Lighting Res. Technol. 2014; **0**: 1–26

Assessing the colour quality of LED sources: Naturalness, attractiveness, colourfulness and colour difference

S Jost-Boissard PhD[a], **P Avouac MSc**[a] and **M Fontoynont** PhD[b]
[a]Laboratoire Génie Civil et Bâtiment, Ecole Nationale des Travaux Publics de l'Etat, Université de Lyon, Lyon, France
[b]Department of Energy and Environment, Danish Building Research Institute, Aalborg University, Aalborg, Denmark

Received 10 June 2014; Revised 26 August 2014; Accepted 23 September 2014

The CIE General Colour Rendering Index is currently the criterion used to describe and measure the colour-rendering properties of light sources. But over the past years, there has been increasing evidence of its limitations particularly its ability to predict the perceived colour quality of light sources and especially some LEDs. In this paper, several aspects of perceived colour quality are investigated using a side-by-side paired comparison method, and the following criteria: naturalness of fruits and vegetables, colourfulness of the Macbeth Color Checker chart, visual appreciation (attractiveness/ preference) and colour difference estimations for both visual scenes. Forty-five observers with normal colour vision evaluated nine light sources at 3000 K, and 36 observers evaluated eight light sources at 4000 K. Our results indicate that perceived colour differences are better dealt by the CIECAM02 Uniform Colour Space. Naturalness is better described by fidelity indices even if they did not give perfect predictions for all differences between LED light sources. Colourfulness is well described by gamut-based indices and attractiveness was found to correlate best with gamut-based indices but also with a preference index or a memory index calculated without blue and purple hues. A very low correlation was found between appreciation and naturalness indicating that colour quality needs more than one metric to describe subjective aspects.

1. Introduction

LEDs will soon become the major source of light due to attractive total cost of ownership, but their spectral properties are quite different from the light sources they are replacing. This leads to possible risks of deterioration as well as improvement of lighting quality. Their different spectral power distributions (SPDs) also raise questions about conventional CIE colorimetry and the calculation of the CIE General Colour Rendering Index (Ra).[1]

Thirty years ago, the CIE introduced the current method for measuring and specifying colour rendering. Since then, the Colour Rendering Index has been widely used and is currently the only internationally recognised indicator for measuring and specifying the colour-rendering properties of light sources. However, over the past few years, there has been increasing evidence of its shortcomings and several attempts have been made to update the method.[2–6] Since the emergence of LEDs,

---

Address for correspondence: Sophie Jost-Boissard, Laboratoire Génie Civil et Bâtiment, Ecole Nationale des Travaux Publics de l'Etat, Université de Lyon, Lyon, France.
E-mail: sophie.jost@entpe.fr

10.1177/1477153514555882

psychological experiments have revealed that the CIE colour-rendering metric correlates poorly with visual appreciation.[7–12] In 2006, after the publication of the CIE report on colour rendering of white LED sources[13] a Technical Committee (CIE TC1-69) was formed to investigate the new methods for assessing the colour rendition of white light sources and new metrics have been proposed.[14–18] The term 'rendition' was employed to mean the influence of light sources on colour appearance of objects. It is different from 'colour rendering' defines by CIE International Lighting Vocabulary as the 'effect of an illuminant on the colour appearance of objects by conscious or subconscious comparison with their colour appearance under a reference illuminant'. (The term rendition will also be preferred within this manuscript).

TC1-69 ended with the recommendation to consider colour fidelity and other aspects of colour quality separately. Two new TCs have been launched: TC 1-90 which evaluates available indices based on colour fidelity for assessing the colour quality of white light sources and TC 1-91 which investigates new methods for evaluating colour quality excluding fidelity.

Colour fidelity in comparison to a reference illuminant is one important aspect of colour quality but other aspects such as preference and naturalness are also of interest in everyday life. One question of interest is the number and nature of non-correlating indices required to obtain a reasonable good prediction of overall colour rendition. To this end, Guo and Houser[19] compared, computationally, several indices and recommended the use of at least two metrics, one reference-based and one area or volume-based index. These two components were validated by Rea and Freyssinier[6,20] who recommend they be used jointly to predict discrimination, vividness and naturalness for general illumination application. Dangol *et al.* and Islam *et al.*[21,22] also found that a reference-based metric and an area-based metric are adequate to predict naturalness, colourfulness, and preference when considered together. Smet *et al.*[23] in their review of various visual studies found those components relevant for naturalness aspects but not optimal for preference/attractiveness. Contrary to Islam *et al.* who found that colourfulness of objects defines the naturalness, Smet *et al.*[24] found that naturalness and colourfulness are non-correlated and that two separate factors are necessary to describe vividness/colourfulness on the one hand and naturalness/fidelity on the other hand and that preference/attractiveness could be considered as a combination of both. Bodrogi *et al.*[25] tested nine different properties of colour quality and their results suggest that six factors are needed to explain those nine properties.

In this paper, we wanted to assess key aspects of colour rendition quality which have been raised by members of the community. We investigate with a psychovisual experiment the objective aspect of colour fidelity dealing with the estimation of colour difference and the subjective aspects related to natural rendering, appreciative viewing and degree of colourfulness. The performances of 19 colour metrics have been evaluated through correlation between metrics' predictions and perceived colour quality judgements by a panel of observers.

2. The experimental set-up

2.1. The triple booth device

An experimental triple booth device was developed. It consists of three identical single booths (600 × 300 × 300 mm) placed side by side (Figure 1). The walls and bottom of the booths are painted in medium grey matt paint (Munsell N5). They are lit by different light sources located in diffusing ceilings. The light passes through a polycarbonate double diffuser which provides a homogeneous

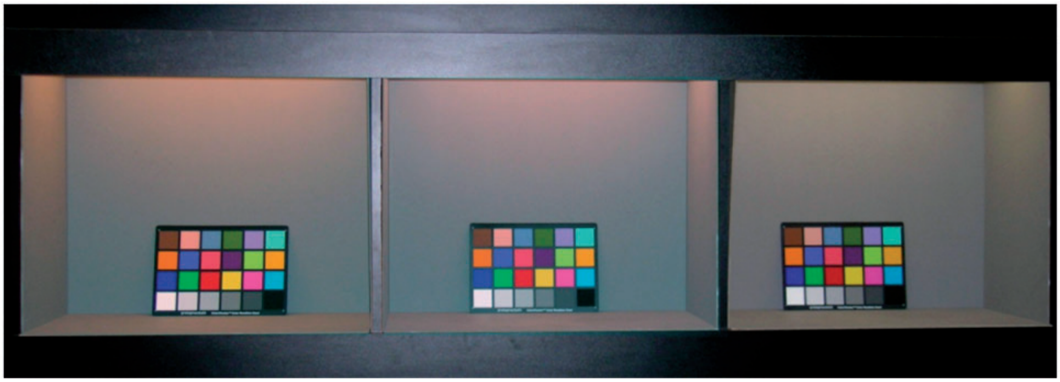

Figure 1 Photograph of the triple booth experimental device

illuminance distribution (difference less than 5%) at the bottom of the booth. The central and the left booths are equipped with five different types of LEDs (white, cyan, green, amber and red), and the right booth is illuminated with halogen or fluorescent sources. The temperatures in the LEDs' booths were checked and stabilised during the experiment. The setup is located in a room with black wall with no daylight or ambient light.

2.2 Measurement instruments

A Minolta CL200 illuminance meter was used to measure horizontal illuminances. The Specbos 1200 spectroradiometer (Number series 319090), was used to measure the spectral power distributions (SPDs) of the light sources and the spectral reflectance of the fruits and vegetables. The spectral range of measurement was 380 to 780 nm with increments of 1 nm. The spectroradiometer was calibrated by SCIENTEC in France before the experiment. The calibration results were performed with a source with spectral lines Helium OL-455-8 N°92201147 calibrated by the LNE (N°G0805291 B.N.M. COFRAC N°2-25). The wavelength verification presented a deviation within −0.002 nm and −0.06 nm and the readings for the luminance and the illuminance differed by ±5%.

SPDs of the light sources were measured in the booths at the surface of a spectral reflectance white standard provided by Gigahertz-Optik (BN-R986SQ2C) placed on the centre of the horizontal plane of the booths, accounting for inter-reflections from booth surfaces and the absorption of the white finish. The SPDs were measured for quasi-perpendicular illumination and a 45° viewing angle, which correspond to the geometry of the visual observation.

2.3 Light sources under study

In this study, we measure the psychophysical responses to the colour rendition of light sources. Two experiments based on the same protocol were done at two correlated colour temperatures (CCT), corresponding to typical indoor lighting in France (3000 K and 4000 K). Some results of the 3000 K experiment have been presented in Jost-Boissard *et al.*[26]

For each CCT, we selected a light source with a good Ra (reference sources) along with a number of LED clusters (test sources).

For both CCTs, the luminance and illuminance of the booths' bottom surface was set to approximately the same value, i.e.

73.3 cd/m^2 ± 1%/230 lux ± 3% at 3000 K and 72.7 cd/m^2 ± 5%/210 lux ± 3% at 4000 K (Table 1). These values were the maximum we could achieve with our LEDs. The other sources were dimmed to these values. Halogen and fluorescent lamps illuminated the right booth, and LED clusters were randomly distributed in the left and central booths.

The 3000 K group consisted of nine light sources:

- Halogen (Ha3K) (five Philips Pro diamondline 35 W),
- Fluorescent (Fl3K) (two Philips Master TL5HO 830 24 W),
- ‘WA3K’ cluster, with white and amber LEDs,
- ‘WR3K’ cluster, with white and red LEDs,
- ‘WAR3K’ cluster, with white, amber and red LEDs,
- ‘WCR3K’ cluster, with white, cyan and red LEDs,
- ‘WGR3K’ cluster, with white, green and red LEDs,
- ‘CRI(WGARC)3K’ cluster with white, cyan, green, amber and red LEDs, which optimised Ra,
- ‘spectrum(WGARC)3K’ cluster with white, cyan, green, amber and red LEDs, which approximated the spectrum of a Planckian radiator at 3000 K.

The 4000 K group consisted of eight light sources:

- Fluorescent (Fl4K) (two Philips Master TL5HO 90 De Luxe 940 24W),
- ‘WGA4K’ cluster, with white, green and amber LEDs,
- ‘WR4K’ cluster, with white and red LEDs,
- ‘WAR4K’ cluster, with white, amber and red LEDs,
- ‘WCR4K’ cluster, with white, cyan and red LEDs,
- ‘WGR4K’ cluster, with white, green and red LEDs,
- ‘CRI(WGARC)4K’ cluster with white, cyan, green, amber and red LEDs, which optimised Ra,
- ‘spectrum(WGARC)4K’ cluster, with white, cyan, green, amber and red LEDs, which approximated the spectrum of a Planckian radiator at 4000 K.

The relative intensities of each LED were set in order to provide the same light level, the

Table 1 Measured photometric and colorimetric properties of the 17 light sources under study

Light Sources	Illuminance (lux)	Luminance (cd/m^2)	CCT (K)	*x*	*y*	duv	Ra
Ha3K	232	73.4	3 107	0.4374	0.4183	0.0058	95
Fl3K	225	73.3	2 996	0.4415	0.4129	0.0025	85
WA3K	236	73.6	2 985	0.4248	0.3782	0.0090	45
WR3K	226	73.6	2 999	0.4076	0.3412	−0.0230	77
WAR3K	236	72.7	3 005	0.4185	0.3669	−0.0130	67
WCR3K	225	72.8	3 074	0.4169	0.3713	−0.0120	34
WGR3K	224	73.7	3 050	0.4371	0.4110	0.0020	39
CRI(WGARC)3K	230	73.8	2 993	0.4434	0.4164	0.0042	89
Spectrum (WGARC)3K	225	72.9	2 982	0.4432	0.4146	0.0028	74
Fl4K	210	71.7	3 935	0.3860	0.3893	0.0047	92
WGA4K	210	73.9	3 889	0.3884	0.3912	0.0042	59
WR4K	205	67.9	3 935	0.3738	0.3452	−0.0150	88
WAR4K	218	73.4	4 001	0.3750	0.3585	0.0069	77
WCR4K	211	73.1	3 881	0.3868	0.3851	0.0022	38
WGR4K	212	72.5	3 894	0.3869	0.3869	0.0033	71
CRI(WGARC)4K	218	74.8	3 972	0.3848	0.3900	0.0052	93
Spectrum (WGARC)4K	214	74.1	3 954	0.3851	0.3887	0.0043	88

same CCT and matched grey backgrounds. The maximum colour difference between grey backgrounds is $\Delta E_{ab} = 0.43$ ($\mu = 0.16$; $\sigma = 0.1$).

