Published paper

Fotios, S.A. and Cheal, C. (2008) *The effect of a stimulus frequency bias in side-by-side brightness ranking tests*. Lighting Research and Technology, 40 (1). 43 - 54. ISSN 1477-1535

http://dx.doi.org/10.1177/1477153507082001

The Effect of a Stimulus Frequency Bias in Side-By-Side Brightness Ranking Tests

SA Fotios CEng, MEI, MSLL, MILE, PhD, BEng(Hons), and **C Cheal** MSc, BSc(Hons)

Fotios SA & Cheal C, The Effect of a Stimulus Frequency Bias in Side-By-Side Brightness Ranking Tests, Lighting Research & Technology, 2008; 40(1); 43-54

Abstract

It was found that a stimulus frequency bias can affect forced-choice brightness ranking tests and is sufficient to affect the conclusions drawn. When a stimulus is compared against a range of comparison stimuli and judgements of relative magnitude are sought, the range of comparison stimuli should be selected to enable all possible responses to be given with approximately equal frequency to avoid a stimulus frequency bias.

1. Introduction

Brightness ranking is one of three categories of assessment technique that have been used to compare the brightness of lighting of different spectral power distribution (SPD).[1] Two stimuli are simultaneously presented and observers identify which one is the brighter – this is a forced choice, observers are not permitted to respond that the two stimuli appear equally bright. To identify the relationship between brightness and illuminance for a particular pair of lamps one stimulus is presented at several different illuminances whilst the second stimulus remains at a constant illuminance, with identification of the brighter stimulus being sought at each step. Subsequent interpolation yields the illuminance match at which the two stimuli would be noted as brighter with equal frequency, this being the illuminance ratio for equal brightness.

There is a potential source of experimental error in this work, that of stimulus frequency bias.[2,3] This refers to the frequency of occurrence of stimuli yielding a particular response: when the frequencies of the stimuli are unequal, observers tend to respond as if the frequencies were more nearly equal. This may arise from a preconception of chance, leading an observer to expect that where a large number of responses are given, each of the permitted responses will be correct on an approximately equal number of occasions.

For brightness ranking tests, stimulus frequency refers to the distribution of comparison illuminances above and below the illuminance at which equal brightness is expected. If this distribution is fairly balanced, meaning the same number of comparison illuminances either side of that which is expected to gives equal brightness, then both stimuli will be identified as being brighter on a near equal frequency. However, if this distribution is not fairly balanced, meaning an unequal number of comparison illuminances either side of that which is expected to gives equal brightness, then one stimulus will tend to be identified as brighter more frequently than the other. At the expected equal brightness presentation, stimulus frequency bias could then cause identification of the brighter stimulus to be unfairly biased to one stimulus, the stimulus which is otherwise be less frequently identified as brighter. This can then suggest a difference between two stimuli when none exists.

This article uses data from two series of brightness ranking tests to investigate the prevalence and impact of the stimulus frequency bias. This provides evidence for the design of further studies and the re-analysis of previous work.

2. Experimental Method

The two series of brightness ranking tests used the side-by-side booths as shown in Figure 1 and the light sources described in Table 1. The viewing chamber of each booth is of dimensions

575mm deep x 680mm wide x 660mm high, presenting a visual field of 38^0 wide by 37^0 high from the seated viewing distance of 1.0 metre in front of the central partition. The interior surfaces were painted matt grey (Munsell N5) and each contained identical coloured objects. The lamps were fitted behind the rear wall of the booths and thus could not be seen directly. Light was directed into the booths using an internally reflective pipe. An iris damper was installed in the pipe to permit mechanical dimming and an integrating chamber at the top of the booths ensured that changes in the type of light source and position of the iris damper did not cause significant differences in luminance distribution in the viewing chamber.

