DOI: 10.2478/v10051-010-0017-y

Factors Affecting Reading Speed Measurements of Coloured Web Pages

Mirko Gradišar[1,2], Tomaž Turk[1], Iztok Humar[3]

[1]University of Ljubljana, Faculty of Economics, Kardeljeva ploščad 17, 1000, Ljubljana, Slovenia,
miro.gradisar@ef.uni-lj.si (corresponding author)
[2]University of Maribor, Faculty of Organizational Sciences, Kidričeva 55a, 4000 Kranj, Slovenia
[3]University of Ljubljana, Faculty of Electrical Engineering, Tržaška 25, 1000 Ljubljana, Slovenia

Most of web-based systems use a fashion-driven graphical user interface design which does not necessarily provide the readers with high reading performance of colour variations of text and background. Many studies addressed this problem but none of them succeeded in offering complete and conclusive results in form of reading performance table which could be used in practice. The aim of this paper is to find reasons for these incomplete results. In our research, we firstly analyzed different experiment designs described in the literature and proposals for further research. Consequently, we tried to find an improved design and carried out an experiment involving 270 students who tested 30 web-safe colour combinations. However, our experiment also did not reveal statistically significant differences in reading speed. Therefore the multidimensional scaling (MDS) method was performed to show that the speed of reading cannot be described as a one-dimensional problem.

Keywords: Colour, Speed of Reading, User Interfaces, Web-Based system

1 Introduction

The acceptance of information when learning, creating, making decisions, and entertaining depends on its presentation (Bostrom and Kaiser, 1981, Sanders, 1993, McDowell et al., 1997, Resinovič et al., 1999, Suh, 1999). The importance of information presentation on a electronic visual display had not become evident until 1973 (Mason and Mitroff, 1973, Dyson, 2004). The systematic research on the role of a colour as an additional dimension of information presentation in computer based information systems started even two years afterwards (Christ, 1975, Teichner, 1979, Gremillion and Jenkins, 1981, Tullis, 1981, Ghani and Lusk, 1981, Silverstein,1982, DeSanctis, 1984, DeSanctis and Jarvenpaa, 1985, Benbasat et al., 1986), which can be supported by at least two reasons: (a) Colour is strongly incorporated into the system of human interactions with the environment. In its aesthetic function colour is much more effective than in the functionally rational category. Therefore there was no need for research into possible effects of colours on the capacity of human information processing for quite a long time. (b) The use of colour depended on the level of information technology development. In the first period the focus of the development was mostly directed towards technical and economic aspects. Ergonomic and personal aspects were neglected.

In the past ten years, due to the intensive development of the Internet, the presentation of information has gained key importance. Most web pages use a fashion-driven graphical user interface design with two main objectives: to attract attention of visitors and to reflect a graphical image of organization. Even though high readability and legibility (Connolly, 1998) of the presented information are rarely treated as important, many studies addressed these aspects, especially in technology enhanced web-based systems (Latchman et al., 1999, Casini et al., 2003) where readability is one of the most critical elements in comparison to printed materials. From the below given review of related work it is evident that the first research on the impact of colour combinations on visual performance was carried out by using printed material. More recent research has mainly focused on the effects of different colour combinations on the information presented on electronic visual displays. Firstly, let us introduce the terms of reading performance: readability, legibility, and reading speed.

1.1 Readability, legibility, and reading speed

Several definitions of readability and legibility exist. Readability, initially defined by Klare (1969), is later addressed in ISO 9241-3 (1992) as the characteristics of text which allows groups of characters to be easily discriminated, recognized and interpreted. Normally, it is concerned with continuous texts. Common measures of readability include identification of misspelled words, searching for pre-specified letters/words within word lists of passages, and reading rate. However,

since readability is considered to be a human psychological response, there are several factors influencing its performance. It is usually difficult to isolate these factors when measuring readability. One of the principal requirements for efficient readability is the legibility of the presented information. Legibility was originally defined by Tinker (1963) as the effect of all relevant text properties, such as type face and colour, on the visual processes involved in reading. ISO 9241-3 (1992) defines legibility in the limited sense as the visual properties of a character or symbol that determine the ease with which it can be recognized. In this sense, legibility is not related with continuous texts.

1.2 Reading performance of subtractive colours

The first research into what particular colour combinations on posters make them most visible from the distance was published by Le Courier, Sheldons Limited House in Leeds (Le Courier, 1912, Luckiesh, 1923), the posters printing company that performed an experiment in which different colour posters were put on wooden signs. Each poster contained two rows of letters. One row had well defined letters, the other had less defined letters like i, j. The posters were exposed to sunlight and a group of people was asked to rank the legibility of the letters, while reading the posters from different distances. Apparently, the most legible poster from the far distance was the poster with black letters on a yellow background. They tested thirteen colour combinations and got the results which were listed from the most legible to the least legible (known as Le Courier legibility table):

(1) black on yellow,
(2) green on white,
(3) red on white,
(4) blue on white,
(5) white on blue,
(6) black on white,
(7) yellow on black,
(8) white on red,
(9) white on green,
(10) white on black,
(11) red on yellow,
(12) green on red,
(13) red on green.

Surprisingly, the most widely used combination of printed text, black letters on white background, was only in the sixth position of legibility. The amount of difference between ranks was not given. Detailed statements regarding the colours and conditions of the experiment such as the number of subjects, kind of ink and paper used, size of type, line width, text used, etc., were omitted as well.

Between 1928 and 1963 Tinker and Paterson carried out a comprehensive research into speed of reading (Tinker and Paterson, 1929, Tinker, 1955, Tinker, 1963). Among other parameters of printed material they also studied the influence of colours. Ten colour combinations were used. Eight of them were comparable to the combinations from Le Courier table while two combinations resulted from the available coloured paper stocks. Students were tested with Chapman-Cook speed of reading test. The obtained results differ from the results of Le Courier table in five out of eight cases. The most important difference is the first place of the combination black on white and the fourth of the combination black on yellow. They stated that speed of reading does not depend on colour as such but on brightness differences. Despite the differences between legibility and readability tables, the main common characteristic is that both generally recommend dark characters on a light background.

1.3 Reading performance of additive colours

The additive colours of electronic visual displays have different optical characteristics than subtractive colours of printed texts: on Cathode Ray Tube (CRT) display an image is produced by an energized beam of electrons bombarding a thin layer of phosphor material. The beam is scanning through all pixels in the image, which results in a flickering picture on a CRT display.

Previous studies (Gould et al., 1987, Dillon, 1992) reported that image quality of additive-colour display was inferior to the subtractive-colour prints. It was shown that the workers performed tasks about 30% slower with CRT display than with paper. The workers also complained about visual fatigue and visual strain. Therefore, visual performances and user preferences of subtractive colours cannot be directly applied to additive colours, which motivates the research into the influence of colour combination on visual performance using electronic visual displays.

Although some of the early work (Radl, 1980, Pace, 1984) failed to identify specific colour combinations that are more readable than others on electronic visual displays, it was evident that colour combination of text and background was an important characteristic of visual stimuli that may affect visual performance. Further studies (Bruce and Foster, 1982, Murch, 1985, Matthews and Mertins, 1987) found that inappropriate use of colour can result in a poor performance and a higher incidence of visual discomfort. They suggested the avoidance of using red, green and blue in combination.

