Report on indoor test and assessment of studied CQIs metrics

EMRP-ENG05-D3.2.1

Version 1.0

D. Renoux
Laboratoire National de Métrologie et d'Essais, France

D. Sabol
Slovenský Metrologický Ústav, Slovakia

A report of the EMRP Joint Research Project

Metrology for Solid State Lightning

www.m4ssl.npl.co.uk

Title	:	Indoor test and assessment of studied CQIs metrics
Reference Version	: :	EMRP-ENG05-D3.1.2 1.0
Date	:	30-11-2012
Dissemination level	:	PUBLIC
Author(s)	:	D. Renoux Laboratoire National de Métrologie et d'Essais, France D. Sabol Slovenský Metrologický Ústav, Slovakia
Keywords	:	Colour rendition metric, field trial, LED lighting
Abstract	:	This report is the second deliverable of JRP ENG05 work package 3. The report is broken down into three parts. The first part presents the subjective experiment conducted with nine lightings (LED-based and traditional). The second part presents the analysis of the subjective results. The third part presents the correlation with metrics predictions studied in the deliverable D3.1.1.
Contact	:	http://www.m4ssl.npl.co.uk/contact

About the EMRP
The European Metrology Research Programme (EMRP) is a metrology-focused European programme of coordinated R&D that facilitates closer integration of national research programmes. The EMRP is jointly supported by the European Commission and the participating countries within the European Association of National Metrology Institutes (EURAMET e.V.). The EMRP will ensure collaboration between National Measurement Institutes, reducing duplication and increasing impact. The overall goal of the EMRP is to accelerate innovation and competitiveness in Europe whilst continuing to provide essential support to underpin the quality of our lives. See http://www.emrponline.eu for more information.

About the EMRP JRP "Metrology for Solid State Lighting"
In the EMRP Joint Research Project (JRP) "Metrology for Solid State Lighting", the following partners cooperate to create a European infrastructure for the traceable measurement of solid state lighting: VSL (Coordinator), Aalto, CMI, CSIC, EJPD, INRIM, IPQ, LNE, MKEH, NPL, PTB, SMU, SP, Trescal, CCR, TU Ilmenau and Université Paul Sabatier. See http://www.m4ssl.npl.co.uk/ for more information.

The research leading to these results has received funding from the European Union on the basis of Decision No 912/2009/EC.

SUMMARY

This report is the second deliverable of JRP ENG05 work package 3. It describes the work achieved for the assessment of the studied of colour rendering metrics for indoor lighting, reviewed in the first deliverable D3.1.1.

A real living room has been built to carry out a subjective experiment in a common and real environment. The test room has been furnished and decorated with many coloured natural and artificial objects. A special lighting system has been designed to only change the spectrum of the illumination in the test room: it enables to uniformly light the test room and to quickly select one lighting out of 12. The experiment has been performed with 43 people and nine lighting sources, including 6 led-based lightings, broken down into two CCT domains centred around 2700 K and 5000 K.

The subjective scores and results of statistical analyses are presented. The averaged results over the first group of panellist (1-22) and the second group of panellists (23-43), the first sequence and second sequence of rating are presented graphically. Similar behaviours, subjective scores versus light sources, are observed, representing few changes in ranking the light sources for the preference attribute. The result of the PCA, over the rated attributes, is that all attributes co-vary in the same direction with the principal component explaining 66% of the total variability.

The correlation with the predictions of fifteen metrics have been computed using the Pearson and Spearman coefficients of correlation and considering different rounding levels of subjective and objective data. These correlations have been computed on the whole set of lightings and the five possible subsets with respect to the CCT (cold/warm) and the technology (all technology/LED). From these correlations, with the subjective scores, a set of best metrics in term of prediction of “preference” has been drawn and the metrics performance is illustrated and commented with the help of radar charts for all subsets of lighting sources. A final conclusion is given with a possible direction of research for an improved metric for colour rendition for interior lighting.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

X	<Definition>

LIST OF ABBREVIATIONS

EMRP	European Metrology Research Programme
NMI	National Measurement Institute
JRP	Joint Research Project
SSL	Solid State Lighting

1. Introduction

The first aim of this work is to get the global preference, for colour rendition, of observers for a representative set of lightings in a typical lighting environment. Along with preference other subjective attributes have been rated, such as the naturalness and the colour vividness, to study the relationships between visual rendering attributes. The second aim of this deliverable is to assess predictions of colour rendering metrics against subjective ratings. For task we have at our disposal of all the metrics, current and proposals, identified and implemented in the deliverable D3.1.1.

2. The subjective experiment

2.1 Principle of the experiment

So far, most of the subjective experiments, for colour rendering, that have been performed used a pair-wise comparison or reference-based method implemented by means of a single lighting booth with fast switching unit or a pair of lighting booths. The use of direct visual comparison or short-term visual memory comparison helps the observer to judge and produces more repeatable subjective judgements. The drawback of fast stimulus or double stimuli is the undetermined visual adaptation condition, and also the drawback of lighting booths is their very limited and specific sceneries, which are somehow abstract colour sceneries unrelated to any colour common environment or visual experience.

To overcome the mentioned drawbacks we set up an experiment in a real and common environment enabling a full immersion and adaptation of the observer with a system to change easily the lighting spectrum of the test room with a large set of lighting sources. Having a large set of lighting sources quickly operated enables to have (1) a training session with few lightings and (2) several sequences, at least two, of all the lightings in different orders for the same observer. All these possibilities were the necessary conditions to conduct a new visual experiment of subjective and absolute colour rendering preference without reference presented in the following sections.

2.2 Description of the subjective room

2.2.1. Construction, interior design and decoration

The goal of the subjective test room is to record the observer's feeling of colour rendition in a familiar environment having only the spectrum of the light source changing. Changing only the spectrum avoids the contribution of other parameters not related to colour rendition and that could modify the subjective scoring such as light distribution. The living room is one of the most familiar places and one of the most versatile places for colour experiment for interior lighting. Therefore we built a real living room with a special design of the lighting system diffusing the light sources from the ceiling.

The internal dimensions of the test room are: 4.5 m (W) x 4 m (D) x 2.85 (H). The test room comes with a double window, an entrance door and an artificial wood floor. The wall, ceiling and door and window frame are coated with white mat paint.

The living room is furnished with casual European furniture and decorated with typical objects as well. Many natural and artificial objects are present in the test room offering a large set of colours with different materials (textile, wool, print, plastic, ceramic, paint, glass, cardboard, ect).

We inserted also many naturals objects or material, providing memory colours for that experiment, beside the complexion tone of the subject:

- a bunch of 5 different plants for the foliage tone,
- a basket of fruit and vegetables : bananas, 4 types of apples, oranges, kiwis, green lemons, tomatoes, red onions, zucchinis, green peppers, carrots.
- wood and cane furniture and objects.

