mann–Whitney U test, $Z = -1.04$, $p = 0.30$).

The Wilcoxon Signed Ranks tests for the full sample showed that the stroboscopic effect detection scores for the three lower SVM values did not differ from one another. However, the stroboscopic effect became

more visible at SVM ~ 1.4 as compared to ~ 0.90, and still more visible for ~ 3.0 as compared to ~ 1.4. There were some differences in the subgroup analysis: At NRC, there was no difference between metronome detection scores at 1.4 and 3.0. For the high-PGS group, there were no differences in metronome detection scores between the levels. Figure 5 displays the descriptive statistics for the full sample as a bar chart.

3.3 Acceptability ratings
An internal reliability analysis of the ratings for 'comfort', 'pleasantness' and 'annoyingness' showed that the annoyingness rating did not correlate well to the other items. Therefore, we formed one scale for 'acceptability' by averaging the ratings for comfort and pleasantness (higher scores meaning better conditions, theoretical range 0 [not at all] to 4 [extremely]). The Cronbach's alpha indicator of internal consistency reliability for this two-item scale was 0.81, which by a common rule of thumb is considered very good. We analysed the annoyingness rating separately as a single rating (see Section 3.4).

Table 6 displays the results for the average acceptability ratings and the comparisons between the SVM conditions and groups. The mean and median overall acceptability for each SVM condition were close to the midpoint of the scale (see Figure 6), and did not differ from one SVM condition to the other. The overall average acceptability (across all SVM conditions) was not different between CSTB and NRC (Mann–Whitney U test, $Z = -1.06$, $p = 0.11$). Looking at the comparisons between locations, we see statistically significant differences between CSTB and NRC for the acceptability ratings of the conditions SVM ~ 0.9 and SVM ~ 1.4. In both cases, the CSTB conditions were rated as slightly more acceptable; however, when considered as a difference in median between 2.0 and 1.5 on a scale of 0–4, the difference is

Table 5 Results of the stroboscopic detection scores for the metronome beating at 208 bpm at both locations (N = 50).

Condition SVM	Descriptive statistics					Wilcoxon signed ranks test			
	1 ~0	2 ~0.4	3 ~0.9	4 ~1.4	5 ~3.0	1 vs. 2 Z (p)	2 vs. 3 Z (p)	3 vs. 4 Z (p)	4 vs. 5 Z (p)
Both sites (N = 50)									
Means	3.12	2.82	2.70	3.34	4.12	−1.66	−0.64	**−2.93**	**−2.28**
(StDev)	(3.10)	(3.17)	(2.78)	(2.95)	(3.15)	(0.10)	(0.53)	**(0.003)**	**(0.02)**
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.75				
50th percentile	2.00	1.00	2.00	3.00	5.00				
75th percentile	6.25	6.25	5.00	6.00	7.00				
Maximum	8.00	8.00	8.00	8.00	8.00				
CSTB (N = 10)									
Means	2.50	2.70	2.50	3.10	5.50	−1.00	−0.11	**−2.45**	**−2.21**
(StDev)	(3.44)	(3.47)	(3.17)	(3.00)	(2.99)	(0.32)	(0.91)	**(0.01)**	**(0.03)**
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.75	2.00				
50th percentile	0.50	1.00	1.50	2.50	7.00				
75th percentile	6.50	7.25	5.00	5.75	8.00				
Maximum	8.00	8.00	8.00	8.00	8.00				
NRC (N=40)									
Means	3.28	2.85	2.75	3.40	3.78	−1.87	−0.64	**−2.48**	−0.90
(StDev)	(3.04)	(3.13)	(2.72)	(2.98)	(3.13)	(0.06)	(0.53)	**(0.01)**	(0.37)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.00				
50th percentile	2.50	1.50	2.00	3.00	3.50				
75th percentile	6.75	6.00	5.00	6.00	7.00				
Maximum	8.00	8.00	8.00	8.00	8.00				
CSTB vs. NRC Z	−0.74	−0.19	−0.46	−0.14	−1.60				
P	0.46	0.85	0.65	0.89	0.11				
Low PGS (N=38)									
Means	2.89	2.58	2.58	3.18	3.97	−1.80	−0.020	**−2.60**	**−2.12**
(StDev)	(3.18)	(3.17)	(2.88)	(3.09)	(3.34)	(0.07)	(0.98)	**(0.01)**	**(0.03)**
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.00				
50th percentile	2.00	1.00	1.50	2.00	4.50				
75th percentile	7.00	6.25	5.00	6.00	7.25				
Maximum	8.00	8.00	8.00	8.00	8.00				
High PGS (N= 12)									
Means	3.83	3.58	3.08	3.83	4.58	−0.65	−1.14	−1.56	−.085
(StDev)	(2.86)	(3.18)	(2.50)	(2.52)	(2.50)	(0.52)	(0.25)	(0.12)	(0.40)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.50	0.25	1.00	2.25	2.25				
50th percentile	4.00	3.50	3.00	3.50	5.50				
75th percentile	6.00	6.75	5.00	5.75	6.75				
Maximum	8.00	8.00	8.00	8.00	8.00				
Low vs. high PGS Z	−0.89	−1.08	−0.92	−0.78	−0.40				
P	0.37	0.28	0.36	0.43	0.69				

