A New Statistical Method for Ranking of Light Sources based on Subjective Points of View

Éva Orbán-Mihálykó[1], László Koltay[1], Ferenc Szabó[2], Péter Csuti[2], Renáta Kéri[2], János Schanda[2] †

[1] University of Pannonia, Faculty of Information Technology, Department of Mathematics, Egyetem u. 10, H-8200 Veszprém, Hungary, email: orbane@almos.uni-pannon.hu, koltay@almos.uni-pannon.hu

[2] University of Pannonia, Faculty of Information Technology, Department of Electrical Engineering and Information Systems, Egyetem u. 10, H-8200 Veszprém, Hungary, e-mail: szabo.ferenc@virt.uni-pannon.hu, csuti.peter@virt.uni-pannon.hu, keri.renata@virt.uni-pannon.hu, schanda.janos@virt.uni-pannon.hu

Abstract: A new method of ranking different light stimuli based on subjective points of view is presented. The objects are pairwise compared and ranked on the basis of the judgments of the observers. More than two options for judgment are allowed. Observers' perceptions are considered as random variables. The differences between the sensations are supposed to be independent, identically distributed random variables with Gauss distribution and equal standard deviations. On the basis of the judgments of the observers the expectations of the random variables are estimated by maximum likelihood method and the maximizations are carried out numerically. Computations were performed by the statistical program package R. The rank of the expectations is considered to be the rank of the observers' perception under the test light stimuli. Equalities of the expectations are tested by likelihood ratio tests. The results are compared to the results derived by the AHP method. Finally, some results in connection with the investigated questions are presented.

Keywords: color quality; LED; lights; ranking; pairwise comparison; statistical evaluation; Thurstone method; AHP

1   Introduction

The application area of Light Emitting Diodes (LEDs) in lighting technology has been growing significantly. LEDs are used nowadays not only as signaling lights but also as functional lighting in public lighting, home [1] and shop lighting [2], museum lighting [3], sport lighting and so on. At the beginning of the LED era, phosphor converted white LEDs had been used as a light source for general lighting. Nowadays, phosphor white LEDs are often supplemented by narrow

band color LEDs in order to tailor the light spectrum for different special applications. This results that one given light at a given correlated color temperature can be realized as an unlimited number of composition of phosphor white and color LEDs.

The human visual system has photoreceptors (rods and cones) at the retina, which have different sensitivities for optical radiation. The photopic (daylight) vision is realized by the cones, containing three different opsins that result in three different sensitivities. Figure 1 shows the different cone sensitivities along the visible wavelength domain (380-780 nm).
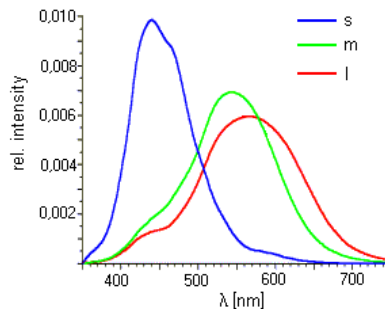

Figure 1
Spectral sensitivity of the human cone photoreceptors (**s**hort, **m**edium, **l**ong)

In order to evaluate the lighting quality of a light source under development, often, human visual experiments are used [4, 5]. During the previous decade, this kind of visual experiments were carried out by using lighting booths with miniature models of real objects or small objects (e.g. fruits, vegetables) which fit into the viewing booth. In order to investigate the lighting quality with total immersion of the observers, two full-scale laboratory rooms have been developed at the Virtual Environments and Imaging Technologies Research Laboratory at the University of Pannonia. One room simulates a living room, while the other simulates a kitchen environment with a dining area. Each room is equipped with 20 channel LED luminaires, where the spectrum of the illumination could be varied on demand.

Figure 2
Part of the living room and the kitchen environment

Figure 2 shows part of the kitchen environment (right) and the view of the living room (left).

Each LED luminaire contained 20 different types of LEDs: 17 channels of colored LEDs and three channels of white LEDs (phosphor converting ones) each with different correlated color temperature. Figure 3 shows the spectra of the different LEDs used. The channel with the shortest peak wavelength has its maximum emission at 414 nm (deep blue), the maximum emission wavelength of the deepest red LED is 691 nm. One channel has serial connected LEDs with the same color which can be set individually via computer control. The control method of the different LED channels is based on the DMX light control technique. One can set the current of each channel in 256 steps up to the present maximal drive current.
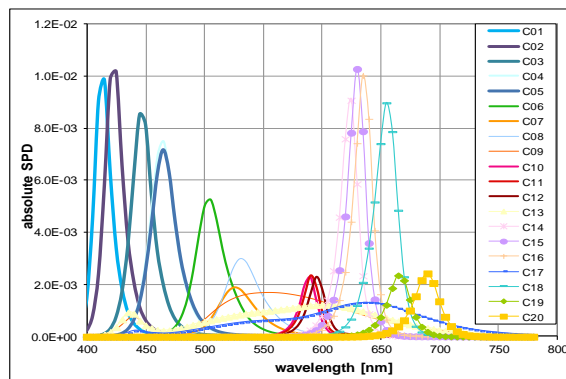

Figure 3
Spectra of different LEDs incorporated in the LED luminaires

By using LEDs with different colors, the spectral power distribution (SPD) of the light can be tailored to fit the requirements of lighting in a particular situation, or to help the human visual system in the actual situation. As the methodology of the human visual system and the process of human vision is very well described and standardized by publications of the International Commission on Illumination (CIE), general trends can be drawn. LED based light sources with different spectral power distribution cause significantly different visual appearance of object colors, even if the color (or correlated color temperature) of the output light is the same.

To perform and evaluate human visual experiments we take into consideration that the sensation of light is based on the observer's perception, therefore, it is always subjective. It depends on the personality of the observer, on the circumstances, moreover, the actual mood and state of mind of the observer and so on. The factors of the sensation can be detailed and examined but presumably all of them cannot be taken, totally into consideration [6, 7]. On the whole, the sensation of light can be regarded as an uncertain phenomenon. Moreover, it is very difficult to measure the sensations. Many questions arise: what is the border-line between equal and different senses, what is the unit, is it the same for all people? Are the differences multiplicative or additive? Is it reasonable to ask the observers to rank the sensations or rather decide between two sensations?