Cool white LEDs were included in all LED clusters, in order to obtain sufficiently high illuminances and also because the production of white light using monochromatic LED sources available to us would not provide light with appropriate colour rendition or colour discrimination capabilities.[10] In a previous study,[11] all possible combinations of two and three LEDs were tested at 3000 K and 4000 K, and the results showed that WCA, WGA and WA were not highly appreciated. WA3K and WGA4K were the ones that came out best, so we kept only them in this experiment.

LED sources with high gamut area and colour rendering indices were also tested. WGR3K and WGR4K were the clusters that optimised Gamut Area Index (GAI) and high Ra values were obtained by optimising the cluster for Ra or approximating the spectrum of a Planckian radiator (CRI(WGARC)3K, CRI (WGARC)4K, spectrum(WGARC)3K and spectrum(WGARC)3K.

Colorimetric properties of the light sources (correlated colour temperature CCT; chromaticity coordinates *x, y*; Delta uv (duv) and CIE colour rendering index Ra) were computed from these SPDs. Table 1 gives the measured photometric and colorimetric properties of the light sources tested. Figure 2 shows the SPDs of the light sources at 3000 K and 4000 K. Note that some LED sources have duv values greater than the CIE13-3 tolerance limit (duv>0.0054) although they are still within the limit of white light (15 MK^{-1}). Even if for these sources Ra values might be less accurate, they were tested because in our previous study they were well appreciated by observers. We wanted to compare them with other LED sources with high Ra values to give an insight concerning observers' preferences.

During the experiments, only two adjacent booths were illuminated at a time, as in a double booth set-up (left-central or central-right depending on which sources were to be compared side-by-side).

2.4 Visual scenes

Observers do not mind very much about the 'real' colours of objects with which they

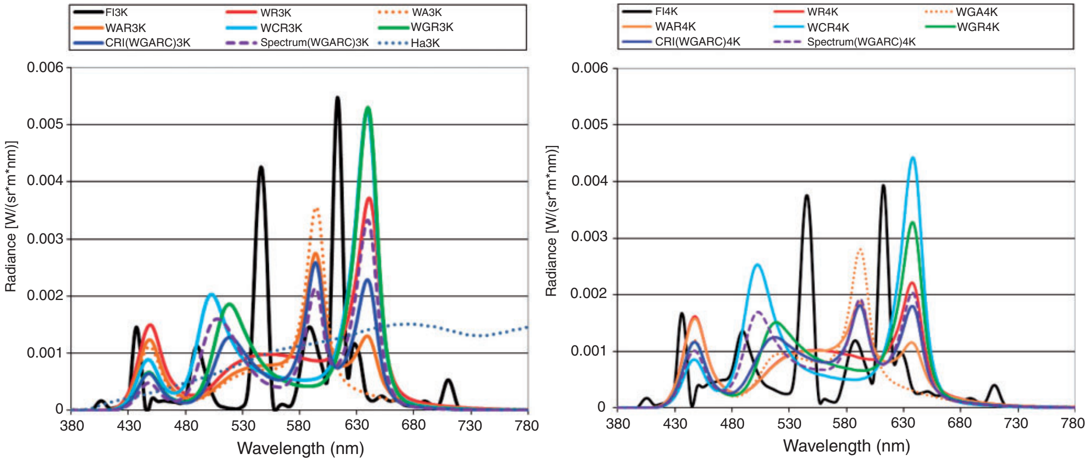

Figure 2 Measured SPDs of the 17 light sources (left: 3000 K; right: 4000 K)

are unfamiliar, but they have clear ideas about what the colours of familiar objects should be.[27,28] That is why the authors believe that an arrangement of real and well-known colourful objects is more adapted to evaluate visual colour quality. Three similar plates of fruits and vegetables containing tomato, red apple, orange, banana, lemon, Belgian endive, leek, green apple and courgette (zucchini) were arranged. Care was taken to arrange the plates as similarly as possible, so that the only variable during the experiment was the SPDs. Figure 3 gives the measured reflectance curves of the fruits and vegetables. The SPDs of the light from the objects were measured for quasi-perpendicular illumination and a 45° viewing angle, which corresponded to the geometry of the visual observation. The SPDs from the reference white were measured in the same conditions, and the reflectance factors were calculated.

A Macbeth Color Checker Chart (MCC) was also presented because it provides a satisfactory coverage of the hue circle and was recommended by TC1-33.[2]

The plates of fruits and vegetables were presented first for all the lighting pairs at the same CCT, followed by the three copies of the 24-sample edition of the MCC.

3. Experimental method

The objective of the experiment is to test the colour-rendition qualities of LED clusters in comparison with standard light sources (halogen or fluorescent) for naturalness, attractiveness, colourfulness and colour difference, using a paired comparison method. The observers were asked to compare simultaneously two booths at a viewing distance of 1.5 m with the same visual scene under different SPDs. Two independent experiments with the same procedure were done at 3000 K and 4000 K. For each CCT, all possible lighting pairs were evaluated by the observers except halogen versus fluorescent at 3000 K because they were only present in the same booth (at the right side of the triple booth).

3.1 Procedure

Before starting the experiment, observers were screened for colour vision deficiencies with the Farnsworth D15 test, and asked to

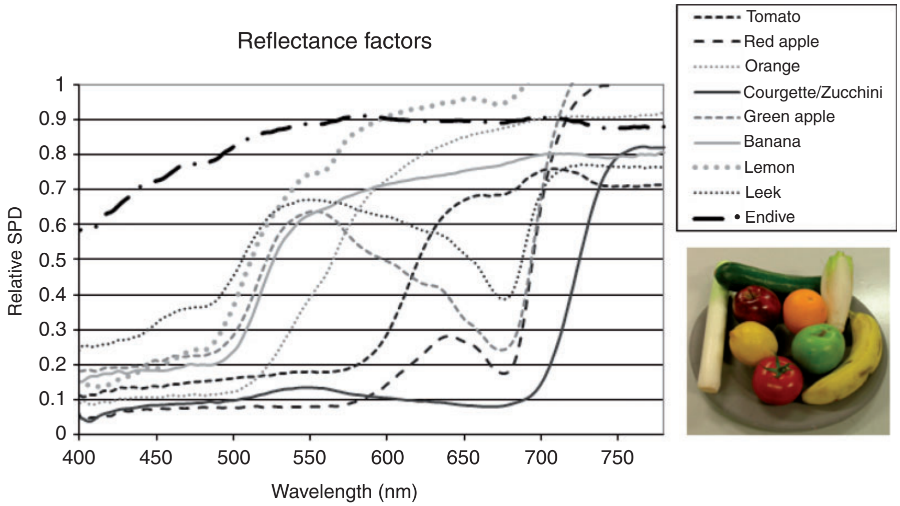

Figure 3 Measured spectral reflectance curves of fruit and vegetables used in the experiment

state their age, sex, eyesight and expertise in lighting and colour. A computer was used to display the questions and record the observers' answers.

Before the start of the tests, observers adapted for five minutes to the dark room and the CCT observing fruits and vegetables in the three booths under three different SPDs. The SPDs were chosen to represent the range of illuminations used. During this time, they were given an oral description of the experiment.

During the experiment, the observers viewed simultaneously two adjacent booths with the same visual scene lit with different light sources (Figure 4). They were instructed to consider the colour rendition of the whole scene and not to focus on the colour of the light or the colour of any particular object.

Fruits and vegetables were presented first. For each pair of light sources (presented in random order), observers assessed three aspects of colour quality:

- Global colour difference, on a 0–6 continuous scale,
- Global appreciation, by choosing in which booth the rendering appeared the most attractive (in French, the word 'joli' was used),
- Naturalness, by choosing in which booth fruits and vegetables appeared most natural ('naturel' in French).

Once all pairs were tested for fruits and vegetables, the plates were removed and the MCC was presented. With this visual scene, the observers estimated:

- Global colour difference, on a 0–6 continuous scale,
- Global appreciation, by choosing in which booth the rendering appeared the most attractive (in French the word 'joli' was used),
- Colour enhancement, by choosing in which booth the MCC appeared most colourful/vivid (in French 'niveau de coloration (couleurs plus vives)' was used).

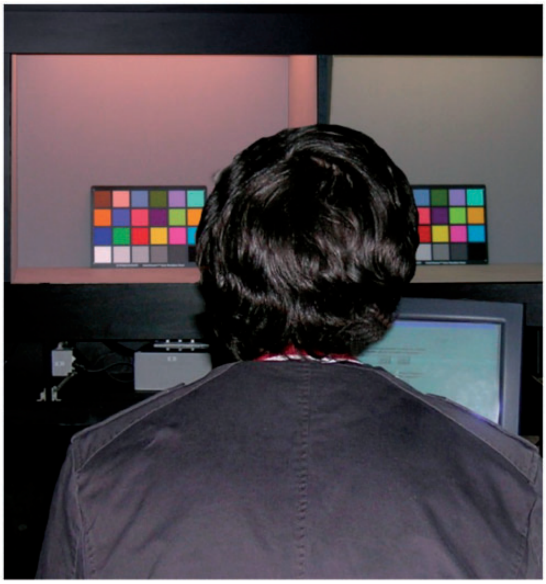

Figure 4 Test setup, showing position of an observer. The observer answered questions using a computer (available in colour in the online version)

3.2 Observers

Forty-five observers (21 females and 24 males) took part in the 3000 K experiment and 36 observers (17 females, 19 males) took part in the 4000 K experiment. All of them passed the Farnsworth D15 test. The observers took as much time as they needed to complete the experiment. The average time taken turned out to be around 45 minutes at 3000 K and 30 minutes at 4000 K.

During the experiment, all possible pairs of light sources were evaluated except halogen versus fluorescence at 3000 K. In an all possible pairs method, if one has n items to test, the total number of possible pairs is: $n(n-1)/2$. The 4000 K group consisted of eight light sources, therefore $(8 \times 7)/2 = 28$ pairs were appraised. The 3000 K group consisted of nine light sources, therefore

(9 × 8)/2 = 36 pairs minus one pair halogen/fluorescent, i.e. 35 pairs were appraised.

The total number of evaluations is:

at 3000 K: 45 observers × 35 lighting pairs × 2 scenes = 3150

at 4000 K: 36 observers × 28 lighting pairs × 2 scenes = 2016.

This means that 5166 situations were evaluated through three aspects of colour quality; thus, around 15,500 data on colour rendition were obtained.

4. Colorimetric calculation

Colorimetric calculations were performed for each of the 17 light sources.

Colour differences were computed for each pair of light sources and each of the test colours (fruits and vegetables and MCC reflectance factors), using five colour difference formulas:

- U*V*W* (ΔE_{U*V*W*}) with von Kries chromatic adaptation as in the calculation of Ra. The calculation steps are given in CIE 13.3-1995.[1]
- CIELAB (ΔE^*_{ab}). The method is described in CIE15:2004.[29]
- And three CIECAM02 formulae developed and explained in Luo *et al.*:[30] Uniform Colour Space (ΔUCS), Small Colour Difference (ΔSCD) and Large Colour Difference (ΔLCD). In these formulae, the following viewing parameters values were used: F = 1.0; c = 0.69; Nc = 1.0.