Note: For all tests, the criterion for rejecting the null hypothesis of no effect (statistical significance, α) was $p < 0.05$. PGS: pattern glare sensitivity.
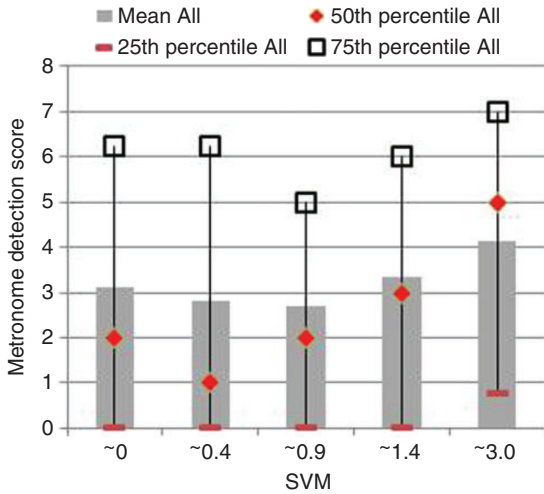
Figure 5 This chart shows the stroboscopic effect detection score for the metronome for all participants at both sites at each SVM level, displaying the means in bars and the 25th, 50th, and 75th percentiles overlaid.

small. There also was one difference between sites in the SVM comparisons; the NRC site showed a drop in acceptability between SVM ~0.40 and SVM ~0.90. This drop is consistent with predictions, but there was no consistent effect across the SVM conditions (higher SVM conditions were not any less acceptable).

There were no differences between the PGS groups in the acceptability of the conditions, shown in the individual SVM tests (Table 6) nor in the overall average scores by PGS groups (Mann–Whitney U test, $Z = -0.62$, $p = 0.54$). Similarly, the two groups showed the same pattern of comparisons across the SVM conditions (Wilcoxon tests).

3.4 Annoyingness ratings

Participants rated the annoyingness of each light source once, on the 10th block of trials. This was a single question rated on a scale from 0 (not at all) to 4 (extremely). Table 7 displays the results for this variable both for the overall sample, and for the subgroups formed by the location (CSTB or NRC) and by the PGS score.

There were differences in responding between CSTB and NRC on the ratings for each SVM level, and in their overall average annoyingness ratings (Mann–Whitney U test, $Z = -4.85$, $p = 0.000$). The CSTB participants rated their light sources as consistently less annoying than the NRC participants. Note that the CSTB sessions took place in French and the NRC sessions took place in English. We cannot rule out the possibility that the difference in annoyingness ratings reflects a translation effect; perhaps the words used in the two languages had subtly different meanings. It is also possible that the difference reflects the fact that the dominant frequency of the TLM was 100 Hz for CSTB and 120 Hz for NRC; however, this seems unlikely to explain the difference for two reasons. First, there was no difference in perception between the two countries, as one might expect if the dominant frequency influenced perception directly.[29] Second, the direction of this effect is opposite to what would be expected; the lower frequency should be more easily detected and, by extension, possibly more annoying. Overall, the language difference seems the most likely explanation.