What can be expected from people, based on their perception, is to answer the question which light is preferred over another light. This can be done several times and independently of the opinions of other people. Then, by summarizing the results of the decisions on pairs, by many people, one may determine the rank of the objects. This approach requires a statistical evaluation method based on pairwise comparisons.

2   A Short Review of the Pairwise Comparison Methods

In statistics, two frequently applied methods can be found for ranking based on pairwise comparisons. One of them is AHP (Analytic Hierarchy Process by Saaty [8]). This method is really widespread and generally accepted in management and its possibility in decision making shows great variety [9].

In the AHP method the objects are pairwise compared, and the results of these comparisons are characterized by positive numbers. The larger the number is, the larger the difference is between the objects. The reverse pair is characterized by the reciprocal. The values are included in a matrix, the maximal eigenvalue of which is positive, its multiplicity equals 1 and the normalized eigenvector belonging to that eigenvalue has only positive coordinates [10]. The rank of the coordinates of the normalized eigenvector is considered to be the rank of the objects compared. The opinion of more than one person can be aggregated in a single matrix taking the geometric mean of the values belonging to the separate opinions. The main advantage of the method is that it can be executed easily, the rank can be determined unambiguously. Unfortunately, it does not provide information about palpability, that is about the measure of the difference in senses which is imperceptible and it is difficult to provide a test whether the weights of the objects can be considered equal or they are significantly different.

Another frequently applied pairwise comparison method is the Thurstone method [11]. Thurstone assumed that the differences of the senses are proportional to certain quantities. His idea can be interpreted as follows: the senses are random variables, therefore so are their differences. Expectation of the differences is the difference of the expectations; the dispersion of the differences can be expressed by the covariance matrix of the random variables characterizing the sense of the objects. Assumptions concerning the covariance matrix characterize the Thurstone method (I, II, ..., V). The most frequently used method is Thurstone V [12], which assumes normal (Gauss) distribution with independent latent random variables with equal dispersions. It provides a closed form for the estimations of the expectations. Its advantage is that the formula can be easily computed, but Thurstone V is very sensitive to those values which are used for $\Phi^{-1}(0)$ instead of $-\infty$ and $\Phi^{-1}(1)$ instead of ∞, where Φ is the standard normal cumulative distribution

function. Another disadvantage is that it permits only two options in the choice, namely worse and better. It is reasonable to allow more than two options, for example, "equal" if there is no too much difference in senses or perhaps "much worse" or "much better", as well.

In this paper a Thurstone-motivated model for ranking lights is introduced. We keep the assumption of normally (Gauss) distributed latent random variables assigned to the sense of lights. More precisely, we suppose that the differences of the senses are charged with random errors and these errors are considered independent. The maximum likelihood method was applied to estimate the expectations of the latent random variables. Results were checked by performing the AHP method as well. A possible link between the AHP weights and the expectations is presented. To test the equality of the expectations the likelihood ratio test was applied. The steps of the evaluation process are shown through a detailed example. The results of a survey for different LEDs tested for shop lighting is also presented.

3   The Model

Let the number of objects (for example light sources, LEDs) investigated be denoted by n. Let the objects be characterized by the latent multidimensional random variable $\underline{\xi}=(\xi_1, \xi_2,..., \xi_n)$ with expectation $E(\underline{\xi})= \underline{m} =(m_1,m_2,...,m_n)$. The random variable ξ_i i=1,2,…,n features the sensation generated by the ith light in the observers. The observers judge about $\eta_{i,j}=\xi_i- \xi_j$, $i<j$, $i=1,2,...,n\text{-}1$, $j=2,...,n$. Now $E(\eta_{i,j})=m_i\text{-}m_j$. Let us assume that $\eta_{i,j}$ are independent identically distributed random variables with equal dispersions and suppose them to be Gaussian. Without any restriction, the dispersion can be fixed at 1. Applying the usual notation $N(m,\sigma^2)$ for Gaussian, $\eta_{i,j}\sim N(m_i\text{-}m_j,1)$. These assumptions were used earlier for economic problems in the case of two options to choose in [13].

In our model the number of options is increased. Five choices are allowed for the observers. First, the two compared objects can be equally preferable if the observer is not able to decide for any of them. Moreover, one of the objects can be more preferable than the other, or it can be less preferable than the other, it can be much more preferable than the other and much less preferable than the other.

The judgment is in connection with the value of the latent random variables as follows. If there is no essential difference between the senses, the observer states that the objects are equally preferable. This happens in the case, when the latent random variable $\eta_{i,j}$ satisfies the inequality $|\eta_{i,j}|<d$, with a positive parameter d. The observer states that the ith object is more preferable than the jth object, if $d\leq\eta_{i,j}<Kd$ with a fixed parameter value $1<K$. The observer states that the ith object is much more preferable than the jth object if $Kd\leq\eta_{i,j}$. On the other hand, the observer states that the ith object is less preferable than the jth object if the

inequality $-Kd \leq \eta_{i,j} \leq -d$ holds. Finally, the observer states that the ith object is much less preferable than the jth object if $\eta_{i,j} < -Kd$ is satisfied. The listed inequalities express that the random variable $\eta_{i,j}$ takes its values from special intervals as follows (Figure 4).
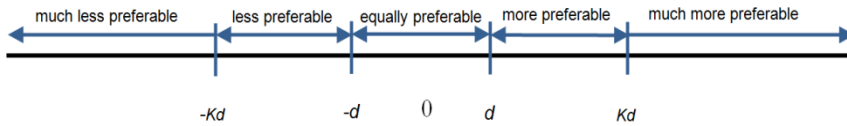

Figure 4

Judgments in the function of the value of the difference of latent random variables

The probabilities of these events can be given by the cumulative distribution function Φ depending on the parameters $\underline{m}$, d and K and these can be seen in Table 1. Due to the continuousness of the random variables the equalities do not change the values of the probabilities.