Figure 1: Views of the subjective room

2.2.2. The lighting system

The best location to uniformly light the room and the objects inside is the ceiling. A removable transparent and diffusing window (1,2 m x 1,2 m) was integrated in the ceiling and a mechanical system was built above the room to place 12 clusters of lamps/modules behind the diffusing window. The mechanical system to support the lamps comprises 3 frames: one fixed frame, located just above the diffuser, and two sliding frames that can be quickly brought above the diffuser and removed. Each cluster is a set of identical lamps, the light level of each cluster inside the room being adjusted by the number of lamps and a dimming system whenever it is applicable. Blocking and black panels are placed in such a way that the light from the 2 sliding frames, when removed, cannot enter the central diffusing window. That enables to warm up lamps on the sliding panels while the fixed frame illuminates the room.

The sliding frames are equipped with guided wheels mounted on ball bearings and actuated by steel cables running along one external wall with the help of pulleys, handles are attached to the cables for easy operation. The frame travel is about one meter and the frame is put into position in few seconds.

An external switching board and a potentiometer board control the power on/off of all lighting and the dimming levels of some lightings. The switching electronic board, controlling the state of the lamps or the electronic drivers of LED, is attached to an external wall of the test room. The electronic box, gathering all the external dimming control signals for power supplies having this functionality (FL tubes, drivers for RGBY LED), has been specifically realised, it enables to switch between two different current adjustments of the RBGY LED clusters. The voltage drivers for the NUV LED have an internal potentiometer to fine tuning the level. All the test room is powered with a very stable voltage regulator of 10 kVA.

As a result less than 10 seconds is needed to change the lighting source, pulling handles and using the switches of the external board. The fast operation allows to design an experiment which can present several times all the lighting sources to the observer and then allows to achieve a more robust experiment.

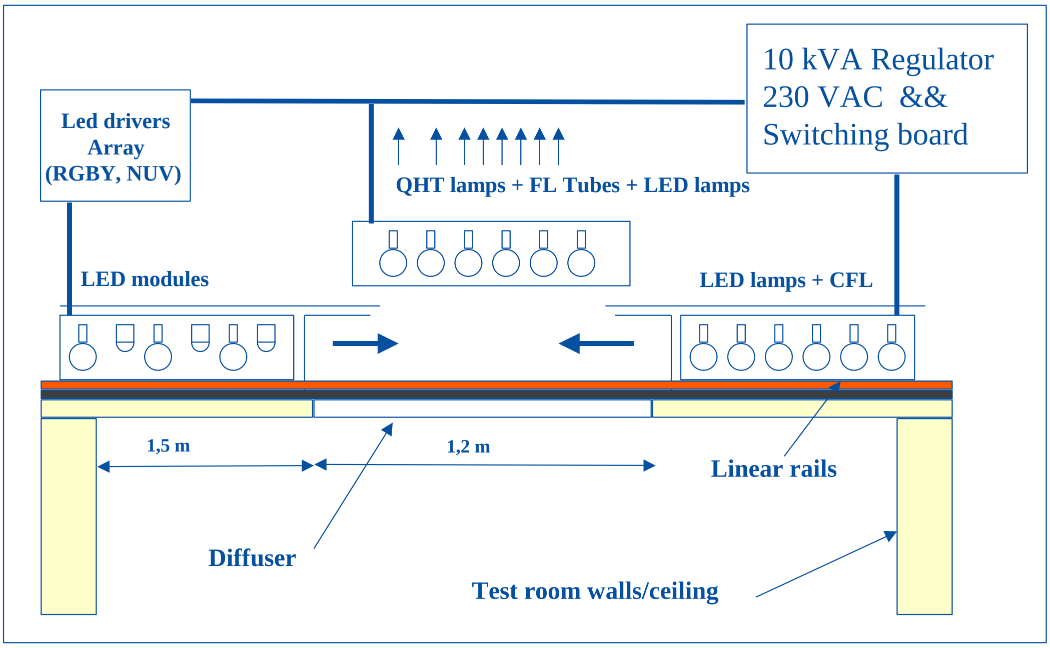

Figure 2: Sketch of the lamps supports and connections

2.3. Specification of the lighting sources

The choice of the lighting sources was governed by two constraints. Firstly to have all the today lighting technologies, including LED, because the development of a new colour rendering index must be established for any type of lighting. Secondly to have two different colour temperatures enabling us to study the metrics on the whole set of lamps or separately on the two lamp groups having the same colour temperature to discount the colour temperature parameter that could contribute to preference in a living room based experiment.

We selected nine different light sources broken down into two groups of colour temperature: around 5000 K and around 2700 K. There are one halogen source (warm), two fluorescent sources (one warm, one cold) and six LED sources: one RGB (cold), one RGBY (warm), two blue excitation with yellow phosphors (one warm, one cold), one blue excitation with yellow phosphors and a red emitter (warm), one purple (NUV) excitation with RGB-phosphors (cold). The LED-based spectra are represented on figure 3.

Lighting sources type and acronyms:

- FL: fluorescent tube – [commercial tubes]
- HAL: Quartz Tungsten Halogen – [commercial lamps]
- CFL: compact fluorescent – [commercial lamps]
- WW / CW: yellow phosphor and blue LED (warm white / cold white) - [commercial lamps]
- WR: yellow phosphor and blue LED with a red LED - [commercial lamps]
- NUV: RGB phosphors – NUV LED – [integrated CMS LED on MCPCB/heat sink]
- RGB / RGBY: LED clusters – [mounted LED OSTAR (Osram) on heat sink with fan]

	CCT (K)	CIE Ra	CQS Qa	D=0,9 MCRI	CRICAM UCS	RCRI
FL 5000K	4745	93.7	96.5	92.3	94.3	100
LED NUV 5000K	5024	98.1	99.1	90.7	98.51	100
LED CW 5000K	5481	70.68	71.3	75.7	71.02	56.1
LED RGB 5000K	5293	35.58	62.9	94.5	49.72	56.1
LED WR 2700K	2906	88.56	90.5	91.2	86.79	98
CFL 2700K	2708	82	75.8	77.9	75.97	74.4
RGBY 2700K	2781	76.2	79.1	90	80.27	80.9
HAL 2700K	2739	99.7	96.9	89.4	99.02	100
LED WW 2700	2624	82.78	79.4	85.2	78.82	74.4

