



A color selection tool for the readability of textual information on Web pages

Silvia Zuffi¹, Giordano Beretta, Carla Brambilla²
HP Laboratories Palo Alto
HPL-2005-216
November 28, 2005*

readability, color,
Web design,
automatic layout

One of the issues in Web page design is the selection of appropriate combinations of background and foreground colors to display textual information. Colors have to be selected in order to guarantee legibility for different devices, viewing conditions and, more important, for all the users, including those with deficient color vision. In this paper we present a tool to select background and foreground colors for the display of textual information. The tool is based on the Munsell Book of Colors; it allows the browsing of the atlas and indicates plausible colors based on a set of legibility rules, which have been defined experimentally.

* Internal Accession Date Only

¹ITC - CNR Italian National Research Council

²IMATI – CNR Italian National Research Council

Published in and presented at the Proceedings 6061 of Internet Imaging VII, 15-19 January 2006, San Jose, CA, USA

Approved for External Publication

© Copyright 2005 Society of Photo-Optical Instrumentation Engineers

1. INTRODUCTION

One of the issues in Web page design is the selection of appropriate combinations of background and foreground colors to display textual information. Colors have to be selected in order to guarantee legibility of the text for many displays and environments, and, more important, for all users, including those with color vision deficiencies (CVD). In general, every color selection task requires taking into account the characteristics of the display device and viewing conditions. At the design stage, color selection must be performed on a calibrated device under controlled viewing conditions; when the Web page is displayed on the user's display, color reproduction issues, such as device characterization and color management based on the ICC standardization, should be considered. A display device is characterized by its color space and gamut, including the luminance level, the white point, and the gamma. The viewing conditions, which affect color appearance, are characterized by the screen background, the surround, the proximal field, the ambient illuminance level and white point, and the veiling glare. Specifically with the purpose to provide an alternative strategy to the adoption of ICC color reproduction solutions on the Internet, all the information on device and viewing conditions listed above have been defined for a generic display of average characteristics under typical viewing conditions, defining the sRGB color space¹. One may use this standard color space for color specification for the Web.

In the case of color selection for textual information display, there is a further issue to consider in addition to hardware and environment characterization — while a user's device characteristics and viewing conditions may be accounted for exploiting ICC solutions or standard color spaces, the problem of legibility for those who have deficient color vision still holds, and may strongly compromise the access to textual information on Web sites. Addressing this issue is of great significance to ensure accessibility; to improve the legibility of a Web page users may try to correct their monitor calibration settings or set better viewing conditions, but cannot discriminate colors if they don't have a normal color vision. CVD, in fact, affects a significant percentage of the male population (about the 8%) including anomalous trichromacy, where the peak sensitivity of a cone type is shifted compared to normal color vision, dichromacy, where a cone type is missing, and achromatopia, where no cones are present. Many countries have accessibility laws requiring readability by users with CVD. For anomalous trichromats it is possible to manipulate the device characteristics so they achieve Farnsworth-Munsell 100 hue (FM-100) test scores similar to those of normal trichromats². Thus, to comply with the accessibility requirements, one can focus on dichromats and achromats.

In this paper we address the problem of selecting colors for the display of textual information accounting for device characteristics, viewing conditions and CVD. The aim of our study is the

definition of readability rules to guide the Web designer in the selection of colors. We use the terms legibility and readability with the following meaning: “legibility” refers to whether it is possible to read a document and it has mostly to do with the tolerances of the device rendering the document; “readability” refers to how comfortable a document is to read and it can be measured in time to read a text without errors or the number of errors in text read at a given speed³.

Many studies have been performed to assess the criteria for comfortable reading of text in many scenarios, conditions, and types of observer. These studies can be traced back to many years ago, when color displays were not yet widely available. Most of them focused on luminance contrast, which is a measure of the difference in luminance between the text and the background. Studies by Legge⁴ and others have observed that luminance contrast is the basis of text readability, even when equiluminant colors with high chromatic contrast can generate legible text^{5,6}. Since visual acuity is much better for lightness changes than for changes in hue and saturation, the lightness and contrast of the color against the background predominates when we try to resolve fine detail. Discrimination on the basis of hue and/or saturation differences alone is not adequate in these situations of fine detail⁷. Since luminance contrast does not affect CVD, for readable text a good strategy is to provide a significant lightness difference between the information and the background. In considering only luminance contrast, one implicitly ignores the issues of “chromo-stereopsis”, where legibility is reduced due to decreased visual acuity, because for example the focus for a blue image is different from the focus for a red image. Modern displays can easily be adjusted to meet reference conditions like for example sRGB. Thus, even when a display is not calibrated and its color is not managed, the luminance for any stimulus presented can be predicted reasonably well⁸. From the luminance, it is possible to predict the perceived lightness.