Different sets of colour matching functions (CMF) were used: CIE standard observers (2° and 10°), the CMF derived from the chromaticity diagram fundamentals[31] and the modified fundamentals presented by Csuti and Schanda.[32]

These calculated colour differences were compared with the visually observed differences.

Nineteen different colour metrics were calculated (Table 2):

- Judd's Flattery Index (Rf)[33]
- Thornton's Color Preference Index (CPI)[34]
- Special and General Colour Rendering Indices (Ri, Ra and Ra14), according to the Test-Color Method, as recommended by the CIE[1]
- Cone Surface Area (CSA) developed by Fotios[35]
- Full Spectrum Colour Index (FSCI) developed by Rea *et al.*[36,37]

Table 2 Colour metric values of the 19 light sources

Light Sources	Ra	Ra14	Ra2012	CQSa	CQSf	CQSp	CQSg	RCRI	FSCI	FCI	GAI	CSA	Sa	Rm	Rf	CPI	CCRI	PS	Gal_Ra
Ha3K	95	94	97	94	94	93	94	100	78	117	100	1.434	0.73	81	84	91	93	96	98
Fl3K	85	75	76	85	84	86	96	86	17	111	98	1.432	0.71	75	84	96	95	87	92
WA3K	45	31	42	49	49	49	73	39	8	62	66	1.440	0.67	54	45	22	63	29	56
WR3K	77	73	80	82	73	102	128	74	41	158	113	1.450	0.78	93	69	108	95	82	95
WAR3K	67	60	61	66	64	70	91	39	27	92	83	1.444	0.71	74	56	51	76	63	75
WCR3K	34	16	44	47	44	62	124	49	32	152	114	1.445	0.69	65	60	90	78	46	74
WGR3K	39	28	52	51	49	65	123	39	16	171	128	1.434	0.72	78	58	83	81	51	83
CRI(WGARC)3K	89	86	86	87	88	86	94	94	14	112	98	1.431	0.72	77	82	88	88	83	94
Spectrum (WGARC)3K	74	65	70	69	67	74	94	71	20	119	96	1.431	0.70	67	75	97	82	94	85
Fl4K	92	86	88	91	91	91	98	94	39	109	100	1.462	0.74	85	87	101	97	95	96
WGA4K	59	45	61	60	62	56	76	62	18	68	76	1.461	0.69	65	61	47	75	26	68
WR4K	88	85	83	86	80	97	114	84	49	133	107	1.471	0.78	93	79	107	96	96	97
WAR4K	77	69	73	74	73	76	93	62	42	98	90	1.470	0.74	84	70	74	88	64	83
WCR4K	38	21	45	41	39	53	111	39	37	137	112	1.462	0.69	65	54	70	75	61	75
WGR4K	71	67	81	81	74	96	117	74	39	149	123	1.462	0.77	91	82	117	93	92	97
CRI(WGARC)4K	93	89	93	93	93	93	97	100	36	110	102	1.463	0.74	85	85	93	96	80	97
Spectrum (WGARC)4K	88	84	87	84	83	85	96	91	36	106	97	1.463	0.73	81	82	92	91	91	92

- Feeling of Contrast Index (FCI) developed by Hashimoto *et al.*[14,38]
- NIST's Colour Quality Scale (CQSa) and its provisional index for fidelity (CQSf), preference (CQSp) and gamut (CQSg)[39] calculated with Excel software version 9.0 on 15 samples
- Rank order Color Rendering Index (RCRI) developed by Bodrogi *et al.*[15] calculated with Excel software developed by the authors
- Memory Color Quality metric (Sa) and its general index (Rm) developed by Smet *et al.*[17]
- Updated Colour Rendering Index (Ra2012) developed by some members of TC1-69[18] calculated with Excel software version 9.0 on the 210 real set
- Categorical Colour Rendering Index (CCRI) based on CIECAM02 developed by Yaguchi and Mizokami[40] calculated with Excel software shared within TC1-91
- Preference of Skin (PS) developed by Yaguchi and evaluated within TC1-91
- Gamut Area Index (GAI) corresponding to the eight MCC colours recommended by TC1-33 and calculated in CIELAB. The calculation of GAI is based on Thornton's method for establishing the Colour Discrimination Index,[41] and is normalised to 100 units for the reference illuminant (Planckian at 3000 K or 4000 K). This normalisation explained why some GAIs are above 100
- GAI_Ra suggested by Rea and Freyssinier-Nova[6] and used by Smet.[23] Note that the GAI calculated by Rea *et al.* is not the same because it is based on the eight Munsell samples, the calculations are performed in CIE u′v′ and the reference source is the equal energy stimulus. We found it more relevant to use an updated GAI with MCC colours because they were evaluated in the experiment.

5. Colour difference results

One of the aims of the experiment was to find the best descriptor (among those presented in Section 4) for the visual colour differences. All the observers assessed the global colour difference for MCC and for fruits and vegetables for every lighting pair. In total, 5166 visual colour difference data were obtained.

5.1. Observer rating analysis

The observers' ratings were first analysed. It is important to test the consistency of the observers and the suitability of the experimental method.

The standard (or z-) scores of the variable were calculated from the mean and the standard deviation of all the data of each observer:

$$z\Delta E_{vis_{ij-k}} = \frac{\Delta E_{vis_{ij-k}} - \Delta E_{vis_i}}{\sigma\left(\Delta E_{vis_i}\right)}$$

where $\Delta E_{vis_{ij-k}}$ is the visual colour difference estimated by the observer i for the lighting pair $j-k$; ΔE_{vis_i} is the mean of all the lighting pairs $j-k$ for the observer i ($N=35$ or $N=28$) and $\sigma\left(\Delta E_{vis_i}\right)$ is the standard deviation of all the lighting pairs $j-k$ for the observer i.

This method was also used by Bodrogi *et al.*[15] and makes it possible to normalise inter-individual differences and avoid scaling effects due to different uses of the scale (some observers may only give scores between 1 and 3 while some will use the entire scale 0–6 and other will scale between 3 and 6).

Z-scores were analysed through an agglomerative hierarchical clustering (AHC) using Ward's method and Euclidean distance. The dendrograms showed that three observers at 3000 K and six observers at 4000 K formed a subgroup. Further analysis of their data showed that these observers did not seem to have understood the process of rating colour difference well. We decided to exclude their data in order to have less variability even though the results in terms of correlation and statistical difference did not change. Therefore, the total number of colour difference data exploited is 42 (observers) × 35 (lighting pairs) × 2 (scenes) + 30 (observers) × 28 (lighting pairs) × 2 (scenes) = 4620.

5.2. Objective evaluation

For each scene and each CCT, we calculated the mean values ($\mu = \Delta E_{vis_{j-k}}$) and the standard deviation (σ) of visual colour difference for all observers ($N = 42$ or $N = 30$) for each lighting pair (Tables 3 and 4). The results concerning MCC are presented above the diagonal line of the tables, and results concerning fruits and vegetables are presented below the diagonal line. The means $\Delta E_{vis_{j-k}}$ for both scenes are highly correlated (R = 0.942, $p < 0.05$, $N = 63$), but a Student *t*-test on the 2310 observations obtained for each scene indicated that the means for the two visual scenes are significantly different (t[2 309] = −13.867, $p < 0.001$). The values obtained with fruits and vegetables are

Table 3 Visual colour differences between each pair of light sources at 3000 K

MCC / Fruits	WA3K		WR3K		WAR3K		WCR3K		WGR3K		CRI(WGARC) 3K		Spectrum (WGARC) 3K		Fl3K		Ha3K	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
WA3K			4.10	0.92	0.94	1.05	4.76	0.93	4.64	0.90	3.06	1.03	3.70	1.23	2.39	1.44	3.24	1.36
WR3K	4.62	0.90			2.90	1.27	1.79	1.25	2.27	1.36	3.46	1.18	2.49	1.41	3.08	1.26	2.39	1.37
WAR3K	1.50	0.96	3.43	1.16			4.13	1.00	3.92	1.16	1.66	1.30	3.11	1.30	1.25	1.05	2.39	1.22
WCR3K	5.17	0.76	2.29	1.25	4.38	1.19			1.10	0.94	3.64	1.27	2.63	1.22	3.34	1.27	2.80	1.36
WGR3K	4.95	0.91	1.94	1.33	4.35	0.82	1.51	1.31			3.04	1.23	1.95	1.16	3.51	1.17	2.20	1.32
CRI(WGARC) 3K	3.59	1.19	3.36	1.16	1.60	1.31	4.06	0.90	3.55	1.17			1.28	1.20	1.53	1.28	0.53	0.79
Spectrum (WGARC)3K	4.50	1.00	2.27	1.16	3.34	1.03	2.79	1.26	2.12	1.32	2.28	1.21			2.27	1.12	0.96	1.02
Fl3K	3.85	1.35	3.80	1.41	2.10	1.56	4.41	1.07	4.27	1.17	2.14	1.30	3.28	1.10				
Ha3K	4.02	1.08	2.77	1.28	2.45	1.19	3.53	1.16	3.33	1.04	0.86	1.30	1.49	1.23				

Note: Mean values (μ) and standard deviation (σ) ($N = 42$ observers). Data for fruit and vegetables are presented below the diagonal line and MCC's data are above the diagonal line.

Table 4 Visual colour differences between each pair of light sources at 4000 K

MCC / Fruits	WA4K		WR4K		WAR4K		WCR4K		WGR4K		CRI (WGARC) 4K		Spectrum (WGARC) 4K		Fl4K	
	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ	μ	σ
WA4K			4.20	1.15	4.22	1.05	3.58	0.90	4.93	1.10	2.84	1.33	2.55	1.35	3.05	1.49
WR4K	4.37	1.24			1.85	1.56	2.36	1.38	3.79	1.38	2.36	1.35	2.23	1.22	2.40	1.54
WAR4K	4.48	0.96	2.11	1.48			2.70	1.21	2.36	1.29	2.45	1.45	3.76	1.08	1.84	1.55
WCR4K	4.05	1.02	1.69	1.31	2.18	1.26			3.82	1.18	4.34	1.06	4.72	0.94	3.73	1.15
WGR4K	5.30	0.68	4.16	1.07	2.77	1.42	4.75	0.86			1.88	1.40	1.33	1.24	1.65	1.46
CRI(WGARC) 4K	3.53	1.29	1.86	1.23	2.43	1.21	4.72	0.99	2.32	1.28			0.85	1.11	0.47	0.62
Spectrum (WGARC)4K	2.84	1.33	2.39	1.38	3.71	1.05	5.03	1.00	1.60	1.41	1.04	0.86			1.06	1.27
Fl4K	3.45	1.51	2.63	1.54	3.14	1.52	4.65	0.97	2.48	1.28	0.91	1.05	1.45	1.16		

Note: Mean values (μ) and standard deviation (σ) ($N = 30$ observers). Data for fruit and vegetables are presented below the diagonal line and MCC's data are above the diagonal line.

significantly higher than those with the MCC. No difference was found between the variances for fruits and vegetables and MCC (Fisher's test: F[1 309] = 0.977, $p = 0.569$). Note that the colour differences computed with the reflectance factors of fruits and vegetables are also significantly higher than those computed with the colours of MCC. This is a hint about the suitability of the experimental method and shows that observers' judgements are relevant to evaluate colour differences in multi-coloured scenes.