Some authors tried to explain the differences in visual performance merely by the luminance contrast. Bruce and Foster (1982) found positive correlation between luminance contrast and the rank order of reading speed. The hypothesis that reading ability is sensitive to luminance contrast and insensitive to chromatic contrast was also supported by the results of Legge and Rubin (1986). In an extensive experiment, Pastoor (1990) analysed a set of 18 colour combinations that were used to measure reading times and preference ratings. However, none of these studies proved statistically significant effect of colour combination on the speed of reading or visual search task.

The luminance contrast was the most important factor in the above mentioned studies, but there are other studies which investigate additional factors, such as chromatic contrast. Apparently, Travis et al. (1990) performed an experiment to investigate the influence of chromatic contrast. They employed 33 subjects to compare reading performance of 36 colour stimuli on white background by detecting given strings among words and nonsense anagrams presented on the screen for a short time. The results show that although the luminance contrast between the alphabetic string and the white background was zero, a near-perfect reading was still possible. This impor-

tant finding means that purely chromatic differences may be sufficient for the visual system to maintain word identification. Again, the results did not show statistical significance.

More recent studies have concentrated on the impact of colour combinations used on the web. Two experiments were conducted by Ling and Schaik (2002) and Pearson and Schaik (2003). In the first study they investigated effect of colour by employing twenty-nine participants rating and performing visual search of information in navigation bar. The combinations were black on white, blue on white, blue on yellow, yellow on blue, red on green and green on red. There was a significant effect of colour combination on accuracy and speed of searching, as well as on preference and perceived display quality. The green/red combination was relatively poor in terms of speed. Regarding the subjective data, blue on white was the best in terms of preference and perceived display quality. Lastly, Hall and Hanna (2004) examined the impact of Web page text-background colour combination on readability, retention, aesthetic and behavioural intention by measuring subjective opinion with questionnaires. Four colour combinations (black on white, white on black, light blue on dark blue and cyan on black) were ranked by thirty-six students answering five questions on a 10-point Likert scale. The major findings were: colours with greater luminance contrast generally lead to greater readability and colour combinations do not significantly affect retention.

With regard to the methods employed for measuring visual performance, the experiments in the mentioned studies can be classified in three groups. The first group consists of the experiments in which visual search tasks were performed (Pace 1984, Ling and Schaik, 2002, Pearson and Schaik, 2003). The results were statistically significant only if a small number (maximum six) of colour combinations was used.

The experiments in the second group (Bruce and Foster, 1982, Pastoor, 1990, Wu and Yuan, 2003) estimated the reading speed by measuring the time needed to read a text. However, the obtained results from this group did not show a significant effect of a colour combination on the speed of reading.

The third group comprises the experiments (Travis et al., 1990, Shieh et al., 1997, Wang and Chen, 2003) in which visual performance was measured as a percentage of correctly recognized characters or words. The tested stimuli were shown to the participants either in a relatively small size or for a very short time. Although these experiments were closer to the measurement of legibility than readability, there were still other psycho cognitive factors influencing the results. This group of experiments also does not offer statistically significant differences.

1.4 The aim of this study

With respect to the effects of colour on visual performance, the available results of all three groups of experiments in the above mentioned studies are inconclusive, as they neither provide a statistically proved and commonly accepted readability table for additive colour combinations, nor the explanation why the results are not statistically significant. Therefore, the aim of our study is to investigate why statistically significant results have not been reached yet.

One possible answer may be the inappropriateness of methods that were used. Consequently, we carefully analyzed different experiment designs described in the literature and proposals for further research. Namely, some authors finished their discussions by giving suggestions on how to improve their research methodology and proposed further research directions. For instance, Lin (2003) suggested further investigation of the visual performance with respect to both chromaticity and luminance contrast. In the guidelines for further work, Hall and Hanna (2004) pointed out that hues should be selected to better represent the wavelengths across the spectrum. On the basis of collected information we developed an improved method. Most of given suggestions were considered. As a reading performance measured in our study a reading speed was selected since reading is the most natural treatment of text. As a reading material, a sequence of meaningless syllables was used in order to minimize the influence of content on reading speed.

Another possible answer lies in the fact that there were not enough participants in an experiment. Pett and Wilson (1996) suggested that contrary to the previously performed research, statistically proved results might be achieved by carrying out an experiment with significantly more subjects. This suggestion was also taken into account. Our experiment involved 270 students who tested thirty most competitive web-safe colour combinations with the highest luminance contrast.

Unless the improved method mentioned above involving 270 participants brings statistically significant results, we need to conclude that a number of participants is still too low. However, it is practically impossible to involve considerably higher number of participants. Therefore we can create a following hypothesis:

The reading speed of a web text in different colour combinations displayed on CRT monitor cannot be described as a one-dimensional problem.

This implies that besides the physical characteristics of colour combinations, such as luminance contrast, colour difference and polarity, which can be controlled and studied separately, there are also many psychological factors influencing the reading speed. These factors differ greatly from a person to person and cannot be neutralized by an improved method and an acceptable number of participants in the experiment.

The remainder of the paper is organized as follows: the description of the experiment is followed by the results, discussion and final remarks.

2 Experiment design

Our study examined the factors which affect the readability of different colour combinations of text and background, presented on CRT display, with the measurement of speed of reading, similar to experiments performed in studies (Tinker and Paterson, 1929, Bruce and Foster, 1982, Pastoor, 1990, and Wu and Yuan, 2003).

2.1 Colour combinations

In their study, Hoadley and Jenkins (1987) found that solid colours without any patterning were the most effective to uti-

lize in multi-colour information presentations on CRT display. In order to be in accordance with this finding and to achieve the same presentation among different monitors and browsers, the colours used in our study were chosen from non-dithering web-safe colour palette (Lehn and Stern, 2000), which consists of 216 different colours. Although a very large number of colour combinations might be utilized in an experiment of this kind, it was necessary to limit the present study to a smaller number of well-defined colours. The colours chosen for the experiment were the elementary colours:

(1) white (hexadecimal red-green-blue (RGB) intensity value is #FFFFFF),
(2) yellow (#FFFF00),
(3) red (#FF0000),
(4) magenta (#FF00FF),
(5) blue (#0000FF),
(6) cyan (#00FFFF),
(7) green (#00FF00),
(8) black (#000000).

Each of these eight colours was combined with all other colours to make the 56 text/background colour combinations.

Having limited number of participants, the experiment was performed with thirty colour combinations of the highest luminance contrast, as it was found to be of a major importance in Foster (1982), Legge and Rubin (1986), and Pastoor (1990). Since evaluating all thirty colour combinations would have been too tiring for our participants, we decided to split the colour combinations ordered by ΔL into three sets of ten combinations. The black on white (B/W) combination was added to all three sets for a reference. B/W combination is also a part of the first set. Therefore, in the first set the B/W combination appeared twice. In the statistical analysis only the second of both results for B/W was taken into account.

Table 1 shows the colour combinations, their colour difference (ΔE) and luminance contrasts (ΔL) which are calculated in accordance with the model of colour space CIE L*a*b* proposed in 1976 by Commission Internationale de l'Éclairage (CIE 1986).