Table 1

Options of judgments and their probabilities

Opinion: i^{th} object is …the jth object	Code f	Probability $p_{i,j,f}$
much less preferable than	1	$P(\eta_{i,j} < -Kd) = \Phi(-Kd-(m_i-m_j))$
less preferable than	2	$P(-Kd \leq \eta_{i,j} \leq -d) = \Phi(-d-(m_i-m_j)) - \Phi(-Kd-(m_i-m_j))$
equally preferable to	3	$P(|\eta_{i,j}| < d) = \Phi(d-(m_i-m_j)) - \Phi(-d-(m_i-m_j))$
more preferable than	4	$P(d \leq \eta_{i,j} < Kd) = \Phi(Kd-(m_i-m_j)) - \Phi(d-(m_i-m_j))$
much more preferable than	5	$P(Kd \leq \eta_{i,j}) = 1 - \Phi(Kd-(m_i-m_j))$

The opinions of the observers are included in the indicator variables $X = (X_{i,j,f,k})$, $i=1,2,...,n-1$, $j=2,...,n$, $i<j$, $f=1,2,3,4,5$, $k=1,2,...,r$ where r is the number of observers. In detail, the values of the sample based on the opinion of the kth observer comparing the ith and j^{th} objects are summarized in Table 2.

Table 2

Values of the sample in case of the possible opinions of the kth observer comparing the ith and j^{th} objects

Opinion	$X_{i,j,1,k}$	$X_{i,j,2,k}$	$X_{i,j,3,k}$	$X_{i,j,4,k}$	$X_{i,j,5,k}$
much less preferable	1	0	0	0	0
less preferable	0	1	0	0	0
equally preferable	0	0	1	0	0
more preferable	0	0	0	1	0
much more preferable	0	0	0	0	1

Obviously, for any fixed value of i, j, k $\sum_{f=1}^{5} X_{i,j,f,k} = 1$. For further computations

let us define the aggregate samples $A_{i,j,1} = \sum_{k=1}^{r} X_{i,j,1,k}$ as the number of observers

who state that the ith object is much less preferable than the jth object;

$A_{i,j,2} = \sum_{k=1}^{r} X_{i,j,2,k}$ as the number of observers who state that the ith object is less

preferable than the jth object; $A_{i,j,3} = \sum_{k=1}^{r} X_{i,j,3,k}$ as the number of observers who

state that the ith object is equally preferable to the jth object; $A_{i,j,4} = \sum_{k=1}^{r} X_{i,j,4,k}$ as

the number of observers who state that the ith object is more preferable than the jth

object; and $A_{i,j,5} = \sum_{k=1}^{r} X_{i,j,5,k}$ as the number of observers who state that the ith

object is much more preferable than the jth object.

4 Likelihood Function of the Sample in the Function of the Parameters

The values of the parameters are estimated on the basis of the sample. In statistics, a frequently used method of estimation is the maximum likelihood method [14, 15]. Its main point is to compute the probability of the actual sample as a function of the parameter(s) and to find those values of the parameters where the probability reaches its maximum. The argument, when the probability has its maximal value is called the maximum likelihood estimation of the parameter(s). The probability of the sample is called the likelihood function.

Let the vector of parameters be denoted by $\underline{\vartheta} = (m_1, m_2, \ldots, m_n, d, K)$, where $m_i \in R$, $0 < d$ and $1 < K$. Assuming the independence of the opinions of the observers, the likelihood function is

$$L(X, \underline{\vartheta}) = \prod_{i=1}^{n-1} \prod_{j=i+1}^{n} p_{i,j,1}^{A_{i,j,1}} \cdot p_{i,j,2}^{A_{i,j,2}} \cdot p_{i,j,3}^{A_{i,j,3}} \cdot p_{i,j,4}^{A_{i,j,4}} \cdot p_{i,j,5}^{A_{i,j,5}}, \qquad (1)$$

when the values $p_{i,j,f}$ are included in Table 1. $L(X, \underline{\vartheta})$ is positive and it has its maximum at the same point when its logarithm has. Let us denote by $T = R^n \times R^+ \times \{y \in R : 1 < y\}$ the domain of parameter $\underline{\vartheta}$. The maximum likelihood estimation of

the parameter $\underline{\vartheta}$ is that $n+2$ dimensional vector in T where the above functions $L(X, \underline{\vartheta})$ and $\log L(X, \underline{\vartheta})$ have the maximum. That is

$$\hat{\underline{\vartheta}} = \arg \max_{\underline{\vartheta} \in T} L(X, \underline{\vartheta}) = \arg \max_{\underline{\vartheta} \in T} \log L(X, \underline{\vartheta}) \qquad (2)$$

Due to the very complicated form of the likelihood function, an explicit form for the argument cannot be given. Consequently, numerical methods were applied to find the argument of the optimum. Unfortunately the uniqueness of the argument where the likelihood function reaches its maximum is not proved, nevertheless, in practice no problem it was experienced performing numerical optimization.

The rank of the estimated values of expectations $\hat{m}_i = \left(\hat{\underline{\vartheta}} \right)_i$ $i=1,2,...,n$ provides

the rank of the objects. The estimated value of d equals $\hat{d} = \left(\hat{\underline{\vartheta}} \right)_{n+1}$. It serves

information about the perceptible border-line. The multiplier which characterizes

the category "much" is estimated by $\hat{K} = \left(\hat{\underline{\vartheta}} \right)_{n+2}$.

5 Likelihood Ratio Test for the Equality of the Expectations

The advantage of the maximum likelihood estimation is the fact that the testing of hypotheses connected to the estimated values is elaborated. To highlight the meaning of that: if some difference among the estimated values of the parameters m_i $i=1,2,..,n$, can be realized, then the question arises if this difference is significant or not. In other words, the differences are due to the randomness of the sample or they are too large to be explained by the random sample. To decide about this question, the likelihood ratio test can be applied [15].

