12-2002

Predicting the readability of transparent text

Lauren F. V. Scharff
Stephen F Austin State University

Albert J. Ahumada Jr.
NASA Ames Research Center

Predicting the readability of transparent text

Lauren F. V. Scharff
Department of Psychology, Stephen F. Austin State University, Nacogdoches, TX, USA

Albert J. Ahumada, Jr.
NASA Ames Research Center, Moffett Field, CA, USA

Will a simple global masking model based on image detection be successful at predicting the readability of transparent text? Text readability was measured for two types of transparent text: additive (as occurs in head-up displays) and multiplicative (which occurs in see-through liquid crystal display virtual reality displays). Text contrast and background texture were manipulated. Data from two previous experiments were also included (one using very low contrasts on plain backgrounds, and the other using higher-contrast opaque text on both plain and textured backgrounds). All variables influenced readability in at least an interactive manner. When there were background textures, the global masking index (that combines text contrast and background root mean square contrast) was a good predictor of search times ($r = 0.89$). When the masking was adjusted to include the text pixels as well as the background pixels in computations of mean luminance and contrast variability, predictability improved further ($r = 0.91$).

Keywords: vision models, luminance contrast, contrast gain, masking, text contrast, word search, transparent displays

Introduction

Many factors influence text readability. Some of the previously studied factors include luminance and/or chromatic contrast (Legge, Rubin, & Luebker, 1987; Knoblauch, Arditi, & Szlyk, 1991), color (Legge & Rubin, 1986; Pastoor, 1990), blur (Legge, Pelli, Rubin, & Schleske, 1985; Farrell & Fitzhugh, 1990), the addition of noise (Parish & Sperling, 1991; Solomon & Pelli, 1994; Regan & Hong, 1994), case (Kember & Varley, 1987), polarity (Legge et al., 1985; Parker & Scharff, 1997), and the use of textured backgrounds (Hill & Scharff, 1999; Scharff, Ahumada, & Hill, 1999; Scharff, Hill, & Ahumada, 2000). The large number of possible combinations of even this noncomprehensive list of factors implies that if a display designer desired to maximize readability, relevant combinations of factors probably would not have previously been examined for readability. Thus, a metric to predict readability would be quite useful.

Scharff et al. (1999, 2000) investigated the ability of two image measures (text contrast and background root mean square [RMS] contrast) and two indices based on image discrimination models (a global masking model and a spatial-frequency-selective model) to predict readability of text on textured backgrounds. They used several textures, some spatial-frequency-filtered textures, and various text contrast and color combinations. For the relatively low luminance text contrasts, spatial frequency content of the background affected readability, as measured using a word search task. Both indices better predicted readability than either of the image measures alone. And, when the different backgrounds included different ranges of spatial frequencies, the frequency-selective index led to slightly better predictability.

How well will the success of these indices generalize to text displays incorporating additional factors? One such factor is text transparency. Transparent text appears naturally in the head-up displays (HUDs) overlaying the external view in some airplanes and automobiles, and is occasionally artificially generated in text displays, such as Web pages and advertisements. Other viewing conditions can also lead to the background being visible through the text: viewing transparent letters on a window or viewing text in a see-through liquid crystal display (LCD) virtual reality display. In HUDs, the text light is added to the background, so the combination is additive and the text appears lighter than the background. On the other hand, a see-through LCD display attenuates the background in a multiplicative fashion, so dark text light is some fraction of the clear background light. Conditions that result in perceived transparency have been studied recently for color images (D'Zmura, Colantoni, Knoblauch, & Laget, 1997) and for moving dots (Mulligan, 1993), but the effect of transparency on text readability has not yet been systematically investigated.

The current work attempted to determine whether the two types of transparency combination would differently affect readability, and to determine how well the global masking index would predict the results. We also wanted to compare the transparent text readability to that in previously tested cases (plain background and one patterned background) using opaque text and several contrast levels (Scharff et al., 2000). Previously, we found background texture effects only when contrast was relatively low, so the transparent text was presented at two relatively low contrasts, 0.30 and 0.45, in three

DOI 10:1167/2.9.7 Received March 5, 2002; published December 26, 2002 ISSN 1534-7362 © 2002 ARVO

backgrounds (a plain background and two periodic textures with different RMS contrast variations).