Table 1: Specification of the lighting sources

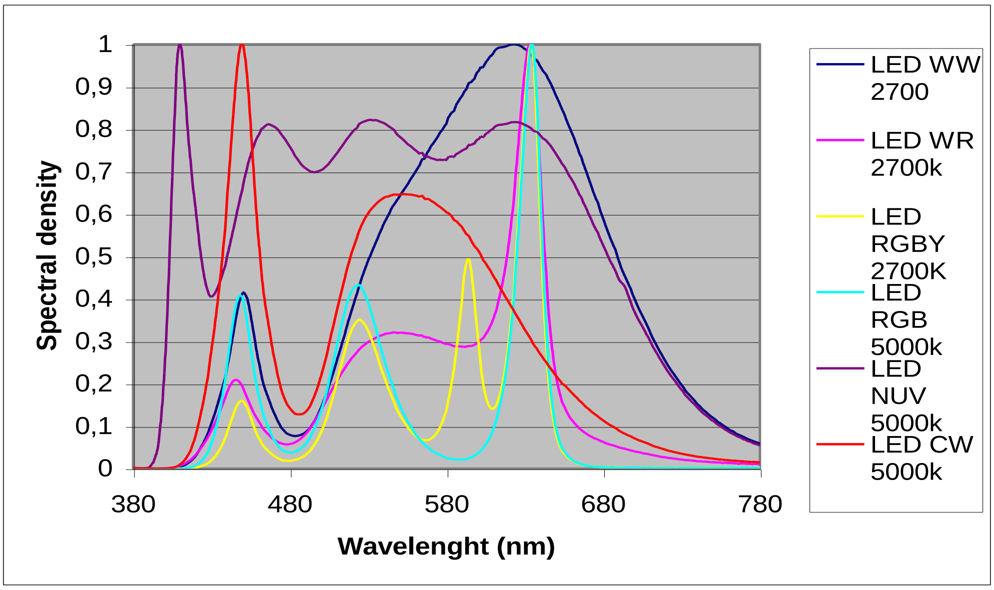

Figure 3: SPDs of LED used in the test room

Each solution was populated with lamps/modules to product the same illuminance level inside the room: 345 lux (+/- 9%) at one metre height at the centre of the room. The lighting sources were chosen and characterized with respect to CCT, CRI Ra, CQS, MCRI, CRICAM UCS, RCRI (see 3.1 for references). The final measurements were performed in-situ with two different set-ups: the first working with a white reflectance diffuser (spectralon) and a luminance spectro-colorimeter and the second working with a cosine corrected head coupled, by mean of a optical fibre, to a different spectro-radiometer. The results are given in the table 1 and show significant differences and ranges of metrics predictions for a valuable experiment.

2700 K :	**HAL**	**LED WW**	**CFL**	**LED WR**	**RGBY**

5000 K:	**FL**	**LED NUV**	**LED CW**	**RGB**

Figure 4: Pictures of the basket of fruits and vegetables under the 9 lightings

2.4. Panellists data and procedure

43 observers volunteered for the experiment, 14 females and 29 males aged from 20 to 62 years. They all successfully passed the 15 Farnsworth Munsell desaturated vision test under a daylight source.

At the beginning of the experiment, observers are instructed of the purpose of the test and of the different phases of the test. Then they are given an explanation of the questionnaire. Then the observers are trained with four lighting sources to build up their own scale, the selected training sources are FL 5000K, HAL 3000K, LED WW 3000K, LED RGB 5000 K.

Finally the observers are asked to rate the lighting sources presented in two sequences. Each sequence presents successively the 9 lighting sources with a different predefined order. The orders are selected among of a set of eight predefined orders. The eight orders were arranged in such way that all lightings belonging to the same group of CCT, are globally visualised by the observers with a different previous lighting of the same group of CCT (see table 2).

The scale is a 5-point quality scale; from 1 (excellent) to 5 (very bad), 3 being the neutral point of the scale. Eight rendering attributes have to be rated. The rendering attributes are listed hereafter, followed by a proposed definition or mean that are told to observers, during the presentation, to explain what they have to judge and help them how to do it :

Attributes to be rated :

1. overall preference - *according to observer's own criterion,*
2. fidelity of colours - *how much colour of objects match their usual colour,*
3. quality of vividness - *how much observer likes the vividness,*
4. overall naturalness - *global perception of naturalness in the room,*
5. naturalness of plants,
6. naturalness of fruits/vegetables,
7. naturalness of complexion – *observer complexion with the help of a mirror for face*
8. rendering quality of the colour chart - *colour balance, saturation, discrimination.*

Test procedure

- I - Oral presentation to observer, reading a written text.
- II - Vision test: 15 Farnsworth Munsell desaturated panel.
- III - Training section with 4 different light sources (2 times).
- IV - Rating of 2 sequences of the 9 light sources (~1hour)

	order A	order B	order C
1	HAL 2700K	RGBY 2700K	CFL 2700K
2	CFL 2700K	CFL 2700K	LED WR 2700K
3	LED WW 2700	LED WW 2700	LED WW 2700
4	LED WR 2700K	HAL 2700K	RGBY 2700K
5	RGBY 2700K	LED WR 2700K	HAL 2700K
6	FL 5000K	LED CW 5000K	LED RGB 5000K
7	LED NUV 5000K	FL 5000K	LED NUV 5000K
8	LED RGB 5000K	LED RGB 5000K	LED CW 5000K
9	LED CW 5000K	LED NUV 5000K	FL 5000K

Table 2: Examples of light sources sequence order (A, B, C)

2.5. Subjective results and statistical analyses

2.5.1. Averaged results

The averaged results of the rated attributes are presented in the figure 5 hereafter. The cold lights are on the left and the warm lights are on the right to better visualize the global rating against the CCT. One can notice that all the attributes co-vary in the same way except the attribute “quality of vividness”.

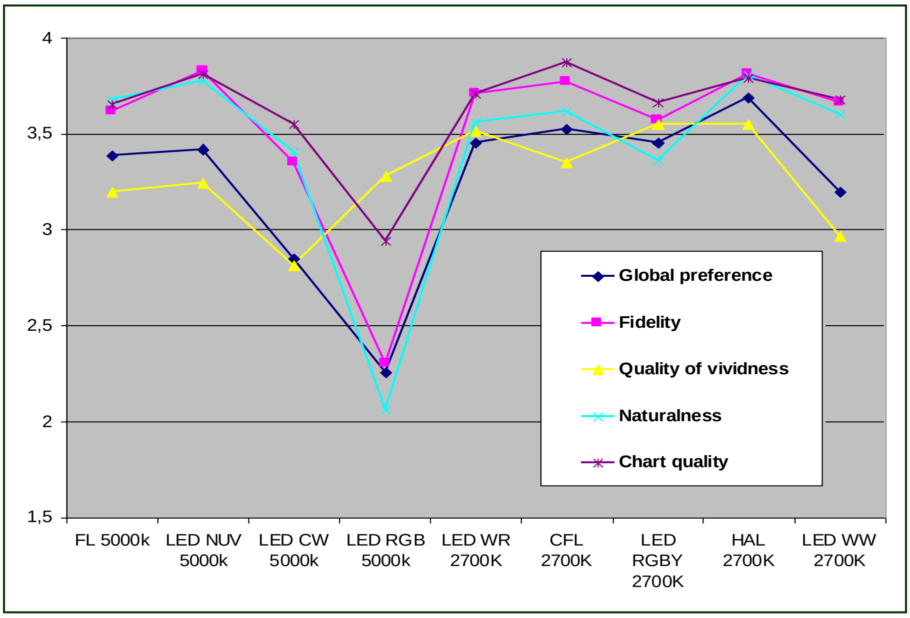

Figure 5: Averaged subjective quality attributes

The subjective scores of preference exhibit a same overall behaviour for the first group of panellists (n°1 to 22) and for the second group of panellists (n°23 to 43) with very small ranking differences (see figure 6) for warm sources, the panellist number is the running order. The subjective scores of preference, averaged over the 43 panellists, exhibit the same behaviour between the first sequence and the second sequence with small differences between the warm sources and then different ranking on the three closest warm sources (see figure 7). For that subjective experiment, excluding RGB, we obtained large standard deviations of the subjective attributes, between 0.60 and 0.98, with mean values varying from 1.8 to 3.9. But as many criteria can enter the observer's judgement in a living room and given that observer preference can vary from one individual to another one we cannot compare to other subjective experiment in term of variability. Statistical developments are underway to better analyse the significance of the mean subjective results.