Looking at the comparisons between SVM levels within the two sites, we see that none of the comparisons showed any statistically significant differences in the CSTB data. Within the NRC data, there was an increase in annoyingness between SVM ~0.40 and SVM ~0.90. The higher level of annoyingness persisted for the NRC participants for SVM ~1.4 and SVM ~ 3.0.

The split of the data set based on PGS scores also showed differences between the groups in the annoyingness ratings, and these were as one would expect, with annoyingness ratings being higher for the high-PGS participants: There was an overall

Table 6 Results of the acceptability ratings of the light sources, both the descriptive statistics and the comparisons.

Condition SVM	Descriptive statistics					Wilcoxon signed ranks test			
	1 ~0	2 ~0.4	3 ~0.9	4 ~1.4	5 ~3.0	1 vs. 2 Z (p)	2 vs. 3 Z (p)	3 vs. 4 Z (p)	4 vs. 5 Z (p)
Both sites (N=85)									
Means	1.90	1.87	1.79	1.85	1.97	−0.40	−0.68	−0.45	−1.15
(StDev)	(0.96)	(0.98)	(0.93)	(1.02)	(0.96)	(0.69)	(0.49)	(0.65)	(0.25)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	1.50	1.00	1.00	1.00	1.50				
50th percentile	2.00	2.00	2.00	1.50	2.00				
75th percentile	2.50	2.50	2.50	2.75	2.50				
Maximum	4.00	4.00	4.00	4.00	4.00				
CSTB (N=27)									
Means	1.78	1.91	2.13	2.22	2.19	−1.08	−1.21	−0.51	−0.08
(StDev)	(0.74)	(0.88)	(0.83)	(0.80)	(0.88)	(0.28)	(0.23)	(0.61)	(0.94)
Minimum	0.50	0.00	0.50	0.50	0.00				
25th percentile	1.00	1.00	1.50	1.50	1.50				
50th percentile	2.00	2.00	2.00	2.00	2.50				
75th percentile	2.00	2.50	3.00	3.00	3.00				
Maximum	3.00	3.50	4.00	3.50	4.00				
NRC (N=58)									
Means	1.96	1.85	1.63	1.67	1.87	−1.34	**−2.01**	−0.24	−1.40
(StDev)	(1.05)	(1.03)	(0.94)	(1.07)	(0.99)	(0.18)	**(0.05)**	(0.81)	(0.16)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	1.50	1.38	1.00	1.00	1.00				
50th percentile	2.00	1.75	1.50	1.50	1.75				
75th percentile	2.63	2.50	2.13	2.50	2.50				
Maximum	4.00	4.00	3.50	4.00	4.00				
CSTB vs. NRC Z	−0.75	−0.21	**−2.42**	**−2.67**	−1.58				
P	0.46	0.84	**0.02**	**0.01**	0.11				
Low PGS (N=59)									
Means	1.81	1.83	1.70	1.87	2.06	−0.22	−1.01	−1.16	−1.81
(StDev)	(0.99)	(1.04)	(1.01)	(1.13)	(0.91)	(0.83)	(0.31)	(0.24)	(0.07)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	1.00	1.00	1.00	1.00	1.50				
50th percentile	1.50	1.50	1.50	2.00	2.00				
75th percentile	2.50	3.00	2.50	3.00	2.50				
Maximum	4.00	4.00	4.00	4.00	4.00				
High PGS (N=26)									
Means	2.12	1.96	1.98	1.79	1.77	−1.00	−0.49	−0.93	−0.58
(StDev)	(0.88)	(0.85)	(0.70)	(0.74)	(1.07)	(0.32)	(0.49)	(0.35)	(0.56)
Minimum	0.50	0.00	0.50	0.50	0.00				
25th percentile	1.50	1.50	1.50	1.00	1.00				
50th percentile	2.00	2.00	2.00	1.50	1.50				
75th percentile	3.00	2.50	2.50	2.50	2.50				
Maximum	4.00	3.50	3.50	3.00	4.00				
Low vs. high PGS Z	−1.54	−0.73	−1.36	−0.40	−1.42				
P	0.12	0.46	0.17	0.69	0.16				