Many legibility criteria have been proposed in the literature. ANSI/HFS 100-1988⁹ recommends using a metric proposed by Lippert for assessing character legibility on visual display terminals (VDTs). Additionally it suggests that the characters and their background should differ by a minimum of 100 error units for adequate legibility. The equation proposed is:

$$\Delta E = \left[\left(155 \frac{\Delta Y}{Y_M} \right)^2 + (367 \Delta u')^2 + (167 \Delta v')^2 \right]^{0.5} \quad (1)$$

where Y_M is the maximum luminance of background or foreground. As reported by Schanda¹⁰, the ISO Committee on “Flat Panel Display Ergonomic Requirements” proposed a similar equation with a different evaluation of luminance contrast:

$$\Delta E = \left[\left(310 \frac{C_m}{(1 + C_m)} \right)^2 + (367 \Delta u')^2 + (167 \Delta v')^2 \right]^{0.5} \quad (2)$$

where C_m is the Michelson luminance contrast:

$$C_m = \frac{(L_H - L_L)}{(L_H + L_L)} \quad (3)$$

These equations are based on an old color space for computing color differences. In 1976 CIE recommended two color spaces, CIELAB and CIELUV, for the specification of color differences, the former based on multiplicative and the latter on additive chromatic adaptation. Today we know that chromatic adaptation is multiplicative and CIELUV is deprecated; (Alman et al.¹¹ provide experimental evidence for the poor performance of CIELUV as a color-difference model). In 1994 CIE recommended an improved color-difference model for industrial color-difference evaluation, with abbreviation CIE94 based on the CIELAB space, sometimes abbreviated as ΔE_{94}^* ¹². More advanced color difference formulae based on the CIELAB space have been recently proposed and compared in a study in which the CIEDE2000 outperformed the other formulae. The study was conducted on colors displayed on a CRT, producing great similarity with results obtained with surface media¹³. Color difference formulae have been designed mainly for accurate color reproduction in surface color industries (textile, plastics, ...), and their typical use is for evaluating differences between very similar colors. In a recent study, different formulae have been evaluated using colors with differences of small, medium and large magnitudes of textile samples and CRT colors¹⁴. Results indicate that for the set with large differences on CRT colors, CIE94 is a good choice.

Currently the World Wide Web Consortium (W3C) is evaluating color design rules for accessibility¹⁵, and an algorithm has been proposed as a W3C guideline for checking the legibility of combinations of foreground and background colors for text¹⁶. The algorithm computes color brightness according to the Equation 4, which corresponds to luminance in the NTSC space, and indicates 125 as range for color brightness difference. Color difference is evaluated as in Equation 5, where the subscripts F and B indicate foreground and background respectively. A suitable range for color difference is indicated in 500¹⁵.

$$Y = (299 \cdot R + 587 \cdot G + 114 \cdot B) / 1000 \quad (4)$$

$$DH = (\max(R_F, R_B) - \min(R_F, R_B)) + (\max(G_F, G_B) - \min(G_F, G_B)) + (\max(B_F, B_B) - \min(B_F, B_B)) \quad (5)$$

Unfortunately, luminance difference does not correlate well with any attribute of human perception. Moreover, the W3C algorithm is based on the antiquated NTSC RGB video standard, which is not representative of modern display devices. The phosphor chromaticities in the NTSC standard are based on ITU-R Recommendation BT.601-4. The current video RGB standard is ITU-R BT.709, and most modern displays are designed to conform to the sRGB standard, which is based on ITU-R BT.709¹⁷.