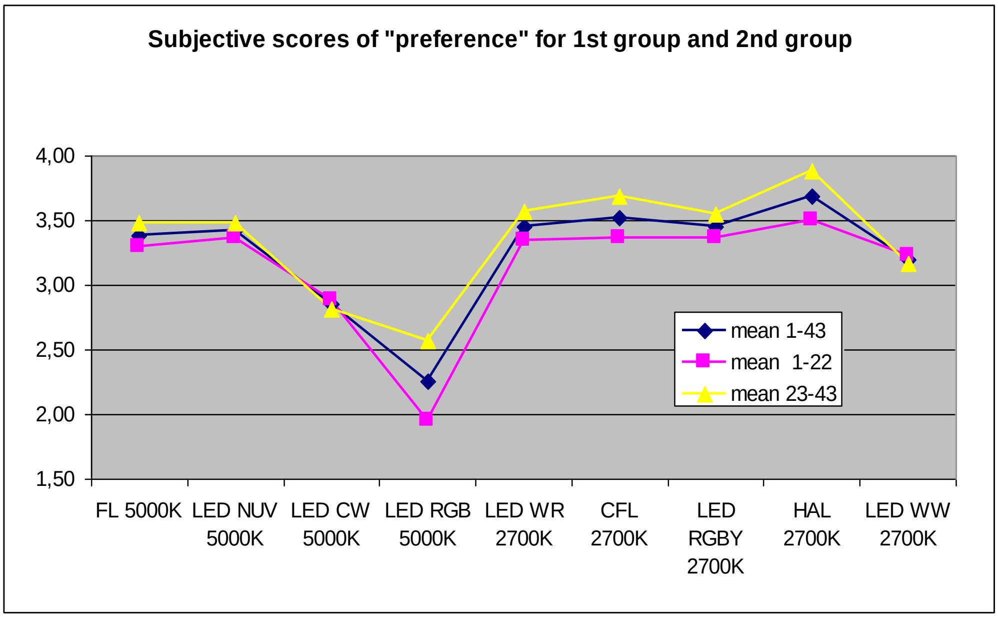

Figure 6: Scores of "preference" for the 1st group and the 2nd group

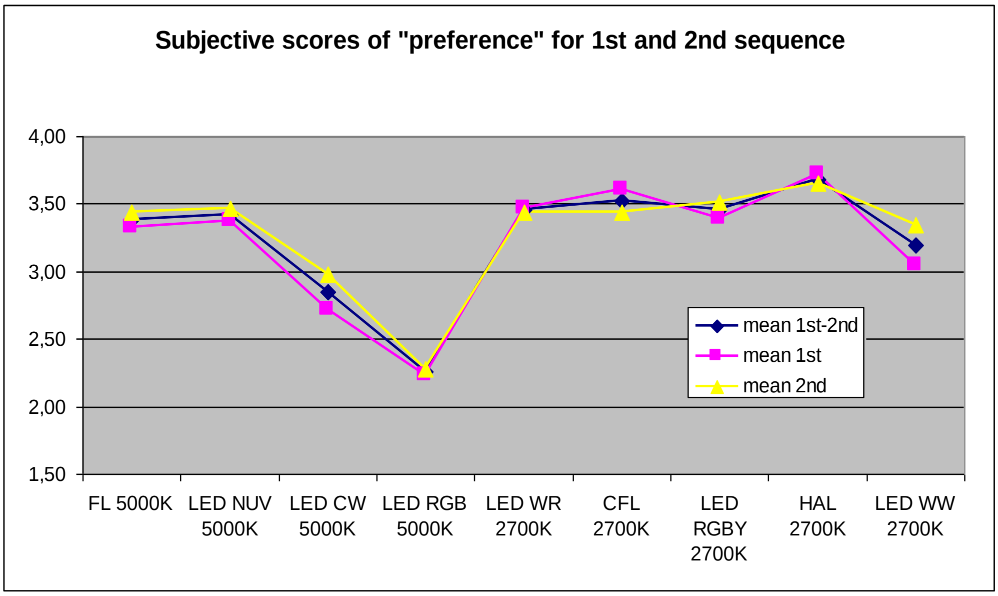

Figure 7: Scores of "preference" for the 1st sequence and the 2nd sequence

2.5.2. Principal components analysis (PCA)

The statistical analyses of the subjective scores has been performed with the help of the software "STATISTICA 6.0". The inputs are the scores of the 8 attributes (variables) for the 9 lighting sources over the 43 individuals. So each variable is represented by 387 scores. For the same individual each score is the average of the score of the first sequence and the score of the second sequence for the same lighting source.

The main result of PCA can be seen on figure 8, the first component explains 66 % of the total variability and the two first components explain 75% of the total variability.

<u>Variables (rated subjective attributes) :</u>

1- overall preference,
2- fidelity of colours,
3- quality of vividness,
4- overall naturalness,
5- naturalness of plants,
6- naturalness of fruits/vegetables,
7- naturalness of complexion,
8- rendering quality of the colour chart.

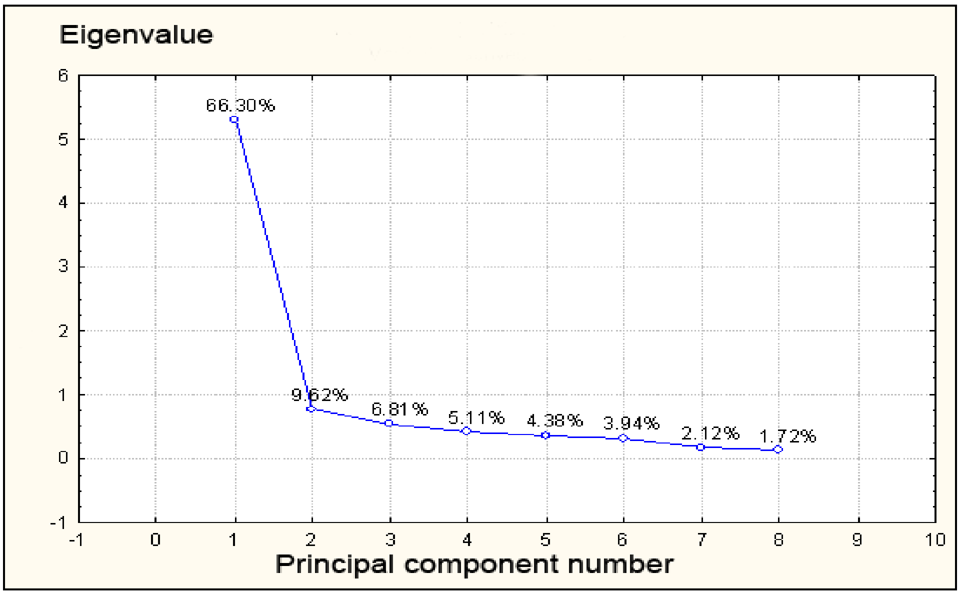

Figure 8: Scree plot of the variability of all lighting sources given by PCA

This analysis was repeated without the RGB 5000K light source, which gives the smallest scores with the greatest dispersions and is the more atypical having a strong chroma enhancement with the highest colour temperature. The results are a bit different but do not change the main figures: correlation for vividness is greater and the correlation of other attributes is slightly smaller.

