Ensuring color legibility

T. H. Nilsson
Department of Psychology, University of Prince Edward Island
Charlottetown, P.E.I., C1A 4P3 (CANADA)
Corresponding author: T. Nilsson (nilsson@upei.ca)

ABSTRACT

There is little systematic, valid, quantitative data on the legibility of colored graphics. Most studies based on reading speed found that hue had no effect because they were insensitive to hue information conveyed by slow conducting parvo-cellular pathways. Measuring legibility by the number of visual pathways needed for legibility reveals the contribution of color. The legibility of 100 combinations of colored words printed on colored backgrounds are presented - based on 48,000 measurements. To ensure legibility of colored graphics, we propose a legibility standard and a standard method similar to the 20/20 system for specifying visual acuity.

1. INTRODUCTION

Color graphics are the requisite front-end of the post-industrial information economy. Yet there is little systematic, valid, quantitative data on the legibility of colored graphics. Most quantitative studies sought to measure the legibility of colored print based on some aspect of reading speed such as tachistoscopic duration thresholds, reaction time, or number of words per second.[1, 2, 3] The rationale that more legible graphics can be processed more quickly seems intuitively obvious. Such studies consistently find that legibility depends solely on the luminance differences of the color of the words and the color of the background - i.e. that color itself has no effect on legibility.

There is a problem, however, with the processing speed methods. They are insensitive to the color information in the display. Luminance differences are conveyed by fast-conducting magno-cellular visual pathways. Hue is conveyed by slow-conducting parvo-cellular pathways. When observers are asked to read as fast as possible, decisions are made on the basis of the information received first, and the color information does not contribute to such legibility measurements. To avoid this confounding effect, we sought a way to measure legibility that did not depend on processing speed.

The 20/20 method of assessing visual acuity provided the basis for our method. The 20/20 system implicitly assumes that persons with normal acuity require a certain number of visual pathways to discern the letters on a Snellen eye chart. Letters on the 20/20 line are about 9.5 mm high. At a distance of 20 feet (6 meters) they have a visual angle of 5'21". The visual field of a single photoreceptor near the fovea subtends about a 25 second visual angle. A single letter in the 20/20 line covers about 12 receptors in height or about 144 receptors within its image area. Persons with poorer acuity require larger letters to stimulate proportionally more receptors. A person with 20/80 acuity can read at 20 feet letters that a person with normal acuity can read at 80 feet (24 m) - i.e. letters 4 times as high, which cover 16 times as many receptors.

The 20/20 method can be reversed to measure legibility based on observations by persons with normal acuity. A letter that must be four times as high as the 9.5 mm high letters which define normal legibility, can be said to be only 1/16th as legible because it requires 16 times as many visual pathways to be read. Correspondingly, a letter which is readable when it is only 6.7 receptors high can be said to twice as legible because it requires only half the number of receptors. Therefore, the legibility of letters as reflected by the number of receptors required to read them is readily determined from their size and the distance at which they can be read.

Initial distance threshold measurements were made under contract with Health Canada to provide legibility data about colored warnings on cigarette packages - data that would withstand legal challenge.[4] For these measurements, we built an 8-meter, automated test track to measure legibility distance. Subsequent work for the Canadian Space Agency enabled systematic measurement of the legibility of all letter /background combinations of the six primary colors (red, yellow, green blue, black and white).[5] These measurements revealed a few color combinations that were significantly more legible than black/white. This disproved the contention that legibility depends solely on lightness contrast. The method was then used to measure the legibility of colored symbols and line drawings.[6, 7] Some differences in the most effective color combinations for symbols, words, and line drawings may be attributable to differences in receptive field size of the various opponent color pathways - further research is in progress.[8] Other research demonstrated that legibility distance thresholds were effective in evaluating image enhancement techniques and consumer packaging designs.[9, 10] A project for NATO demonstrated that legibility distance measurements could distinguish more types of camouflage than reaction time measurements.[11]

We present here the results of measuring the legibility of 100 letter/background color combinations that include the six primary colors, intermediate colors such as orange and purple, and a few other common colors such as brown and pink. These measurements were intended to identify the legible color combinations, so combinations that were obviously difficult to read (like brown/black) were not tested. To avoid complexities of electronic displays, to date our color measurements have focused on printed media.