It should also be noted that initially all 56 colour combinations were tested with a small group of participants. Consequently, it was obvious that some colour combinations deviated significantly from the average performance. These combinations consist of colour pairs with low luminance contrast (white & yellow, cyan & green, red & magenta, and blue & black) and were thus not included in our study. In accordance with the conclusions of previous studies (e.g. Matthews and

Table 1: Colour combinations sorted by decreasing ΔL (coulours are visible in the internet version of the journal, http://versita.metapress.com/content/121156/)

no.	sample	text/bckg	lum. contrast Δ L	colour diff. Δ E
1	sample	white/black	100,00	100,00
2	sample	black/white	100,00	100,00
3	sample	yellow/black	98,00	136,05
4	sample	black/yellow	98,00	136,05
5	sample	cyan/black	91,00	105,39
6	sample	black/cyan	91,00	105,39
7	sample	green/black	88,00	143,34
8	sample	black/green	88,00	143,34
9	sample	white/blue	70,00	148,55
10	sample	blue/white	70,00	148,55
11	sample	yellow/blue	68,00	231,74
12	sample	blue/yellow	68,00	231,74
13	sample	cyan/blue	61,00	165,20
14	sample	blue/cyan	61,00	165,20
15	sample	magenta/black	60,00	126,63
16	sample	black/magenta	60,00	126,63
17	sample	green/blue	58,00	249,44
18	sample	blue/green	58,00	249,44
19	sample	red/black	54,00	119,90
20	sample	black/red	54,00	119,90
21	sample	white/red	46,00	116,52
22	sample	red/white	46,00	116,52
23	sample	yellow/red	44,00	108,97
24	sample	red/yellow	44,00	108,97
25	sample	white/magenta	40,00	118,47
26	sample	magenta/white	40,00	118,47
27	sample	magenta/yellow	38,00	192,23
28	sample	yellow/magenta	38,00	192,23
29	sample	cyan/red	37,00	161,30
30	sample	red/cyan	37,00	161,30

Mertins, 1987, Hall and Hanna, 2004), such colours perform significantly low and thus it is suggested to avoid their use for presentations on electronic visual displays.

2.2 Participants

In response to advertisements at the introductory course of Informatics in the first-year of studies at the University of Ljubljana, 300 students were recruited as volunteers. They consisted of 121 males and 179 females. The mean age of participants was 19 (ranging from 18 to 21). All participants had normal or corrected to normal visual acuity, and were tested with the Ishihara test for colour blindness to identify the participants with colour vision deficiencies (protanopia, deuteranopia, and tritanopia). Nine participants (six males and three females) failed on this test. The data were collected from them, but are not considered in this paper. All participants had at least basic computer experience. After collecting 270 valid results the experiment was terminated.

2.3 Apparatus, materials, environment and viewing conditions

To assure an adequate and equal testing environment for all participants, the viewing conditions were arranged in conformance with the ISO 9241-3 (1992) and ISO 12646 (2004) standards. The experimental tasks were presented on 21 Dell CRT display. The screen resolution was 1280 x 1024 pixels without interpolation and the refresh rate was 85 Hz (non-interlaced). Chromatic resolution was 32 bit. The chromaticity of white point was set to D50, gamma value to 2.0, and the luminance level of white point was greater than 120 cd/m2. The display was calibrated with the X-Rite Colour Monitor Optimizer (2004).

Following the standard, the ambient was neutral (light brown) with no areas causing glare or reflections on the monitor screen. The mean ambient illumination was below 300 lux. The only source of light in the room was a shielded lamp on the ceiling, while other sources of light had been curtained. The luminance values of the ambient were measured with a digital lux-meter.

The participants were seated in a position where the distance between the screen and the participants' eyes was 1 m. This is upper level of interval suggested by (Kroemer, 1993). The screen centre was slightly below the participants' eye-level, forming viewing angle of approximately 15°. The inclination of the monitor was 105°.

2.4 Procedure

We employed a very similar method to the one presented in Tinker and Paterson (1929). However, the colours used in their study were not well defined. The colour names such as green or red can incorporate a great variety of colour casts – from a light green to a dark one. The Chapman-Cook Speed of Reading tests (Tinker and Paterson, 1929-1946) had been slightly adapted for the measurement of reading speed from a electronic visual display. Instead of measuring a number of paragraphs and words read in a certain amount of time, the participants were requested to silently read a single fixed length paragraph (Figure 1). The time of reading was collected under the supervision of a tutor. Silent reading was selected after initial testing of the procedure with a small group of participants. It turned out that loud reading of unusual words may cause significant pronunciation problems.

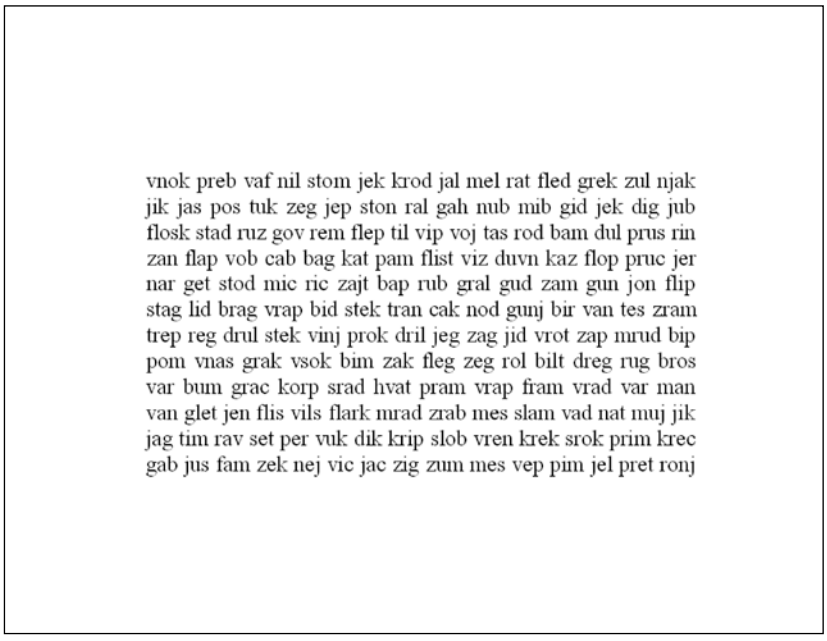


Figure 1. An example of the reading speed test page

The participants were divided in three groups of 90 people. Each group performed experiments by reading 10 colour combinations as well as the B/W combination for reference. A single participant tested a slightly different paragraph in each colour combination. All paragraphs were of the same length and consisted of the same collection of words but the word order was different in order to prevent the participants from memorizing the text after reading it several times and sharing the text content with future participants. A 10 x 10 Latin square determined what paragraph was read in what colour combination and in what sequence. In each group of 90 people, the Latin square was used nine times. With three groups, all thirty colour combinations were tested.

Before the experimental session, participants had been thoroughly explained the system and had performed a practice session under the supervision of a tutor. The participants were then asked to read the paragraph as quickly and as accurately as possible.

The experiment, developed as a web-based application with java on client side and database on the server side (Humar and Gradišar, 2003), consisted of two parts. The first part was a colour vision test. The second part consisted of reading colour combinations starting with B/W for each participant. Before the paragraph was read in a particular colour combination, the whole electronic visual display was coloured in the colour of the combination background. After pressing a key, the timer was started and the paragraph was displayed in the middle of the screen. The participants started to read the text as soon as it was displayed. Having finished reading, they pressed the key again to stop the timer. The time needed for reading the paragraph in each colour combination was collected automatically for all users and stored into the database. After finishing one test and before starting a new one, the participants were exposed to a screen with a relaxing non-glaring grey colour (10 cd/m^2) to neutralize the effect of the previous colour combination. When ready for the next test, they pressed a key. The participants took approximately 15 min to complete the experiment. Following completion, the participants were thanked and then fully debriefed.

