Analysis of Experimental Data Used for Development of CIE DE 2000 Color Difference Formula

Pavel Bourov, UniqueIC's, Saratov, Russia; Sergey Bezryadin, KWE Int. Inc, San Francisco, USA.

Abstract

Analysis of experimental data used for development of CIE DE 2000 color difference formula discloses several problems:

- *Visual differences recorded by KLAUS WITT and by M. R. LUO for very similar pairs usually differ several times*
- *Within data recorded by one group there are examples of similar pairs for which Visual difference varies about two times*
- *Visual difference between two coincide stimuli is not zero.*

To become useful for CDF development the data has to be thorough extra processed.

Introduction

In spite of the effort to design a color difference formula (CDF) which adequately reflects peculiarities of human color perception, its current versions, including CIE DE 2000, do not completely satisfy this demand and a search for a better option does not stop. The evaluation of quality of a color difference formula has to be founded on experimental data of color discrimination.

One of the known criteria for comparison CDF quality is the shape of MacAdam's ellipses [1], but it is based on only two-dimensional data. Other method [2] uses more comprehensive data provided by Wyszecki & Fielder. However, it also cannot be counted as the only criterion because the experiment contains a limited number of color-matching ellipsoids (only for 28 stimuli).

The founders of CDF CIE DE 2000 [3] use a large pool of experimental data obtained by experimentalists in various countries [4]. The data sheet contains 3813 pairs. The first in the pair is a stimulus named color center or standard of this pair and the second is color difference between a batch and standard in a pair. For each pair there is a corresponding visual result dV. Color centers are distributed among the Gamut, so the data might serve as a base for comprehensive criterion of CDF quality.

The authors intended to examine the degree of conformity of various CDFs to experiment with experimental data [4], and, as the first step, they conduct an analysis of the data making consistency check and testing data reliability with methods routinely used in experimental physics. For easy reference, a list of data was numbered.

The test discloses a number of conflicting data. Some examples are presented in Table.1. For instance, in two pairs 334 and 3317 corresponding stimuli coincide with a fair accuracy, while the dV (Visual difference) value varies nearly five times. The large difference in dV value suggests that at least one of these two near-coincident pairs contains incorrect data and, thus, makes both pairs useless for CDF testing. Unfortunately, this defect is not unique. There are scores of similar examples in the data sheet.

Table.1: Examples of conflicting data

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
334	62.9	69.5	30.2	-1	-1	-0.92	1.36
3317	62.9	69.5	30.3	0	-2	0.88	0.29
450	8.9	8.8	23.2	0	-3	0.55	2.10
2411	9.1	8.9	23.5	-1	9	0.52	0.48

The mentioned difference in dV value might have a natural cause, if involved experimentalists use different scales for their measurement. (It is like some person measure weight in grams and the other does it in ounce). In this case, there has to be a conversion factor between scales. To check this possibility, the authors attempt to determine these factors and convert data obtained by various experimentalists to a single scale.

Description of the method

Visual difference dV may be described by a function with six variables f(X, Y, Z, dx, dy, dY). However, for a group of pairs having the same color center (which means X, Y, and Z values are fixed) and neglible small chromatic coordinates change (so we can assume dx = dy = 0), a Visual difference dV might be represented by a function $f_{X,\ Y,\ Z,\ dx=0,\ dy=0}$ (dY) with only one variable. The function takes zero value when dY=0, because in this case a batch coincide with the standard of a pair.

When dY values small, $f_{X,Y,Z,dx=0,dy=0}$ (dY) may be approximated by a linear function with Tailor formula. Graphically it would be represented by a straight line.