The correlation table 3 shows that all variables (attributes) co-vary, in mean value, in the same direction. The greatest correlations are obtained between "Overall naturalness" and the following attributes: "colour fidelity" , "naturalness of fruits and vegetables", "naturalness of plants".

variable	overall preference	fidelity of colours	quality of vividness	overall naturalness	naturalness of plants	Naturalness of fruits/vegetables	naturalness of complexion	quality of the colour chart
overall preference	1,00000	0,71041	0,54934	0,70728	0,63650	0,61846	0,57321	0,54586
fidelity of colours	0,71041	1,00000	0,47767	0,83073	0,81088	0,68825	0,63577	0,60878
quality of vividness	0,54938	0,47767	1,00000	0,44224	0,35654	0,40689	0,42684	0,39550
overall naturalness	0,70728	0,83073	0,44224	1,00000	0,85041	0,75473	0,64740	0,58249
naturalness of plants,	0,63650	0,81088	0,35665	0,85040	1,00000	0,71736	0,66019	0,56121
naturalness of fruits/vegetables,	0,61847	0,68825	0,40689	0,75473	0,71736	1,00000	0,60745	0,57146
naturalness of complexion	0,57321	0,63000	0,42684	0,64740	0,66019	0,60745	1,00000	0,58805
quality of the colour chart	0,54585	0,60878	0,39550	0,58249	0,56121	0,57146	0,58805	1,00000

Table 3: correlation of subjective attributes

Other statistical analyses

The specificity of the test procedure is the two sequences rating. We analysed the distributions of the difference of all possible pairs of scores (1st sequence - 2nd sequence) for all attributes and lighting sources. There is no difference, statistically significant, between the first sequence and the second sequence for all attributes except for the attribute "quality of vividness". We used the Wilcoxon signed-rank test at a level of confidence of 95%.

Mainly other results are available from the statistical analysis and are not yet exploited like the homogeneity of variances, with and without the RGB 5000K light source which presents the greatest variances.

3. Correlation of subjective scores with metrics predictions

3.1. Metrics predictions

The metrics predictions, i.e. the numerical results of the implemented metrics have been computed using a program written in C++. The predictions for the lighting sources of the test room are presented in the table 4 hereafter with the subjective ratings of the global preference. The detailed index/metric descriptions are given in the deliverable D3.1.1 and identified bellow:

- CRI Ra : current CIE CRI 13.3 general index [1],
- CQS Qa : proposal Colour Quality Scale general index [4],
- MCRI Sa : Memory Colour rendering Index general index [10] ,
- CRI-CAM02UCS : updated CRI Ra with the CAM02-UCS [3],
- CRI Ra96 : CIE proposal to update the CRI Ra [2],
- RCRI : Ranking colour rendering Index [11],
- CCRI : Colour Category Rendering Index [8],
- HRI : Harmony Rendering Index [9],
- FCI : Feeling of Contrast Index [6],
- GAI : Gamut Area Index [5],
- CQS Qg : Gamut Area Scale [4],
- CFI stat : Colour Fidelity Index – statistical [7],
- Combination “X“ + “Y”: combination using the mean value of the indices “X” and “Y”.

Some indices, like the gamut indices, are intended to be combined with a fidelity or main index as the CRI Ra or CQS Qa. No specific rule being given for the combination we use the average values of indices for that analysis.

More combinations and many alterations as well of original metrics/indices, are possible using the computing program. The set of investigated metrics/indices covers the proposals, recommended by CIE for a new index of colour rendering, we added a statistical method (CFI) updated from the original implementation using the CAM02-UCS colour space.

	FL 5000K	LED NUV 5000K	LED CW 5000K	LED RGB 5000K	LED WR 2700K	LED WW 2700	RGBY 2700K	HAL 2700K	CFL 2700K
CRI Ra	93,70	98,10	70,68	35,58	88,56	82,78	76,20	99,70	82,00
CQS Qa	96,45	99,10	71,33	62,89	90,48	79,40	79,06	96,91	75,78
MCRI Sa	92,71	91,90	75,87	95,36	94,38	88,36	95,22	94,32	82,90
CRI-CAMUCS	94,30	98,51	71,02	49,72	86,79	78,82	80,27	99,02	75,97
RCRI	100,00	100,00	56,09	56,09	98,00	74,40	80,90	100,00	74,40
CCRI	91,96	94,63	73,92	55,06	82,75	77,07	71,93	86,06	77,58
CRI Ra96	95,21	98,52	71,30	47,50	88,23	82,02	80,79	99,28	79,49
HRI	97,92	99,36	104,51	66,93	95,16	101,01	92,16	100,11	99,32
FCI	111,09	108,59	86,23	178,79	138,31	118,83	148,20	123,00	116,31
GAI	97,71	90,14	78,39	134,56	63,44	48,35	57,57	48,93	49,02
CQS Qg	103,41	100,84	88,54	139,64	109,28	97,29	110,69	97,36	98,20
CRI + GAI	95,71	94,12	74,54	85,07	76,00	65,57	66,89	74,32	65,51
CQS Qa + Qg	99,93	99,97	79,94	101,27	99,88	88,35	94,88	97,14	86,99
Ra96+ CCRI	93,59	96,58	72,61	51,28	85,49	79,55	76,36	92,67	78,54
CFI stat	94,25	100,00	16,71	11,11	52,72	33,57	24,90	100,00	30,42
Subj. Preference	***85,35***	***86,74***	***63,95***	***40,23***	***88,14***	***77,91***	***88,14***	***97,44***	***90,93***

Table 4: values of metrics predictions and subjective preference ratings

3.2. Correlations of metrics with subjective rating

3.2.1. Correlation tables

We computed two correlation coefficients, the Pearson correlation coefficient and the Spearman correlation coefficient to analyse the relationship between the metrics predictions and the subjective scores of the rated preference. The Pearson coefficient represents the strength of the linear relationship between two variables of a data set, and the Spearman coefficient represent the strength of the monotonic relationship between two variables of a data set and is computed as the Pearson correlation on the ranks of the data.

Depending on the data distribution these coefficients deliver different information that should be carefully examined and interpreted. We considered the ability of the metrics to rank correctly the lighting sources as the main criterion to assess the metrics; this criterion is better represented by the Spearman coefficient.

Other point to be considered is when there are very close values of a variable with respect to the psychophysical scale: very small variations can yield to very different assigned ranks and thus yield to a very different Spearman correlation coefficient for almost equal subjective scores or psychometric prediction. On other hand Pearson coefficient is un-sensitive to that effect but has no monotone relationship with the correlation of ranks.