5.3 Comparisons between perceived and calculated colour differences

The Pearson correlation coefficient between visual colour difference ($\Delta E_{vis_{i,j-k}}$) and calculated colour difference was computed from the whole dataset of 4620 observations. The CIECAM02 colour space and the UCS formula yield higher correlations with the perceived differences (Table 5). Note that, as expected, the correlation coefficients are higher for $z\Delta E_{vis_{ij-k}}$ than for $\Delta E_{vis_{ij-k}}$ indicating the effect of inter-individual differences. Note also the high correlation coefficient (R = 0.824) between ΔUCS and the visual mean on all observers $\Delta E_{vis_{j-k}}$ (for all lighting pairs and both scenes $N = 126$).

Table 5 underlines that the use of CMF derived from the fundamentals improves the correlations, and that the modified CMFs proposed by Schanda *et al.* improve them still further. The Steiger, Williams and Hotelling *t*-test was used to evaluate the significance differences between the Pearson correlations coefficients.[42]

ΔUCS performed significantly better than ΔE_{U*V*W*} and ΔE_{ab} ($p < 0.005$).

The use of CMFs derived from the fundamentals and the modified CMFs increase significantly the correlation for the whole dataset ($N = 4620$) but not significantly for the visual mean values ($N = 126$).

Recently, the standardised residual sum of squares (STRESS) has been recommended to measure the strength of the relationship between perceived and calculated colour differences.[43] Table 6 shows the values of STRESS for the colour difference formulae. The colour difference formula using CIECAM02 performed best. The performance between the formulae was compared by F-tests ($P = 0.95$; $N = 4\,620$; $F_c = 0.9\,527$; $1/F_c = 1.0\,496$). Table 6 shows the values of F when the performance of the formulae was compared with ΔUCS and with ΔSCD with the modified CMFs from the fundamentals. Comparing these F values to the value of $1/F_c$ shows that ΔE_{U*V*W*} and ΔE_{ab} performed significantly worse and that there is no statistical difference between ΔUCS, ΔSCD and ΔLCD and no statistically significant improvement when CMFs derived from the fundamentals are used.

Our results show that the calculated differences using CIECAM02 with the UCS formula were a better predictor of perceived colour differences than U*V*W* used in the calculation of the CIE Ra, or than the

Table 5 Pearson's correlation coefficients between visual colour differences $\Delta E_{vis_{ij-k}}$, their z-scores $z\Delta E_{vis_{ij-k}}$ and calculated values for the whole dataset of observations $N = 4620$ and for the means over the observers $\Delta E_{vis_{ij-k}}$ $N = 126$ $\Delta E_{vis_{j-k}}$ $N = 126$

Pearson	ΔE_{U*V*W*}	ΔE_{ab}	ΔUCS	ΔLCD	ΔSCD	ΔUCS	ΔUCS modified
						Fundamentals	Fundamentals
$\Delta E_{vis_{ij-k}}$	0.506	0.517	0.579	0.577	0.5778	0.581	0.583
$z\Delta E_{vis_{ij-k}}$	0.525	0.559	0.626	0.628	0.625	0.628	0.629
$\Delta E_{vis_{j-k}}$	0.721	0.736	0.824	0.822	0.824	0.828	0.830

Note: All values for statistical significance are less than 0.05 ($p < 0.05$).

CIELAB formula, as recommended by TC 1-33. This result has also been found by others.[15,44,45]

Consequently, we recommend the use of the CIECAM02 model in colour fidelity metrics (where a light source is compared to a reference source) as in the calculation of Ra2012.[18]

6. Comparative judgments

For each comparative judgment (natural fruit, attractive fruit, attractive chart and colourful chart), we obtain a paired comparison matrix. The paired comparison matrix A is an observer-lighting pair matrix. For each observer i for each lighting pair $j-k$, $a_{i,j-k} = 1$ if stimulus j was judged higher than stimulus k; $a_{i,j-k} = -1$ if $j < k$ and $\Delta E_{vis_{ij-k}} = 0$. Four 45×35 matrices were obtained at 3000 K and four 36×28 matrices were obtained at 4000 K.

6.1. Analysis method

Paired comparison data are often analysed using the Thurstone law of comparative judgment.[46,47] This law provides a method to convert subjective paired comparison judgments into one-dimensional scores. We applied Thurstone's case V model and rescaled the scores between 0 and 100 to correspond to the range of colour metrics. Thurstone values for the four comparative judgments and their 95% confidence intervals are given in Table 7.

Tables 8 and 9 give Pearson and Spearman correlation coefficients between Thurstone visual values and the nineteen calculated metrics. Bold values indicate a significance level $p < 0.05$; values in italic indicate a correlation which is significantly higher among the significant ones using the Steiger, Williams and Hotelling *t*-test.

Paired comparison matrices can be transformed into 'first-scores' matrices S (for each observer i and each light source j; $s_{i,j} = \sum_k \left(a_{i,j-k}\right)$). The 'first-scores' matrix is an observer-light source matrix. The sum of the rows gives the global score for all the observers for each light source j. The resulting scores give the rank order of preference for the stimuli.

Variance analyses (ANOVAs) (significance level $p = 0.05$) were performed on the scores to investigate the statistical significance of the observers mean ratings between the light sources for each comparative judgment.

ANOVA results are summarised in Tables 10 and 11. They show that for all criteria, there is a significant effect of light source.

To interpret these findings, a *post-hoc* analysis was performed using the Ryan, Einot, Gabriel, Welch Q test procedure (REGWQ) (Figures 5 and 6). The REGQW is recommended by Howell[42] because it appears to be the most powerful in most cases.

Table 6 Values of STRESS for visual colour differences ($\Delta E_{vis_{ij-k}}$) and calculated values for the whole dataset ($N = 4620$)

STRESS	ΔE_{U*V*W*}	ΔE_{ab}	ΔUCS	ΔLCD	ΔSCD	ΔSCD	ΔSCD modified
						Fundamentals	Fundamentals
ΔE_{vis}	0.4455	0.4588	0.4269	0.429	0.4265	0.426	0.4257
F/ΔUCS	1.089	1.1553	1	1.0099	0.9983	0.996	0.9942
F/ΔSCD modfund	1.0953	1.1621	1.0059	1.0458	1.0041	1.0018	1

Note: Values of F by comparing CIECAM02-UCS (F/ΔUCS) and CIECAM02-SCD with the modified fundamentals (F/$\Delta SCDmodfund$) with the other formulae in F-tests.

Table 7 Thurstone scale values (rescaled between 0 and 100) with their 95% confidence intervals, for the four comparative judgments

Light sources		Naturalness of fruit	Attractiveness of fruit	Attractiveness of MCC	Colourfulness of MCC
3000 K	Ha3K	100 ± 10	81 ± 7	78 ± 10	49 ± 7
	Fl3K	91 ± 10	55 ± 7	61 ± 12	33 ± 8
	WA3K	23 ± 16	0 ± 12	0 ± 18	0 ± 7
	WR3K	86 ± 12	100 ± 8	100 ± 13	98 ± 7
	WAR3K	74 ± 10	39 ± 8	58 ± 12	26 ± 7
	WCR3K	0 ± 14	76 ± 10	88 ± 13	100 ± 7
	WGR3K	25 ± 10	91 ± 8	84 ± 14	98 ± 6
	CRI(WGARC)3K	72 ± 9	55 ± 7	63 ± 11	30 ± 5
	Spectrum (WGARC)3K	56 ± 10	73 ± 7	78 ± 10	53 ± 7
4000 K	Fl4K	100 ± 10	65 ± 7	71 ± 11	41 ± 11
	WGA4K	40 ± 15	0 ± 12	0 ± 22	0 ± 15
	WR4K	75 ± 9	83 ± 7	89 ± 13	77 ± 9
	WAR4K	72 ± 10	37 ± 10	57 ± 15	34 ± 8
	WCR4K	0 ± 12	69 ± 13	50 ± 19	100 ± 15
	WGR4K	57 ± 11	100 ± 9	100 ± 15	85 ± 12
	CRI(WGARC)4K	84 ± 8	66 ± 8	67 ± 11	36 ± 8
	Spectrum (WGARC)4K	83 ± 9	71 ± 9	62 ± 12	39 ± 8

6.2. Naturalness

6.2.1 Visual assessment

Figures 5 and 6 show that no LED cluster ranks better than the standard light sources concerning the natural aspect. The fruits and vegetables appeared more natural under halogen or fluorescent light at both CCTs. Among the LED combinations, WCR3K, WCR4K, WA3K, WGA4K, WGR3K and WGR4K significantly ranked worse. At 3000 K, there are no statistically significant differences between Ha3K, Fl3K and WR3K, and between WR3K, WAR3K and CRI (WGARC)3K. At 4000 K, there are no statistically significant differences between CRI(WGARC)4K, Spectrum(WGARC)4K, WR4K and WAR4K.

6.2.2 Comparing perceived and calculated values

Fidelity indices (Ra, Ra14, Ra2012, CQSa and CQSf) correlate best with the judgment of naturalness (Tables 8 and 9). Their *p*-values are below 0.05 and the Steiger, Williams and Hotelling test rates them in the highest categories for both correlation (Pearson and Spearman) at both CCTs (3000 K and 4000 K). This indicates that Ra, Ra14, Ra2012, CQSa, and CQSf are better descriptors than the other indices. Although they give satisfactory correlations, some rankings are significantly different between predictions and visual estimations, especially among LEDs. For example, at 3000 K:

- CRI(WGARC)3K has the highest fidelity metrics (mean over the five fidelity indices = 87) but is in the same visual category as WR3K (mean over indices = 77) and WAR3K (mean over indices = 64). CQSa seems to be the index which gives the lowest deviation.
- WAR3K (mean over indices = 64) is perceived significantly more natural than Spectrum(WGARC)3K (mean over indices = 69).

One possible reason is the reduced accuracy of CRI because the sources have duv values above the limit. But if we consider the comparison between fluorescent (Fl3K,

Table 8 Pearson correlations between visual scales and the metrics predictions at 3000 K and 4000 K

	Fruit naturalness	Fruit attractiveness	Chart attractiveness	Chart colourful
Pearson3K				
Ra	***0.940***	0.148	0.157	−0.328
Ra14	***0.944***	0.175	0.176	−0.305
Ra2012	***0.893***	0.387	0.369	−0.083
CQSa	***0.936***	0.286	0.278	−0.176
CQSf	***0.916***	0.175	0.169	−0.295
CQSp	**0.834**	0.604	0.598	0.230
CQSg	−0.197	**0.838**	***0.837***	***0.972***
RCRI	**0.738**	0.327	0.299	−0.092
FSCI	0.467	0.482	0.450	0.262
FCI	−0.151	*0.917*	*0.882*	*0.969*
GAI	−0.112	*0.902*	*0.860*	**0.916**
CSA	−0.175	0.078	0.145	0.352
Sa	0.599	**0.716**	***0.684***	0.508
Rm	0.653	**0.742**	***0.720***	0.482
Rf	**0.716**	0.436	0.439	0.006
CPI	0.368	***0.876***	***0.880***	0.647
CCRI	**0.742**	**0.693**	***0.680***	**0.680**
PS	**0.823**	0.486	0.500	0.500
GAI_Ra	**0.723**	**0.730**	***0.710***	0.336
Pearson4K				
Ra	***0.977***	0.257	0.460	−0.322
Ra14	***0.965***	0.326	0.525	−0.260
Ra2012	**0.940**	0.379	0.549	−0.238
CQSa	***0.951***	0.364	0.559	−0.235
CQSf	***0.973***	0.237	0.430	−0.368
CQSp	**0.798**	0.663	**0.824**	0.142
CQSg	−0.117	***0.918***	**0.848**	***0.949***
RCRI	**0.904**	0.310	0.418	−0.302
FSCI	0.327	**0.736**	**0.846**	0.632
FCI	−0.170	***0.920***	**0.823**	***0.955***
GAI	−0.134	***0.930***	**0.821**	**0.905**
CSA	0.274	0.106	0.343	0.139
Sa	0.646	0.686	**0.871**	0.298
Rm	**0.731**	0.645	**0.859**	0.216
Rf	**0.904**	0.510	0.652	−0.098
CPI	0.533	***0.902***	***0.963***	0.512
CCRI	**0.877**	0.582	**0.766**	0.050
PS	0.606	***0.875***	***0.919***	0.469
GAI_Ra	**0.725**	***0.788***	**0.887**	0.293

Note: Bold values indicate the significance level $p < 0.05$. Values in italic indicate the correlations which are significantly higher among the significant ones.