In this paper we present a design tool for the selection of background and foreground colors. The tool is based on the Munsell Book of Color and allows browsing the atlas, indicating plausible colors on the basis of a set of legibility/readability rules. The tool takes into account the problem of the display gamut by implementing a proofing preview, and allows to model different degrees of veiling glare. Most of the rules for the selection of appropriate colors are based on luminance information, therefore they are valid for color deficient users. We report our experimental results for the validation of a heuristic rule proposed by one of the authors¹⁸, who indicated in 27 CIELAB units the lowest bound for discriminability of a pair of background and foreground colors. The software tool described in this paper allows the visual assessment of the readability rule in different configurations (device profiles, proofing profile, and veiling glare). The method can also be compared with other previously proposed criteria for legibility. The validation was performed for the sRGB reference display conditions. The experiments were performed in environments conforming to the sRGB reference viewing conditions, for a set of six observers having normal color vision and corrected visual acuity when necessary. Our evaluation of readability is based on the time the user needs to count the occurrence of a letter in a text.

2. THE COLOR SELECTION TOOL

The color selection tool allows choosing combinations of foreground and background colors on the basis of readability rules. The tool is based on the Munsell Book of Color, a color system commonly used for specifying colors. It defines a perceptually uniform color space, the dimensions of which are Munsell hue, value and chroma. There are five principal hues, red, yellow, green, blue and purple, together with five more that are mixtures of these. The value is specified by an integer in the range of 0 (black) and 10 (white). The chroma is the radial distance from the achromatic axis.

The main functionalities of the tool are: the implementation of parameterized readability rules; the support of device and proofing profiles (exploiting the library littleCMS¹⁹); veiling glare simulation; or CVD simulation^{20,21}. Figure 1 shows the layout of the system.

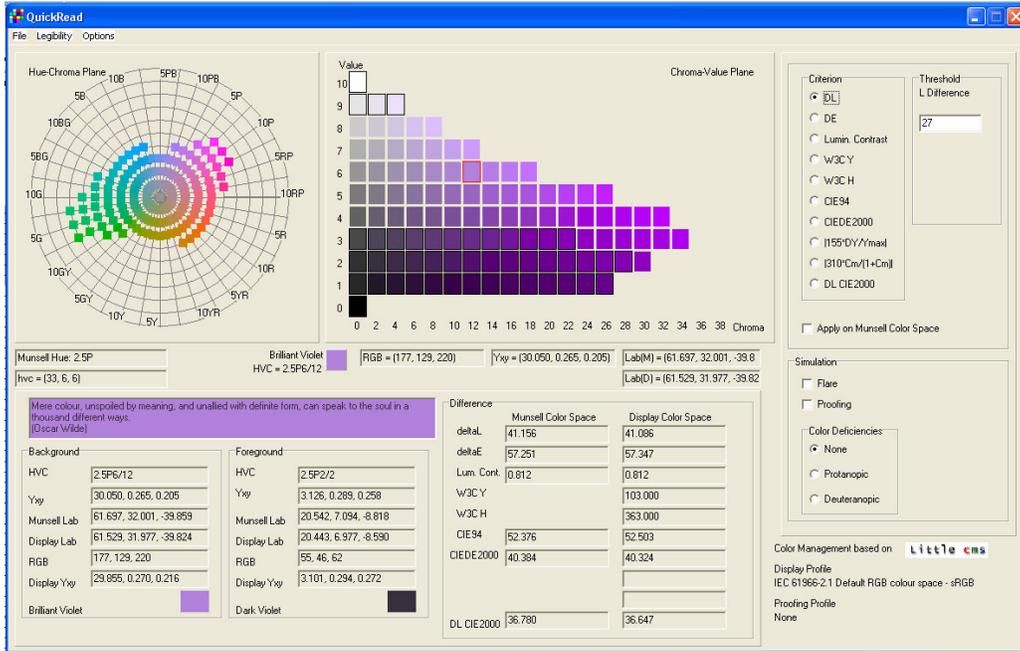


Figure 1. Layout of the color selection tool

On the top-left side is displayed the hue-chroma plane of the atlas. This plane can be used to select a chroma-value plane with a mouse click in correspondence of a hue. The color selection can be performed on the chroma-value plane. When the user clicks on a color sample on this plane, a set of colors are selected (see Fig. 2). These colors are all the admissible colors on the basis of the current discriminability criterion. This criterion can be one of: lightness difference; color difference ΔE_{ab}^* ; Michelson contrast (Eq. 3); W3C luminance difference; W3C hue difference (Eq.5); ΔE_{94}^* ; CIEDE2000; the luminance contrast derived from the ANSI/HFS 100-1988 formula (Eq. 1); the luminance contrast derived from the ISO formula (Eq. 2). The user can set a threshold at his choice for each of the criterion selected.