Note: For all tests, the criterion for rejecting the null hypothesis of no effect (statistical significance, α) was $p < 0.05$. PGS: pattern glare sensitivity.
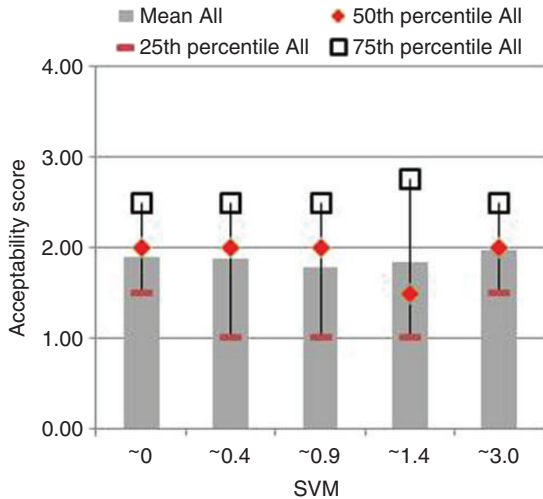
Figure 6 The acceptability of the light conditions in these short exposures did not differ by SVM condition.

difference in average annoyingness between PGS groups (Mann–Whitney U test, $Z = -2.12$, $p = 0.03$; low-PGS median $= 0.80$ and high-PGS median $= 1.5$). For SVM ~1.4 and SVM ~3.0, annoyingness was higher for the people higher in PGS. The high-PGS group showed no statistically significant differences in the comparisons between the SVM conditions, but the low-PGS group showed a small increase for SVM ~0.90 in comparison to SVM ~0.40.

Figure 7 displays the overall result graphically, for the full sample. Overall, the annoyingness ratings were low, but there was a small increase from SVM ~0.40 to SVM ~0.90.

3.5 Inter-relationships

As a final step to understand and to validate the pattern of responses, we explored the correlations between six variables: pattern 2 sensation scores; discomfort ratings of pattern 2; average stroboscopic effect detection of the rotating disc; average stroboscopic effect detection of the metronome (high speed only); average acceptability rating; and average annoyingness rating. These are reported in Table 8.

People with higher scores on PGS pattern 2 also tended to experience greater discomfort while looking at that pattern ($r = 0.31$, $p = 0.004$, N $= 85$). Given that the viewing time for each pattern was short, this relatively low correlation is to be expected because of restricted variability in discomfort among people who were less sensitive. People who tended to see the stroboscopic effect for the rotating disc also tended to see it for the metronome ($r = 0.40$, $p = 0.004$, N $= 50$). Average acceptability was negatively correlated with average annoyingness ($r = -0.22$, $p = 0.04$, N $= 85$) and with the discomfort experienced in response to pattern 2 ($r = -0.26$, $p = 0.02$, N $= 85$). Given the short exposure times to each condition when the ratings were made, the correlations are best interpreted as providing guidance for new hypotheses to be developed. The directions of these correlations are as expected, which adds evidence for the validity of these outcome measures and suggests that they could be useful in studies with longer exposure times, from which guidance about the effects of SVM in application could be derived.

4. Summary and conclusions

In the preparations for this experiment, both laboratories acquired a variety of commercially available LED replacement lamps that are available on the North American and European markets in 2018. In the laboratory, the lamps were measured under clean sinusoidal power supply conditions and found to exhibit a wide range of TLM characteristics, from nearly none to very high SVM. Each laboratory selected five lamps for this experiment, based on their SVM characteristics, taking care to seek similar performance at each chosen level to permit the data to be combined. The primary visual perception task

Table 7 Results of the annoyingness ratings of the light sources, both the descriptive statistics and the comparisons.