Very low-contrast text is obviously detrimental to readability, so most (but not all) display designers know to avoid it. However, in displays with textured backgrounds and in HUDs, there may be very low-contrast regions, simply because the background may show large variations in luminance. Thus, we wanted to determine if the global masking index would predict the readability of a previously collected, very low-contrast data set. Because the index predicts that the effect of a textured background is to lower the effective contrast of the text, low-contrast text on a plain background is needed to see if the index is working.

All experiments (the transparent text experiment, the opaque text experiment, and the low-contrast experiment) used the same basic procedure to measure readability (Scharff et al., 1999, 2000): text excerpts were placed on backgrounds and participants performed a three-alternative forced-choice search for a hidden target word. Texts that are more readable are assumed to lead to faster search times.

We correlated these readability measures with the global masking index. Although the index did a relatively good job of predicting readability (r = 0.89 for the combined data), it failed to predict the improved performance with dark text relative to that for light text. We then adjusted the global masking index by dropping the simplifying assumption from detection models that the signal effect on masking and adaptation is negligible. Text pixels were not uniformly distributed across the stimulus, but in text areas, they comprised approximately 20% of the pixels. When this percentage of text pixels was used to calculate image contrasts, the light versus dark text effect was predicted and overall predictability of readability was improved (r = 0.91). The moderate improvement using the adjusted measure suggests that even when the background is uniform, the percentage of text pixels should be taken into account when measuring text contrast.

Methods

Three Experiments Measuring Readability

Macintosh Power PC 7200/120 computers were used to create and run all experiments. The stimuli were created using MATLAB, and B/C Power Laboratory (an experiment presentation application) was used to present the stimuli and collect the data. A chin-rest-controlled viewing distance (475 mm) resulted in a viewing angle of 0.04 deg for each pixel.

Experiment I: Measuring the Readability of Transparent Text

This experiment employed a 2 (text transparency type) x 2 (text contrast) x 3 (background) within-participants design. Text transparency conditions were blocked, and their presentation order was counterbalanced across participants. The text contrast and background combinations were randomly presented within each block.

Apparatus and stimuli

The three backgrounds used in this experiment were a plain (uniform) background and two periodic textures taken from a Web page dedicated to supplying free graphical backgrounds to designers (Schorno, 1996). These textures were two of those used by Scharff et al. (1999), one of which was used and filtered in Scharff et al. (2000). These textures were originally chosen because they had obvious pattern differences with respect to the size of the texture elements, and because text placed on top of them was still readable, although less so than the plain backgrounds. Thus, we felt that they would generalize in some ways to textures that designers might actually choose to use in a Web site. Because of their appearance, the two textures will be referred to as the “culture” pattern and the “wave” pattern. The textures had a period of 72 pixels horizontally and vertically. The final, textured background size was created by tiling six of the periodic textures horizontally and vertically, and then chopping them so the final background was a textured rectangle 14 cm in height and 15 cm in width (17.2 deg x 18.4 deg). The plain background was matched in size. See Figure 1 for examples of single 72 x 72 pixel tiles of the three backgrounds.

Seven newspaper excerpts presented in 12 point (6 vertical pixels per letter times 0.25° at our viewing distance) Times New Roman font were used to create the text arrays. The single font and size were chosen based on favorable readability results from Hill and Scharff (1997), and so that the results could be more directly compared to our previous results. The text excerpts were the same as those used in the previous Scharff et al. (1999, 2000) experiments. The text blocks to be read (the middle paragraph of each excerpt) contained 99-101 words. A target word (“triangle,” “circle,” or “square”) was placed in a counterbalanced manner for each trial within each text block. Thus, there were 12 of each of the text excerpts (one for each of the 12 conditions), and the target word was systematically placed in a different location in each one of them. Further, there was an equal number of target words in each roughly defined paragraph area (top left, top right, middle left, middle right, bottom left, bottom right) for each condition. Therefore, no one condition would have more target words in a particular location (e.g., top left) that would give a search advantage over another condition.

Figure 1. From left to right, tiling elements of the three backgrounds used in the transparent text experiment: culture, wave, and plain.