First, the equality of all the expectations can be tested. In this case the null-hypothesis is H_0: $m_i = m_j$ for all values of i and $j, i \neq j$, the alternative hypothesis looks $H_1 : m_i \neq m_j$ for at least one pair of different i and j. One can check that the probabilities in Table 1 depend only on the difference of expectations, hence one can fix $m_1 = 0$. In this case we can simplify for $H_0 : m_i = 0$ for all values of $i=2,3,...,n$, and $H_1 : m_i \neq 0$ for some value of $i=2,3,...,n$.

For the sake of simplicity we use notation $T_0 = \{0\}^n \, x R^+ \, x \{y \in R : 1 < y\}$ and $T_1 = \{0\} x R^{n-1} x R^+ x \{y \in R : 1 < y\}$. The test statistics for testing H_0 is

$$D_t = -2\log(\frac{\sup\limits_{\vartheta \in T_0} L(X, \vartheta)}{\sup\limits_{\underline{\vartheta} \in T_1} L(X, \vartheta)}) \, . \tag{3}$$

If H_0 is fulfilled, then $D_t \sim \chi^2_{n-1}$ asymptotically, where χ^2_{n-1} denotes the Pearson chi-square distribution with degree of freedom n-1 [15]. Consequently, if the value of the test statistics is above the critical value of the chi-square distribution belonging to the significance level α, then there is a significant difference among the expectations at the significance level α, consequently H_0 is rejected.

Moreover, we can test if there exists a significant difference between the ith and j^{th} object for fixed values of i and j, $i \neq j$. To formulate the statistical problem, now the null-hypothesis is H_0: $m_i = m_j$ for fixed values of i and j, $i \neq j$, the alternative hypothesis is H_1: $m_i \neq m_j$ for these fixed values of i and j. First we notice that the hypothesis is symmetric in i and j, so we can assume that $i<j$. Now, fix $m_i = 0$, then $H_0 : m_j = 0$ versus $H_1 : m_j \neq 0$ for the fixed value of j. Apply the notations for the parameter spaces $T_i = R^{i-1} x \{0\} x R^{n-i} x R^+ x \{y \in R : 1 < y\}$ and $T_{i,j} = R^{i-1} x \{0\} x R^{j-i-1} x \{0\} x R^{n-j} x R^+ x \{y \in R : 1 < y\}$, $i<j$. The test statistics is

$$D_{i,j} = -2\log(\frac{\sup\limits_{\vartheta \in T_{i,j}} L(X, \underline{\vartheta})}{\sup\limits_{\vartheta \in T_i} L(X, \underline{\vartheta})}) \, . \tag{4}$$

Now, asymptotically, $D_{i,j} \sim \chi^2_1$. If the value of the test statistics is above the critical value of the chi-square distribution with degree of freedom 1 belonging to the significance level α, then there exists a significant difference between the expectations m_i and m_j at the significance level α.

6 Statistical Evaluation Process

In this section the steps of the maximum likelihood estimation (MLE) and the likelihood ratio test presented in Section 4 are followed on an example. For the next part of this chapter, neither the physical meaning of the objects, nor the question itself is taken into account, we only concentrate on the consequences of the data.

For ranking purposes, 5 LED based lights (having the same correlated color temperature but different spectral content) were compared pairwise by 48 observers. Each of the observers made 10 observations and decided about the questions: which one of the two objects is more preferable. The judgments of the observers in connection with the questions are included in Table 3.

Table 3

Aggregate opinion of the observers

$A_{i,j,f}$	f=1	f=2	f=3	f=4	f=5
i=1,j=2	5	11	21	11	0
i=1,j=3	4	21	11	6	6
i=1,j=4	3	16	18	8	3
i=1,j=5	8	7	7	13	13
i=2,j=3	11	14	19	2	2
i=2,j=4	3	16	23	6	0
i=2,j=5	4	9	13	7	15
i=3,j=4	3	13	18	14	0
i=3,j=5	4	9	5	16	14
i=4,j=5	4	6	20	11	7

Altogether 480 data are contained in this aggregate table. Recall, that the number $A_{1,2,1}$ presents that 5 observers think that the 1th object is much less preferable than the 2nd one. The numbers $A_{1,2,2}$, $A_{1,2,3}$, $A_{1,2,4}$ and $A_{1,2,5}$ present how many observers state that the 1st object is less preferable, equally preferable, more preferable, and much more preferable than the 2nd one, respectively. We can see that there is a large variety in the opinions. Using these sample values, the likelihood function (1) and its logarithm, the log-likelihood function are given in the function of the parameter $\underline{\vartheta}$. Fix $m_1 = 0$, and we search the point when the log-likelihood function has its maximum. After performing numerical optimization, we obtain the point

$\hat{\underline{\vartheta}} = (0, -0.007, 0.287, 0.197, -0.245, 0.453, 2.76)$ .This means, that the estimated values of the parameters are $\hat{m_1} = 0$, $\hat{m_2} = -0.007$, $\hat{m_3} = 0.287$, $\hat{m_4} = 0.197$, $\hat{m_5} = -0.245$, $\hat{d} = 0.453$, $\hat{K} = 2.76$. The rank of expectations and that of the objects (decreasing) is: 3, 4, 1, 2, 5. The average border-line when the observers are able to distinguish objects is characterized by the number 0.453, and the factor between more preferable and much more preferable is about 2.76. This means that if the difference between two sensations is between -0.453 and 0.453, then an average observer cannot distinguish sensations. The sign refers to the priority. If the difference in the values of the two latent random variables is more than 0.453·2.76≈1.250, then the difference in sensations is characterized by category "much" (Figure 5).
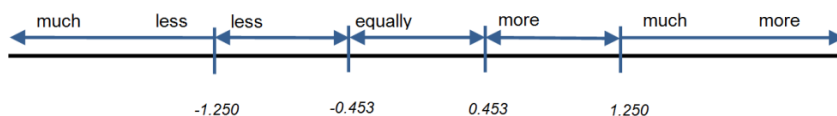
Figure 5
Estimated values of the judgment bounds

The calculations and numerical optimization were performed by the statistical program package R applying the function MLE [16]. In this function the initial values and the multivariate methods for finding the optimum can be changed. In practice we experienced stability concerning the initial values and methods.