Condition SVM	Descriptive statistics					Wilcoxon signed ranks test			
	1 $\sim$0	2 $\sim$0.4	3 $\sim$0.9	4 $\sim$1.4	5 $\sim$3.0	1 vs. 2 Z (p)	2 vs. 3 Z (p)	3 vs. 4 Z (p)	4 vs. 5 Z (p)
Both sites (N=85)									
Means	1.11	1.06	1.27	1.25	1.31	−0.55	**−2.40**	−0.29	−0.55
(StDev)	(1.01)	(0.94)	(1.07)	(1.11)	(1.22)	(0.58)	**(0.02)**	(0.77)	(0.59)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.00				
50th percentile	1.00	1.00	1.00	1.00	1.00				
75th percentile	2.00	2.00	2.00	2.00	2.00				
Maximum	3.00	3.00	4.00	4.00	4.00				
CSTB (N=27)									
Means	0.52	0.41	0.59	0.59	0.59	−0.80	−1.39	−0.04	0.00
(StDev)	(0.80)	(0.50)	(0.80)	(0.89)	(0.89)	(0.43)	(0.17)	(0.97)	(1.00)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.00				
50th percentile	0.00	0.00	0.00	0.00	0.00				
75th percentile	1.00	1.00	1.00	1.00	1.00				
Maximum	3.00	1.00	2.00	3.00	3.00				
NRC (N=58)									
Means (StDev)	1.38	1.36	1.59	1.55	1.64	−0.11	**−1.96**	−0.31	−0.68
	(0.99)	(0.95)	(1.04)	(1.08)	(1.64)	(0.91)	**(0.05)**	(0.76)	(0.50)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	1.00	1.00	1.00	1.00	1.00				
50th percentile	1.00	1.00	2.00	1.00	2.00				
75th percentile	2.00	2.00	2.00	2.25	3.00				
Maximum	3.00	3.00	4.00	4.00	4.00				
CSTB vs. NRC Z	**−3.83**	**−4.46**	**−4.02**	**−3.87**	**−3.78**				
P	**0.000**	**0.000**	**0.000**	**0.000**	**0.000**				
Low PGS (N=59)									
Means	1.02	1.00	1.24	1.07	1.08	−0.16	**−2.45**	−1.34	−0.12
(StDev)	(1.03)	(0.91)	(1.10)	(1.14)	(1.13)	(0.87)	**(0.01)**	(0.18)	(0.90)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.00	0.00	0.00	0.00	0.00				
50th percentile	1.00	1.00	1.00	1.00	1.00				
75th percentile	2.00	2.00	2.00	2.00	2.00				
Maximum	3.00	3.00	4.00	4.00	4.00				
High PGS (N=26)									
Means	1.31	1.19	1.35	1.65	1.81	−0.69	−0.89	−1.33	−0.68
(StDev)	(0.97)	(1.02)	(1.02)	(0.94)	(1.27)	(0.49)	(0.37)	(0.18)	(0.50)
Minimum	0.00	0.00	0.00	0.00	0.00				
25th percentile	0.75	0.00	0.00	1.00	1.00				
50th percentile	1.00	1.00	1.50	2.00	2.00				
75th percentile	2.00	2.00	2.00	2.00	3.00				
Maximum	3.00	3.00	3.00	3.00	4.00				
Low vs. high PGS Z	−1.37	−0.74	−0.56	**−2.53**	**−2.47**				
P	0.17	0.46	0.58	**0.01**	**0.01**				

Note: For all tests, the criterion for rejecting the null hypothesis of no effect (statistical significance, α) was $p<0.05$. PGS: pattern glare sensitivity; SVM: stroboscopic visibility measure.

in this experiment was carefully chosen to replicate the rotating disc task with which the SVM was developed.[15,20,27] The metric is defined such that the average person ought to detect the stroboscopic effect 50% of the time if SVM = 1.

The results of this experiment showed that, under these experimental conditions, the average rotating disc detection was lower than expected. For our SVM~0.9 condition, the average score was 1.26 (out of 8) and only 25% of the people scored above 2 out of 8. The definition of the SVM would lead to the

Figure 7 The annoyingness of the light conditions in these short exposures showed a small increase between SVM ~0.40 and SVM −0.90 in the full sample.

expectation that the average should be closer to 3, or perhaps close to 4, when the SVM approaches 1.0. There were two individuals who scored 8 when the SVM was ~0.90. The 90th percentile was 5.4 at the SVM ~0.90, indicating that 10% of the population (eight people) scored higher than this. The rotating disc stroboscopic detection performance is strongly non-linear with increasing SVM (see Figure 4), and that it appears that even small increases in the SVM above 0.9 can dramatically increase the disc detection score. For SVM ~1.4, half of the sample scored more than the threshold detection (4 out of 8) and 25% of the sample scored 7 or 8.