To take into account this effect we computed and considered the correlation coefficients for different levels of rounding (0.1 %; 0.5%, 1.0%, 2.0%) of subjective data and objective data (metrics predictions/results).

To get the most information as possible we performed the correlation analysis on the whole set of lighting sources and several subsets considering the technology (all type / only LED) and the colour temperature domain (warm/cold). We are well aware that the number of samples is too low to correctly state on some subsets, but the goal here is to see the trend.

To summarize some results of table 5 and 6 the fidelity metrics are well correlated to “cold sources”, the current metric CRI Ra fails for “warm sources” and new fidelity metrics, including CQS Qa, perform better for “warm source”. The gamut metrics fails for “cold sources” and “warm sources” but correlate very well for “warm LED”. The MCRI also correlates very well for “warm LED” but fails for all other source subsets. Other metrics like CCRI or HRI, based on specific properties of rendering, produce interesting results for some cases but globally do not perform better than the current CRI. Statistical methods, like CFI based on fidelity, can also give interesting results and better results than the current index for some cases.

Then we computed the mean of correlation coefficients over all subsets for each metrics and the corresponding ranking of the metrics. We considered this ranking to sort the best metrics. Those means comprise 2/3 of “warm” and “cold” lighting source subsets reducing a colour temperature effect that could account for the global preference. The detailed results are given hereafter in the tables 5,6 for data rounded at 1%.

We present in table 7 and table 8 the averaged results of metrics for all subsets of lighting sources for 0.1%, 0.5%, 1.0 % and 2.0 % data rounding and the related ranking with respect to that mean. We draw from these tables the best metrics for our experiment in term of correlation coefficients.

Metrics	all sources	cold sources	warm sources	LED sources	cold LED	warm LED	mean	rank
CRI Ra	0.912	0.998	0.576	0.923	0.998	-0.044	0.727	6
CQS Qa	0.762	0.961	0.581	0.839	0.948	0.500	0.765	4
MCRI	0.130	0.016	0.235	0.162	-0.159	0.991	0.229	12
CRI-CAMUCS	0.862	0.996	0.634	0.921	0.996	0.596	0.834	1
RCRI	0.783	0.894	0.602	0.842	0.860	0.725	0.784	3
CCRI	0.776	0.999	0.531	0.812	0.999	0.052	0.695	7
CRI Ra96	0.892	0.998	0.612	0.937	0.998	0.381	0.803	2
HRI	0.699	0.782	-0.075	0.682	0.791	-0.945	0.322	9
FCI	-0.413	-0.712	0.020	-0.379	-0.731	0.941	-0.212	13
GAI	-0.761	-0.690	0.006	-0.797	-0.757	0.945	-0.342	15
CQS Qg	-0.628	-0.715	-0.041	-0.577	-0.740	0.991	-0.285	14
CRI + GAI	-0.191	0.613	0.486	-0.222	0.463	0.577	0.287	10
CQS Qa + Qg	0.056	0.106	0.430	0.067	-0.054	0.910	0.252	11
Ra96+ CCRI	0.851	0.999	0.594	0.885	0.999	0.064	0.732	5
CFI stat	0.609	0.916	0.654	0.618	0.889	0.202	0.648	8

Table 5: Pearson correlation for data rounded at 1%

Metrics	all sources	cold sources	warm sources	LED sources	cold LED	warm LED	mean	rank
CRI Ra	0.577	1.000	0.359	0.667	1.000	0.000	0.601	4
CQS Qa	0.517	1.000	0.289	0.750	1.000	0.500	0.676	3
MCRI	0.110	-0.400	-0.026	0.250	-0.500	0.866	0.050	10
CRICAM UCS	0.580	1.000	0.359	0.812	1.000	0.866	0.769	1
RCRI	0.575	0.894	0.526	0.794	0.866	0.866	0.754	2
CCRI	0.435	1.000	0.667	0.464	1.000	0.000	0.594	6
CRI Ra96	0.521	1.000	0.205	0.667	1.000	0.000	0.565	8
HRI	0.017	0.400	-0.051	-0.116	0.500	-0.866	-0.019	13
FCI	0.109	-0.400	-0.154	0.058	-0.500	0.866	-0.004	11
GAI	-0.555	-0.400	0.105	-0.522	-0.500	0.866	-0.168	15
CQS Qg	-0.092	-0.400	-0.158	0.116	-0.500	0.866	-0.028	14
CRI + GAI	-0.353	0.600	0.263	-0.145	0.500	0.866	0.289	9
CQS Qa + Qg	-0.162	-0.316	0.103	-0.015	-0.500	0.866	-0.004	12
Ra96+ CCRI	0.427	1.000	0.359	0.667	1.000	0.000	0.575	7
CFI stat	0.546	1.000	0.359	0.667	1.000	0.000	0.595	5

Table 6: Spearman correlation for data rounded at 1%

Average Pearson correlation coefficients						
rounding	0.1 %		1 %		2 %	
metric	mean	rank	mean	rank	mean	rank
CRI Ra	0.732	6	0.727	6	0.750	5
CQS Qa	0.761	4	0.765	4	0.776	4
MCRI D=1	0.219	12	0.229	12	0.195	12
CRICAM UCS	0.847	1	0.834	1	0.868	1
RCRI	0.785	3	0.784	3	0.784	3
CCRI	0.689	7	0.695	7	0.663	7
CRI Ra96	0.805	2	0.803	2	0.799	2
HRI	0.333	9	0.322	9	0.336	9
FCI	-0.208	13	-0.212	13	-0.198	13
GAI	-0.351	15	-0.342	15	-0.351	15
CQS Qg	-0.276	14	-0.285	14	-0.280	14
CRI + GAI	0.291	10	0.287	10	0.246	10
CQS Qa + Qg	0.244	11	0.252	11	0.223	11
Ra96+ CCRI	0.756	5	0.732	5	0.741	6
CFI stat	0.651	8	0.648	8	0.642	8

Table 7: Pearson correlation coefficients

Average Spearman correlation coefficients						
rounding	0.1 %		0.5 %		1.0 %	
metric	mean	rank	mean	rank	mean	rank
CRI Ra	0.601	3	0.601	3	0.601	4
CQS Qa	0.560	8	0.560	8	0.676	3
MCRI D=1	-0.005	13	0.020	10	0.050	10
CRICAM UCS	0.775	1	0.775	1	0.769	1
RCRI	0.754	2	0.754	2	0.754	2
CCRI	0.594	5	0.594	5	0.594	6
CRI Ra96	0.571	7	0.571	7	0.565	8
HRI	-0.024	14	-0.019	13	-0.019	13
FCI	-0.004	12	-0.004	11	-0.004	11
GAI	-0.178	15	-0.168	15	-0.168	15
CQS Qg	0.015	10	-0.028	14	-0.028	14
CRI + GAI	0.255	9	0.298	9	0.289	9
CQS Qa + Qg	0.008	11	-0.004	12	-0.004	12
Ra96+ CCRI	0.575	6	0.575	6	0.575	7
CFI stat	0.595	4	0.595	4	0.59534127	5

Table 8: Spearman correlation coefficients

Rank = 1 to 5 for all cases	Rank change at 0.5 %
Rank = 1 to 5 for some cases	Rank change at 1.0 %
	Rank change at 2.0 %

3.2.2. Charts of the best metrics

We notice from table 7 and 8 that rounding the data does not change the Pearson coefficients, but change significantly the Spearman coefficients; linear behaviour is little affected while ranks of close scores or predictions are significantly changed.