For the analysis authors select only those pairs, in which dY/Y ratio is less than 0.1 and an angle between a vector representing the standard of a pair and a vector representing a color difference is less than 5˚. (Here and further lengths and angles are in Cohen metric [5])

Analysis of Grey stimuli

For the first approach authors choose those pairs whose standard of a pair makes with Day Light D_{65} not more than 5˚. Name these stimuli as Grey. KLAUS WITT data has 13 pairs satisfying stated condition. Because selected pairs have the same X, Y, and Z values for their color centers (with accuracy 5%), a visual difference for all these pairs may be described with a single function $f_{X,\ Y,\ Z,\ dx=0,\ dy=0}$ (dY). In order to be described with the same function, the selection of Grey stimuli data provided by other experimentalists was narrowed further to only those pairs which have Y ≈ 30, the same as KLAUS WITT pairs have. Within stated restrictions there are 31 pairs in the examined data pool:

- 8 experimental pairs from DONG-HO KIM;
- 13 experimental pairs from KLAUS WITT;
- 10 experimental pairs from M. R. LUO.

Unfortunately, the number of ROY S. BERNS's Grey pairs is insufficient for statistical analysis and therefore was not considered in further discussion.

DONG-HO KIM data

DONG-HO KIM pairs satisfied selection restriction are presented in the Table 2. Here and further X, Y, and Z values are rounded to three digits. Figure 1 displays relationship between Visual difference dV and |dY|, an absolute value of a deviation between the standard of the pair and a batch.

Table.2: Grey stimuli. DONG-HO KIM

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
91	27.2	28.7	30.8	-1	-2	1.69	1.13
92	26.6	28.2	30.2	-2	-2	1.98	1.39
99	26.9	28.4	30.6	2	1	0.84	0.70
100	28.1	29.7	31.9	1	1	1.25	0.96
101	27.7	29.2	31.4	0	0	1.59	1.29
102	27.0	28.5	30.7	0	-1	1.97	1.48
242	27.7	29.2	31.4	0	0	1.59	1.25
243	27.0	28.5	30.7	0	-1	1.97	1.78

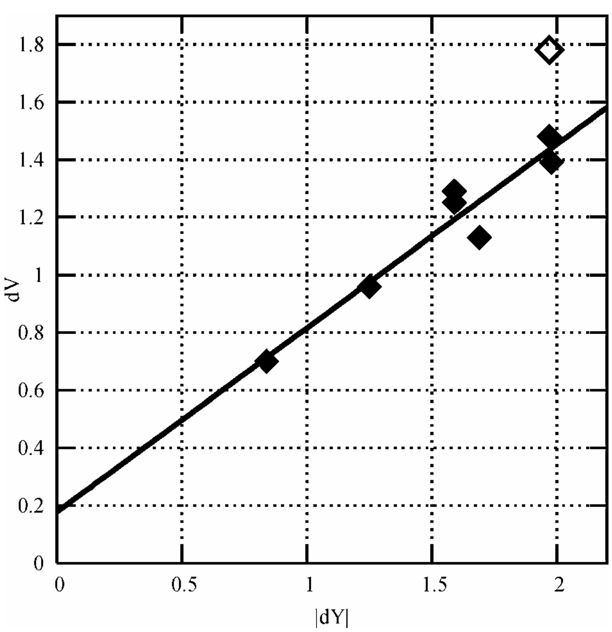

Fig.1: Grey stimuli. DONG-HO KIM

As it could be seen from the picture, DONG-HO KIM data is in a good concordance with an assumption, that with fixed standard of a pair and zero chromatic coordinates change dV linearly depends on |dY|. All points might be well approximated with a straight line except one (line 243 in Tab.2). Although the pip might be caused by some very specific peculiarities of human color perception, the authors suppose that it is more likely due to an experimental error.

Linear approximation

$$dV = 0.64 \cdot |dY| + 0.18 \quad (1)$$

made with least-squares method is based on 7 points with point 243 set aside. It has to be mentioned, that the line does not pass zero.