The metronome task replicated Bullough and Marcus,[29] who examined a metronome moving at 208 bpm. They reported that a logarithmic function of the SVM values of their range of experimental conditions explained 79% of the variance in average metronome detection at 100 Hz, which compares very well to the 78% explained variance for the data reported here (excluding our SVM = 0 condition, for which a log function cannot be fit). This consistency gives us greater confidence in the interpretation of the metronome detection results. Most notably, the increase in metronome detection scores between SVM ~0.9 and SVM ~1.4, and between SVM ~1.4 and SVM ~3.0 were statistically significant. Note that the metronome task was based on a negative contrast, whereas the disc detection task had a positive
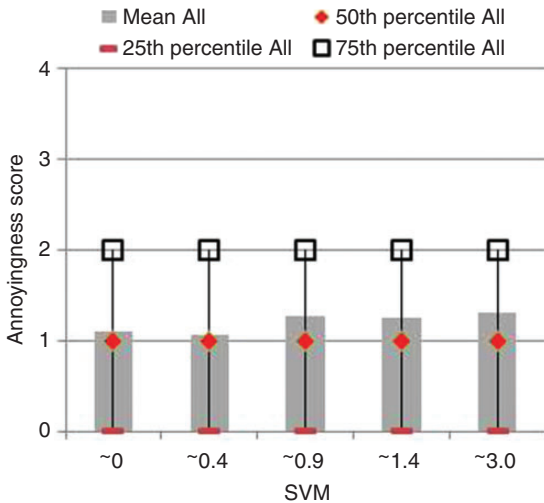
Table 8 Pearson correlation coefficients between the principal variables.

	PGS 2 discomfort rating	PGS 2 Score	Disc detection score	Metronome detection score	Acceptability
PGS 2 Score	0.31**				
Disc detection score	0.06	−0.04			
Metronome detection score	0.16	0.08	0.40**		
Acceptability	−0.26*	0.04	−0.05	−0.19	
Annoyance	0.09	0.20	0.13	0.23	−0.22*

Note: Correlations involving the metronome detection score have N = 50; all others have N = 85. **$p < 0.01$. *$p < 0.05$. PGC: pattern glare sensitivity.

contrast; contrast polarity is another variable that might influence stroboscopic visibility and could be worthy of further investigation.

The judgements of acceptability (pleasantness and comfort average) did not show any consistent relation to the SVM, but over the whole sample, annoyingness did increase from SVM ~0.4 to SVM ~0.9 and stayed at that level for higher levels of the SVM. Annoyingness was higher for the NRC participants than the CSTB participants, a difference that could have been caused unintentionally in the translation from the English used at the NRC to the French used at the CSTB.

Visual stress is a syndrome characterised by headache and the occurrence of unwanted visual illusions.[21] We had hypothesised that stroboscopic effect detection would be higher for people with greater risk of visual stress, which are those with high sensation scores when looking at pattern 2 of the Wilkins and Evans[23] Pattern Glare Sensitivity test. However, there were no differences in stroboscopic detection scores for either the horizontal rotating disc or the vertical metronome. The acceptability ratings also did not differ between the groups based on this indicator of sensitivity; however, the more sensitive group (high-PGS) did report greater annoyance with the SVM ~1.4 and SVM ~3.0 conditions than the less sensitive (low-PGS) group. Note that these ratings occurred following a very short exposure to the conditions. Longer exposures could be expected to lead to different outcomes.