Each transparent text stimulus was centered at the top of the screen, and heavy black lines on the left and right separated each textured background from the surrounding white background. At the bottom of each screen, there were three black geometric shapes (circle, square, and triangle) that corresponded to each of the three possible target words. These 1 cm x 1 cm shapes were spaced 3.5 cm apart and centered below the textured area. One text excerpt was used for the four practice trials; the remaining six were each presented once in each of the 12 conditions. (Links to several actual stimuli can be found at http://hubel.sfasu.edu/research/tt_stim/extransstim.html. For proper rendering, they need to be displayed with an effective gamma of 1.262.) Figure 2a shows opaque (or multiplicative with unity contrast) text on the culture background together with the response choices, while Figure 2b shows additive text on the wave background.

Using a screen calibration function with a gamma of 1.262, the background images B were adjusted to have the same mean luminance ($L_B = 47$ cd/m^2), but they did have different background RMS contrasts ($C_{RMS} = 0.0$, 0.15, and 0.27 for the plain, culture, and wave backgrounds, respectively).

The background RMS contrast was defined as

$$C_{RMS} = E[(L_i - L_B)^2]^{0.5}/L_B = ((\mathrm{Sum}(L_i - L_B)^2)/n)^{0.5}/L_B, \quad (1)$$

where E[.] is the averaging operator, the summation (□□□) is over all background image pixels, L_B is the average background luminance, L_i is the luminance of the *ith* pixel, and n is the number of pixels.

Prior to combining the text and background, a white buffer was added to the text samples so that they would be the same size as the backgrounds. (Digital text arrays had a value of zero where there was text, and a value of 1 where there was no text.) For both the additive and the multiplicative transparency conditions, text contrast (C_T) was defined as

$$C_T = (L_T - L_B)/L_B = L_T/L_B - 1, \quad (2)$$

where L_T is the average text luminance.

The additive transparency stimuli T_A were created by first scaling the luminance of the text arrays so that they

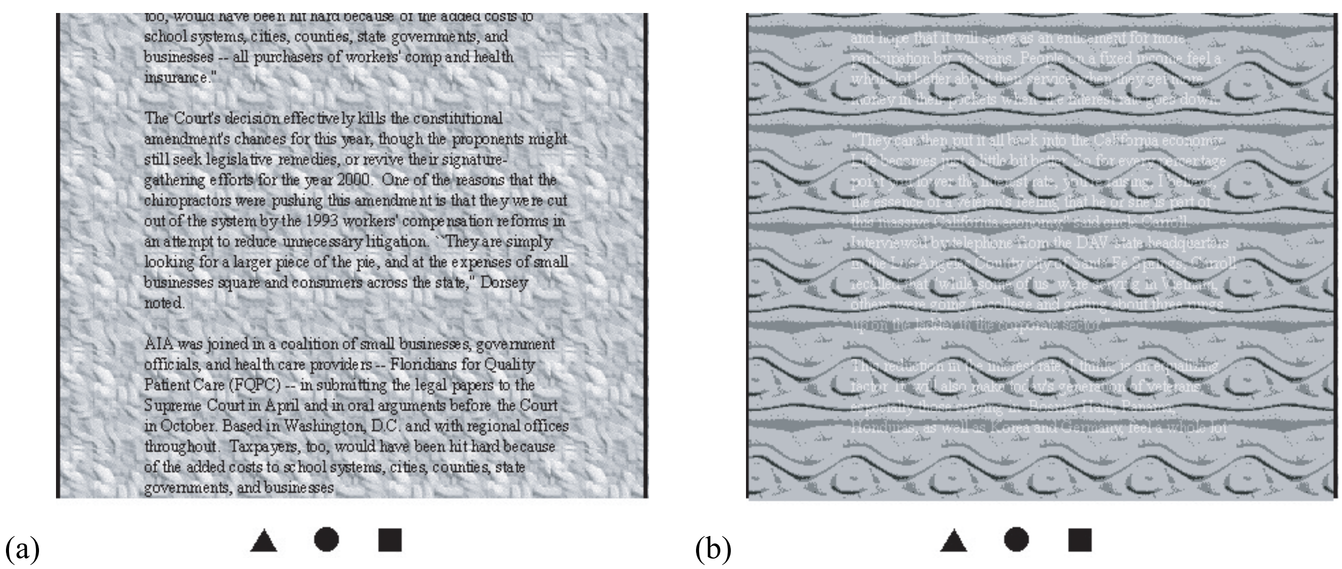

Figure 2. (a) An example text display with opaque text or a multiplicative contrast of 1.0 on the culture background. On each trial, a target word (square, triangle, or circle) would be placed somewhere in the middle paragraph of text. The participant was instructed to find the target word and click on the corresponding shape at the bottom of the screen as quickly as possible. In this example, the correct response was to click on the square. (b) An example text display using an additive text contrast of 0.45 on the wave background. In this example, the correct response was to click on the circle.