KLAUS WITT data

Table.3: Grey stimuli. KLAUS WITT

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
359	29.6	31.2	33.4	2	2	0.61	2.31
361	29.6	31.2	33.4	0	1	1.04	3.65
362	29.6	31.2	33.4	-1	2	1.37	4.45
363	29.6	31.2	33.4	2	3	0.76	1.49
364	29.6	31.2	33.4	-1	-2	-0.46	1.95
365	29.6	31.2	33.4	2	2	-0.54	2.30
366	29.6	31.2	33.4	3	4	-0.74	2.92
367	29.6	31.2	33.4	-1	-2	-1.13	3.75
673	29.6	31.2	33.4	3	-1	0.73	1.50
709	30.2	31.8	34.0	-3	-4	-1.07	3.49
714	30.3	31.9	34.4	7	7	-1.28	4.45
719	30.6	32.2	34.5	3	3	-1.78	5.44
723	30.9	32.5	34.8	0	-4	-2.50	6.96

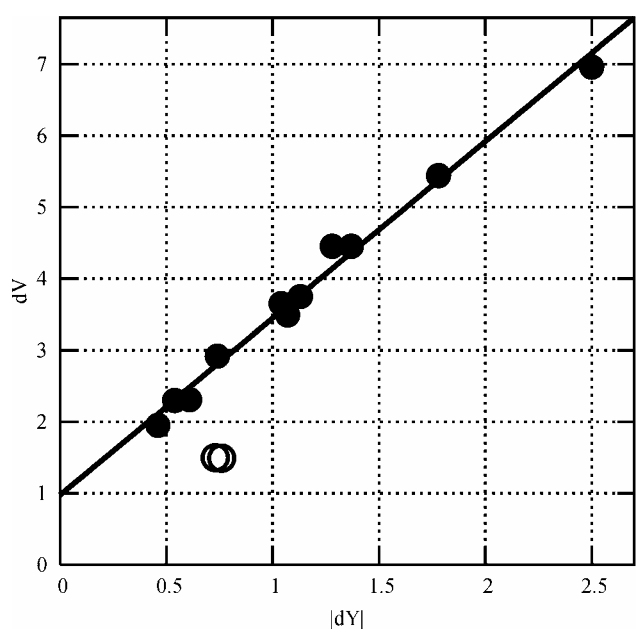

Fig.2: Grey stimuli KLAUS WITT

Table 3 represents KLAUS WITT Grey stimuli. Figure 2 displays the correspondence between Visual difference dV and an absolute value of a deviation |dY| for 13 experimental pairs selected from KLAUS WITT data.

This picture also confirms the thesis stated above. All points are situated on a straight line with a fair accuracy except two (lines 363 and 673 in Tab.3). As in previous case, authors suppose, that pips are due to an experimental error and do not consider them in approximation

$$dV = 2.47 \cdot |dY| + 0.97 \qquad (2)$$

Point out, that the line does not pass zero.

M. R. LUO data

Table 4 displays M. R. LUO data that satisfy criterion stated above. Relationship between Visual difference dV and an absolute value of a deviation |dY| for Grey stimuli recorded in M. R. LUO experiments depicted in Figure 3.

Table.4: Grey stimuli M. R. LUO

N	X	Y	Z	10^4·dx	10^4·dy	dY	dV
2124	28.2	29.6	31.5	10	0	3.07	2.01
2125	28.2	29.6	31.5	7	-8	2.00	2.20
2142	28.2	29.6	31.5	0	1	-1.97	1.53
2144	28.2	29.6	31.5	3	-2	2.44	1.70
2147	28.2	29.6	31.5	-4	6	2.06	1.76
2148	28.2	29.6	31.5	7	16	2.66	1.84
2150	28.2	29.6	31.5	1	7	1.38	1.17
2163	28.2	29.6	31.5	-6	9	-2.96	2.08
2164	28.2	29.6	31.5	1	2	-2.38	1.56
3248	29.6	31.2	33.4	0	0	-0.16	0.16