Table 9 Spearman correlations between visual scales and metrics predictions at 3000 K and 4000 K

	Fruit naturalness	Fruit attractiveness	Chart attractiveness	Chart colourful
Spearman3K				
Ra	*0.850*	0.033	−0.167	−0.333
Ra14	*0.850*	0.033	−0.167	−0.333
Ra2012	*0.833*	0.317	0.150	−0.033
CQSa	*0.867*	0.150	−0.067	−0.217
CQSf	*0.850*	0.033	−0.167	−0.333
CQSp	*0.833*	0.450	0.300	0.067
CQSg	0.017	*0.831*	*0.881*	*0.865*
RCRI	0.678	0.254	0.153	0.034
FSCI	0.467	0.600	0.533	0.467
FCI	−0.133	*0.917*	*0.950*	*0.933*
GAI	−0.017	*0.867*	**0.817**	*0.883*
CSA	−0.200	0.283	0.333	0.283
Sa	0.683	*0.700*	0.450	0.283
Rm	0.683	*0.700*	0.450	0.283
Rf	0.667	0.267	0.150	0.133
CPI	0.467	0.617	0.617	0.500
CCRI	*0.733*	0.583	0.433	0.283
PS	*0.767*	0.233	0.067	0.000
GAI_Ra	*0.817*	0.517	0.317	0.133
Spearman4K				
Ra	*0.952*	0.048	0.524	−0.190
Ra14	*0.952*	0.048	0.524	−0.190
Ra2012	*0.952*	0.167	0.571	−0.119
CQSa	*0.929*	0.190	0.643	−0.071
CQSf	*0.952*	0.167	0.571	−0.119
CQSp	0.524	0.619	**0.929**	0.238
CQSg	−0.024	*0.833*	**0.762**	*0.905*
RCRI	*0.934*	0.156	0.551	−0.132
FSCI	0.120	0.349	0.566	0.398
FCI	−0.143	*0.810*	0.619	*0.905*
GAI	−0.143	*0.810*	0.619	*0.905*
CSA	0.548	0.214	0.429	−0.095
Sa	0.500	0.500	**0.929**	0.238
Rm	0.482	0.566	**0.940**	0.301
Rf	*0.857*	0.214	0.714	0.024
CPI	0.476	0.714	*1*	0.476
CCRI	*0.881*	0.214	**0.762**	0.048
PS	0.619	0.619	**0.905**	0.429
GAI_Ra	0.500	0.714	**0.952**	0.405

Note: Bold values indicate the significance level $p < 0.05$. Values in italic indicate the correlations which are significantly higher among the significant ones.

Fl4K) and the mixings which optimise Ra (CRI(WGARC)3K, CRI(WGARC)4K), they are within the limit, and they have equivalent fidelity indices but fluorescent is always judged as significantly more natural.

To inspect the correspondence between the metric and the visual appreciation and take into account the inter-observer variability, Ra and CQSa values have been plotted with

Table 10 Summary of the ANOVA tests with a significance level of $p = 0.05$ at 3000 K

3000 K	Df	F	p	R^2
Attractiveness of the MCC	8	24.9	< 0.0001	0.334
Colourfulness of the MCC	8	262.7	< 0.0001	0.841
Attractiveness of the fruit/veg	8	72.4	< 0.0001	0.594
Naturalness of the fruit/veg	8	46.7	< 0.0001	0.485

Table 11 Summary of the ANOVA tests with a significance level of $p = 0.05$ at 4000 K

4000 K	df	F	p	R^2
Attractiveness of the MCC	7	16.1	<0.0001	0.287
Colourfulness of the MCC	7	90	<0.0001	0.692
Attractiveness of the fruit/veg	7	47.5	<0.0001	0.543
Naturalness of the fruit/veg	7	40.5	<0.0001	0.503

Thurstone values for naturalness (Figure 7). To make a visual comparison easier, Thurstone scale values have been linearly rescaled from their 0–100 range to the range occupied by the metric values (linear regression). Figure 7 shows that Ra corresponds well to visual values at 4000 K but at 3000 K

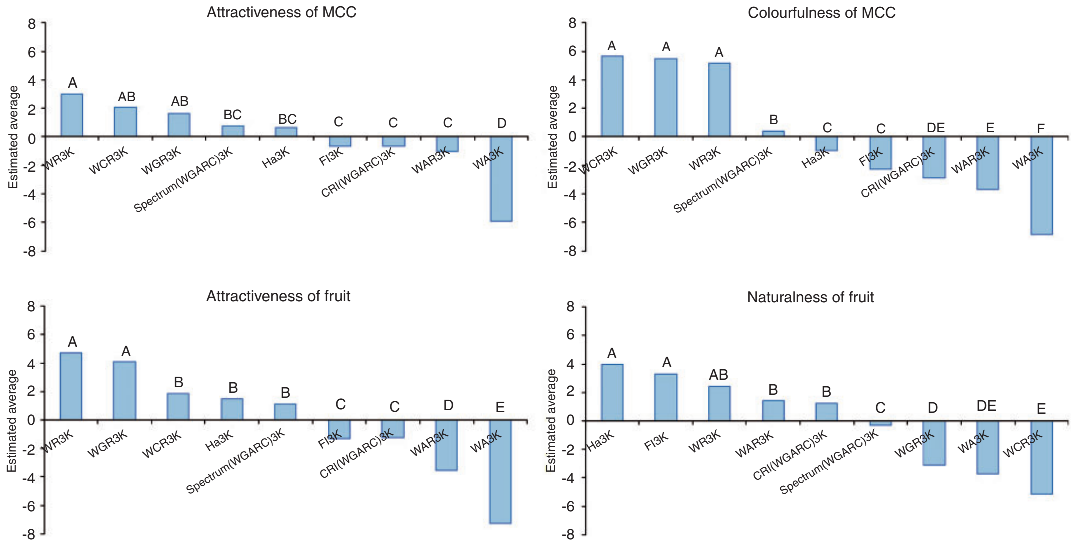

Figure 5 Results of the REGWQ test on the ANOVA values at 3000 K

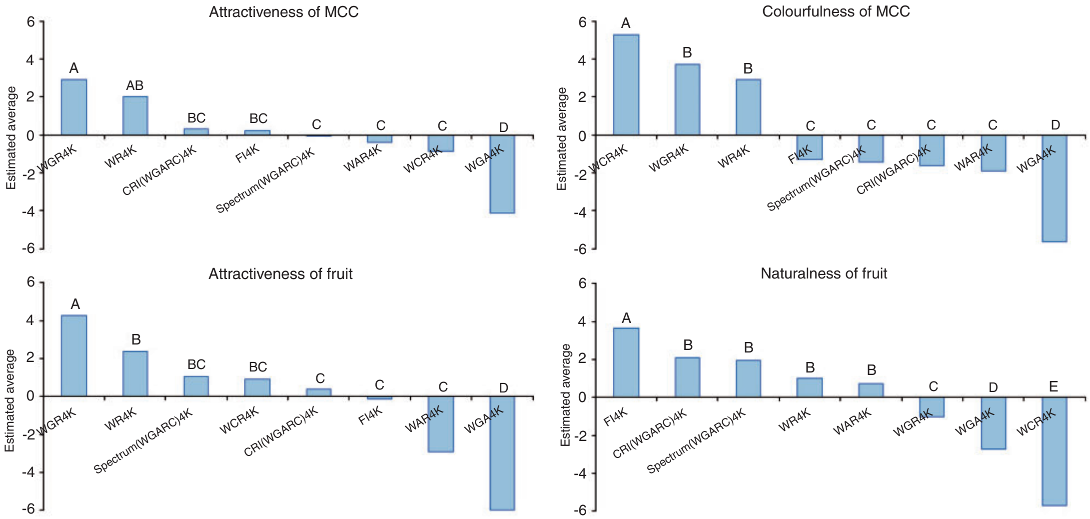

Figure 6 Results of the REGWQ test on the ANOVA values at 4000 K

some discrepancies occurred. At 3000 K, CQSa seems to improve the correspondence even if some discrepancies persist (for CRI(WGARC)3K and WAR3K whose confidence intervals do not cross the predicted metric).

Although fidelity indices correlate well with naturalness, they did not give perfect predictions for all differences between LED light sources. This fact possibly indicates that a high colour fidelity score does not necessarily mean a natural rendition

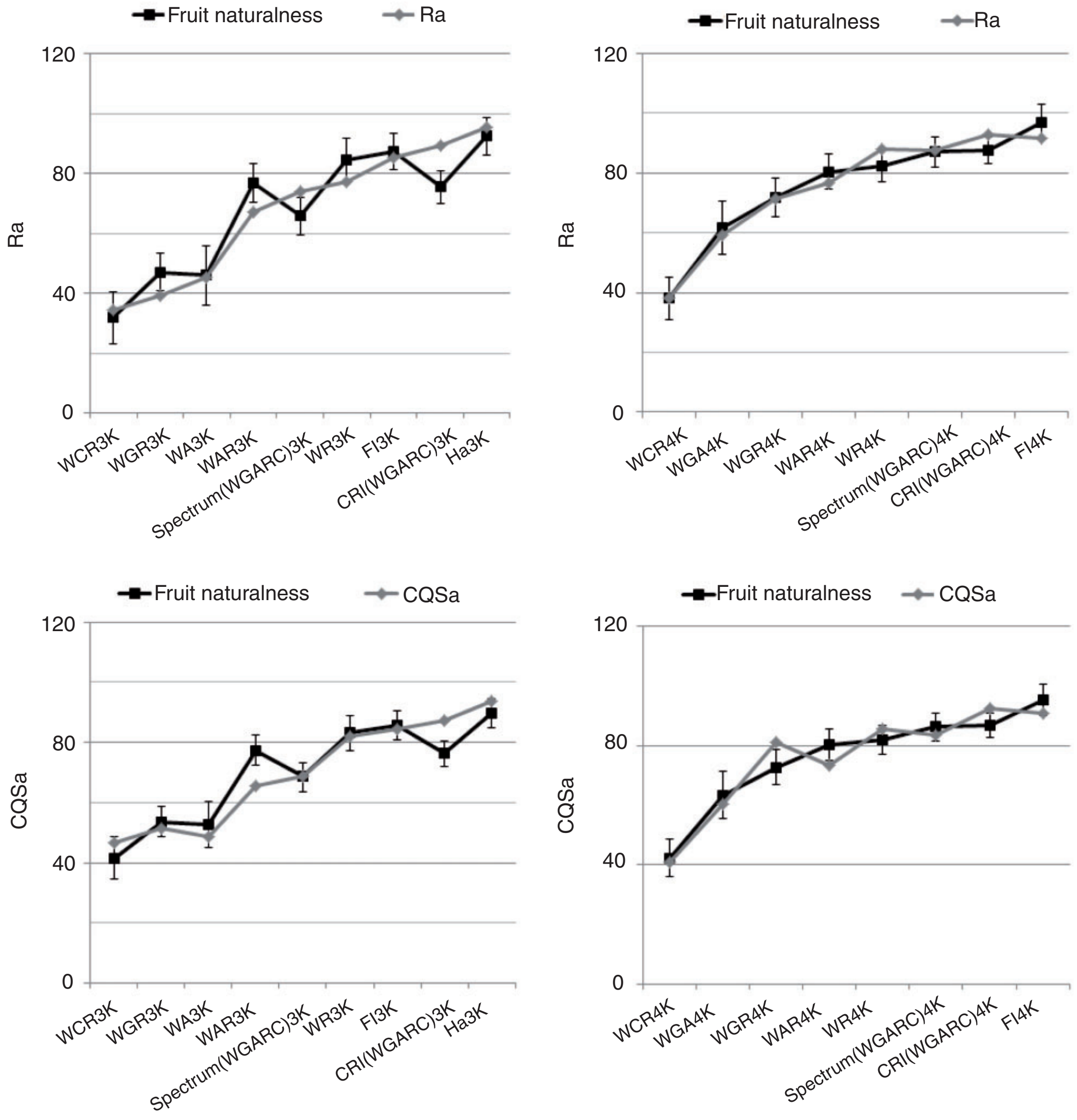

Figure 7 Correspondence of Ra and CQSa metrics with the visual appreciation of naturalness. Thurstone scale values have been linearly rescaled to the range occupied by the metrics (linear regressions)

of an object. Fidelity represents a deviation from a reference illuminant and this illuminant may not be considered as the most natural one.