Beneath the panel for the discriminability criteria selection (Fig. 3a), the user may access the parameters that allow to simulate a given percentage of flare; to make a proofing of the selected combination of colors (e.g., simulate how the colors will appear on a projector); to simulate how the colors will be seen by an observer with CVD (Fig. 3b).

The user can set a pair of colors as background and foreground, and then visualize a proof of the combination in the display of a short text, together with all the information on color coordinates,

color distances, contrast measures, etc. (Fig. 4). By default the display color space is sRGB. The user can load any display device ICC profile, and also load a proofing profile.

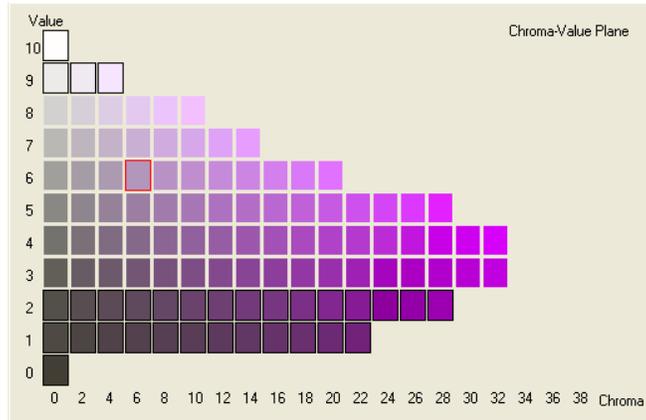


Figure 2. Chroma-value plane for color selection

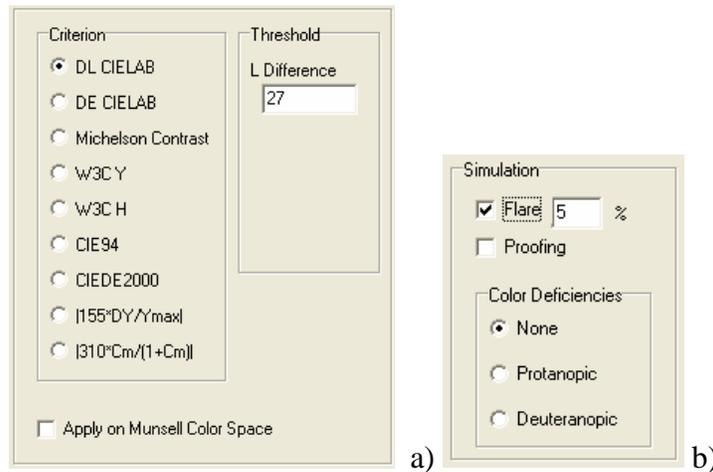


Figure 3. (a)The discriminability criteria and (b) simulation settings.

Although the system may apply many color difference criteria, we focused on lightness difference as our preferred readability metric. We performed a readability experiment in order to gather information about an appropriate value for its threshold.

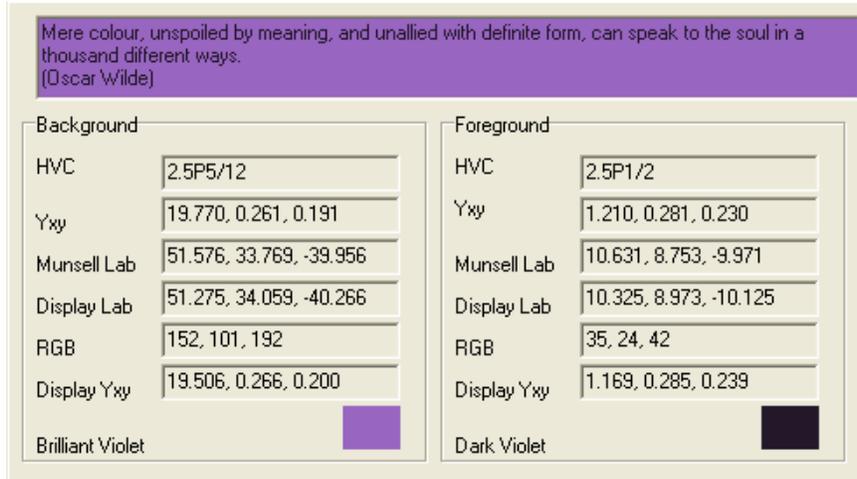


Figure 4. Proofing and information about background and foreground selected colors.