To test whether all the expectations can be considered equal or not, the hypothesis is $H_0 : m_1 = m_2 = ... = m_5$ Performing the likelihood ratio test, the value of the test statistics (3) equals 35.93. It is so large that H_0 is rejected at any level of significance.

On the basis of the estimated expectations one can realize that the expectations of the 1st and 2nd objects are close to each other. The small difference between the estimated expectations may be due to the randomness of the sample. To decide about it, perform the pairwise likelihood ratio test.

First, test the hypotheses $H_0 : m_1 = m_2$ with the alternative hypothesis $H_1 : m_1 \neq m_2$. The value of the test function (4) equals $D_{1,2} = 0.005$. The critical values belonging to the chi-square distribution with degree of freedom 1 are denoted by $\chi^2_{1,\alpha}$. At $\alpha = 0.1$, $\alpha = 0.05$, $\alpha = 0.01$ these values are $\chi^2_{1,0.1} = 2.74$, $\chi^2_{1,0.05} = 3.84$, and $\chi^2_{1,0.01} = 6.63$, respectively. The value of the test function (4) equals 0.005. It is less than all of the above critical values, consequently the equality of the 1st and 2nd objects is accepted at any level of significance.

Generally, test the hypotheses $H_0 : m_i = m_j$ versus $H_1 : m_i \neq m_j$ for fixed values of i and j, $i \neq j$. Making these pairwise tests the values $D_{i,j}$ of the test functions by (4) and the levels of significance at which the equality of expectations is accepted are included in Table 4. Deciding on the usual significance level α=0.05, the results of the test $H_0 : m_i \neq m_j$ versus $H_1 : m_i \neq m_j$ can be seen in Table 5.

Table 4
The values of the test statistics $D_{i,j}$ given by (4) and significance levels to the test values

	j=1		j=2		j=3		j=4		j=5	
i	$D_{i,j}$	α	$D_{i,j}$	α	$D_{i,j}$	α	$D_{i,j}$	α	$D_{i,j}$	α
1	-	-	0.005	0.943	8.987	0.003	4.296	0.038	6.495	0.013
2	0.005	0.943	-	-	9.320	0.002	4.534	0.033	6.083	0.014
3	8.987	0.003	9.320	0.002	-	-	0.896	0.345	30.01	4e-8
4	4.296	0.038	4.534	0.033	0.896	0.345	-	-	21.10	4e-6
5	6.495	0.013	6.083	0.014	30.01	4e-8	21.104	4e-6	-	-

The results of the experiments can be summarized as follows: The first object in the rank is the 3rd one, but there is no significant difference between it and the following object in the rank, the 4th one. They are followed by the 1st and 2nd objects. There is no difference between them according to the opinion of the observers. The last object is the 5th object, it is the least preferable of the five objects and significantly worse than the others.

Our decisions may be refined as follows: we accept H_0 if the level of significance is more than 0.1, reject it if the level of significance is less than 0.01 and the cases when the level of significance is between 0.1 and 0.01 are considered border-line cases. These are also included in Table 5.

Table 5

Decisions at the standard significance level α=0.05 (A-H_0 is accepted, R-H_0 is rejected) and the refined version (A-H_0 is accepted, R-H_0 is rejected, B-Border-line case)

Decision	*j=1*		*j=2*		*j=3*		*j=4*		*j=5*	
i=1	-	-	A	A	R	R	R	B	R	B
i=2	A	A	-	-	R	R	R	B	R	B
i=3	R	R	R	R	-	-	A	A	R	R
i=4	R	B	R	B	A	A	-	-	R	R
i=5	R	B	R	B	R	R	R	R		-

The refined evaluation of the results is the following: two main groups of the objects can be distinguished. The first group consists of the 3rd and 4th objects, the second one consists of the 1st, 2nd and 5th objects. The 3rd and 4th one are equal, and so are the 1st and the 2nd at any level of significance according to the opinions of the observers. The difference between the pairs 2nd and 5th, moreover 1st and 5th are on the border-line of the sensation. The same situation holds for the pairs 4th and 1st and 4th and 2nd. If the sample had been larger than the border-line cases could be decided about.

7   Comparison with the Results of AHP

As no theoretical justification of the existence of latent random variables with normal distributions is proved theoretically, we try to justify the results experimentally. For this purpose the results were compared with the ones coming from the other used statistical method. In order to check the rank, we performed the AHP method generally accepted in pairwise comparison. Number 1 is applied for the opinion "equally preferable", number 3 for the opinion "more preferable" and number 5 for the opinion "much more preferable", and inversely number 1/3 is used for the opinion "less preferable" and 1/5 for the opinion "much less preferable" in case of separate observers. The separate opinions are aggregated by taking geometric means. Using the matrix consisting of aggregated opinions, we

determined the normalized eigenvector belonging to the maximal eigenvalue and we obtained $\underline{v}$=(0.184, 0.181, 0.253, 0.231, 0.150). The components of $\underline{v}$ provide the AHP weights and it can be seen that they produce the same rank that was presented by our method. The vector $\underline{v}$ has only positive components and their sum equals 1. To be able to compare them and the appropriate components of $\hat{\underline{\vartheta}}$, take the normalized vector

$$(\frac{\exp(\hat{m}_1)}{\sum\limits_{i=1}^{5}\exp(\hat{m}_i)}, \frac{\exp(\hat{m}_2)}{\sum\limits_{i=1}^{5}\exp(\hat{m}_i)}, \frac{\exp(\hat{m}_3)}{\sum\limits_{i=1}^{5}\exp(\hat{m}_i)}, \frac{\exp(\hat{m}_4)}{\sum\limits_{i=1}^{5}\exp(\hat{m}_i)}, \frac{\exp(\hat{m}_5)}{\sum\limits_{i=1}^{5}\exp(\hat{m}_i)}) \tag{5}$$

Its coordinates are close to the above AHP weights. This suggests that the AHP method may be in close connection with our presented method.