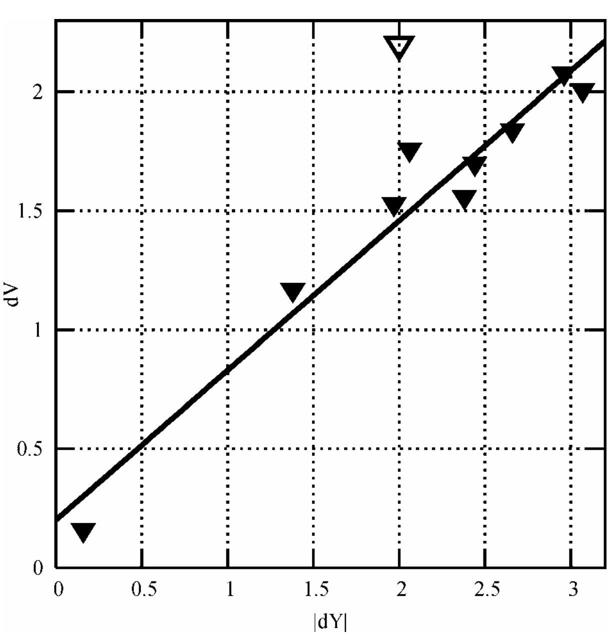

Fig.3: Grey stimuli. M. R. LUO

M. R. LUO data selected according to stated criterion has more spread in values, than data provided by DONG-HO KIM or KLAUS WITT. Point 2125 is not considered in linear approximation (3) because it is more likely an experimental error. As in previous cases, the line does not pass zero.

$$dV = 0.63 \cdot |dY| + 0.20 \qquad (3)$$

Point 2125 draw a special attention. According to Wyszecki & Fielder [1] experiment, Grey stimuli differing less than 2% are not discriminated. However, at point 2125 M. R. LUO registers less, than 0.5% luminance change (dx and dy are even zero) with six (!) significant digits (Tab. 4 displays rounded data). This fact adds concerns about data reliability.

Conversion of combined data to a single scale

If all three experimentalists use the same scale for their measurement, approximation lines for their data have to coincide. What one can see comparing (1), (2) and (3) is that slopes of DONG-HO KIM and M. R. LUO lines are in a close agreement, while a slope of KLAUS WITT line is about 4 times bigger. It might happen if KLAUS WITT uses smaller unit for visual difference measurement than two other experimentalists. DONG-HO KIM and M. R. LUO data was converted to KLAUS WITT's unit and results presented in Figure 4.

Basically, all three scales are in equal status and any of them might be chosen as the one. The choice in favor of KLAUS WITT scale was due to authors' opinion, that his data looks more accurate and less spread in values.

Although for Grey stimuli conversion factor is equal 3.9, and its use brings better visual result for this particular color group, figure displaying them is omitted. Instead, because of limited space for illustrations, authors display graphs with a unified conversion

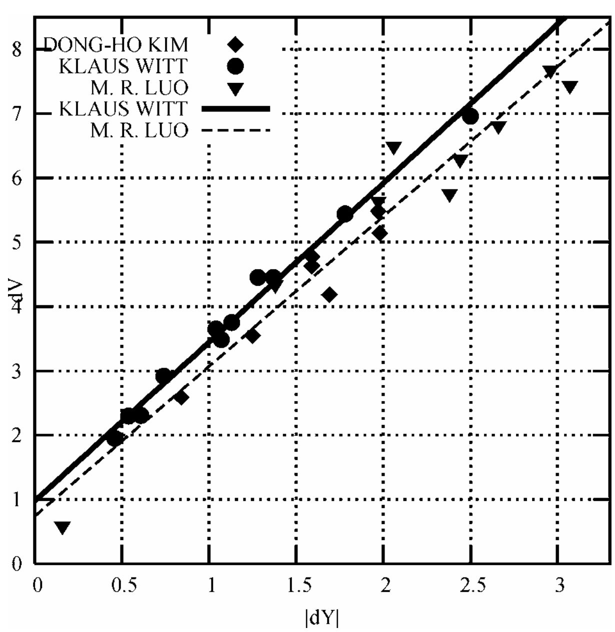

Fig.4: Grey stimuli. All experimentalists. DONG-HO KIM and M. R. LUO data is taken with a factor 3.7.

factor 3.7 (Fig.4), which might be used for any color group of M. R. LUO data, and which will be derived further, at the very end of the paper. For the same reason, figures 5-8 display graphs with the unified factor. This factor may be used for DONG-HO KIM's Grey stimuli. Colored DONG-HO KIM's stimuli do not participate in this research.