2.5 Data Analysis

In order to test the hypothesis, multidimensional scaling was used (MDS; Davidson 1983). This technique is similar to the principal components analysis - its goal is to detect meaningful underlying factors that allow exploring a chosen phenomena. The basis for factor exploration with MDS is just one variable which is the "distance", or dissimilarity between stimuli. In our case, the colour combinations can be regarded as stimuli, and distances between them are estimated as differences in measured reading speeds between colour combinations. The MDS algorithm tries to arrange the stimuli in a multidimensional space in such a way that distances are preserved as much as possible. The obtained dimensions in this kind of configuration can be seen as factors that influence (and explain) the ordering of stimuli within space. However, there are two obstacles of this analysis in comparison to the principal components analysis. The first is that the axes are, in themselves, meaningless and the second is that the orientation of the picture is arbitrary. Unfortunately, principal components analysis is possible when we have a reach set of observed variables.

In this way, the number of statistically significant factors can be obtained. Since other non-significant factors could be present in every experiment, the fit of measured distances to statistically significant factors is usually not ideal. The quality of the fit can be estimated in different ways. The most common way is to use STRESS measures (like a Phi value or a coefficient of alienation), which are calculated as a sum of squared deviations of the observed distances from the reproduced distances. (For instance, the raw stress value Phi is estimated by firstly transforming the measured distances by a monotone transformation function.) If these measures show that the fit is relatively poor, one can increase the number of dimensions in the space and engage MDS to arrange colour combinations to a better fit.

One can decide on the number of factors by using the STRESS value. For instance, when increasing the number of factors, when one achieves a Phi value smaller than 0.05, additional factors are usually regarded as nonsignificant. A scree plot can also help when deciding about the number of significant factors. In the scree plot, the stress value is plotted against different numbers of dimensions. The cut-off point is normally chosen where the smooth decrease of stress values appears to level off to the right of the plot.

Besides a goodness of fit, the MDS technique allows one to use Shepard diagram which shows the reproduced distances (on vertical axis) for a particular number of dimensions, against the observed input data (measured distances, shown on horizontal axis). Shepard diagram also shows a step-function from D-hat values, which are monotone transformations of the input data. If all reproduced distances fall onto the step-line of D-hat values, then the rank-ordering of distances would be perfectly reproduced by the respective solution (dimensional model).

3 Results

The average reading times in seconds reduced by the average reading time of the referential B/W combination (Tinker and Paterson, 1929) are shown in Table 2. The analysis of variance was performed without giving any statistically significant differences in the average reading times between all thirty tested colour combinations.This implies that despite an improved method and an increased number of participants the readability table cannot be offered to the practice. Consequently, we performed a MDS analysis of the obtained results which explained why statistically significant differences were not found.

For the MDS analysis, we firstly need to define distances between colour combinations. The distances between two colour combinations in our case are the differences in speed of reading between these colour combinations. The measure of dissimilarity (distance) between i-th and j-th colour combination is d_{ij}. d_{ij} is the absolute difference in reading times for a stimulus pair (i, j). Obviously $d_{ij} = d_{ji}=0$ when i = j. The obtained distance was then averaged on the participants, thus establishing the n-dimensional diagonal square matrix of data

Table 2: The average reading times in seconds reduced by average reading time of the referential B/W combination

rank	sample	text/bckg	avg time - - avg time ref B/W
1	sample	blue/white	-0,692
2	sample	white/black	-0,437
3	sample	red/white	-0,236
4	sample	black/magenta	-0,034
5	sample	white/blue	-0,022
6	sample	black/white	0,000
7	sample	green/black	0,061
8	sample	cyan/blue	0,107
9	sample	magenta/black	0,170
10	sample	magenta/white	0,468
11	sample	blue/yellow	0,585
12	sample	black/green	0,787
13	sample	cyan/black	0,909
14	sample	blue/cyan	1,006
15	sample	black/cyan	1,244
16	sample	white/red	1,555
17	sample	black/yellow	1,575
18	sample	white/magenta	1,587
19	sample	yellow/blue	1,785
20	sample	red/cyan	1,824
21	sample	red/black	1,968
22	sample	green/blue	2,319
23	sample	yellow/red	2,390
24	sample	yellow/black	2,407
25	sample	yellow/magenta	2,488
26	sample	red/yellow	2,589
27	sample	blue/green	2,945
28	sample	magenta/yellow	3,179
29	sample	black/red	3,832
30	sample	cyan/red	6,411

where each cell represented the average absolute difference in reading times for a colour combination (i, j), and where n is in our case the number of different colour combinations. Since we have experiments in three sets, there are three separate sets of results. As mentioned above, the first og both results for B/W in the first set was not taken into account, Therefore n = 10 for the first group and n = 11 for the second and third. The number of distances/dissimilarities when n = 10 is thus 45 and 55 in second and third group.

Firstly, we calculated STRESS values Phi and coefficient of alienation as fit measures for different dimension setups. According to Table 3, for all three groups of experiments we chose at least five-dimensional setup to reduce Phi below 0,05, which helped us achieve relatively low values for the coefficient of alienation, too. Both measures for all dimension setups are shown in Table 3. Figure 2 shows the graphic presentation of Phi for different dimension setups (the scree plot).

Figures 3 to 5 represent Shepard diagrams of five dimensional setups for three groups of experiments. Most of the points are clustered around the step-line. In our case, the reproduced distances are close to the step-line, which suggests a relatively good fit but with five dimensions.

Tables 4, 5 and 6 show the results from five-dimensional setups, for each experiment group separately. The dimensions

Table 3: Coefficients of alienation and Phi values for all dimension setups and sets of experiments

dim.	colour set 1		colour set 2		colour set 3	
	coeff. of alien.	phi	coeff. of alien.	phi	coeff. of alien.	phi
1	0,546	0,432	0,525	0,425	0,516	0,415
2	0,297	0,226	0,322	0,247	0,302	0,237
3	0,211	0,148	0,215	0,149	0,201	0,134
4	0,132	0,078	0,132	0,060	0,129	0,083
5	0,080	0,047	0,034	0,034	0,065	0,039