Further investigation might be necessary to understand better naturalness but unlike Rea and Freyssinier,[6,20] Islam *et al*. and Dangol *et al*.,[21,22] we found no evidence to suggest that a Ra_GAI metric is adequate to predict naturalness.

6.3 Attractiveness

6.3.1 Visual assessment

Figures 5 and 6 show that some LED clusters rank better than standard light sources concerning the attractiveness of fruits and vegetables and MCC.

The WR and WGR mixings are in the highest categories whatever the CCT and the scene presented. The amount of red in a LED cluster seems to have a positive influence on the attractive aspect (the more red, the more attractive the rendering of the scenes are).

Conversely, LED combinations with a large proportion of amber were not considered attractive. In particular, for the attractiveness of fruits and vegetables, WA3K, WAR3K, WGA4K and WAR4K are significantly less attractive than the other sources. In a previous study,[11] we found that WA, WCA and WGA LEDs were not appreciated in terms of attractiveness, and we may conclude that this would be the case, in general, for combinations containing large proportion of amber.

Looking at Figure 2, the amount of amber and red in the SPD may also explain why CRI(WGARC)4K and Spectrum (WGARC)4K are in the same category, whereas CRI(WGARC)3K is less attractive than Spectrum(WGARC)3K.

If we compare the results of attractiveness for the two visual scenes, we notice that there is less spread with fruits and vegetables than with MCC. *Post hoc* tests (Figures 5 and 6) are more discriminating between light sources for fruits and vegetables and confidence intervals (Table 7) are smaller. A possible explanation is that it is easier for observers to answer subjective questions on real objects with familiar colours than on a patch of colours making discrimination finer. Nevertheless, there is a high correlation between attractiveness for both visual scenes (Table 12) and the rank order is about the same. We can find differences only for WGR3K, which is significantly more attractive than WCR3K for fruits and vegetable but not for MCC. This is may be related to the excessive chroma enhancement of WCR3K (see Section 6.4).

Table 12 Pearson and Spearman correlations between Thurstone visual scales at 3000 K and 4000 K

Spearman Pearson		Fruit naturalness	Fruit attractiveness	Chart attractiveness	Chart colourful
Fruit naturalness	3000 K	**1**	0.200	−0.100	−0.217
	4000 K	**1**	0.000	0.476	−0.190
Fruit attractiveness	3000 K	0.184	**1**	**0.900**	**0.850**
	4000 K	0.150	**1**	0.714	**0.762**
Chart attractiveness	3000 K	0.187	**0.965**	**1**	**0.933**
	4000 K	0.403	**0.932**	**1**	0.476
Chart colourful	3000 K	−0.264	**0.872**	**0.850**	**1**
	4000 K	−0.396	**0.791**	0.652	**1**

Note: Bold values indicate the significance level $p < 0.05$. Pearson correlations are presented below the diagonal line and Spearman correlations are above the diagonal line.

6.3.2 Comparing perceived and calculated values

We can notice that colour fidelity indices (Ra, Ra14, Ra2012) do not provide good predictions for the attractiveness of light sources (Tables 8 and 9). WCR and WGR, which have very low indices, are perceived attractive, whereas LEDs with an optimised CRI are not. Moreover, there is a very low correlation between the visual rankings of attractiveness and the calculated metrics. As a result, and as expected, the colour fidelity indices are not good indicators of the attractiveness of light sources.

Strict colour fidelity metrics penalise any deviation from reference illuminant and it is clear with our data that not all deviation should be penalised for preference because some light sources score visually better than standard light sources (halogen or fluorescent).

Judgment of attractiveness correlates best with gamut-based indices (Tables 8 and 9). GAI, FCI and CQSg are better descriptors than the other indices for fruits and vegetables for both CCTs and whenever we consider Pearson or Spearman correlations (tested with the Steiger, Williams and Hotelling *t*-test). They are also good descriptors for MCC but at 4000 K, the really high correlation of visual values with CPI makes it the optimal index. CPI is also a good descriptor for fruits and vegetables if we consider the Pearson correlation but not for the Spearman correlation.

To investigate the correspondence between metric and visual appreciation and take into account the inter-observer variability, FCI and CPI values have been plotted with rescaled Thurstone values for attractiveness (Figure 8). CPI values clearly correspond better for the attractiveness of MCC at 4000 K. For fruit attractiveness, FCI seems to correspond better but even if correlations are high some problems occurred. WCR seems difficult to predict for both CCTs and both scenes. Some sources are well predicted by CPI while others are better predicted with

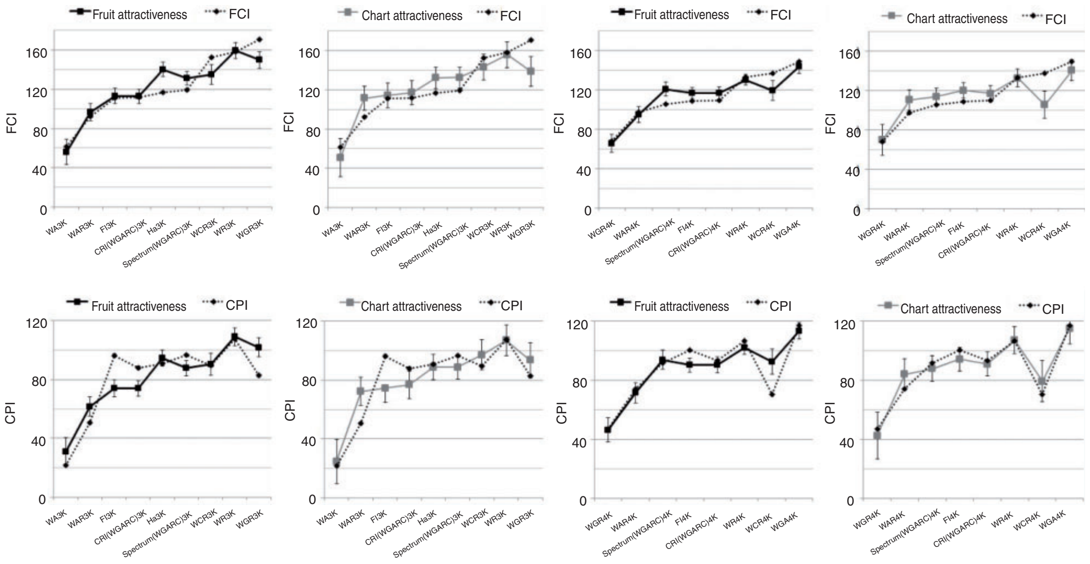

Figure 8 Correspondence of FCI and CPI metrics with the visual appreciation of naturalness. Thurstone scales values have been linearly rescaled to the range occupied by the metrics (linear regressions)

FCI. Note also that too high a FCI seems to be detrimental.

6.4 Colourfulness

When observers were asked to make comparisons between LEDs and halogen light, they often said that with LEDs there was an increase in colour contrast and that is why the object appeared more attractive to them, which suggests that the attractiveness of fruits and vegetables is linked to the saturation of their colours. Indeed, colourfulness is well correlated with attractiveness (Table 12). We can notice that the colourfulness of MCC is even more correlated to the attractiveness of fruits and vegetables than to the attractiveness of MCC and this difference is significant at 4000 K (Steiger, Williams and Hotelling test).

The rank order for attractiveness and colourfulness is about the same. The only difference is for WCR4K which is significantly the most colourful source but is not found attractive (Figure 6). This might be explained by the fact that the colours with this source are very saturated which leads to abnormal/unusual colours. WCR4K enhances chroma but it is not acceptable and appreciated especially with fruits and vegetables. The colours are unpleasant and unnatural (WCR4K is the source that provide a less natural aspect to fruits and vegetables).

To a lesser extent, this could explain the difference of judgment for WCR3K concerning MCC but also the difference between attractiveness of fruits and vegetables and MCC noticed in Section 6.3.1. WCR3K gives the most colourful rendering but it does not produce the most attractive rendering on MCC and it makes colours of fruits appear unnatural. That is why fruits appeared less attractive under this source.

We are here at the limit between colourfulness and preference. More colourful does not always mean more appreciated which underlines the limitation of gamut indices for appreciation.

As expected, for MCC, colourfulness correlates well with the gamut-based indices. GAI, FCI and CQSg are significantly better descriptors (Tables 8 and 9).

Colourfulness of objects influences the observers' appreciation, and they are important parameters to consider for a future Colour Appreciation Index.

7. Discussion

7.1 Importance of graphical method

Chroma enhancement or increase in colourfulness seems to be a reason for visual appreciation. This was pointed out by Judd[33] 40 years ago, and by Thornton with the Color Preference Index[34] and Color Discrimination Index.[41] More recently, some visual studies involving LEDs have underlined this point.[11,21,22,48] CQS[16] metrics takes this into account and it is clear that such an approach had a positive impact on the performance of the metric.[49] Gamut area-based metrics (GAI, FCI) underline this because an increase in chroma of test samples is often accompanied by a global increase in the gamut area.

To investigate that point, we compare the chromaticity of eight MCC colours in CIELAB space (Figure 9). Those eight colours were chosen because they represent the hue coverage of our colours.