Colored stimuli analysis

Authors select groups of data satisfying one-variable-function-condition (X, Y, and Z are constant for group members; dx=dy=0). Although the data pool is pretty large, only KLAUS WITT and M. R. LUO data sheets have enough intersections for colored stimuli analysis.

The same approach as was used for Grey stimuli analysis is applied to colored stimuli to test data reliability and check whether it is possible to convert data obtained by various experimentalists to a single scale.

Tables 5-12 present data which was selected for analysis and figures 5-8 show a correspondence between dV and |dY| for various colors.

To determine a unified conversion factor between M. R. LUO and KLAUS WITT scales, authors first calculate a scope ratio for Red, Yellow, Green and Blue stimuli. Results are presented in Table 13. Their average value is used as a unified factor.

Point 3193 (Tab.6) is more likely an experimental error, so it is excluded from the data used for linear approximation in Figure 5. The scope ratio for red stimuli data is 3.40.

KLAUS WITT data (Tab.7) is in good conformity with an approximation line (Fig.6). As for M. R. LUO data for yellow stimuli (Tab.8), corresponding points heavily spreads in values. However, using least-squares method it is possible to get an approximation line for them. When all 7 points from table 8 are taken into account, the slope ratio for yellow stimuli data is 3.91.

The slope ratio for green stimuli data is 3.50 and the slope ratio for blue stimuli data is 3.50.

Table 13 represents conversion factors for visual difference measurement between M. R. LUO and KLAUS WITT scales for each examined color. The factor value varies from 3.5 to 3.91 with the average number 3.7, which can serve as a unified factor. As it could be seen on Figures 4-8, in every examined color group both approximating lines become very close when M. R. LUO data is taken with factor 3.7. However, even then, deviation of KLAUS WITT's data is essentially smaller, then M. R. LUO's.

The conducted research confirms authors' supposition, that participating experimentalists use different scales for their measurement. The fact, that none of approximating lines passes the zero points out on some methodological error.

Table.5: Red stimuli, KLAUS WITT

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
416	20.08	14.24	7.24	0	-3	0.18	0.85
417	20.08	14.24	7.24	-1	0	0.35	1.40
418	20.08	14.24	7.24	-1	-2	0.63	2.12
419	20.08	14.24	7.24	4	-2	0.79	2.39
420	20.08	14.24	7.24	0	0	0.96	2.83
423	20.08	14.24	7.24	0	-4	-0.61	1.92
424	20.08	14.24	7.24	-3	1	-0.79	2.40
425	20.08	14.24	7.24	-2	1	-1.01	2.78
711	20.57	14.59	7.42	-5	4	-0.67	2.15
716	20.98	14.87	7.58	1	-2	-1.24	3.23

Table.6: Red stimuli, M. R. LUO

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
2207	20.55	14.40	7.28	12	-2	-1.02	1.05
2223	20.55	14.40	7.28	3	-15	-1.19	1.26
2225	20.55	14.40	7.28	2	5	-1.13	0.97
3183	20.38	14.34	7.28	-2	2	0.43	0.53
3193	20.38	14.34	7.28	11	22	-1.18	1.56
3275	20.29	14.35	7.24	-1	-1	0.23	0.44
3280	20.29	14.35	7.24	0	0	0.31	0.67
3286	20.29	14.35	7.24	3	4	0.49	0.84