Future research will be necessary in order to address the inevitable limitations of this experiment. The limitations, and suggestions for studies that could address each, include:

- A greater number of female than male participants overall – future studies should aim for a better balance between the sexes.
- Only young participants – other research will need to extend the work to older people, who might be less sensitive, but who might also find conditions less acceptable and more annoying.
- Similarly, a limited range of eye colour and ethnicity in the sample might have excluded some sensitive individuals – replication in a broad sample of varied ethnicity and eye colour is needed for the best population characterisation.
- Short viewing times – effects that develop over a longer time course than visual perception (e.g. health effects) will require longer exposures.
- Only an overhead view of the horizontal task and a straight-on view of the vertical task – other geometries of moving objects should be investigated.
- Only five TLM levels, leaving gaps where information is lacking – future studies could include conditions between the SVM levels studied here, or could characterise the light sources using other indicators of TLM (e.g. frequency; modulation depth; flicker index; etc.).
- Non-immersive surroundings (which had been the setting for prior research[15]) – future investigations could vary the TLM of ambient room lighting, perhaps also including multiple luminaires in the field of view; and,
- Only one visual perception outcome investigated, the stroboscopic effect – thus, these data cannot inform concerning possible effects of TLM (whether characterised by the SVM or any other metric) on detection of the phantom array, nor on complex phenomena like eyestrain, headache, reading or cognitive performance, which are outcomes that have also been identified as being of interest for the writers of standards and regulations related to TLM.[10]

This report provides the first set of population data concerning stroboscopic detection in response to commercially available lamps, characterised specifically by the SVM, and including consideration of the possible sensitivity of the observers. This experiment was

conducted to contribute data for discussions concerning limit values for TLM where, at present, very limited evidence exists.[16] There remains discussion concerning which metrics would be the best predictors of the various effects of TLM,[10,11] questions which this experiment was not designed to address. Regardless of which metric or combination of metrics are chosen based on the evidence, the decision to place a limit on any metric involves two choices: (1) The acceptable frequency of the outcome occurring; and (2) the acceptable proportion of the population who might experience this outcome.[10] These choices are value judgements that research can inform but cannot determine.

Acknowledgements

Declaration of conflicting interests

The authors declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

ORCID iD

JA Veitch   https://orcid.org/0000-0003-3183-4537

References

1 Wilkins AJ, Veitch JA, Lehman B. *LED lighting flicker and potential health concerns*: *IEEE Standard PAR1789 update. Proceedings of the Energy Conversion Congress and Exposition (ECCE)*, New York, NY, 12–16 September, 2010: pp. 171–178.
2 IEEE Power Electronics Society. *IEEE Recommended Practices for Modulating Current in High-Brightness LEDs for Mitigating Health Risks to Viewers*. New York, NY: Institute for Electrical and Electronics Engineers, 2015: pp. S1789–2015.
3 Wilkins AJ. A physiological basis for visual discomfort: application in lighting design. *Lighting Research and Technology* 2016; 48: 44–54.
4 Wilkins AJ. Intermittent illumination from visual display units and fluorescent lighting affects movement of the eyes across text. *Human Factors* 1986; 28: 75–81.
5 Veitch JA, McColl SL. Modulation of fluorescent light: flicker rate and light source effects on visual performance and visual comfort. *Lighting Research and Technology* 1995; 27: 243–256.

6   Veitch JA, Newsham GR. Lighting quality and energy-efficiency effects on task performance, mood, health, satisfaction and comfort. *Journal of the Illuminating Engineering Society* 1998; 27(1): 107–129.

7   Wilkins AJ, Nimmo-Smith I, Slater AI, Bedocs L. Fluorescent lighting, headaches and eyestrain. *Lighting Research and Technology* 1989; 21: 11–18.

8   Behar-Cohen F, Carré S, Enouf O, Falcon J, Gronfier C, Hicks D, et al. *Effets sur la santé humaine et sur l'environnement (faune et flore) des diodes électroluminescentes (LED)*. [2014-SA-0253]. Maisons-Alfort, France: Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES), 2019.

9   Genet R. Opinion of the French Agency for Food, Environmental and Occupational Health and Safety on ''The effects on human health and the environment (fauna and flora) of systems usng light-emittng diodes (LEDs)''. ANSES Opinion Request No 2014-SA-0253 Maisons-Alfort, France: Agence Nationale de Sécurité Sanitaire de l'Alimentation, de l'Environnement et du Travail (ANSES), 2019.

10   Commission Internationale de l'Eclairage. *Final Report CIE Stakeholder Workshop for Temporal Light Modulation Standards for Lighting Systems*. [CIE TN 008:2017]. Vienna, Austria: CIE, 2017.