2. METHOD

Observations were made by 12 hired students 19-24 years old, who had normal or corrected-to-normal acuity and normal color vision based on the Dvorine Test. Each observer made 500 practice measurements before commencing a total of 40 distance threshold measurements for each color.

Five, color neutral, four-letter words of similar familiarity and imagery were printed in 36 Pt. Helvetica Medium font on 3.5 by 6 cm colored backgrounds and mounted on 13 X 18 cm flat-black cards. These materials were illuminated at 78 cd/m^2 with a 2530^0 K color temperature from 45^0 on both sides by two General Electric well-diffused "Soft White" incandescent bulbs using lenseless projectors. A 1 X 1 X 0.6 m enclosure lined with black velvet eliminated stray light from the test cards and from the observer's view. The colors were selected from the Natural Color System. A professional graphics technician matched the printed colors to NCS standard plates. Table 1 shows the CIE L*u*v* characteristics of the printed colors.

The psychophysical Method of Limits was used to measure legibility distance thresholds with the observer initiating each measurement. A hierarchy of computers managed the measurement procedure, regulated display box movements to a constant velocity of 14 cm/sec along the elevated track, and recorded distances. Color combinations were tested in a pseudo-random order which changed the word, word color, and background color after each set of 8 measurements. The order was counterbalanced across subjects.

Table 1. Natural Color System code and CIE L*u*v* coordinates of the colors tested.

COLOR	NCS	L*	u*	v*	COLOR	NCS	L*	u*	v*
black	9500	18.7	-11.5	-0.3	lime	1080-G50Y	89.4	-84.3	25.8
white	0500	96.3	-73.2	3.1	green	1070-G	54.3	-119	22.5
pink	0050-R	70.3	40	1.4	dark-green	5050-G	37.8	-81.4	11.2
red	1090-R	54.9	90	-2.7	turquoise	2060-B50G	46.9	-99.5	3.7
orange	0090-Y50R	68.6	60	6.7	light-blue	0040-B	71.4	-110	28.8
yellow	0080-Y	91.6	-41	29	blue	1070-B	48	-111	-20.4
brown	5040-Y	59.2	-22.3	16	purple	2050-R50B	53.1	-11.2	-10.6

3. RESULTS

The legibility results are shown in Table 2. To check reliability, each observer's data were correlated with the average data of all observers. The mean of these correlations was >0.90. Generally, two color combinations whose distance thresholds differ by more than from 15 cm down to 10 cm (for the closer distances) differ significantly at the 5% level. Legibility of all colors down to white/blue do not differ significantly from the most legible, black/pink. Colors down to black/light-blue do not differ in legibility from black/white.

Table 2. Legibility distance (MEAN), standard error (se) and "relative legibility" based on ratio of retinal image area to that of black/white (REL LEG) for 100 word/background colors.