The aims of the two studies were to compare the brightness of lighting from high pressure sodium lamps (HPS) against lighting from lamps of poorer efficacy but higher colour rendering index (CRI). This is because the British Standard[4] for lighting in subsidiary streets now permits a trade-off between CRI and design illuminance, with lighting of high CRI (R_a>60) being able to adopt a reduced design illuminance. The lamps of high CRI, two types of metal halide (MH1 and MH2) and a compact fluorescent (CFL), were thus individually compared against the HPS. The CFL, MH1 and MH2 lamps are hereafter collectively referred to as the *white* lamps.

The booth illuminated by the HPS lamp was presented at one of three reference illuminances, 2.0 lux, 7.5 lux and 15.0 lux, these being the bottom, middle and top classes of the S-series for lighting in subsidiary streets.[4,5] This gave mean luminances in the range of 0.09 cd/m^2 to 0.66 cd/m^2 on the rear wall of the booth. Illuminances are horizontal illuminances measured at the centre of the floor of the booths.

The booth illuminated by one of the white lamps was set to one of several steps of illuminance as shown in Table 2. In Table 2 the illuminance at which the white light is expected[†] to yield equal brightness with the HPS is shown in bold font, and the ranges of illuminances either side of this point reveals where a stimulus frequency bias would be expected and its likely direction. For the tests with a *biased* stimulus frequency these were simply the full range of S-series illuminances, i.e. 2.0 lux, 3.0 lux, 5.0 lux, 7.5 lux, 10.0 lux and 15.0 lux.[4,5] For tests carried out with the HPS at 2.0 lux an additional comparison illuminance of 1.0 lux was used. It can be seen in Table 2 that when the HPS is presented at either 2.0 or 15.0 lux the allocation of expected brighter and dimmer comparisons is heavily biased, less so at 7.5 lux, and therefore this may unfairly influence the response given to the equal brightness presentation. For the tests with a *balanced* stimulus frequency the white lamps were presented at equal illuminance to the HPS and at four dissimilar illuminances equivalent to three S-class steps below and one step above the HPS illuminance,

[†] Equal brightness was predicted from the results of parallel work using brightness matching work.

thus offering two presentations on each side of expected equal brightness. This required extrapolation of the S-series to create further classes of illuminance at each end of the series.

Light source location was counterbalanced between the left-hand and right-hand booths. The order of presentation of the white lamps was balanced across the observers but, due to apparatus limitations, when a white lamp was introduced it remained in continuous use for tests with all three reference illuminances. The order of presentation of reference illuminances was balanced across the observers but again when a reference illuminance was set this was continued whilst the whole range of comparison illuminances were presented. Test participants were confirmed colour normal using the Ishihara[6] test and were dark adapted for 20 minutes prior to the commencement of tests.

The two series of tests (i.e. biased and balanced stimulus frequencies) were carried out independently and hence there are slight differences in experimental design. The *biased* stimulus frequency tests employed four groups of twenty one observers, these different groups being employed to compare effects of observer age and interior colourfulness. The 84 observers were in the age range 18 to 85 years old with an approximate mean age of 42 years old, and 52 of them were female. The *balanced* stimulus frequency tests employed twenty one observers, these being in the age range 18 to 54 years old with an approximate mean age of 31 years, and 14 were female.

3.1 Null condition tests: *Biased* stimulus frequency

Null-condition tests were carried out using the same method as described above but with identical HPS lamps in both booths. The same three reference illuminances were used, 2.0 lux, 7.5 lux and 15.0 lux. The second booth was presented at the full range of S-series illuminances, from 2.0 lux to 15.0 lux, with the additional 1.0 lux level used when the reference illuminance was 2.0 lux. Assuming equal brightness at equal illuminance, these stimulus frequencies are highly biased, with one stimulus being presented at the higher illuminance in the majority of cases. Eighty-four participants were used, these being the same participants who participated in the main brightness ranking tests. For half of the trials the lamps were swapped between the left-hand and right-hand booths to counterbalance unforeseen bias.