3. EXPERIMENTS

A common technique to evaluate the readability of displays is to measure the user's performance in reading a text in terms of speed, as proposed by Legge^{4,6}. An operational measure may be defined as the probability with which characters, or clusters of characters, will be recognized within a fixed time period²². In our experiment we chose to consider a task of characters counting. The purpose of the experiment was to contribute to the definition of a good choice for the lightness difference threshold. We considered the set of five Robertson zones, which correspond to the hues 5Y, 5YR, 5R, 10G, 5PB in the Munsell Book. We considered combinations of colors within the same hue (that is, foreground and background differ only in value and chroma) and combinations where colors belong to opponent hues, with a total of 10 hue combinations. For each of the 10 hue combinations, we considered three background colors at average lightness, but with different chroma. If C_{\max}^B is the maximum chroma on the Munsell atlas for value at 5 and for the hue considered, the three background colors we selected were those having chroma at zero, half of C_{\max}^B , and C_{\max}^B . For these background colors, the foreground colors were chosen, respectively, with chroma at zero, at C_{\max}^F and at zero. The values of the value for the foreground colors were set depending on the threshold considered, and this also holds for the value of C_{\max}^F . In summary, for each hue combination considered, and for a given threshold, we considered 6 color combinations, as summarized in Table 1.

Background			Foreground		
H	V	C	H	V	C
H_B	5	0	H_F	$>(5+thr)$	0
H_B	5	0	H_F	$<(5-thr)$	0
H_B	5	$0.5 \cdot C_{max}^B$	H_F	$>(5+thr)$	C_{max}^F
H_B	5	$0.5 \cdot C_{max}^B$	H_F	$<(5-thr)$	C_{max}^F
H_B	5	C_{max}^B	H_F	$>5+thr$	0
H_B	5	C_{max}^B	H_F	$<5-thr$	0

Table 1. Munsell coordinates of background and foreground colors selected, for a given threshold, for each of the ten hue combinations considered in the experiments.

Three different thresholds, 8, 22, and 36, were considered. The procedure described allowed the definition of 180 color combinations spanning uniformly the whole atlas. Due to the fact that we adopted a discrete color space, the thresholds considered do not involve the selection of colors at distances equal to the thresholds' values. The Munsell Book of Colors includes samples that are uniformly spaced out in a perceptually uniform color space, and the differences in lightness are approximately equivalent to 10 units. If we plot the value of lightness differences for all the pairs of foreground and background colors considered in the experiments, it can be noticed that the threshold of 8 determines the selection of colors with a difference of about 10 units, and the threshold of 22 determines the selection of colors at distances of about 30 units.

Six observers having normal color vision and corrected visual acuity (when necessary) participated to the experiments, which took place in a room at fixed viewing conditions. The experiments were performed using a Barco monitor having a gamma value of 2.2 that was properly placed in order to reduce flare and user adaptation to the ambient illumination. For the experiments, we used the sRGB color space ICC profile as display profile. The color combinations described previously were randomly presented to the observers. For each combination, the software generated a text composed of random characters grouped in words

having a mean length of 5 letters; the text consisted of five rows of 40 characters each. The software also chose a character to count, and set about 10% of the letters in the text equal to this character. The observers had to count the occurrence of the character by clicking on a button. At the end of the counting task, they also had to give a judgment about the legibility of the text. The text was presented in a square in the middle of the monitor screen (Fig. 6). The observers were asked to perform the experiment at the distance at which they usually read Web pages. They were also asked to interrupt the experiment and take a break in case of fatigue. The text font was set to Arial, with a size of 10 points. The monitor screen diagonal was 21 inch, with resolution set at 1280x1024.