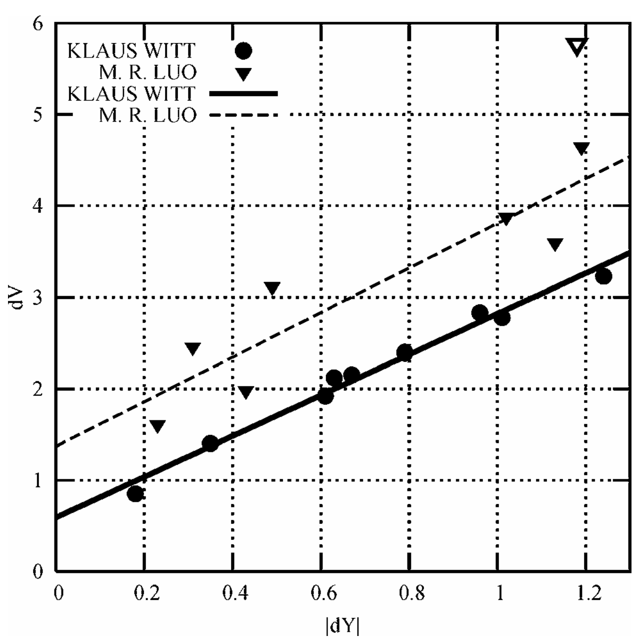

Fig.5: Red stimuli. M. R. LUO data is taken with a factor 3.7.

Table.7: Yellow stimuli. KLAUS WITT

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
330	62.89	69.53	30.22	-5	-7	2.35	2.88
331	62.89	69.53	30.22	1	1	3.58	3.65
332	62.89	69.53	30.22	-4	-3	4.33	4.55
334	62.89	69.53	30.22	-1	-1	-0.92	1.36
335	62.89	69.53	30.22	2	2	-1.57	1.93
336	62.89	69.53	30.22	-1	-2	-2.35	2.51
337	62.89	69.53	30.22	4	4	-3.22	3.36
708	64.51	71.34	31.19	4	3	-2.73	2.87
713	65.04	71.88	31.49	7	9	-3.92	4.28
718	66.13	73.11	31.73	-2	-3	-5.93	5.49

Table.8: Yellow stimuli. M. R. LUO

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
2253	63.71	69.89	29.47	-1	2	6.26	1.38
2254	63.71	69.89	29.47	12	-6	6.31	1.71
2280	63.71	69.89	29.47	6	24	-5.40	1.63
2281	63.71	69.89	29.47	4	7	-4.41	0.81
2282	63.71	69.89	29.47	1	-3	-5.78	1.26
2284	63.71	69.89	29.47	3	21	-6.34	1.14
3317	62.88	69.53	30.28	0	-2	0.88	0.29

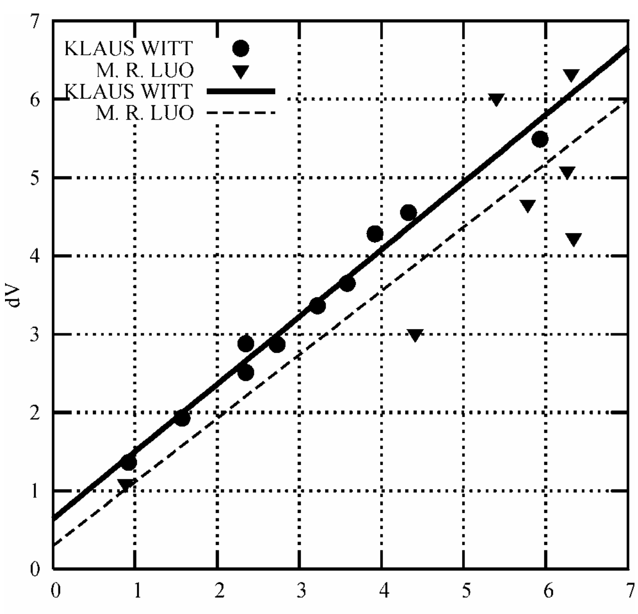

Fig.6: Yellow stimuli. M. R. LUO data is taken with a factor 3.7.