11   Commission Internationale de l'Eclairage. *Visual Aspects of Time-modulated Lighting Systems – Definitions and Measurement Models*. [TN 006-2016]. Vienna, Austria: CIE, 2016.

12   DiLaura DL, Houser KW, Mistrick RG, Steffy GR. *The Lighting Handbook*. 10th ed. New York: Illuminating Engineering Society of North America, 2011.

13   International Electrotechnical Commission. *Equipment for General Lighting Purposes – EMC Immunity Requirements – Part 1: An Objective Light Flickermeter and Voltage Fluctuation Immunity Test Method*. [IEC TR 61547-1:2017]. Geneva, Switzerland: IEC, 2017.

14   Alliance for Solid-State Illumination Systems and Technologies (ASSIST). *ASSIST recommends . . . Evaluating light source flicker for stroboscopic effects and general acceptability*. Vol. 11. Troy, NY, USA: Lighting Research Center, Rensselaer Polytechnic Institute, 2017.

15   Perz M, Vogels IMLC, Sekulovski D, Wang LL, Tu Y, Heynderickx IEJ. Modeling the visibility of the stroboscopic effect occurring in temporally modulated light systems. *Lighting Research and Technology* 2015; 47: 281–300.

16   National Electrical Manufacturers Association Lighting Systems Division. *Standard for Temporal Light Artifacts: Test Methods and Guidance for Acceptance Criteria*. [NEMA 77-2017]. Rosslyn, VA: NEMA, 2017.

17   Bullough JD, Sweater Hickcox K, Klein TR, Narendran N. Effects of flicker characteristics from solid-state lighting on detection, acceptability and comfort. *Lighting Research and Technology* 2011; 43: 337–348.

18   Bullough JD, Sweater Hickcox K, Klein TR, Lok A, Narendran N. Detection and acceptability of stroboscopic effects from flicker. *Lighting Research and Technology* 2012; 44: 477–483.

19   Perz M, Sekulovski D, Beeckman P. *Acceptability criteria for the stroboscopic effect visiblity measure*: Proceedings of the Conference ''Smarter Lighting for Better Life'' at the CIE Midterm Meeting, Jeju, South Korea, 23–25 October, 2017: pp. 453–459.

20   Vogels IM, Sekulovski D, Perz M. *Visible artefacts of LEDs*: Proceedings of the 27th Session of the Commission Internationale de l'Eclairage, Sun City, South Africa, 10–15 July, 2011: pp. 42–51.

21   Wilkins AJ. *Visual stress*. Oxford, UK: Oxford University Press, 1995.

22   Veitch JA, Martinsons C. *Detection of the stroboscopic effect under low levels of the stroboscopic visibility measure*: Proceedings of the 29th CIE Quadrennial Session, Washington, DC, USA, 16–22 June, 2019: pp. 682–691.

23   Wilkins AJ, Evans BJW. *I.O.O. Pattern Glare Test*. London, UK: I.O.O Sales Ltd, 2012.

24   International Electrotechnical Commission. *Equipment for General Lighting Purposes – Objective Test Method for Stroboscopic Effects of Lighting Equipment*. [IEC TR 63158:2018]. Geneva, Switzerland: IEC, 2018.

25 Brundrett GW. Human sensitivity to flicker. *Lighting Research and Technology* 1974; 6: 127–143.

26 Wilkins AJ, Allen PM, Monger LJ, Gilchrist JM. Visual stress and dyslexia for the practicing optometrist. *Optometry in Practice* 2016; 17: 103–112.

27 Perz M, Sekulovski D, Vogels IM, Heynderickx IEJ. Stroboscopic effect: contrast threshold function and dependence on illumination level. *Journal of the Optical Society of America A: Optics and Image Science, and Vision* 2018; 35: 309–319.

28 Whissell CM. Sex, eye-color, and aesthetic ratings of colors. *Perceptual and Motor Skills* 1980; 51(Pt 1): 1012–1014.

29 Bullough JD, Marcus D. Influence of flicker characteristics on stroboscopic effects. *Lighting Research and Technology* 2016; 48: 857–870.