would have contrasts $C_T = 0.30$ or 0.45 with respect to the average luminance of the backgrounds and then adding them to the background image, B.

$$T_A = B + C_T L_B T, \quad (3)$$

where T is the text array with text pixels having a value of one and nontext pixels having a value of zero. These manipulations resulted in text that was brighter than the background.

The multiplicative transparency stimuli T_M were also scaled to have the given text contrasts with respect to the average background luminance. Their combination rule was

$$T_M = B * (1 + C_T T), \quad (4)$$

where the contrast values were $C_T = 0.30$ and 0.45 and the * operator indicates pixel-by-pixel multiplication of the background image and the scaled text image. These manipulations resulted in text that was dimmer than the background.

Procedure

Fifty-eight undergraduates participated in the experiment; however, data were not analyzed from four of the participants (two participants could not finish the experiments within the allotted time of 1 hr, and two had high error rates and patterns of behavior during the experiment, which indicated that they did not attend to the task). All participants were naive to the hypothesis and had self-reported 20/20 or corrected-to-20/20 vision. At Stephen F. Austin State University. the majority of undergraduate students are aged 18 to 21 years.

Participants were instructed to scan the middle paragraph of text and find a target shape word ("triangle," "square," or "circle"). When they found the target word, they clicked (using the mouse pointer) on the corresponding shape at the bottom of the screen. The start of each trial was self-paced by clicking a button icon on the screen, and each trial ended when the participant clicked the target-word shape. Participants were instructed to respond as quickly and accurately as possible. When the participants finished the first block of trials, they were instructed to raise their hands; the experimenter then started the second block of trials. Total time to complete the experiment varied between 30 and 60 min.

Experiment II: Measuring the Readability of Opaque Text

Design and stimuli

This experiment (summarized from Scharff et al., 2000) originally employed a 6 (background) x 3 (text contrast) design minus two conditions that were not readable. Three text shades (medium gray, dark gray, and black) resulted in three text contrast levels (0.15, 0.35, and 0.95) given the average background luminance of 62.5 cd/m^2. There were six background textures: plain, a periodic texture (the culture pattern described above), and four spatial-frequency-filtered textures created from the periodic texture. Pilot testing revealed that for all conditions the text was detectable on the background textures. It was not readable for two conditions: those using the lowest contrast text placed on the periodic texture containing all frequencies and the band 3 filtered texture. Thus, these two conditions were excluded from the experiment.

For the purpose of this work, however, only the results from the plain and the periodic texture containing all original spatial frequencies will be summarized. The text excerpts and the layout of the stimuli were the same as those described above for the transparent text experiment (although the hidden words were inserted in different counterbalanced places).

Procedure

Sixty undergraduate participants completed the experiment. All participants were naive to the hypothesis and had self-reported 20/20 or corrected-to-20/20 vision. Each readable condition was repeated three times, for a total of 48 trials. There were also three practice trials to familiarize participants with the procedure. Otherwise, the procedure was identical to that described above, except there was one block of trials rather than two.

Experiment III: Measuring the Readability of Low-Contrast Opaque Text

Design and stimuli

This experiment (summarized from Hill, 2001) originally employed a 3 (background luminance levels: 70, 80, and 90 cd/m^2) x 6 (text luminance contrast levels) x 2 (foreground/background color combinations) within-participants' design. Background luminance conditions were blocked, and their presentation order was counterbalanced across participants. The text contrast and foreground/background combinations were randomly presented within each block. There were 6 trials per condition, leading to a total of 180 trials, plus 6 practice trials.