WORD	BKGD	MEAN (cm)	se	REL LEG	WORD	BKGD	MEAN (cm)	se	REL LEG	WORD	BKGD	MEAN (cm)	se	REL LEG
black	pink	517	5.3	1.03	purple	yellow	489	5.0	0.92	brown	turq	456	4.7	0.80
black	yellow	516	5.0	1.03	lt-blue	black	489	5.0	0.92	purple	dk-grn	455	4.7	0.80
black	orange	512	5.2	1.01	orange	blue	488	4.9	0.92	turq	brown	455	4.4	0.80
dk-grn	yellow	512	5.2	1.01	purple	white	487	4.9	0.92	black	turq	454	4.8	0.80
black	red	511	5.1	1.01	white	red	486	4.8	0.91	red	blue	453	5.1	0.79
dk-grn	white	509	5.0	1.00	orange	turq	486	5.2	0.91	turq	red	452	5.0	0.79
BLACK	**WHITE**	**509**	**5.1**	**1.00**	pink	turq	486	4.6	0.91	brown	blue	452	4.3	0.79
white	purple	507	5.1	0.99	orange	dk-grn	486	4.8	0.91	purple	orang	448	4.1	0.77
blue	yellow	505	4.9	0.98	lime	turq	486	4.4	0.91	red	turq	446	5.0	0.77
orange	black	504	5.2	0.98	pink	dk-grn	484	5.0	0.90	green	brown	441	4.6	0.75
pink	black	503	5.1	0.98	yellow	black	484	5.4	0.90	brown	green	439	4.4	0.74
black	lime	503	5.1	0.98	green	pink	483	4.6	0.90	lime	purple	438	4.1	0.74
blue	white	502	5.0	0.97	green	orang	483	5.3	0.90	red	green	436	4.6	0.73
white	blue	502	4.8	0.97	yellow	red	483	4.8	0.90	turq	purple	432	4.5	0.72
yellow	purple	499	4.8	0.96	pink	blue	483	4.7	0.90	purple	lime	426	4.4	0.70
green	yellow	499	4.8	0.96	lime	blue	482	0.0	0.90	red	lt-blue	416	5.0	0.67
dk-grn	pink	499	4.6	0.96	turq	orang	482	4.7	0.90	purple	turq	414	4.7	0.66
white	dk-grn	499	4.5	0.96	brown	white	481	4.8	0.89	red	lime	413	5.2	0.66
yellow	turq	499	5.0	0.96	black	green	480	4.9	0.89	green	purple	397	4.5	0.61
green	white	498	4.7	0.96	lt-blue	dk-grn	478	4.4	0.88	lime	red	396	4.8	0.60
yellow	blue	498	5.0	0.96	dk-grn	lt-blue	478	4.8	0.88	lt-blue	red	394	4.5	0.60
yellow	dk-grn	497	4.7	0.95	red	white	478	5.9	0.88	brown	lt-blue	380	3.6	0.56
lime	black	497	4.9	0.95	blue	lime	474	4.6	0.87	purple	green	378	4.4	0.55
dk-grn	orang	497	5.1	0.95	red	yellow	474	5.1	0.87	lt-blue	brown	375	3.7	0.54
white	green	497	4.7	0.95	pink	green	473	5.1	0.86	pink	lime	367	3.8	0.52
yellow	green	496	4.9	0.95	green	black	472	5.0	0.86	orange	lime	364	3.9	0.51
turq	yellow	496	4.9	0.95	orang	green	470	4.8	0.85	lime	pink	363	4.2	0.51
blue	pink	496	5.2	0.95	dk-grn	purple	469	4.6	0.85	lime	orang	354	4.2	0.48
blue	orang	495	5.1	0.94	orang	purple	466	4.4	0.84	brown	purple	353	3.4	0.48
red	black	495	5.3	0.94	lt-blue	green	464	4.6	0.83	purple	brown	344	3.2	0.46
black	lt-blue	494	4.9	0.94	blue	brown	463	4.8	0.83	lime	lt-blue	200	2.5	0.16
white	black	492	5.0	0.93	green	lt-blue	460	4.2	0.82	lt-blue	lime	184	2.4	0.13
turq	pink	492	4.8	0.93	turq	black	458	5.0	0.81					
white	brown	491	4.7	0.93	blue	red	458	5.4	0.81					

Subjectively, the ease of reading is more closely related to image area than distance. Table 2 compares the legibility of each color combination with that of black/white in terms of the ratio of their image areas at threshold. This ratio, "relative legibility" is calculated from their distance thresholds by: d_t^2 / d_s^2 , where d_t is the distance threshold of the colors being tested, and d_s is the distance threshold of the black/white standard. In these terms a color combination, such as black/yellow that is more legible than black/white has a relative legibility greater than 100%, namely 103%; while a less legible combination such as lime/light blue is only 16% as legible as black/white.