The AHP method is able to filter out those opinions which are inconsistent. Consistency is an important concept in AHP and there are different approaches in this area [17]. The advantage of this filtering is the usage of more reliable opinions, the disadvantage is the smaller sample size. In order not to reduce the sample extremely, we used those observations which have consistency index less than 0.3. The number of observers taken into account decreased to 34. The opinions, after aggregating them, are in Table 6. This table contains 340 judgments as opposed to the original 480, but these judgments are from reliable observers.

Table 6
Aggregate opinions of the reliable observers

$A_{i,j,f}$	f=1	f=2	f=3	f=4	f=5
i=1,j=2	5	9	15	5	0
i=1,j=3	4	16	6	6	2
i=1,j=4	2	12	15	4	1
i=1,j=5	5	4	7	9	9
i=2,j=3	8	9	14	1	2
i=2,j=4	3	11	17	3	0
i=2,j=5	3	7	9	6	9
i=3,j=4	2	8	15	9	0
i=3,j=5	3	6	4	13	8
i=4,j=5	4	2	15	8	5

After the computations based on these data, the AHP weights are $\underline{v}^*$= (0.170, 0.183, 0.257, 0.240, 0.151). The estimated value of the parameter $\underline{\vartheta}$ is $\hat{\underline{\vartheta}}^* = (0, 0.095, 0.404, 0.331, -0.136, 0.491, 2.595)$, which serves the same rank (3,4,2,1,5) as AHP weights. We computed the normalized exponential weights by

(5) and we obtained the vector (0.171, 0.188, 0.255, 0.237, 0.149), which is really close to $\underline{v}^*$, the maximum difference in the coordinates is 0.005. We can realize that this rank is slightly different from the one which was produced from all observations, actually the 2nd object is before the 1st one, but the equality of the two expectations is accepted in both cases. The levels of significance of the equality of expectations based on reliable opinions and the decisions can be summarized as follows:

Table 7

Significance levels using only reliable opinions and decisions using only reliable opinions (A-H_0 is accepted, R-H_0 is rejected, B-Border-line case)

	j=1		j=2		j=3		j=4		j=5	
	A	*dec.*	*α*	*dec.*	*A*	*dec.*	*A*	*dec.*	*α*	*dec.*
i=1	-	-	0.402	A	0.0004	R	0.004	R	0.230	A
i=2	0.402	A	-	-	0.007	R	0.038	B	0.042	B
i=3	0.0004	R	0.007	R	-	-	0.516	A	3e-6	R
i=4	0.004	R	0.038	B	0.516	A	-	-	4e-5	R
i=5	0.230	A	0.042	B	3e-6	R	4e-5	R	-	-

Comparing them to the decisions arising from the data without filtering in Table 5 we can realize that the results are very similar. Only 2 border-line cases remained. Two border-line cases become definite answers: from pairs 1st and 4th, 1st and 5th confirming the two groups mentioned in Section 6. Otherwise, there is no principal difference in the results whether we filter the opinions or not.

8   Experimental Results of the Survey

During the European research project (SSL4EU – Solid State Lighting for Europe) visual experiments were conducted at the University of Pannonia, where the optimal spectral power distribution of LED light sources was developed for home environment and fashion stores. In this part of the paper, the application of the mathematical method is demonstrated by the evaluation of a visual experiment on shop lighting. In order to present a realistic shop environment for the observers, the experimental room was equipped with four mannequins which were dressed with different clothes. Figure 6 shows the setup of the experimental room. Mannequins 1 and 2 can be seen on the left and mannequins 3 and 4 on the right of Figure 6. The different clothes worn by mannequins are the following: 1: green t-shirt, brown trousers of suit, 2: red t-shirt, blue jeans, violet belt, blue dress shirt, 3: white blazer, black t-shirt, turquoise trousers, yellow hat, 4:purple sweat-shirt, dark blue t-shirt, brown cord trousers.

Figure 6
Setup of experimental room

During the realization of light sources several photometric and colorimetric parameters and color quality metrics were used such as the Color Rendering Index (CRI-R_a) [18], the Color Quality Scale (CQS) [19], Feeling of Contrast Index (FCI) [20], Gamut Area Index, luminous efficiency or the Correlated Color Temperature (CCT) of the exited light. In addition, the R_9 value means the color rendering of saturated red color. The aim of the shop lighting experiments was to find an optimal light source spectrum for shop lighting by optimizing spectral distribution according to the FCI index. During the investigations five test light sources were developed. In Table 8 the first column (Test light ID) shows the code of the test lights.

Table 8
Summary of colorimetric parameters of the realized test lights investigated in the shop environment

Test light ID	LED selection criteria	FCI	CCT [K]	Δuv	CRI R_a	R_9	CQS Q_a
SS1	Incandescent mimic (R_a=97)	123	2900	-0.008	97	93	94
SS2	High CQS Q_p value (Q_p=96, R_a=80)	142	2947	-0.014	84	32	87
SS3	FCI = 110, R_a=82	107	2941	-0.014	78	15	75
SS4	FCI = 140, R_a=80	134	2920	-0.014	84	54	78
SS5	FCI = 150, R_a=80	151	2983	-0.009	85	-5	89

In order to find the preferred spectral power distribution (SPD) the presented paired comparison method has been applied during the visual experiments. Two SPDs (called "light source pair") were changed five times at regular intervals before the observers evaluated their impressions. For the evaluation random lamp pairs had been determined from the set of five different test spectra and each spectrum was compared to the four other SPDs in the same group. SS1 is considered the 1st object, SS5 the 5th one. So the five optimized spectra resulted in ten paired comparisons. The observers' task was to compare the two lights presented to them alternately every 20 seconds for 3 minutes, rate their visual experience for the given objects (different clothes) and fill in the questionnaire based on their visual impression.