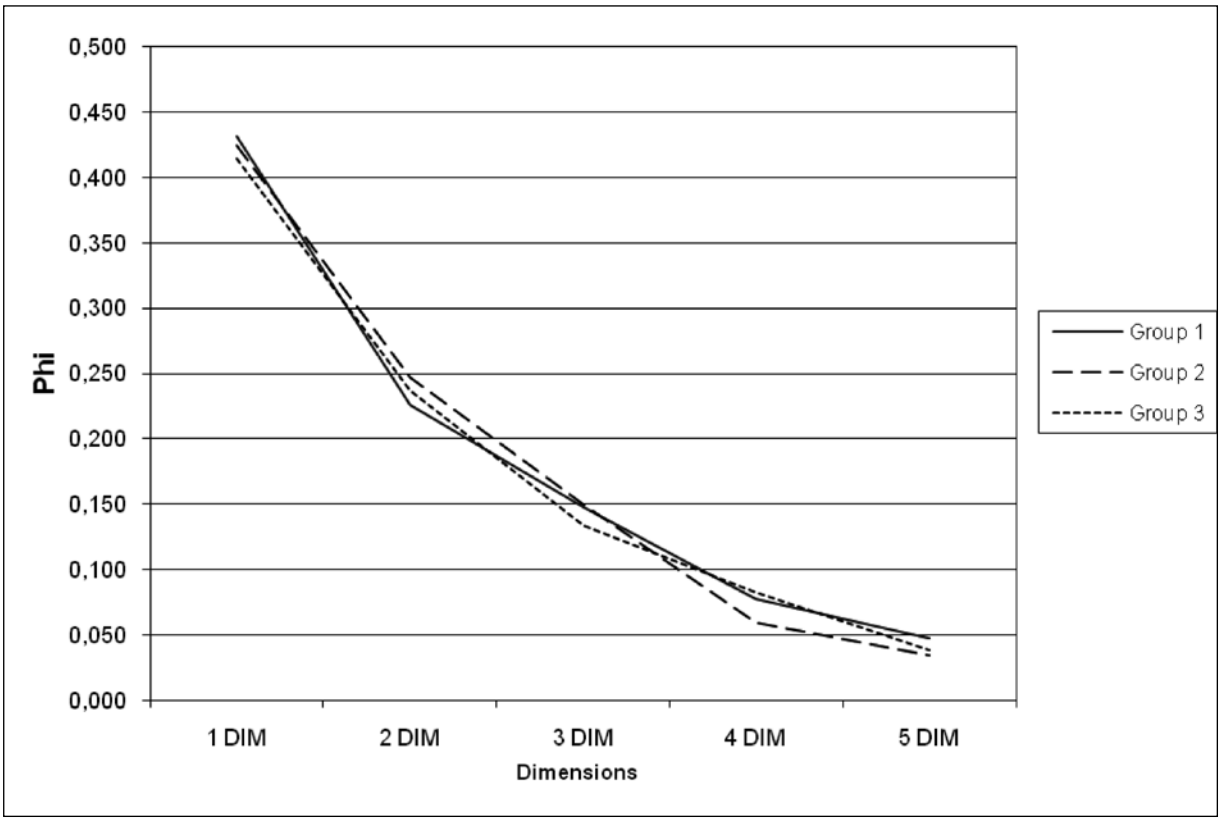


Figure 2. Phi values and number of dimensions for the MDS of colour combinations processing times

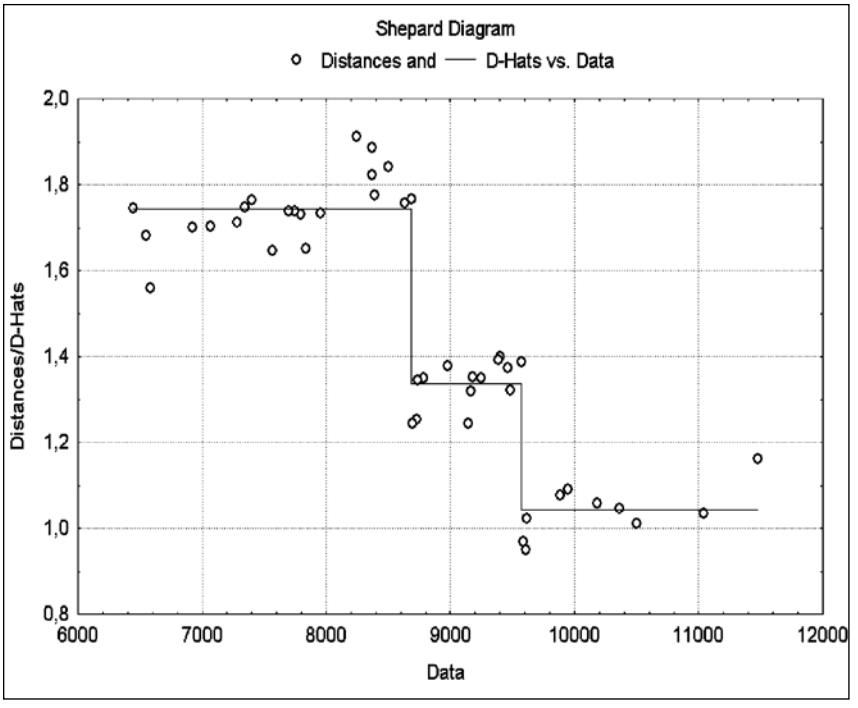


Figure 3. Shepard diagram for five-dimensional MDS solution for the first group of experiments

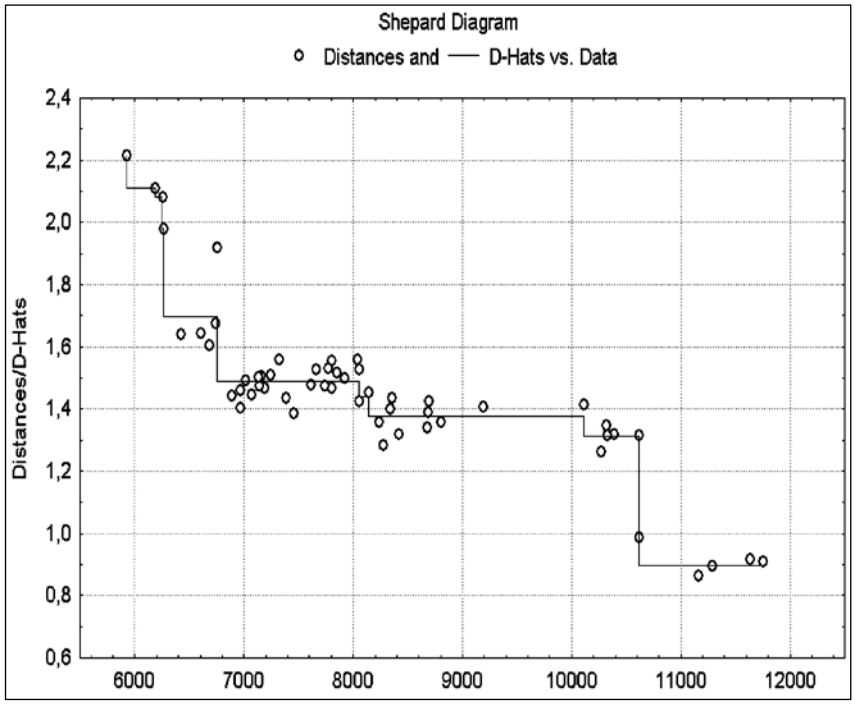


Figure 4. Shepard diagram for five-dimensional MDS solution for the second group of experiments