Table.9: Green stimuli, KLAUS WITT

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
390	16.64	24.06	25.60	4	7	1.33	2.69
392	16.64	24.06	25.60	3	0	-0.38	1.05
393	16.64	24.06	25.60	6	5	-1.08	2.32
394	16.64	24.06	25.60	-4	1	-1.33	2.70
395	16.64	24.06	25.60	14	1	-1.72	3.70
683	16.73	24.23	26.08	1	16	-1.50	3.45
704	16.91	24.51	25.88	4	-4	-0.76	1.63
705	16.91	24.51	25.88	3	-8	-1.53	3.04
710	16.96	24.52	25.85	5	-3	-2.18	4.29

Table.10: Green stimuli. M. R. LUO

N	X	Y	Z	$10^4 \cdot dx$	$10^4 \cdot dy$	dY	dV
2309	16.58	24.17	25.81	-5	-1	2.24	1.37
2311	16.58	24.17	25.81	17	1	2.38	1.29
2327	16.58	24.17	25.81	2	4	-1.91	1.16
2329	16.58	24.17	25.81	-3	-11	2.26	1.19
2331	16.58	24.17	25.81	0	13	2.23	1.43
2345	16.58	24.17	25.81	-5	17	-2.09	1.43
2348	16.58	24.17	25.81	-4	-7	-1.99	1.19
3348	16.79	24.30	25.70	-2	-1	-0.34	0.21
3375	16.79	24.30	25.70	1	-2	0.66	0.41
3405	16.79	24.30	25.70	6	-1	-1.00	0.67
3417	16.79	24.30	25.70	0	-2	-0.48	0.36
3420	16.79	24.30	25.70	0	-2	0.36	0.45

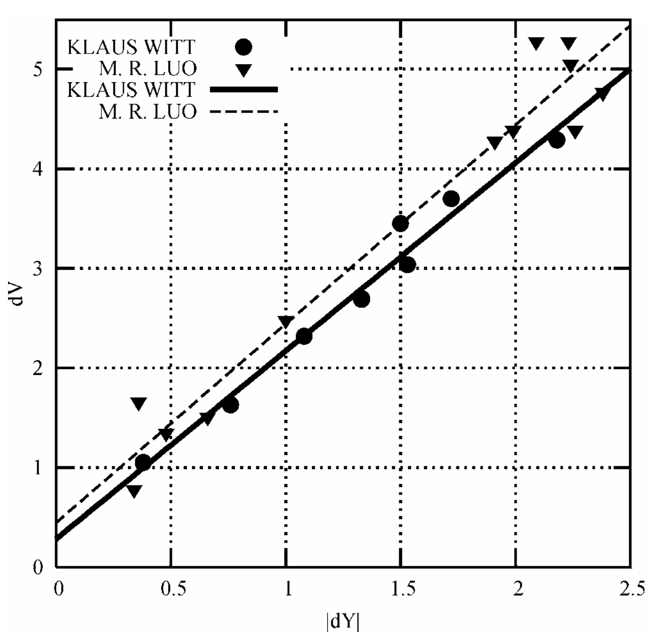

Fig.7: Green stimuli. M. R. LUO data is taken with a factor 3.7.

Table.11: Blue stimuli. KLAUS WITT

N	X	Y	Z	10^4·dx	10^4·dy	dY	dV
448	8.91	8.77	23.23	-3	-7	0.21	1.05
449	8.91	8.77	23.23	-2	-3	0.44	1.89
450	8.91	8.77	23.23	0	-3	0.55	2.10
451	8.91	8.77	23.23	-2	-2	-0.11	0.60
452	8.91	8.77	23.23	0	-2	-0.27	1.28
453	8.91	8.77	23.23	0	-2	-0.36	1.63
454	8.91	8.77	23.23	0	-3	-0.48	2.02
455	8.91	8.77	23.23	-4	-5	-0.61	2.29
707	9.01	8.86	23.48	-4	-1	-0.20	0.92
717	9.14	8.98	23.91	3	5	-0.57	2.31