At the end of the experiment sixty-three (63) observations were available to draw the conclusions from. The observers made their observations separately. Before the visual experiment the visual acuity of the observers was tested with the Kettesy reading test [21] (Hungarian version of the Snellen chart), and color vision was tested with the help of the Munsell Farnsworth 100 Hue test [22]. Only color normal observers were involved in the experiments.

Based on the observers' judgment preference orders of test light sources for the different testing conditions were obtained. Besides the preference orders the similarity of the test lights has also been evaluated based on the significance value achieved by different light pairs. Three cases were distinguished: Two spectra were regarded as different, equal or border-line.

Although during the experiment the observers' task was to evaluate thirteen different clothes under the light source pairs, but in this paper only results for five garments will be presented. These are the following: red t-shirt, yellow hat, green t-shirt, violet belt and blue jeans.

Based on the observers' opinion the following sequences were obtained in case of these garments. At the ranking, the estimations of the m_i values were taken into account. Table 9 shows the ranking orders based on the values and diagrams in Table 10 summarize the sequences.

Table 9

Summarized table from the ranking orders

Clothes	*Preference*				
red t-shirt	*5*	*2*	*1*	*4*	*3*
yellow hat	*2*	*1*	*4*	*5*	*3*
green t-shirt	*2*	*1*	*4*	*3*	*5*
violet belt	*4*	*2*	*3*	*5*	*1*
blue jeans	*4*	*3*	*2*	*1*	*5*

Table 10
Estimated values of expectations
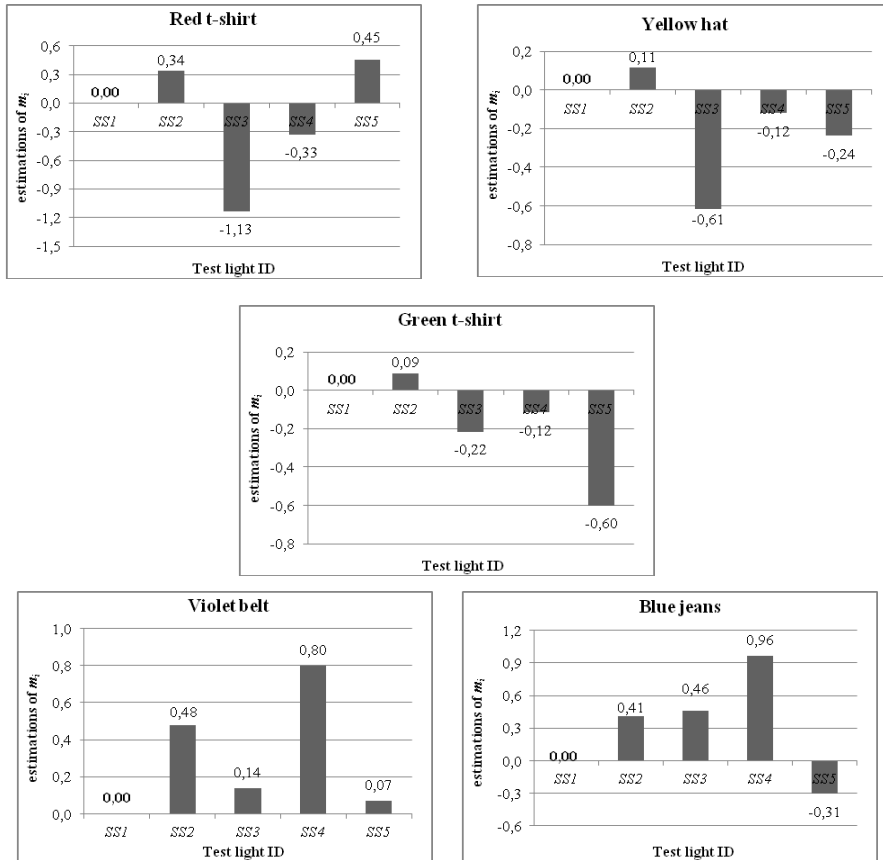
It can be seen that the preference of LED based lights depends on the color of the clothes. The first diagram in Table 10 shows that in case of red t-shirt the 5th illuminant was the most preferable one, but there is no significant difference between this light and the 2nd light source. Hence in case of the red – yellow and green clothes rather the 2nd test light source was perceived as the best. In case of the violet and blue garments rather the 4th was the best. In addition the observers least preferred the 3rd illuminant for the red and yellow clothes, while for the violet and blue garments the 1st and 5th light sources were preferred at least. The results were the same as those when the questions concerning vividness were evaluated.

Table 11

Decisions (A – equality accepted, R – rejected, B – border-line case)

Shop lighting experiment																		
red t-shirt						yellow hat						green t-shirt					
SS	*1*	*2*	*3*	*4*	*5*	*SS*	*1*	*2*	*3*	*4*	*5*	*SS*	*1*	*2*	*3*	*4*	*5*
1	-	R	R	R	R	*1*	-	A	R	B	R	*1*	-	A	R	B	R
2	R	-	R	R	A	*2*	A	-	R	R	R	*2*	A	-	R	B	R
3	R	R	-	R	R	*3*	R	R	-	R	R	*3*	R	R	-	A	R
4	R	R	R	-	R	*4*	B	R	R	-	A	*4*	B	B	A	-	R
5	R	A	R	R	-	*5*	R	R	R	A	-	*5*	R	R	R	R	-

Shop lighting experiment										
violet belt						blue jeans				
SS	*1*	*2*	*3*	*4*	*5*	*SS*	*1*	*2*	*3*	*4*	*5*
1	-	R	B	R	A	*1*	-	R	R	R	R
2	R	-	R	R	R	*2*	R	-	A	R	R
3	B	R	-	R	A	*3*	R	A	-	R	R
4	R	R	R	-	R	*4*	R	R	R	-	R
5	A	R	A	R	-	*5*	R	R	R	R	-

Considering the similarity of the light sources in case of the red t-shirt the 2nd and 5th light sources were perceived the same and the others were different based on the observers' opinion. For the yellow hat next to the 1st and 2nd light illuminants the 4th and 5th show similarity, while between the 1st and 4th light source the difference moves at the border of the demonstrability. In case of the green t-shirt less difference can be seen: the 1st illuminant shows similarity with the 2nd test light and the 3rd illuminant shows similarity with the 4th light source. The difference between the 1st – 4th and 2nd – 4th illuminants is on the borderline.