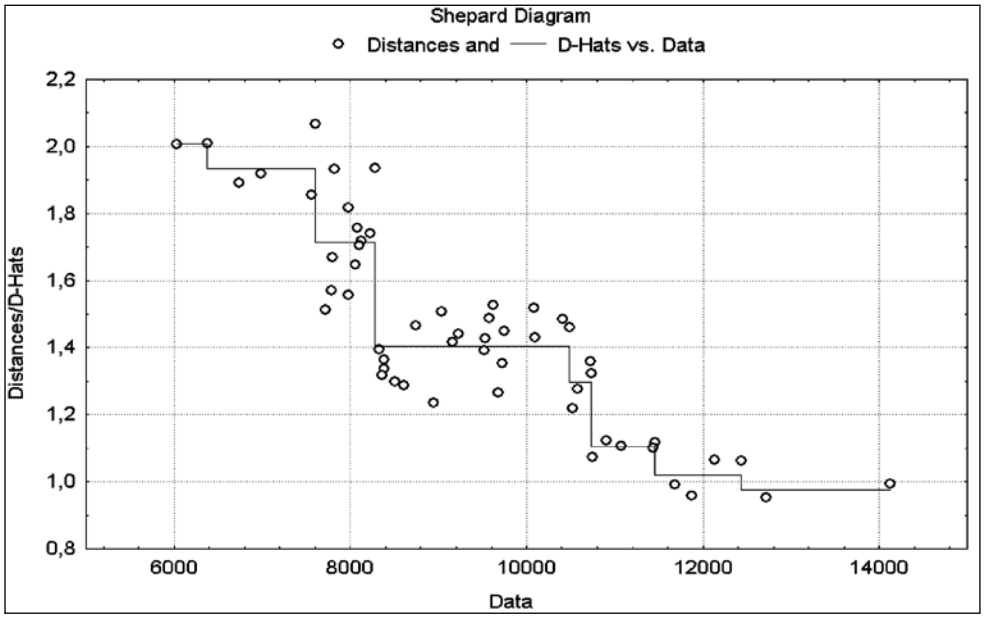


Figure 5. Shepard diagram for five-dimensional MDS solution for the third group of experiments

are ranked from 1 to 5 according to their strength of distinguishing between different colour combinations. Colour combinations are arranged according to the values of the first dimension which is the strongest factor (it distinguishes the colour combinations better than other four dimensions).

A power analysis was performed for this setup. The average reading time was estimated at 16.18 seconds with 4.21 seconds standard deviation. If we assume that the maximum difference between colour combinations is around two seconds, the sample size of 90 participants in each group is large enough, since the analysis shows that the power of 0.8953 can be expected for the effect size of 0.475 with a = 0.05.

Table 4: Five-dimensional colour combinations scale: first group of experiments

dimensions					text/bckg	sample
1	2	3	4	5		
-0,817	-0,346	-0,176	0,251	-0,693	black/yellow	sample
-0,709	0,297	-0,028	0,506	0,247	blue/white	sample
-0,667	-0,374	0,131	-0,377	0,447	cyan/black	sample
-0,209	0,699	-0,178	-0,695	0,118	white/black	sample
0,013	-0,119	-0,275	-0,017	0,092	black/white	sample
0,345	0,355	0,685	-0,371	-0,554	black/green	sample
0,389	-0,581	0,676	-0,069	0,246	white/blue	sample
0,412	0,368	0,464	0,720	0,143	yellow/black	sample
0,591	0,434	-0,758	0,123	0,031	green/black	sample
0,653	-0,734	-0,540	-0,071	-0,078	black/cyan	sample

Table 5: Five-dimensional colour combinations scale: second group of experiments

dimensions					text/bckg	sample
1	2	3	4	5		
-1,041	-0,114	-0,333	0,182	-0,072	yellow/blue	sample
-0,341	-0,284	0,765	-0,154	0,403	black/magenta	sample
-0,280	0,209	0,024	-0,105	0,220	black/white	sample
-0,181	0,865	0,271	0,302	-0,199	cyan/blue	sample
-0,135	0,287	-0,074	-0,863	-0,220	black/red	sample
0,001	-0,336	0,466	0,019	-0,783	green/blue	sample
0,081	0,120	-0,389	0,014	0,872	blue/green	sample
0,255	-1,022	-0,325	-0,257	-0,002	blue/cyan	sample
0,321	0,217	-0,940	0,197	-0,383	magenta/black	sample
0,333	-0,249	0,230	0,933	0,074	red/black	sample
0,986	0,306	0,305	-0,269	0,089	blue/yellow	sample

Table 6: Five-dimensional colour combinations scale: third group of experiments

dimensions					text/bckg	sample
1	2	3	4	5		
-1,100	0,211	0,050	-0,097	0,131	red/yellow	sample
-0,303	-0,798	-0,165	0,515	-0,097	red/white	sample
-0,236	-0,015	-0,655	-0,685	-0,371	magenta/white	sample
-0,193	0,283	0,936	0,441	0,086	yellow/magenta	sample
-0,084	-0,104	-0,433	-0,048	0,477	cyan/red	sample
0,001	-0,632	0,619	-0,649	0,110	red/cyan	sample
0,005	0,149	-0,260	0,689	-0,666	magenta/yellow	sample
0,041	0,003	0,052	-0,055	-0,016	black/white	sample
0,160	0,867	-0,295	0,067	0,500	white/red	sample
0,689	0,436	0,341	-0,419	-0,493	yellow/red	sample
1,021	-0,399	-0,190	0,241	0,340	white/magenta	sample

4 Discussion

This study investigated the impact of colour combinations on the reading speed of web page text presented on a CRT display. Reading speed was measured through the time needed to read a fixed length paragraph, as it has been performed in previous studies (Tinker and Paterson, 1929, Bruce and Foster, 1982, Pastoor, 1990, and Wu and Yuan, 2003). The experiment was designed in conformance with the ISO 9241-3 (1992) and ISO 12646 (2004) standards. We tried to improve the design also by using the experience from similar previous studies to avoid some of their drawbacks and by taking into account the suggestions for future research especially regarding the number of participants, the number of colour combinations and their selection in such a way that they were distributed over the entire spectrum. In spite of that we did not get the results with statistically significant differences in reading speed between colour combinations.

The results presented in Table 2 are in agreement with previous studies, which investigated smaller sets of colour combinations. The results support the suggestion of Bruce and Foster (1982) not to use green and blue or red and cyan in

combination. They are also in agreement with study of Pearson and Schaik (2003), who preferred blue-on-white combination against red-on-white. The results mostly support the succession of four colour combinations, tested by Hall and Hanna (2004), although black & white combinations are in inversed order.

Many studies addressed the problem of measuring the reading performance. Untill now none of them has succeeded in offering complete and conclusive results in form of readability table which would include a large set of colour combinations and where the differences between them would be statistically significant. The goal of this research is not to compare the results of our experiments with other studies into more depth to find differences and similarities. Our intention was either to get the results with statistically significant differences between colour combinations on the basis of the improved experiment design and higher number of participants or in other case to use obtained results for further statistical investigation in order to find the reasons for being unsuccessful.

The speed of reading depends on a colour combination, font selection, type size, type rendering technology, etc. (Le Courrie, 1912, Luckiesh, 1923, Tinker and Paterson, 1929, Boyarski et al., 1998). Since our goal was to explore the association between the speed of reading and colour combination, we prepared the experiment where other factors were fixed. However, there are many psychological factors influencing the speed of reading, e. g. aesthetics or possible different interpretation of the instructions for the participants. Variations of psychological factors could be theoretically neutralized by an increased number of participants. In our case there were 270 participants. In comparison to other studies, this number can be considered very high, but it was nevertheless not high enough to provide significant results. Therefore the intention of further statistical analysis was to determine the presence of other factors besides colour combinations which affect reading speed and which were not neutralized by a given number of participants.

MDS was selected as the most appropriate method since the goal of MDS is to detect meaningful underlying factors that allow a researcher to explore observed dissimilarities between stimuli which are colour combinations in our case. MDS attempts to arrange colour combinations in a space with a particular number of dimensions so as to reproduce the observed distances. Since the experiment in our case was carried out in three groups, the MDS analysis was performed three times.