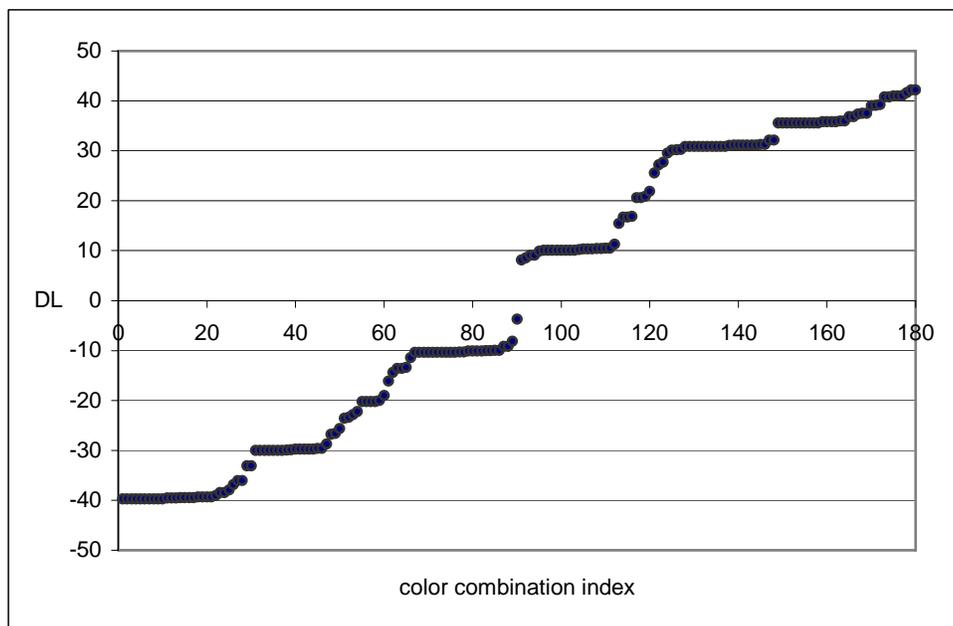


Figure 5. Plot of the L differences between foreground and background colors for all the 180 color combination considered in the experiments. The abscissa is the index of the combination, the ordinate the lightness difference.

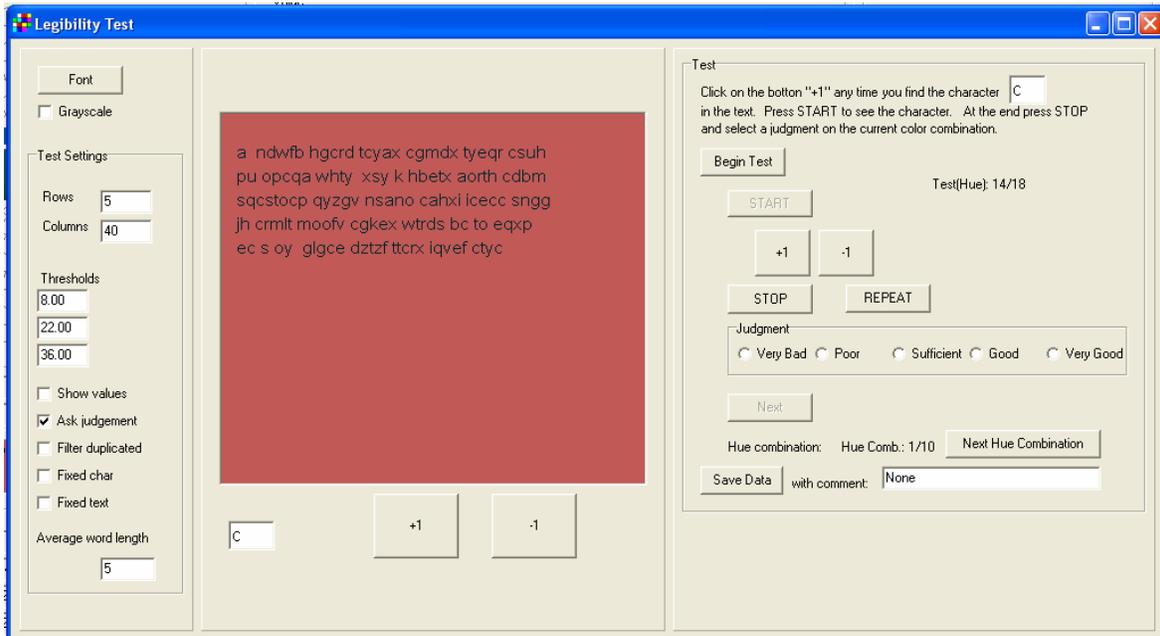


Figure 6. Form for the experiment of characters counting.