For the purpose of this study, however, the results from only one color combination (gray on gray) and one background luminance level (70 cd/m^2, which most closely matches the backgrounds in the first experiment) will be summarized. The six text contrasts were 0.30, 0.25, 0.20, 0.15, 0.10, and 0.05. See Hill (2001) for the RGB values for each condition.

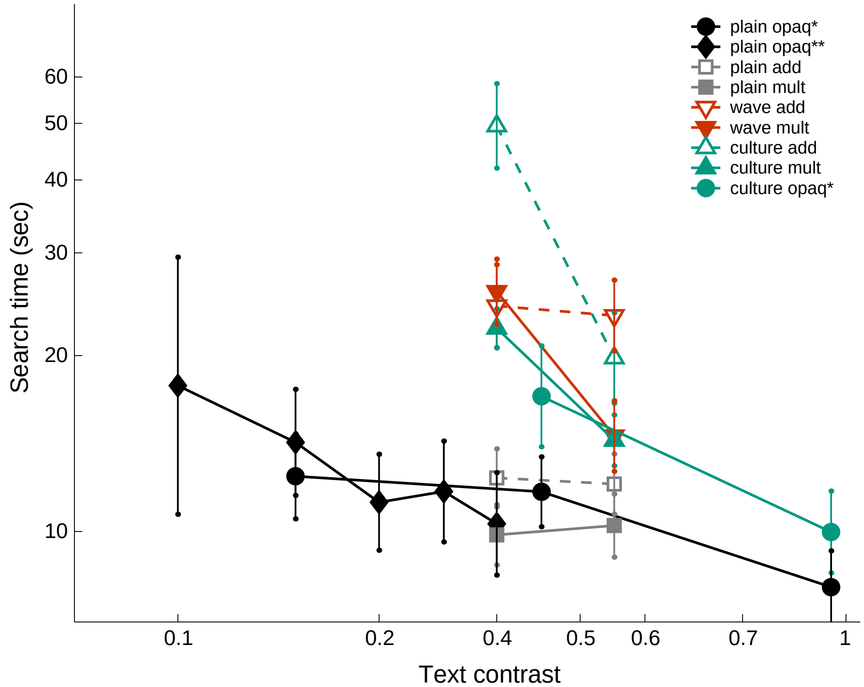

Figure 3. Geometric average search times (seconds on a log scale) and 95% confidence intervals plotted as a function of the nominal text contrast (Equation 2), for the transparent text data of Experiment I, the opaque text data of Experiment II (*), and the low-contrast data of Experiment III (**). All dark-text-on-a-plain-background conditions (solid black symbols) approximately lie on a single monotonic curve. The textured background conditions (red and green symbols) are slower than the plain background conditions (black symbols). And, the additive text search times (open symbols) are slower than the multiplicative times (corresponding closed symbols). The text excerpts and the layout of the stimuli were the same as those described above for the transparent text experiment (although the hidden words were inserted in different counterbalanced places).

Procedure

Sixteen participants between the ages of 18 and 51 years participated in this experiment. All participants were naive to the hypothesis and had self-reported 20/20 or corrected-to-20/20 vision and normal color vision (screened using the Ishihara color plates). The procedure was identical to that described for the transparent text experiment, except there were three blocks of trials rather than two.

Results of the Three Experiments

For all experiments and for each participant, the search time data were summarized by the median search time of the correct trials for each condition and were included as long as the participant performed above chance. (For example, with a three-alternative task and six trials per condition, at least three correct was needed to perform above chance.) The search time medians were then transformed by a logarithm transformation prior to analysis.

For the transparent text experiment, 28 participants had complete search time data sets. Data from all participants were used to analyze the error rate data. For the opaque text experiment, search time data were included from only 47 participants whose overall error rates were less than 10%. For the low-contrast text experiment, there were no participants who performed above chance for the 0.05 contrast gray-on-gray conditions. Therefore, these conditions were not included in the analysis. Several participants did not perform above chance for a small number of the other contrast conditions. Because of the small N, rather than dropping them or the conditions, an ANOVA with unequal N was performed.