The results of the MDS analysis for each group of experiments show that there are at least five factors having influence on the speed of reading. The reader should note that the results for all three groups are very similar. Also the reported values of fit measures are similar and not in favour of a one-dimensional setup where the speed of reading would depend only on the colour combination. It seems that in spite of a relatively high number of participants, reading speed still depends on a mixture of factors. According to the given results, a hypothesis that the speed of reading web text in different colour combinations presented on electronic visual display cannot be described as a one-dimensional problem is supported because a drastic increase in the number of participants from several hundred to maybe several thousand would be practically almost impossible.

These results can also be used to explain the reason why previous studies (Bruce and Foster, 1982, Pastoor, 1990, Wu and Yuan, 2003) failed to find statistical differences in reading speed for larger groups of additive colour combinations distributed over the entire spectrum.

5 Final remarks

Reading speed of thirty most competitive colour combinations with the highest luminance contrast, selected out of 56 combinations composed of eight elementary web-safe colours was tested. The selected colour combinations differ in the luminance contrast, colour difference and polarity. Luminance contrast was used as a selection criterion because it affects reading speed more than colour difference.

The aim of the study was to propose a readability table with statistically significant differences between colour combinations or to find out why this was not possible.

The obtained results show that despite the improved experiment and the higher number of used subjects there are no statistically significant differences in reading speed between thirty colour combinations. To find out why, the MDS method was used. We noticed at least five factors which simultaneously and differently affect reading speed of a coloured text. It would be very difficult if not impossible to identify them and to design a new experiment in such a way that these factors would be neutralized and statistical significance would be reached within an acceptable number of participants. Such a result is not in accordance with some findings of previous authors which suggest that statistically proved results might be achieved by carrying out an experiment with significantly more subjects.

Even though we are not able to find out what variable individual factors represent, we can at least create a hypothesis on the meaning of these factors:

Besides the physical characteristics of colour combinations, such as luminance contrast, colour difference and polarity, which can be controlled and studied separately, there are also many psychological factors influencing the speed of reading. These factors are:

- different understanding of instructions especially the part which says: read the text thoroughly and as fast as possible
- psychological stress caused by fear that the participant will not be able to complete the task properly
- unconscious attempts in trying to understand the meaning of the text
- unconscious attempts in trying to figure out the context
- different perceptions of aesthetics of the text.

One possible and approximate solution to the described problem would be in limiting the research to visibility/legibility of colour combinations. It can be assumed that visibility is the most important common factor which influences reading speed and is independent of aesthetics, content, context etc. of text.

Acknowledgements

The authors are grateful to Professor Milton A. Jenkins from the University of Baltimore for his help and support.

This research was supported by the Ministry of Higher Education, Science and Technology of the Republic of Slovenia under grant No. SLO-USA-2002/38 and grant No. P2-0037.

References

Benbasat, I., Dexter, A. S. and Todd, P., 1986, The Influence of Colour and Graphical Information Presentation in a Managerial Decision Simulation, *Human-Computer Interaction*, 2(1), pp. 65-92, DOI: 10.1207/s15327051hci0201_3.

Bostrom, R.P. and Kaiser, K.M., 1981, Personality Differences within System Project Teams – Implications for Designing Solving Centers, in *Proceedings of the XVIII. Computer Personnel Research Conference*, New York, ACM, DOI: 10.1145/800051.801855.

Boyarski, D., Neuwirth, C., Forlizzi, J. and Harkness, S.R., 1998, A Study of Fonts Designed for Screen Display, in *CHI 98*, 18 – 23 April, pp. 87–94, DOI: 10.1145/274644.274658.

Bruce, M. and Foster, J.J., 1982, The Visibility of Coloured Characters on Coloured Back-grounds in Viewdata Displays, *Visible Language*, **16**(4), pp. 382–390.

Casini, M., Prattichizzo, D. and Vicino, A., 2003, The Automatic Control Telelab: A User-Friendly Interface for Distance Learning, *IEEE Transactions on Education*, 46(2), pp. 252–257, DOI: 10.1109/TE.2002.808224.

Christ, R. E., 1975, Review and analysis of colour coding research for visual displays, *Human Factors*, 17, pp. 542-570.

CIE Pub. 15.2., 1986. *Colourimetry*, 2nd Ed. Vienna: CIE Central Bureau.

Connolly, K., 1998, *Legibility and Readability of Small Print. Effects of Font, Observer Age and Spatial Vision*, A Master of science Thesis, Department of Psychology, Calgary Alberta.

Davidson, M.L., 1983, *Multidimensional Scaling*, John Wiley & Sons, New York.

DeSanctis, G., 1984, *Computer Graphics as Decision Aids*, Working paper, University of Minnesota, Minneapolis.

DeSanctis, G. and Jarvenpaa, S.L., 1985, An Investigation of the Tables Versus Graphs Controversy, in *Proceedings of the VI. International Conference on Information Systems*.

Dillon, A., 1992, *Reading from paper versus screens, a critical review of empirical literature*, Taylor and Francis Ltd.

Dyson, M.C., 2004 , How physical text layout affects reading from screen, *Behaviour & Information Technology*, 23(6), pp. 377–393, DOI: 10.1080/01449290410001715714.

Ghani, J. and Lusk, E.J., 1981, *The Impact of Information Presentation and Modification on Decision Performance*, Working paper, The Sloan School, MTT.

Gould, .J.D., Alfaro, L., Barnes V., Finn, R., Grischkowsky, N. and Minuto, A., 1987, Reading is slower from CRT displays than from paper: attempts to isolate a single-variable explanation, *Human Factors*, **29**(3), pp. 269–299.

Gremillion, L.L. and Jenkins, A.M., 1981, *The Effects of Colour Enhanced Information Presentations*, Discussion Paper #173, Indiana University, Bloomington, Indiana.

Hall, R. and Hanna, P., 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. *Behaviour & Information Technology*, **23**(3), 183–195, DOI: 10.1080/01449290410001669932.

Hoadley, E. and Jenkins, A.M., 1987, *The effects of colour on performance in an information extraction task using varying forms of information presentation: Pilot studies*, IRMIS Working Paper #W713, Bloomington, Indiana University Institute for Research on the Management of Information Systems, Graduate School of Bussiness.

Humar, I. and Gradišar, M., 2003, Colour test, Available online at: http://spin.fe.uni-lj.si/colours/ (accessed 1. 4. 2004)

ISO 9241-3., 1992, *Ergonomic requirements for office work with visual display terminals (VDTs), Part 3: Visual display requirements*. International Organization for Standardization.

ISO 12646, 2004, *Graphic technology – Displays for colour proofing – Characteristics and viewing conditions*. International Organization for Standardization.

Klare, G.R., 1969, *The Measurement of Readability*, Ames, Iowa: The Iowa State University Press, pp. 1–2.

Kroemer, K.H.E., 1993, Locating the Computer Screen: How High, How Far?, *Economics in Design*, pp. 7–8.

Lehn, D. and Stern, H., 2000, Death of the Web-safe Colour Palette?, Available online at: http://webmonkey.wired.com/webmonkey/00/37/index2a.html?tw=desin (accessed 15. 10. 2005)

Latchman, H.A., Salzmann, Ch., Gillet, D. and Bouzekri, H., 1999, Information Technology Enhanced Learning in Distance and Conventional Education, *IEEE Transactions on Education*, 42(4), 247–254, DOI: 10.1109/13.804528.