The results are shown in Table 3, this being the percentage frequency by which the test booth was reported to be brighter. The test booth is that which was illuminated by the HPS lamp at the range of several illuminances, and the reference booth is that which was illuminated by the HPS lamp to one of the three reference illuminances. The results shown are only those where the illuminances of the two booths were equal. When the booths were not presented at equal illuminance, there was a frequency of almost 100% for the booth of higher illuminance to be noted as brighter – in only three out of 420 observations was the booth of lower illuminance reported to be brighter. A frequency of 50% would result from the test booth and reference booth being noted as brighter on an equal number of occasions.

The results were analysed using Dunn-Rankin variance stable rank sums[7] as has been applied previously to analyse similar data.[8] For the equal-illuminance presentations at 7.5 lux and 15.0 lux the test booth was reported to be brighter on a significantly greater number of occasions than was the reference booth ($p<0.05$ at 7.5 lux; $p<0.01$ at 15.0 lux) despite the equality of the two stimuli. It is suspected that this is due to the stimulus frequency bias. At 2.0 lux the test booth was noted as brighter on fewer than 50% of observations, the trend expected from the stimulus frequency bias, but not significantly so.

3.2 Null condition tests: *Balanced* stimulus frequency

In the balanced stimulus frequency null-condition tests, two sets of identical lamp pairs were compared, HPS and CFL. The same three reference illuminances were used, 2.0 lux, 7.5 lux and 15.0 lux, but the comparison booth was presented at only three levels, these being equal illuminance to, and one S-class step above and below, the reference illuminance. Eighteen participants were used, aged 18-54 years old (of which only one was in the 45-54 age band, approximate mean 29 years old) and 11 were female. At each illuminance comparison, each of the participants provided four brightness assessments, counterbalancing presentation in the left-hand and right-hand booths and which lamp was nominated as the test or reference source. Hence there are 72 observations at each condition and the test was carried out by all participants under both the HPS and CFL lamps.

The results are shown in Table 3. Again, this is only the results of the equal illuminance condition since when presented at dissimilar illuminances the booth of higher illuminance was identified as brighter in 100% of the observations. The percentage frequency with which the test booth was noted as brighter is now much closer to the expected 50% than was found with the biased stimulus frequency. Analysis using Dunn-Rankin variance stable rank sums identifies no significant difference between the test and reference booths at equal illuminance.

Comparison of null-condition results from the *biased* and *balanced* stimulus frequency tests shows that stimulus frequency can significantly affect the observer's response. Therefore the results of the main brightness ranking tests were examined to determine whether the bias was present and whether it affected conclusions drawn from the results.

4. Evidence of a stimulus frequency bias

Results of the *balanced* stimulus frequency tests are shown in Table 4. These results give the percentage frequency with which the booth with the white lighting was reported to be brighter, and were analysed using Dunn-Rankin variance stable rank sums. At equal illuminance the white lamps are significantly brighter than the HPS ($p<0.001$). When the white lamps are presented at an illuminance one class of the S-series lower than the HPS the two booths are ranked equally bright, any difference is not statistically significant. When the white lamps are presented at an illuminance two classes of the S-series lower than the HPS then the HPS is found to be brighter in six of the nine cases ($p<0.01$), brighter in one case ($p<0.05$) and close to the critical value ($p\approx0.05$) in the remaining two cases.

Results of the *biased* stimulus frequency tests are shown in Table 5. At equal illuminances, booths lit by the white lamps are significantly brighter than booths lit by the HPS lamp ($p<0.001$). When the white lamps are presented at one S-series illuminance lower than the HPS, these booths are now noted as brighter than the HPS booth on a significantly greater number of observations for five of the nine cases, these being the CFL at 15.0 lux ($p<0.05$), the MH2 at 15.0 lux ($p<0.01$), and all three white lamps at 2.0 lux ($p<0.01$), but for the other four cases there is no significant difference in brightness (MH1 at 15.0 lux, all three lamps at 7.5 lux). When the white lamps are presented at two S-series illuminances below the HPS then the HPS booth is brighter for all combinations of lamp type and reference illuminance ($p<0.001$).