For the transparent text experiment, a three-way ANOVA for search times showed significant main effects for all variables and all interactions, except the interaction between transparency type and contrast. Appendix A has a summary table of the ANOVA results, and Table A2 in Appendix A gives the mean log search times for each condition. In general, additive transparency search times were slower than multiplicative. The plain background led to significantly faster search times than the wave background, and the wave background led to significantly faster search times than the culture background. For the third main effect, the higher contrast led to faster search times. Figure 3 shows the three-way interaction from the transparent text experiment, along with the opaque and the low-contrast data. Except for the lack of the additive/multiplicative effect on the low-contrast wave background, the interactions can be explained as floor effects in the log search times.

Error rates showed results similar to those of the search times. The three-way ANOVA for error rate showed significant main effects for all variables, and all interactions were significant. There were more errors when using additive transparency, the low contrast, and the culture then the wave and then the plain background. The directions of these main effects indicate that the search time effects were not simply the result of speed–accuracy trade-offs. The pattern of the interactions was the same as with the search times, except there was also a significant effect of contrast for the wave pattern with the additive transparency. (Appendix A has a summary table of the ANOVA results and a graph showing the number of errors for each condition.)

For the opaque text data summarized from Scharff et al. (2000), a single-factor ANOVA with five levels was used on the nonfactorial design (there were three readable contrast levels for the plain background (0.15, 0.35, and 0.95) and only two readable contrast levels for the culture pattern containing all frequencies (0.35 and 0.95). Post hoc comparisons indicated that on the plain background, the small decrease in search times as the contrast increased from 0.15 to 0.35 was not significant, but on both backgrounds, the search time decrease as the contrast increased to 0.95 was significant. At the 0.35 contrast, the culture pattern significantly increased the search time relative to that of the plain background, but at the 0.95 contrast level, it did not. (Appendix A has the ANOVA summary table.)

For the low-contrast text data from Hill (2001), there was a significant effect of contrast. The 0.1 contrast level led to significantly slower search times when compared to all other contrast levels. There were no other significant differences among the remaining contrast levels. (Appendix A has the ANOVA summary table.)

Predicting Readability

Now we will look at how our previously developed index (based on the Global Masking Model, Scharff et al., 2000) predicts readability in the conditions of the three experiments. We then describe an adjustment to the global masking index in which the text and the background are used to compute text contrast and masking RMS contrast. The original global masking index was based on signal detection models, where the small effect of the signal on masking and adaptation can be ignored. Because the text comprises a relatively large part of the stimulus (~20%), we hoped that readability would be better predicted if the text was also included in the contrast calculations, and we knew that the adjustment would allow the index to predict a difference for light and dark text.

The Original Global Masking Index

As described in Scharff et al. (1999, 2000), this index, modified from a global masking model of signal detection, combines the influence of text contrast and background RMS contrast with a single parameter, the masking contrast threshold. Although the reading search task is different from a target detection task, we felt that a measure of text detectability on the background might predict search times. As explained in our previous work, the index derivation assumes a flat contrast sensitivity function, and the readers sat close enough to the display that the frequencies relevant to reading were in the optimal visual range (about 6 cpd) or lower. The global masking model also assumes that the masking contrast energy is uniform over the target background and is similar to the target in spatial frequency content.

This original readability index is defined as the equivalent text contrast on a plain background having the same discriminability. As derived in Scharff et al. (1999) for binary text, the equivalent contrast C_M of the masked text is

$$C_M = C_T/(1+ (C_{RMS}/C_2)^2)^{0.5} , \quad (5)$$

where C_2 is the masking contrast threshold. (When $C_{RMS} = C_2$, the masked contrast is obtained by dividing the unmasked contrast by $2^{0.5}$, and then C_M is 3 dB lower than C_T.) In all predictions the masking threshold was set to $C_2 = 0.05$.

Figure 4 plots the mean search times for the conditions in three experiments as a function of the global masking index of Equation 5. For the transparent text data alone (Experiment I), the global masking index provides good predictability of search times (r = 0.83) because the index's equivalent contrast for the textured backgrounds is now lower than the nominal contrast. However, this index predicts no effect of transparency type, and it predicts more masking by the wave pattern than the culture pattern, because the wave pattern background has a larger RMS contrast. For the opaque text data (Experiment II) and the low-contrast data (Experiment III), the global masking index also leads to good predictability of search times (r = 0.9 in both cases). When the three data sets are combined, the global masking index results in a Spearman rank correlation value of r = 0.89.