For the violet belt the 1st – 5th and the 3rd – 5th light sources were perceived as similar and the difference between the 1st and 3rd illuminants moves at the border of the demonstrability. In case of the blue jeans significant difference can be seen between the light sources. Only the 2nd test light shows similarity with the 3rd illuminant. A more detailed presentation of the results was published in [2].

Conclusion

A statistical method was elaborated herein for ranking spectral power distributions (SPD) based on subjective points of view. Pairwise comparisons were performed with five options in judgment. The evaluation method was improved and a maximum-likelihood estimation was performed by applying the R statistical package. Not only ranking has been developed and compared to the results of the AHP method, but the equality of the object characteristics has also been tested.

Differences are often traceable on the basis of the opinions of 60 observers. A possible connection between the AHP weights and the estimated expectations is presented. The method was successfully applied in choosing light sources for home lighting and shop lighting [1, 2] and can be applied in the case of further choices based on subjective points of view. The results of the experiments depend on the dominant color of the illuminated objects.

Acknowledgment

References

[1]   F. Szabó, R. Kéri, J. Schanda, P. Csuti, É. Mihálykó Orbán: A Study of Preferred Colour Rendering of Light Sources: Home Lighting, Lighting Research and Technology, 2015 47: 165-182, Online ISSN: 1477-0938 Print ISSN: 1477-1535

[2]   F. Szabó, R. Kéri, A. Wilm, J. Schanda, P. Csuti, E. Baur: A Study of Preferred Colour Rendering of Light Sources: Shop Lighting, Lighting Research and Technology, in press, Online ISSN: 1477-0938 Print ISSN: 1477-1535

[3]   J. Schanda, P. Csuti, F. Szabó: Colour Fidelity for Picture Gallery Illumination, Part 1: Determining Optimum LED Spectrum, Lighting Research and Technology, in press, Online ISSN: 1477-0938 Print ISSN: 1477-1535

[4]   J. Schanda , P. Csuti, F. Szabó, P. Bhusal, L. Halonen: Introduction to a Study of Preferred Colour Rendering of Light Sources, Lighting Research and Technology, 47:28-35, 2015

[5]   M. S. Islam, R. Dangol, M. Hyvärinen, P. Bhusal, M. Puolakka, L. Halonen: User Acceptance Studies for LED Office Lighting: Lamp Spectrum, Spatial Brightness and Illuminance Level, Lighting Research and Technology, 47: 54-79, 2015

[6]   K. R. Gegenfurtner, L. T. Sharpe: Color Vision from Gene to Perception, Cambridge University Press, 1999

[7]   W. G. K. Backhaus, R. Kliegl, J. S. Werner: Color Vision Perspectives from Different Disciplines, Walter de Gruyter, 1998

[8]   L. Saaty: The Analytic Hierarchy Process: Planning, Priority Setting, Resource Allocation, McGraw-Hill International Book Co., New York, 1980

[9]   O. S. Vaidya, S. Kumar: Analytic hierarchy process: An Overview of Applications, European Journal of Operational Research, 169: 1-29, 2006

[10] O. Perron: Zur Theorie der Matrices, Mathematische Annalen, 64(2):248-263, 1907

[11] L. L. Thurstone: A Law of Comparative Judgment, Psychological Review, 101(2): 266-270, 1994

[12] F. Mosteller: Remarks on the Method of Paired Comparison I. The Least Squares Solution Assuming Equal Standard Deviations and Equal Correlations, Psychometrica (16)1: 3-9, 1951

[13] T. Pfeiffer, X. A. Gao, A. Mao, Y. Chen, and D. G. Rand: Adaptive Polling for Information Aggregation, In the Proceedings of the 26th AAAI Conference on Artificial Intelligence (AAAI '12) 2012 http://www.eecs.harvard.edu/econcs/bib/author.html

[14] S. S. Wilk: Mathematical Statistics, New York, John Wiley, 1962

[15] F. T. Móri, J. G. Székely: Multivariate Statistical Analysis, Műszaki Könyvkiadó, Budapest, 1986 (In Hungarian)

[16] R Core Team: R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria, ISBN 3-900051-07-0, URL http://www.R-project.org/ 2012

[17] J. A. Alonso, M. T. Lamata: Consistency in the Analytic Hierarchy process: a New Approach. International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems, 14: 445-459 2006

[18] Commission Internationale de l'Eclairage (CIE): Method of Measuring and Specifying Color Rendering Properties of Light Sources, CIE Publication 13.3, Vienna: CIE, 1995

[19] W. Davis, Y. Ohno: Color Quality Scale, Optical Engineering 49: 033602, 2010

[20] K. Hashimoto, T. Yano, M. Shimizu, Y. Nayatani: New Method for Specifying Color-Rendering Properties of Light Sources based on Feeling of Contrast, Color Research and Application, 32: 361-371, 2007

[21] Webbeteg. Testing visual acuity. Retrieved 14 January 2015 from http://www.webbeteg.hu/cikkek/szem_betegsegei/3244/lataselesseg-vizsgalat

[22] D. Farnsworth: The Farnsworth-Munsell 100-Hue Test for the Examination of Color Discrimination, Revised 1957; Baltimore, MD: Munsell Color Company