Le Courier du Livre, 1912, Lisibilite des affiches en couleurs, Sheldons Limited House, *Cosmos*, Sept. 5, pp. 255.

Legge, G.E. and Rubin, G.S., 1986, Psychophysics of reading: IV. Wavelength effects in normal and low vision, *Journal of the Optics Society of America*, A(3), 40–51.

Lin, C.C., 2003, Effects of contrast ratio and text colour on visual performance with TFT-LCD, *International Journal of Industrial Ergonomics*, 31, pp. 65–72, DOI: 10.1016/S0169-8141(02)00175-0.

Ling, J. and Schaik, P., 2002, The effect of text and background colour on visual search of Web pages, *Displays*, 23, pp. 223–230, DOI: 10.1016/S0141-9382(02)00041-0.

Luckiesh, M., 1923, *Light and Colour in Advertising and Merchandising*, H. Van Nostrand Co., New York, pp. 246–251.

Mason, R.O. and Mitroff, L.L., 1973, A Program for Research on Management Information Systems, *Management Science*, 19(5), DOI: 10.1287/mnsc.19.5.475.

Matthews, M.L. and Mertins, K., 1987, The influence of colour on visual search and subjective discomfort using CRT displays. In *Proceedings of the Human Factors Society 31st Annual Meeting*, Santa Monica, CA: HFES, pp. 1271–1275.

McDowell, D., Kodak, E. and Warter, L., 1997, Viewing Conditions, *The Prepress Bulletin*.

Murch, G.M., 1985, Colour graphics – Blessing or ballyhoo?, *Computer Graphics Forum*, 4, pp. 127–135.

Pace, B.J., 1984, Colour combinations and contrast reversals on visual display units, In *Proceedings of the Human Factors Society 28th Annual Meeting*, Santa Monica, CA: HFES, pp.326–330.

Pastoor, S., 1990, Legibility and subjective preference for colour combinations in text, *Human Factors*, 32(2), pp. 157–171.

Pearson, R. and Schaik, P., 2003, The effect of spatial layout of and link colour in web pages on performance in a visual search task and an interactive search task, *International journal on Human-Computer studies*, 59, pp. 327–353, DOI: 10.1016/S1071-5819(03)00045-4.

Pett, D. and Wilson, T., 1996, Colour research and its application to the design of instructional materials, *Educational Technology Research & Development*, 44(3), pp. 19-35, DOI: 10.1007/BF02300423.

Radl, G.W., 1980, Experimental investigations for optimal presentation-mode and colours of symbols on the CRT-screen, In *Ergo-*

nomic aspects of visual display terminals, E. Grandjean & E. Vigliani (Ed.), London: Taylor & Francis, pp. 127–135.

Resinovič, G., Jenkins, M.A., Gradišar, M. and Jaklič, J., 1999. *Legibility and visibility issues in colour enhanced information presentation*, Working paper No. 88, University of Ljubljana, Faculty of Economics.

Sanders, M.S. and McCormick, E.J., 1993, *Human factors in engineering and design*. Singapore: McGraw-Hill.

Shieh, K.K., Chen, M. and Chuang, J.H., 1997, Effects of Colour Combinations and Typography on Identification of Characters Briefly Presented on VDT's, *International Journal of Human-Computer Interaction*, 9(2), pp. 169–181, DOI: 10.1207/s15327590ijhc0902_5.

Silverstein, L. D., 1982, Human factors for colour CRT displays. In Proceedings of the *Society for Information Display*: Seminar Lecture Notes, 2, pp. 1-41.

Suh, K.S., 1999, Impact of Communication Medium on Task Performance and Satisfaction: An Examination of Media-richness Theory, *Information & Management*, 35(5), 10.1016/S0378-7206(98)00097-4.

Teichner, W. R., 1979, Colour and visual information coding. In Proceedings of the *Society for Information Display*, 20(1), pp. 3-9.

Tinker, A.M., 1963, *Legibility of Print*, Ames, IA: Iowa State University Press.

Tinker, A.M., 1955, *Speed of Reading Test*, University of Minnesota Press, Minneapolis.

Tinker, A.M. and Paterson, G.D., 1940, *How to make type readable: A manual for typographers, printers, and advertisers, based on twelve years of research involving speed of reading tests given to 33,031 persons*, Harper & Brothers, New York.

Tinker, A.M. and Paterson, G.D., 1929, Studies of typographical factors influencing speed of reading: VI. Black type versus white type, *The Journal of Applied Psychology*, 13(2), pp. 241–247.

Tinker, A.M. and Paterson, G.D., 1929, Studies of typographical factors influencing speed of reading, VII. Variations in colour of print and background, *The Journal of Applied Psychology*, 13(2), pp. 471–479.

Travis, D.S., Bowles, S., Seton, J. and Peppe, R., 1990, Reading from colour displays: A psychophysical model, *Human Factors*, 32(2), pp. 147–156.

Tullis, T. S., 1981, An evaluation of alphanumeric, graphic and colour information displays. *Human Factors*, 23(5), pp. 541-550.

Wang, A. H. and Chen, C. H., 2003, Effects of screen type, Chinese typography, text/background colour combination, speed, and jump length for VDT leading display on users' reading performance, *International Journal of Industrial Ergonomics*, 31(4), pp. 249-261, DOI: 10.1016/S0169-8141(02)00188-9.

Wu, J. H. and Yuan, Y. F., 2003, Improving searching and reading performance: The effect of highlighting and text colour coding, *Information & Management*, 40(7), pp. 617-637, DOI: 10.1016/S0378-7206(02)00091-5.

–, Monitor Opimizer Software Guide, X-Rite Colourimeter, Available online at: http://www.xrite.com/Products/Product.asp?Show=Description&id=11 (accessed 14. 1. 2005)

Mirko Gradišar is full professor at the University of Ljubljana, Faculty of Economics. His areas of expertise include development of information systems, human-computer interactions, optimization of business processes and operations research. As author and coauthor he published 50 scientific articles and seven textbooks.

Iztok Humar received B.Sc., M.Sc. and PhD degrees in the field of Telecommunications at the Faculty of Electrical Engineering, University of Ljubljana, Slovenia, in 2000, 2003, and 2007 respectively. He also received PhD in the field of Information Management at the Faculty of Economics, University of Ljubljana, Slovenia, in 2009. Currently, he is an Assistant professor at the Faculty of Electrical Engineering, Slovenia. As a part of his research work, he analyses the impact of color combinations on legibility for different modern types of information systems' displays. He is involved in many research and national industry projects as well in projects funded by EU FP and structural and Cohesion EU funds. Dr. Humar is author and coauthor more than 50 journal and conference papers.

Tomaž Turk is an economist and holds a Ph.D. in Information Sciences. He is Associate Professor and researcher at the Faculty of Economics of the University of Ljubljana. He holds courses on Development of Information Systems, Economics of Information Technology, Economics of Telecommunications, and Business Simulations. Currently his research work includes themes like information technology adoption, economics of information technology, communication networks management and Internet society issues. He has participated in several national and international projects and published over 50 papers/book chapters, including papers in Technology Forecasting & Social Change, Computer Standards and Interfaces, Mathematics and Computers in Simulation, International Journal of Industrial Ergonomics, Computer Communications, and Telecommunications Policy. He is Vice Chair of the European Commission funded research project COST Action 298 'Participation in the Broadband Society'.