We can draw, from table 7 and 8, with the corresponding colour codes explained above, the six first metrics considering the 0.1% and 1.0 % rounding for the Pearson coefficients and the 0.1%, 0.5% and 1% rounding for the Spearman coefficients. From the mean value of the correlation coefficients the following performance order is obtained: CRI-CAMUCS (1^{st}), RCRI (2^{nd}), CRI Ra96 (3^{rd}), current CRI Ra (4^{th}), CQS Qa (5^{th}), and "CRI Ra96+ CCRI" (6^{th}). CFI Stat (7^{th}).

We represent the best metrics with a "radar chart" with the different subsets as axes for the Pearson coefficients (figure 9) and Spearman coefficients (figure 10) obtained with data rounded at 0,5 % to visualize the difference with the standard CRI Ra 13.3.The new metrics, in plain area, can be compared to the current CRI Ra area bounded by a red solid line.

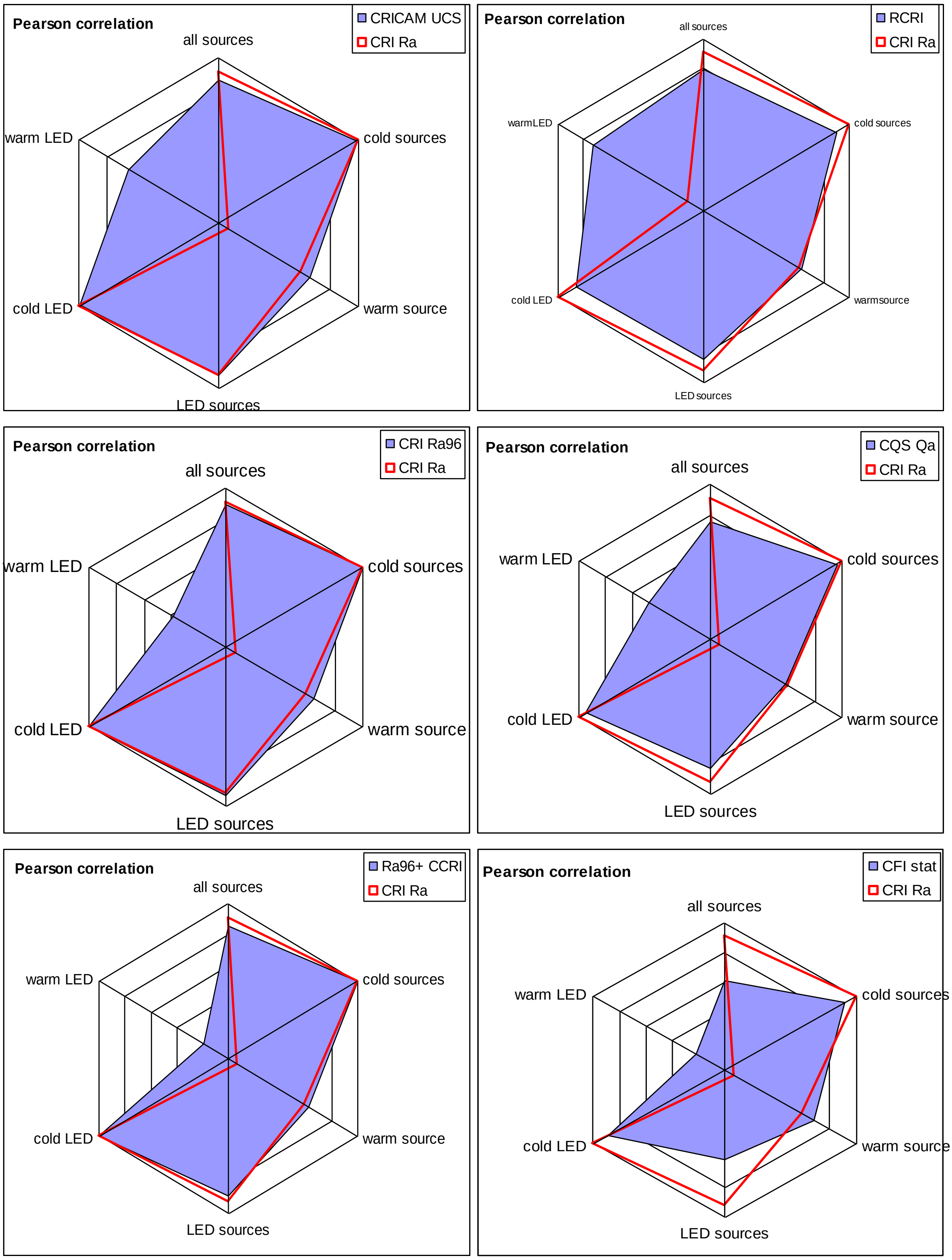

Figure 9: Radar charts of Pearson correlation coefficients with data rounded at 0,5%