The observers involved in the experiment performed a total of 460 tests on different background and foreground combinations randomly chosen between the 180 color combinations previously described. For each of them they carried out a character counting task. Statistics of the time employed are reported in Table 2 and depicted in Figure 7.

Threshold	Min	1 st Qu.	Median	Mean	3 rd Qu.	Max
DL > 8	17.00	27.00	35.00	33.32	40.00	52.00
DL > 22	13.00	25.00	31.00	29.93	35.00	49.00
DL > 36	14.00	24.00	29.00	28.69	33.00	51.00

Table 2. Statistics of the time employed to perform the character counting task for the three thresholds of 8, 22 and 36 on lightness difference between foreground and background (DL). Times are in seconds.

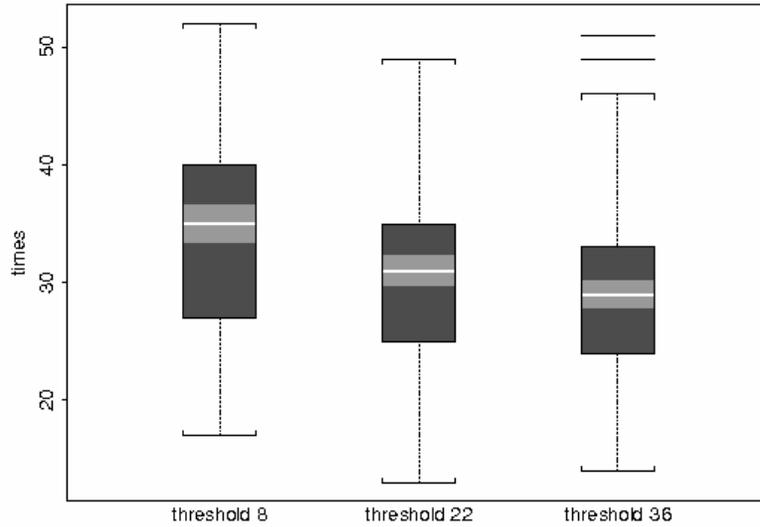


Figure 7. Box-plots of the times employed to perform the character counting task with the three thresholds.

The median times employed to perform the counting task with the thresholds 8 and 22, and with the thresholds 22 and 36 respectively, were compared by using the Mann-Whitney test (or Wilcoxon rank sum test), a nonparametric test for comparing locations of two different populations²³. We performed a one-side test, assuming as alternative hypothesis that the median time employed with the threshold 8 was greater than the one employed with the threshold 22 in the former case, and that the median time employed with the threshold 22 was greater than the one employed with the threshold 36 in the latter. The results of the test are shown in Table 3.

Comparison	Null hypothesis	Alternative hypothesis	p value
thr.8 and thr.22	no difference in median time	median(thr.8) > median(thr.22)	<<0.001
thr.22 and thr.36	no difference in median time	median(thr.22)>median(thr.36)	<0.05

Table 3. Results of the Wilcoxon rank sum test.

4. CONCLUSIONS

In this work we presented a tool allowing the selection of background and foreground colors for the display of textual information satisfying readability rules. The color selection is based on a readability measure that can be selected from a set of criteria. In order to take into account color vision deficiencies, we focused on color selection based on luminance. In particular, we performed experiments to evaluate readability of many color combinations selected at three different levels of lightness difference by means of a character counting task. The statistical analysis shows that there is a very significant difference in readability between the lightness thresholds 8 and 22, and that the difference is much less significant between the thresholds 22 and 36.