Statistical analysis of the results obtained when the white lighting was presented one S-series illuminance below the HPS leads to different conclusions being drawn from the *biased* and *balanced* stimulus frequency tests. A comparison of Tables 4 and 5 further reveals the stimulus frequency bias. Consider the results recorded with the HPS at 15.0 lux and the white lamps at 10.0 lux, a comparison at which approximately equal brightness was expected. With the balanced stimulus frequency, the nomination of brighter booth is allocated to both booths with almost equal frequency, but with the biased stimulus frequency the nomination of brighter booth is given more frequently to the booth lit by white lighting. Of the six matches made by an observer in the biased stimulus frequency tests at 15.0 lux, with four matches the HPS would tend to be reported as brighter but at only one match would the white lighting be reported to be brighter. Hence at the expected equal brightness presentation the allocation of brighter stimulus was biased towards the white lighting in an attempt to balance allocation of ‘brighter stimulus’ more equally between the two stimuli.

Consider also the results recorded with the HPS at 7.5 lux and the white lamps at 5.0 lux. With the balanced stimulus frequency the allocation of brighter booth is given to the white lighting by a higher percentage than it is with the biased stimulus frequency, and this is consistent for all three white lamps, although for both series of tests the difference in brightness between the white lamps and HPS lamps is not significant. Of the six matches made by an observer in the biased stimulus frequency tests at 7.5 lux, with three matches the white lighting was reported to be brighter and at two matches the HPS would be reported to be brighter. Hence, at the expected equal brightness presentation, the allocation of brighter stimulus was biased towards the HPS in an apparent attempt to balance allocation of 'brighter stimulus', and this has reduced the frequency by which the white lighting is reported to be brighter compared to the tests using a balanced stimulus frequency.

At other conditions the results from the two sets of results tend to be in good agreement. This suggests that the stimulus frequency bias is not strong enough to bias observers' responses when there is a clearly noticeable difference in brightness between the two booths, it is only when the decision is made difficult by the absence of an 'equally bright' response option that the bias has significant effect.

5. Summary

Two series of brightness ranking tests were carried out, in one of which the distribution of comparison stimuli were balanced around the expected equal brightness condition and in the other they were not – the biased stimulus frequency. Examination of null condition data demonstrates that this can affect the observers' responses: when the two stimuli have equal illuminance, and are hence expected to appear equally bright, results from the biased stimulus frequency tests show a significant bias toward one stimulus whereas with the balanced stimulus frequency the allocation of brighter booth is more equally distributed. The bias is also identifiable in the main tests, those comparing HPS lighting with lighting from the *white* lamps, and this was of sufficient impact to affect conclusions drawn from the results. Therefore the results from tests using a biased stimulus frequency must be considered unreliable. Stimulus frequency bias can be avoided in further work by ensuring that the stimulus range is equally balanced about the region of expected equal brightness. Illuminances for equal brightness can be estimated from brightness matching tests or from appropriate models – for brightness response at mesopic levels the model from Sagawa[9] has been found to give a good prediction.[10]

References

1 Fotios SA, Experimental conditions to examine the relationship between lamp colour properties and apparent brightness, Lighting Research & Technology, 2002; 34(1); 29-38

2 Senders, V.L., and Sowards, A., 1952. Analysis of response sequences in the setting of a psychophysical experiment. American Journal of Psychology. 65(3); 358-374

3 Poulton EC, Bias in quantifying judgements, Psychology Press, 1989

4 British Standards Institution (BSI) BS5489-1:2003, Code of practice for the design of road lighting —Part 1: Lighting of roads and public amenity areas, London: BSI, 2003

5 British Standards Institution (BSI) BS EN 13201-2:2003, Road lighting - Part 2: Performance requirements, London: BSI, 2003

6 Ishihara's Tests for Colour Deficiency, 24 Plates Edition, 2002. Tokyo; Kanehara Trading Inc.

7 Dunn-Rankin, P., Knezek, G.A., Wallace, S., and Zhang, S., Scaling methods. 2nd Edition, 2004; Lawrence Erlbaum Associates, Mahwah, New Jersey.