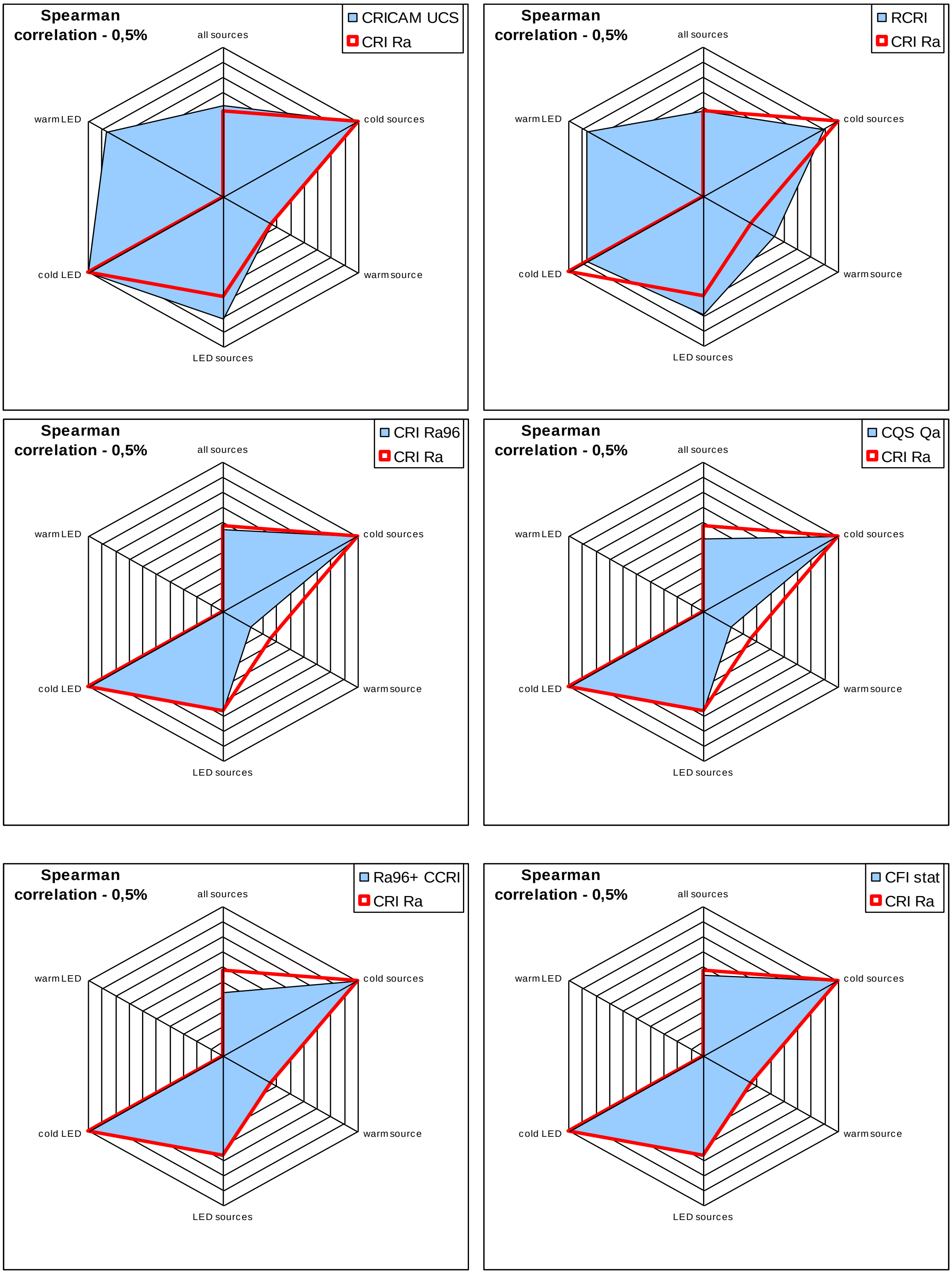

Figure 10: Radar charts of Spearman correlation coefficients with data rounded at 0,5%

Comments on Pearson correlation charts

We can notice that the new metrics or proposals obtain greater correlations for the “warm LED” and equal or slightly greater correlations for the “warm sources” than the current CRI Ra. But correlations with “all sources”, “cold sources” and “cold LED” are globally smaller.

Comments on Spearman correlation charts

We can observe that the CRI-CAMUCS is the only metric which better ranks than the current CRI Ra for all cases, and that CCRI better ranks than the current CRI except for the subsets “cold sources” and “cold LED”. Other metrics do not rank better than the current CRI Ra for all subsets.

4. Conclusion

The construction of the test room with its lighting system represents an important work and is today a unique facility to test lighting source spectra effect in a real interior environment. The test room enabled to carry out a large subjective experiment with nine traditional and LED-based lighting sources to get the observer preference without reference lighting.

The subjective experiment did not yield to specific relationship between the subjective attributes : “global preference” , “fidelity of colours”, “global naturalness”, “naturalness fruit/vegetables” and “naturalness of complexion, “quality of vividness”, and “colour chart quality”. All the attributes appear to co-vary in the same way.

We found a big difference of correlation (metrics predictions /subjective scores) between “cold lights sources” and t “warm light sources” of the experiment. Globally the assessed proposals, relying on more recent colorimetric spaces and colour shift computations, provide better predictions against subjective preference scores, especially for “warm LED” where the current CRI Ra completely fails. We found very high correlation for the current CRI Ra with “cold sources” and for the new fidelity metrics as well, namely CRI-CAMUCS, CRI Ra96, and CQS Qa. The gamut metrics provide very high correlation with the “warm LED”, and also the MCRI, which seems driven by the chroma of the lighting sources.

Globally fidelity metrics, as well as CQS, perform well on “cold sources” and perform moderately on “warm sources” and gamut metrics perform well on “warm LED” but cannot improve each other by simple combination.

One possible direction of future research, suggested by the results, is the combination of an improved fidelity metric with an improved gamut metric. Improvement of a fidelity metric can be made from the CRI-CAMUCS using a different and broader set of TCS (Test Colour Samples). The gamut metric can be improved not considering the area of the bounding TCS in a chromatic space but the gain/drop of chroma for a set of TCS uniformly distributed in the chromatic space. The combination may be improved using a more complex combination than a linear combination and also modulated/weighted by the chromaticity of the light source. Other and deeper direction should be considered but are out of reach for this study.

The subjective assessment of some metrics seems quite different of those obtained by other experiments, more experiments are needed to consolidate and complement this study. LNE will continue to work on colour rendering metrics and plan to conduct another experiment using the test room, within the two next years, on a set of lighting sources with CCT ranged in a limited domain, 2700 K to 3500 K, and focused on preference with an improved method using pair-wise comparison or more rating repetition.

References

More references can be found in the deliverable D3.1.1, the following references covers only the proposals of colour rendering metric assessed in this study.

[1] CIE 13.3: 1995, Publ. No. 109-1995 - Method of measuring and specifying colour rendering properties of light sources.

[2] 1999: CIE Collection 1999 research note: colour rendering, TC 1-33 closing remarks

[3] C.Li et Al. "Assessing Colour Rendering Properties of Daylight Sources part II: a new colour rendering index: CRI-CAM02UCS – University of Leeds, 2011

[4] W. Davis, Y. Ohno, 2010, "Color quality scale," Opt. Eng. **49**(3), 033602, 16 pages.

[5] M.S. Rea, J.P. Freyssinier-Nova, 2007, "Color Rendering: a tale of two metrics", Col. Res. Appl. **33**, pp 192-202.

[6] K. Hashimoto, T. Yano, M. Shimizu, Y. Nayatani, "New method for specifying color-rendering properties of light sources based on feeling of contrast", Col. Res. and Appl. **32**, 361-371, (2007).

[7] A.Zukauskas et al., "Statistical Approach to colour quality of solid-state lamps", IEEE journal of selected topics in quantum electronics, vol n°6 (2009).

[8] H. Yaguchi, Y. Takahashi, S. Shioiri, "A proposal of color rendering index based on categorical color names," Int. Lightning Congress Istanbul, Vol. II, (2001)

[9] F. Szabó, P. Bodrogi, J. Schanda, "A colour harmony rendering index based on predictions of colour harmony impression," Lighting Res. Technol. **41**, pp 165-182, (2009).

[10] K. Smet,W.R. Ryckaert, M.R. Pointer, G. Deconinck, P. Hanselaer, "Colour Appearance rating of familiar real objects," ," Col. Res. Appl. **36**, pp 192-200, (2011).

[11] P. Bodrogi, S. Bruckner, T. Q. Khanh, "Ordinal scale based description of Colour Rendering," Col. Res. Appl. **36**, pp. 272-285, (2010).