The lateral changes on the left and right (in the direction of red and green/blue) were larger with some LED clusters than with halogen or fluorescent at both CCTs. Moreover, considering the gamut area of WR and WCR, we see that an overall increase (WR) was more appreciated, in visual terms, than a large increase in specific colours (WCR) at both CCTs. WCR4K really modifies red and blue/green colours; that may be a reason why it is rejected for the attractiveness of fruits and vegetables. If we compare the gamut areas of Spectrum(WGARC)3K, CRI(WGARC)3K

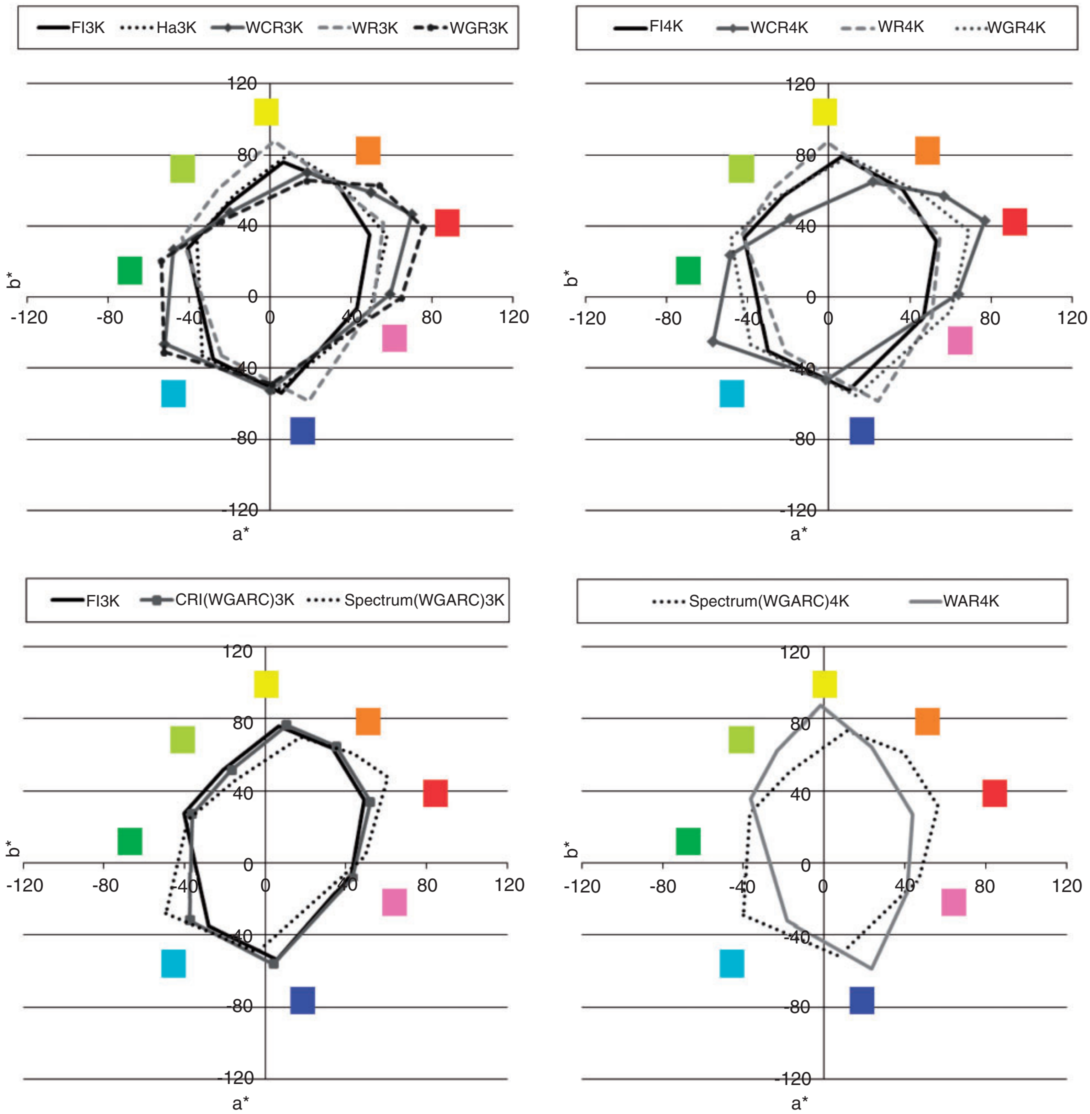

Figure 9 Colour coordinates of eight MCC colours plotted in the a*b* plane for different light sources (available in colour in the online version)

and Fl3K which have similar gamut indices, we see that a reduction in the chroma of red-orange and blue-green is prejudicial for attractiveness, whereas a gain in those colours seems to compensate for a loss in the yellow-green and blue-pink regions. The same result was found at 4000 K with the comparison of the gamut areas of Spectrum(WGARC)4K and WAR4K.

The red–blue/green contrast seems especially important to overall satisfaction for colour rendition. Gamut metrics oversimplify saturation changes and that is why they can be inaccurate predictors. A graphical representation is required to observe simultaneously specific colour variations (and their directions) and global change. These kinds of representations are very expressive for expert and non-expert users. We would therefore highly recommend graphical representations as van der Burgt *et al.* suggested[50] to explain colour-rendition properties of light sources.

7.2 Effect of visual scene

Pearson's correlation coefficients between visual colour differences $\Delta E_{vis_{ij-k}}$ and calculated values for each visual scene ($N = 2310$) show that the correlation is higher with fruits and vegetables than with MCC. These results suggest that fruits and vegetables are judged better than the MCC and that colour difference evaluations might be easier to make on real coloured objects in comparison with patches of colours. Note that observers were asked to judge the overall difference on a multicoloured scene and not on a single colour. More colours were evaluated on MCC than on the plate of fruits and vegetables and that may also explain the lower correlation.

We remind the reader that the questions concerning attractiveness for each scene were not asked simultaneously (fruits were presented first for all the lighting pairs and then MMC was evaluated) and that order of presentation of sources was randomised within observer and scene. Therefore, no influence on one scene on the other was possible during the judgments.

As we mentioned in Section 6.3, we notice that there is less spread with fruits and vegetables than with the MCC. Comparing the coefficients of consistency for the two scenes, we see that for attractiveness, fewer transitivity errors were made with the fruits and vegetables than with the MCC, at both CCTs. The coefficient of consistency measures, for each observer and each question, the errors of transitivity made in a triad (if A > B and B > C then A should dominate C, otherwise there is an error of transitivity).

From our results, it seems easier for observers to answer subjective questions on real objects with familiar colours. We would therefore suggest using natural objects of familiar colours in tests on perceived colour quality to facilitate the observer's tasks. However, the use of patches of colour such as those of the MCC give correlated results and they could be used to make some judgment of quality.

In their article, Smet *et al.*[23] suggest adjusting metric calculations to correct for the lack of blue and purple hue in judgment of fruits and vegetables. The Memory Colour Metric (Sa) was recalculated with the blue and purple samples (Smurf and lavender) omitted (Sa restricted). The results concerning attractiveness of fruits and vegetables are highly correlated with Sa restricted (Table 13)

Table 13 Pearson and Spearman correlations between Thurstone visual scales and Memory Colour Metrics Sa and Sa restricted (calculated with blue and purple samples omitted) at 3000 K and 4000 K

		Fruit naturalness	Fruit attractiveness	Chart attractiveness	Chart colourful
Pearson3K	**Sa**	0.599	**0.716**	**0.684**	0.508
	Sa restricted	0.234	**0.982**	**0.937**	**0.822**
Spearman3K	**Sa**	0.683	**0.700**	0.450	0.283
	Sa restricted	0.150	**0.983**	**0.917**	**0.833**
Pearson4K	**Sa**	0.646	0.686	**0.871**	0.298
	Sa restricted	0.268	**0.993**	**0.949**	**0.732**
Spearman4K	**Sa**	0.500	0.500	**0.929**	0.238
	Sa restricted	0.333	**0.905**	**0.905**	0.619

Note: Bold values indicate the significance level $p < 0.05$.

and the results for MCC are also significantly higher when all the colours were evaluated. MCRI is colour sample-dependent and the calculation without blue and purple hue is more correlated to our visual results.

Figure 10 represents Sa and Sa restricted with rescaled Thurstone visual values for attractiveness. Sa restricted clearly corresponds better. In particular, WCR and WR sources which do not correspond at both

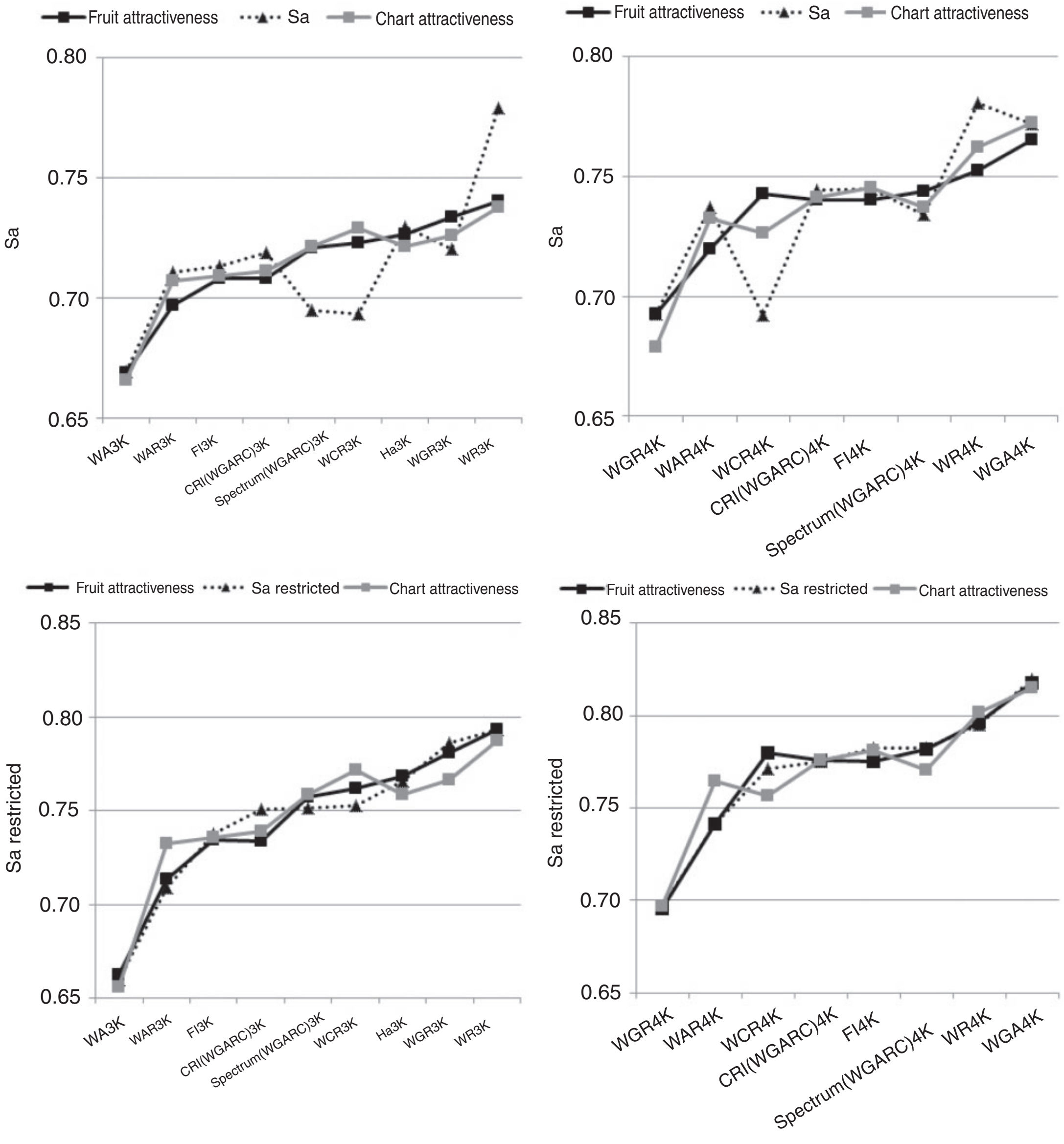

Figure 10 Correspondence of Sa and Sa restricted metrics with the visual appreciation of attractiveness. Thurstone scales values have been linearly rescaled to the range occupied by the metrics (linear regressions)

CCTs to Sa are well described with Sa restricted. A comparison of Figures 10 and 7 underlines that Sa restricted seems to be better descriptor than FCI or CPI for our visual values. For future colour quality experiments, we would therefore suggest to use hues covering the entire space to be sure that metric calculations will correspond to actual colours present in a scene.

7.3 Influence of the visual objectives

Three different aspects of colour rendition were assessed: naturalness for fruits and vegetables, attractiveness for fruits and vegetables and MCC, and colourfulness for MCC. As mentioned in Section 6.4, colourfulness is well correlated with attractiveness and we can notice that colourfulness of MCC is even more correlated to the attractiveness of fruits and vegetables than to the attractiveness of MCC. There was better discrimination for colourfulness than for attractiveness. The confidence intervals are smaller and the differences are more significant with *post hoc* tests (Figures 5 and 6 and Table 7). Transitivity errors were more numerous for the attractiveness of MCC, possibly because it was easier for people to assess its colourfulness than its attractiveness, in the sense that they did not have any 'feelings' for the chart. This problem of interpretation concerning the attractiveness of chart could also explain why the colourfulness of MCC is more correlated to the attractiveness of fruits and vegetables than the attractiveness of the MMC (Table 12).