8 Quellman EM, Boyce PB. The light source color preferences of people of different skin tones. Journal of the Illuminating Engineering Society, 31(1) Winter 2002, 109-118.

9 Sagawa K, Toward a CIE supplementary system of photometry: brightness at any level including mesopic vision, Ophthalmic and Physiological Optics, 2006; 26; 240-245.

10 Fotios SA & Cheal C, Lighting for subsidiary streets: investigation of lamps of different SPD. Part 2 – Brightness, Lighting Research & Technology, 2007; 39(3); in press

Acknowledgement
This work was carried out with support from EPSRC, grant references GR/S25074/01 and GR/T09385/01. Lamps and luminaires were supplied by Philips Lighting and Urbis Lighting; side-by-side booths were procured through a CIE-UK teaching and research grant.

Lamp type		CCT (K)	CRI (R_a)
HPS	70W SON-T	2000	25
CFL	55W PL-L	3000	82
MH1	70W CDO-TT	2800	83
MH2	70W CDM-T	4200	92

Table 1

Summary of lamps used in the brightness ranking tests. CCT and CRI are as reported in manufacturers' literature.

Stimulus sequence	HPS illuminance (lux)	White lighting illuminances (lux)										
Biased	2.0				1.0	**-**	2.0	3.0	5.0	7.5	10.0	15.0
	7.5			2.0	3.0	**5.0**	7.5	10.0	15.0			
	15.0	2.0	3.0	5.0	7.5	**10.0**	15.0					
Balanced	2.0			0.6	1.0	**1.4**	2.0	3.0				
	7.5			2.0	3.0	**5.0**	7.5	10.0				
	15.0			5.0	7.5	**10.0**	15.0	22.0				

Table 2

Steps of illuminance used in brightness ranking tests. The illuminances shown in **bold** are those at which equal brightness with the HPS would be expected.

Illuminance of reference booth (lux)	**2.0**	**7.5**	**15.0**
Biased stimulus frequency (HPS)			
Percentage preference for the test booth	42%	69%	75%
Difference in votes for the test and reference booths	n.s.	$p<0.05$	$p<0.01$
Balanced stimulus frequency (HPS)			
Percentage preference for the test booth	57%	47%	56%
Difference in votes for the test and reference booths	n.s.	n.s.	n.s.
Balanced stimulus frequency (CFL)			
Percentage preference for the test booth	53%	54%	49%
Difference in votes for the test and reference booths	n.s.	n.s.	n.s.

Table 3

Results of null-condition tests: percentage frequency with which the test booth was voted to be brighter.

White lamp	HPS booth @ 2.0 lux Illuminance of white lamp (lux)				
	0.6	**1.0**	**1.4**	**2.0**	**3.0**
CFL	*0*	*0*	*48*	*95*	*100*
MH1	*0*	*10*	*67*	*100*	*100*
MH2	*0*	*19*	*52*	*95*	*100*
	HPS booth @ 7.5 lux Illuminance of white lamp (lux)				
	2.0	**3.0**	**5.0**	**7.5**	**10.0**
CFL	*0*	*5*	*71*	*100*	*100*
MH1	*0*	*0*	*38*	*100*	*100*
MH2	*0*	*0*	*57*	*95*	*100*
	HPS booth @ 15.0 lux Illuminance of white lamp (lux)				
	5.0	**7.5**	**10.0**	**15.0**	**22.0**
CFL	*0*	*14*	*52*	*100*	*100*
MH1	*0*	*5*	*52*	*100*	*100*
MH2	*0*	*5*	*57*	*95*	*100*

Table 4 Results of the brightness ranking tests using a *balanced* stimulus frequency – percentage frequency with which the booth with white lighting was reported to be brighter.