As first result, our data suggest that, assigning a lightness difference of about 30 in the color selection tool could be a good strategy to address legibility issues. This is in accordance with a previous study on colored text on colored background, where a difference of 27 CIELAB units was suggested as a threshold for discriminability¹⁸. Our results, even if interesting, require further extensions and comparisons, which we plan to perform in the future. Currently the tool has a plethora of instrumentation parameters; once the optimal parameter values are determined through psychophysics experiments, we will streamline the graphical user interface so it can easily be used by Web designers.

REFERENCES

1. M. Nielsen, M. Stokes, "The Creation of the sRGB ICC Profile," Proc. of IS&T Sixth Color Imaging Conference: Color Science. Systems and Applications, 1998.
2. Seunji Yang, Yong Man Ro, Edward K. Wong, and Jin-Hak Lee, "Quantification of Color Compensation for Anomalous Trichromats Based on Error Score on the Computerized FM-100 Hue Test", to be published, 2006.
3. G. McKinsty, "Some typesetting conventions", TUGboat 9(3), Nov. 1988.
4. Legge, G. E. And G.S. Rubin. Psychophysics of reading: IV. Wavelength effects in normal and low vision. Journal of the Optical Society of America A, 3, pp. 40-51, 1986.
5. K. Knoblauch, A. Arditi, and J. Szlyk, "Effects of chromatic and luminance contrast on reading," J Opt Soc Am A. 8(2), pp. 428-39, 1991.

6. Legge, G. E., D.H. Parish, A. Luebker and L.H. Wurm. Psychophysics of reading: XI. Comparing color contrast and luminance contrast. *Journal of the Optical Society of America*, A, 7, pp. 2002–2010, 1990.
7. R. Jackson, L. MacDonald, K. Freeman, *Computer Generated Color: A Practical Guide to Presentation and Display*, Wiley, Chichester, 1994.
8. B. Macintyre, W.B. Cowan, “A Practical Approach to Calculating Luminance Contrast on a CRT”, *ACM Trans. on Graphics*, 11(4), pp. 336-347, Oct 1992.
9. American National Standard for Human Factors Engineering of Visual Display Terminal Workstations, The Human Factors Society, Inc., ANSI/HFS 100, 1988.
10. J. Schanda, “Colour and the Visual Display Unit”, AIC, Budapest, 1993.
11. D.H. Alman, R.S. Berns, G.D. Snyder, W.A. Larsen, “Performance testing of color-difference metrics using a color tolerance dataset”, *Color Research and Applications*, 14, pp.139-151, 1989.
12. CIE, “Industrial Colour-Difference Evaluation,” *Technical Report 116 – 1995*, 1995.
13. G. Cui, M. R. Luo, B. Rigg, W. Li, “Colour-Difference Evaluation Using CRT Colours. Part I: Data Gathering and Testing Colour Difference Formulae”, *Color Research and Applications* 26(5), Oct. 2001.
14. K.M.R. Ho, G. Cui, M.R. Luo, and B. Rigg, “Assessing colour differences with different magnitudes”, AIC 2004 Color and Paints, Interim Meeting of the International Color Association, 2004.
15. <http://www.w3.org/TR/AERT#color-contrast>
16. C. Ridpath, J. Treviranus, and P. L. Weiss, “Testing Readability of Web Page Colors”, (<http://www.aprompt.ca/WebPageColors.html>) , 2003
17. S. Süsstrunk, R. Buckley, S. Swen, “Standard RGB Color Spaces”, The Seventh Color Imaging Conference: Color Science, Systems, and Applications, Nov. 1999.
18. G. Beretta, “Color Aspects of Variable Data Proofing”, IS&T/SPIE Electronic Imaging, San José, Jan 2005.
19. <http://www.littlecms.com/>
20. H. Brettel, F. Viénot, J.D. Mollon, “Computerized simulation of color appearance for dichromats”, *J. Opt. Soc. Am. A*, 14(10), Oct. 1997.
21. F. Viénot, H. Brettel, J.D. Mollon, “Digital Video Colourmaps for Checking the Legibility of Display by Dichromats”, *Color Research and Applications*, 24(4), Aug. 1999.
22. L.W. MacDonald, A.C. Lowe, *Display Systems*, Wiley, Chichester , 1997, pag. 98.
23. M. Hollander, D. Wolfe, *Nonparametric Statistical Methods*, 2nd Edition, Wiley, New York, 1999.