The naturalness of fruits and vegetables was more difficult to assess than their attractiveness, and there were more circular errors and less discrimination (Figures 5 and 6). This is consistent with the results we had previously found.[11] Attractiveness seems easier to judge than naturalness, no doubt due to the lack of predetermined reference points.

We also notice that there is a negative correlation between the naturalness of fruits and the colourfulness of MCC meaning that the objects appeared less natural when there is chroma enhancement.

7.4 Other effects

The combinations of colours in the LED clusters really influence the judgments in our experiment. High proportions of amber LEDs seem detrimental to attractiveness while an increased proportion of red seems beneficial. Moreover, a multiplicity of colours did not necessarily improve the rendition. WR, for example, was considered to be both attractive and natural at both CCTs.

Sources of different CCTs were not compared together, so we cannot conclude anything about the influence of CCT on the observers' judgments. Nevertheless, we show that at 3000 K or at 4000 K, the conclusions were the same. Colour difference formulae using CIECAM02 performed best (Tables 5 and 6), fidelity metrics are well correlated to naturalness of fruits and vegetables, gamut-based indices and memory rendering index calculated without blue hues are well correlated with attractiveness and colourfulness.

As emphasised in Section 2.3, some sources were above the CIE duv limit. We are conscious that for these sources the Ra value might be less accurate and that some results might be lowered. But WR sources are below the Planckian locus (in the red region) and are well appreciated at both CCTs. Further investigations are necessary, but if duv has an impact on subjective appreciation a new colour rendition metric should take it into account (within the limit of white light). The impact of duv also suggested in Dangol *et al.*[21]

8. Conclusion

The results obtained for the perception of LED-based lighting lead to suggestions for optimal use of such a light source, according to the application. Colours were found more attractive under WR, WCR and WGR LED clusters than with standard light sources, and this seems mainly related to the contribution of red LEDs. Concerning naturalness, the standard light sources ranked higher, and this could be related to a cultural memory. However, some LED combinations need to be avoided. For example, in the present experiment, a high proportion of amber LEDs was found to be neither attractive nor natural while WR mixing is good for attractiveness and naturalness. The composition of LED combinations is important and we need to take care, in setting out guidelines, not to introduce any artificial obstacles to the emergence of new LED technologies.

Concerning the central question of colour rendition quality, the data presented here and the arguments put forward over the last 50 years indicate that the CRI cannot provide a satisfactory rating of overall colour quality rendition on its own. But any other of the 19 metrics tested are adequate for the overall characterisation of colour rendition. Our experiment suggests that naturalness is better characterised by colour fidelity metrics while attractiveness is correlated with gamut area indices, Colour Preference Index or restricted Memory Colour Rendering Index. Moreover, some visual objectives (naturalness and colourfulness) seem not achievable together making a unified index hard to develop.

It can be seen from the above remarks that if we define colour rendering according to the CIE definition, we should adopt an updated colour fidelity index with appropriate colour space and CMFs; in this case, we would suggest the use of CIECAM02. But if one considers colour rendition from the viewpoint of appreciation, we should give interest to gamut diagrams. We highly recommend a graphical representation which is very expressive and could help the understanding and the interpretation not only for colour scientists but also for end users.

The question of colour discrimination or duv which was not studied here may raise other concerns. TC1-91 will have to be vigilant on the multiple aspects of colour quality.

The colour rendition quality of a light source is a complex question. And of course, the issues involved are not interesting only for professionals – metrics and descriptors should be understandable by all.

Funding

This work was financially supported by Philips Lighting.

Acknowledgements

The authors gratefully acknowledge the observers who gave their time for this research and Philips Lighting's contribution for the development of the experimental apparatus.

References

1 Commission Internationale de l'Eclairage. *Method of Measuring and Specifying Colour Rendering Properties of Light Sources*. Vienna: CIE, 1995.
2 Commission Internationale de l'Eclairage. *Vision and Colour*. *Physical Measurement of Light and Radiation*. Vienna: CIE, 1999.
3 Worthey JA. Color rendering: A calculation that estimates colorimetric shifts. *Color Research and Application* 2004; 29: 43–56.
4 Li C, Luo MR, Li C. *Evaluation of LED sources based on colour appearance data: Proceeedings of the 26th Session of the CIE* Beijing, China, 4–11 July 2007 162–165.
5 Pointer M. *Measuring Colour Rendering – A New Approach II*. Leeds: National Physics Laboratory, 2004.

6 Rea MS, Freyssinier-Nova JP. Color rendering: A tale of two metrics. *Color Research and Application* 2008; 33: 192–202.

7 Narendran N, Deng L. Color rendering properties of LED light sources. *Second International Conference on Solid State Lighting, SPIE Congress*. Seattle, USA, 9–11 July 2002, pp. 61–67.

8 Bodrogi P, Csuti P, Szabó F, Schanda J. Why does the CIE Colour Rendering Index fail for white RGB LED light sources?: *Proceedings of the CIE Expert Symposium on LED Light Sources: Physical Measurement and Visual and Photobiological Assessment*. CIE Publication x026:2005, Vienna: CIE, 7–8 June 2004, pp. 24–27. Tokyo, Japan.

9 Schanda J, SÃndor N. Colour rendering, past - present - future: *International Lighting and Colour Conference*, Cape Town, South Africa, Nov 2–5: 2003.

10 Mahler E, Ezrati J-J, Viénot F. Testing LED lighting for colour discrimination and colour rendering. *Color Research and Application* 2009; 34: 8–17.

11 Jost-Boissard S, Fontoynont M, Blanc-Gonnet J. Perceived lighting quality of LED sources for the presentation of fruit and vegetables. *Journal of Modern Optics* 2009; 56: 1420–1432.

12 Boissard S, Fontoynont M. Optimization of LED-based light blendings for object presentation. *Color Research and Application* 2009; 34: 310–320.

13 Commission Internationale de l'Eclairage, *Colour Rendering of White LED Light Sources*. CIE Publication177. Vienna: CIE, 2007.

14 Hashimoto K, Yano T, Shimizu M, Nayatani Y. A new method for specifying color rendering properties of light sources based on feeling of contrast. *Color Research and Application* 2007; 32: 361–371.

15 Bodrogi P, Brückner S, Khanh TQ. Ordinal scale based description of colour rendering. *Color Research and Application* 2010; 36: 272–285.

16 Davis W, Ohno Y. Color quality scale. *Optical Engineering* 2010; 49: 033602–033616.

17 Smet K, Ryckaert W, Pointer MR, Deconinck G, Hanselaer P. A memory colour quality metric for white light sources. *Energy and Buildings* 2012; 49: 216–225.

18 Smet KA, Schanda J, Whitehead L, Luo RM. CRI2012: A proposal for updating the CIE colour rendering index. *Lighting Research and Technology* 2013; 45: 689–709.

19 Guo X, Houser K. A review of colour rendering indices and their application to commercial light sources. *Lighting Research and Technology* 2004; 36: 183–199.

20 Rea MS, Freyssinier J. Color rendering: Beyond pride and prejudice. *Color Research and Application* 2010; 35: 401–409.

21 Dangol R, Islam M, Hyvärinen M, Bhusal P, Puolakka M, Halonen L. User acceptance studies for LED office lighting: Preference, naturalness and colourfulness. *Lighting Research and Technology* 2013; 45: 666–688.

22 Islam M, Dangol R, Hyvärinen M, Bhusal P, Puolakka M, Halonen L. Investigation of user preferences for LED lighting in terms of light spectrum. *Lighting Research and Technology* 2013; 45: 641–665.

23 Smet K, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Correlation between color quality metric predictions and visual appreciation of light sources. *Optics Express* 2011; 19: 8151–8166.

24 Smet KA, Ryckaert WR, Pointer MR, Deconinck G, Hanselaer P. Memory colours and colour quality evaluation of conventional and solid-state lamps. *Optics Express* 2010; 18: 26229–26244.

25 Bodrogi P, Bruckner S, Khanh TQ, Winkler H. Visual assessment of light source color quality. *Color Research and Application* 2013; 38: 4–13.

26 Jost-Boissard S, Fontoynont M, Blanc-Gonnet J. Colour rendering of LED sources: visual experiment on difference, fidelity and preference. *CIE Conference on Light and lLghting with Special Emphasis on LEDs and Solid State Lighting*. Budapest, Hungary, May 27–29.

27 Boust C, Brettel H, Viénot F, Berche S, Alquié G. Color enhancement of digital images by experts and preference judgments by observers. *Journal of Imaging Science and Technology* 2006; 50: 1–11.

28 Yendrikhovskij SN, Blommaert FJJ, Ridder H. Color reproduction and the naturalness constraint. *Color Research and Application* 1999; 24: 52–67.

29 Commission Internationale de l'Eclairage. *Colorimetry*. CIE Publication 15: 2004. Vienna: CIE, 2004.

30 Luo MR, Cui G, Li C. Uniform colour spaces based on CIECAM02 colour appearance model. *Color Research and Application* 2006; 31: 320–330.

31 Commission Internationale de l'Eclairage. *Fundamental Chromaticity Diagram with Physiological Axes – Part 1*. Vienna: CIE, 2006.

32 Csuti P, Schanda J. A better description of metameric experience of LED clusters. *Light and Lighting Conference with Special Emphasis on LEDs and Solid State Lighting*. Budapest, Hungary May 27–29, 2009.

33 Judd D. A flattery index for artificial illuminants. *Illuminating Engineering* 1967; 62: 593–598.

34 Thornton WA. A validation of the color-preference index. *Journal of the Illuminating Engineering Society* 1974; 4: 48–52.

35 Fotios S. *The perception of light sources of different colour properties*. PhD Thesis. Manchester UK: UMIST,1997.

36 Rea MS, Deng L, Wolsey R. *Full-Spectrum Light Sources*. Troy, NY: Rensselaer Polytechnic Institute, 2003 (Revised March 2005).

37 Rea MS, Deng L, Wolsey R. *Light Sources and Color*. Troy, NY: Rensselaer Polytechnic Institute, 2004.

38 Yano T, Hashimoto K. Preference for Japanese complexion color under illumination. *Color Research and Application* 1997; 22: 269–274.

39 Davis W, Ohno Y. Development of a color quality scale. *Sixth International LRO Lighting Research Symposium*, Orlando, FL, February 5–8: 2006.

40 Yaguchi H, Mizokami Y. Categorical colour rendering index based on the CIECAM02. *12th Congress of the International AIC Colour Association*. Newcastle Upon Tyne, UK, July 8–12: 2003.

41 Thornton WA. Color-discrimination index. *Journal of the Optical Society of America* 1972; 62: 191–194.

42 Howell D. *Statistical Methods for Psychology*. Belmont, CA: Duxbury Press, 1997.

43 Garcia PA, Huertas R, Melgosa M, Cui G. Measurement of the relationship between perceived and computed color differences. *Journal of the Optical Society of America A* 2007; 24: 1823–1829.

44 Sándor N, Schanda J. Visual colour rendering based on colour difference evaluation. *Lighting Research and Technology* 2006; 38: 225–239.

45 Li C, Luo MR, Li C. Assessing colour rendering properties of daylight sources. Communication within TC1-69, 2009.

46 Thurstone LL. A law of comparative judgment. *Psychological Review* 1927; 34: 273–286.

47 Thurstone LL. The method of paired comparisions for social values. *The Journal of Abnormal and Social Psychology* 1927; 21: 385–400.

48 Wei M, Houser KW, Allen GR, Beers WW. Color preference under LEDs with diminished yellow emission. *Leukos* 2014; 10: 119–131.

49 Ohno Y. Spectral design consideration for white LED color rendering. *Optical Engineering* 2005; 44: 111302.

50 van der Burgt P, van Kemenade J. About color rendition of light sources: The balance between simplicity and accuracy. *Color Research and Application* 2010; 35: 85–93.