White lamp	HPS booth @ 2.0 lux						
	Illuminance of white lamp (lux)						
	1.0	**2.0**	**3.0**	**5.0**	**7.5**	**10.0**	**15.0**
CFL	*5*	*94*	*100*	*100*	*100*	*100*	*100*
MH1	*3*	*87*	*100*	*100*	*100*	*100*	*100*
MH2	*14*	*92*	*100*	*100*	*100*	*100*	*100*
	HPS booth @ 7.5 lux						
	Illuminance of white lamp (lux)						
		2.0	**3.0**	**5.0**	**7.5**	**10.0**	**15.0**
CFL		*0*	*0*	*49*	*99*	*100*	*100*
MH1		*0*	*0*	*35*	*98*	*100*	*100*
MH2		*0*	*1*	*50*	*95*	*100*	*100*
	HPS booth @ 15.0 lux						
	Illuminance of white lamp (lux)						
		2.0	**3.0**	**5.0**	**7.5**	**10.0**	**15.0**
CFL		*0*	*0*	*1*	*18*	*71*	*100*
MH1		*0*	*0*	*0*	*8*	*49*	*100*
MH2		*0*	*0*	*0*	*18*	*74*	*99*

Table 5 Results of the brightness ranking tests using a *biased* stimulus frequency – percentage frequency with which the booth with white lighting was reported to be brighter.

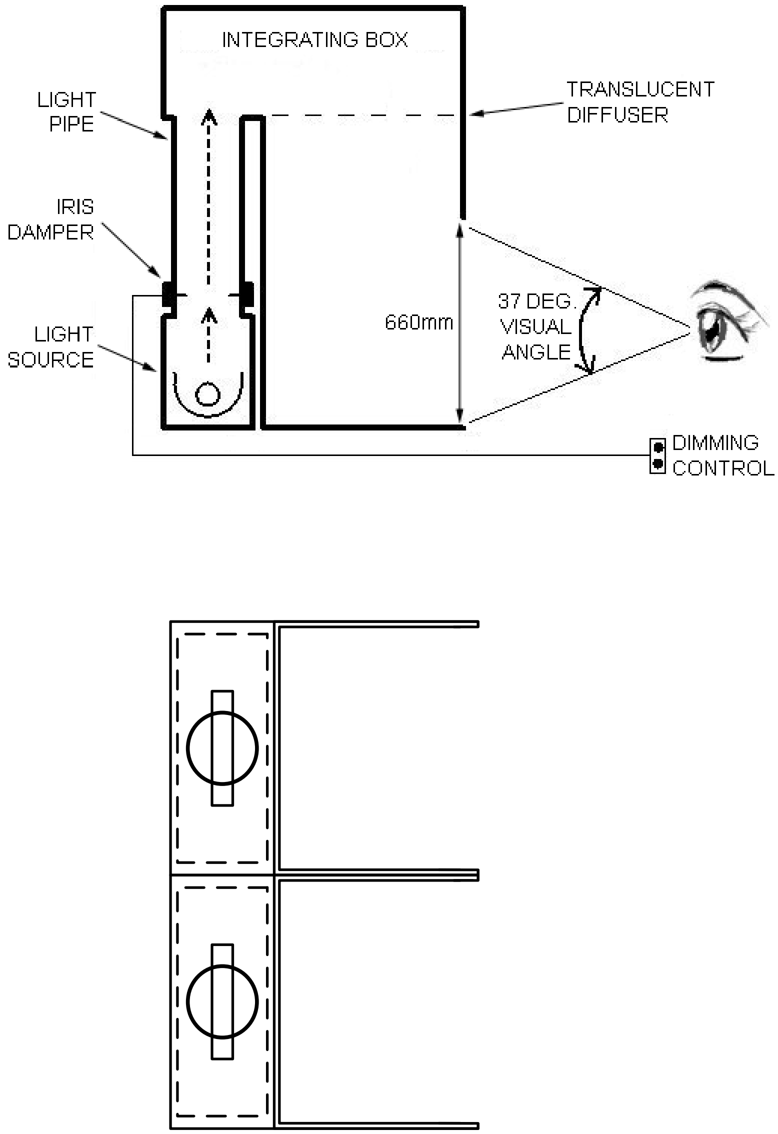

Figure 1

Vertical and horizontal sections through the side-by-side booths.