4. DISCUSSION

The legibility of the six primary color combinations can be predicted with better than 80% accuracy based on chromaticity, and lightness contrast.[6] However, with 100 color combinations, it has not been possible to account for even 50% of the variance. This indicates that there is no colorimetric solution to specifying the legibility of colored graphics. The 100 color combinations tested represent only a tiny fraction of the color possibilities for graphics. Clearly, legibility standards can not be based on such tables either. A practical way to ensure the legibility of colored graphics requires a legibility standard and a standard method of comparing the legibility distance of any color combination to that standard.

For a legibility standard, a common message like DO NOT PASS WHEN LIGHTS ARE FLASHING from the back of school buses would be appropriate. Printing this message in black letters on a white background would be the easiest color combination to specify and reproduce. In principle, any size font could be used, but our experiences indicate that a 24 to 36 Pt font provides a good compromise for reproducible graphics and measurement precision, both of which become more a difficult the smaller the font. Larger fonts require correspondingly larger measurement distances and put more demands on space, illumination, and mechanics.

Standardization will require certain specifications of the measurement apparatus in terms of illuminance and geometry. A standard procedure would describe the measurement steps for comparing the legibility of a test target with that of the standard. Nevertheless, a comparative approach is robust in that digressions in apparatus or procedure that favor or hinder the legibility of the test target will tend to have the same effect on the standard and have minimal effect on the relative legibility ratio.

Distance threshold measurements can be directly related to20/20 specifications of visual acuity. This enables setting minimal levels of legibility to meet the needs of persons with lower visual acuity.

References

1. Knoblauch, K., Arditi, A. and Szlyk, J., "Effects of chromatics and luminance contrast on reading," J. Optical Soc. Amer., A, 8, 428-439, (1991).

2. Legge, G.E. and Rubin, G.S., "Psychophysics of reading IV. Wavelength effects in normal and low vision," J. Optical Soc. Amer., A, 3, 40-51, (1986).

3. Tinker, M.A., Legibility of print (Iowa State University Press, Ames, Iowa, 1963).

4. Nilsson, T.H. and Kaiserman, M., "The legibility of warnings in color," in Handbook of Human Factors in Litigation, I. Noyes and W. Karwowski, eds. (New York: Taylor & Francis, 2004) chap 32.

5. Nilsson, T.H. and Connolly, G.K., "Color (not just contrast) affects legibility of printed words," presented at Human Factors and Ergonomics Society, Seattle, (1993).

6. Nilsson T.H. and Connolly G.K., "Chromatic contribution to shape perception revealed in a non-temporal detection task using distance." in John Dalton's Colour Vision Legacy, C. Dickinson, I. Murphy, & D. Carden, eds., (Taylor & Francis, Bristol, Pa., 1997), pp. 197-206.

7. Nilsson, T.H. and Connolly, G.K., "Chromaticity contributes substantially to legibility and figure detection when parvo pathways are used," presented at Association for Research in Vision and Ophthalmology, Fort Lauderdale, Invest. Opthal. & Vis. Sci. 38, S640, abs., (1997).

8. Nilsson, T.H., "Effects of color and target-to-background size on recognition of letter shapes," York International Conference on Processing of Spatial Form Defined by Colour, Motion, Texture, and Binocular Disparity. York University, Toronto, (2003).

9. Nilsson, T.H.; Connolly, G.K. and Ireland, W., "Evaluating the effectiveness of image enhancement techniques in terms of visibility distance," European Conference on Visual Perception, Eindhoven, Netherlands, Perception, 23, sup. 12, (1994).

10. Nilsson, T.H., Evaluation of visual effectiveness. Invited address to the Space Technologies Forum - Atlantic Canada 93, Halifax, (1993).

11. Nilsson, T.H., "Evaluation of target acquisition difficulty using distance to measure required retinal area," Optical Engineering, 40, 1827-1834 (2001).