Table.12: Blue stimuli. M. R. LUO

N	X	Y	Z	10^4·dx	10^4·dy	dY	dV
2387	9.12	8.94	23.50	4	1	0.75	1.04
2390	9.12	8.94	23.50	16	1	-0.78	0.73
2410	9.12	8.94	23.50	6	20	0.93	1.33
2411	9.12	8.94	23.50	-1	9	0.52	0.48
3456	8.84	8.68	22.93	1	-2	0.11	0.32
3459	8.84	8.68	22.93	1	-2	0.14	0.48
3471	8.84	8.68	22.93	3	3	-0.11	0.28
3479	8.84	8.68	22.93	2	-1	-0.17	0.59
3481	8.84	8.68	22.93	2	1	-0.13	0.20

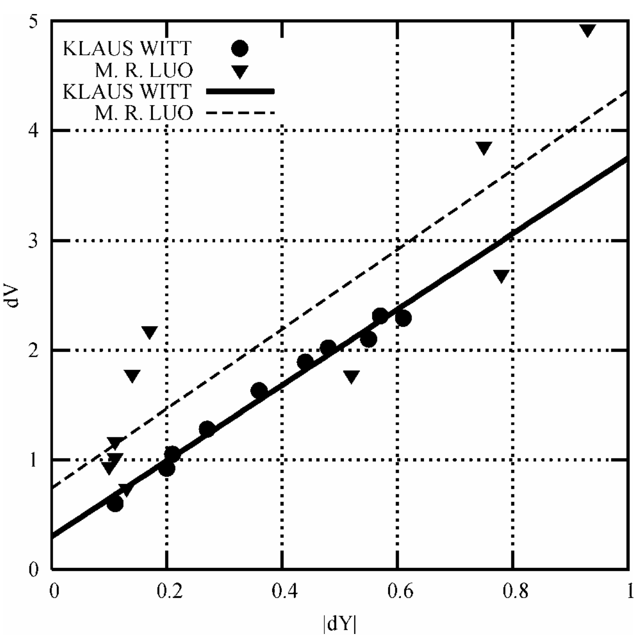

Fig.8: Blue stimuli. M. R. LUO data is taken with a factor 3.7.

Table.13: conversion factor for visual difference units of measurement

Color	$dV_{KLAUS\ WITT}/dV_{M.\ R.\ LUO}$
Gray	3.89
Red	3.40
Yellow	3.91
Green	3.50
Blue	3.50

Conclusion

Repeatability and reproducibility are the main requirements for experiment results and the data used for development of CIE DE 2000 CDF does not meet this criterion.

Visual differences recorded by KLAUS WITT and by M. R. LUO for very similar pairs usually differ several times. As it was demonstrated in the paper, data spread might be scientifically reduced if Visual difference recorded by DONG-HO KIM or M. R. LUO multiplied by 3.7.

However, even within data recorded by one group there are examples of similar pairs for which Visual difference varies about two times, for instance, pairs 363 and 366 (Table 3), or pairs 3459 and 3481 (Table 12). Usually, an experimentalist carefully filters out such contradictory results.

All examined data carry the same systematic bias. With vanishing dY the limit of Visual difference is significantly more, then zero, it is about 10% of the value that Visual difference takes for dY = 0.1·Y. Having no access to complete information describing the experiment method, authors only can presume, that this fact is due to a methodological error.

Thus authors can state, that data [4] cannot be used for CDF quality testing without thorough extra processing, which, in order, requires extra information regarding experiment method.

References

1. MacAdam, D. L., "Visual sensitivities to color differences in daylight", J. Opt. Soc. Am. 32, 247, 1942.
2. S. Bezryadin and P. Bourov, "Local Criterion of Quality for Color Difference Formula", Proceedings this conference
3. CIE, "Improvement to industrial colour-difference evaluation", CIE Publications No. 142-2001, Central Bureau of the CIE, Vienna, 2001.
4. http://cie.kee.hu/newcie/doc/Data_CIE_DE2000.pdf
5. Cohen J. B., "Visual Color and Color Mixture: The Fundamental Color Space", Univ of Illinois Pr, (2000).

Author Biography

Pavel Bourov received MS in mathematics from the Novosibirsk State University (2000). Since 2005 he has worked in the UniqueIC's, Saratov, Russia.