The Global Masking Index with Adjusted Contrast

Because the global masking index computes average luminance and RMS contrast from the background alone, it predicts that the two transparency conditions will lead to the same readability, contrary to the results shown in

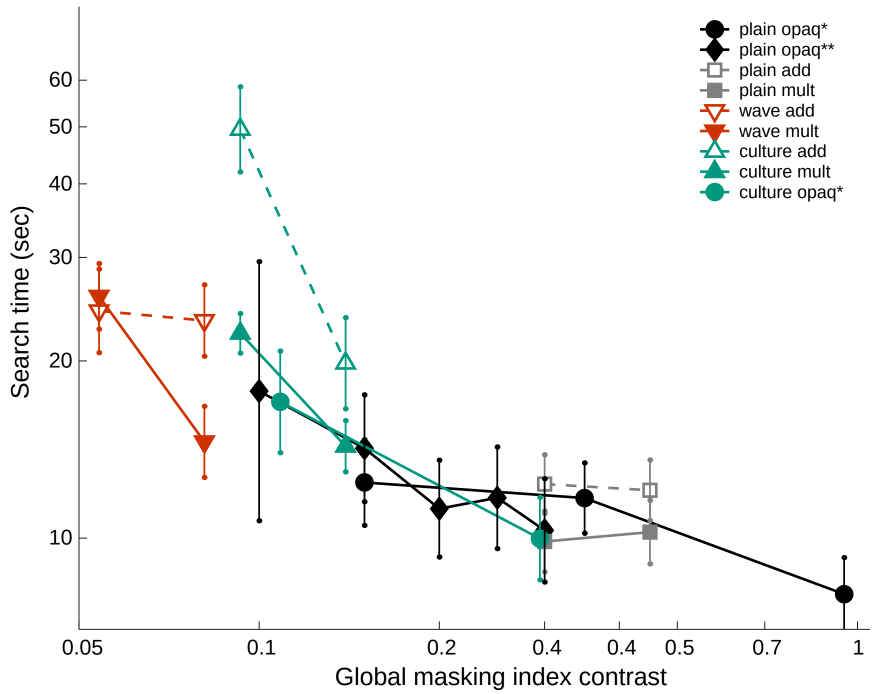

Figure 4. The same search times shown in Figure 3 plotted as a function of the global masking index (Equation 5) for the transparent text data of Experiment I, the opaque text data of Experiment II (*), and the low-contrast data of Experiment III (**). As compared with Figure 3, where the latencies are plotted as a function of nominal text contrast, the index assigns low equivalent contrasts to the textured background conditions (red and green symbols) so that they more nearly fall on one monotonic curve with the plain background conditions (black symbols). Unfortunately, the additive text search times (open symbols) are still above the multiplicative times (corresponding closed symbols) illustrating that the index does not predict the difference between the two transparency conditions.

Table 1. Original and Adjusted Text Contrast Values

Conditions	Text contrast	
	Original	Adjusted
Additive transparency	0.45	0.330
	0.30	0.226
Multiplicative transparency	–0.45	–0.396
	–0.30	–0.255
Opaque text	–0.95	–0.938
	–0.35	–0.301
	–0.1	–0.124
Low-contrast opaque text	–0.30	–0.255
	–0.25	–0.211
	–0.20	–0.167
	–0.15	–0.124
	–0.10	–0.082

Negative numbers indicate that the text was darker than the background. Text contrast values for each text type do not depend on the background pattern.

Figure 4 above, which show worse performance for the additive condition even though the effective contrasts are the same. In an effort to improve our index, we decided to remove the approximation borrowed from signal detection models that the signal, or text, would have no effect on masking and adaptation, and instead compute the average luminance and RMS contrast from the combined text and background image. Unfortunately, this decision generates a dilemma. While the backgrounds are relatively uniform, the text is not (i.e., some areas of the stimulus have text whereas others do not). An area limited to a word contains ~23% text pixels, a text area with line spaces contains ~17% text pixels, and the entire stimulus including the border areas contains only ~8% text pixels. We assumed a proportion of 20%, as though the participants were able to keep their eyes mainly on the text areas. When using the combined data and plotting the rank correlation as a function of the proportion of text pixels, any proportion from 0.08 to 0.28 gave the same value. Appendix B contains derivations for the equations for the adjusted average luminance and the adjusted RMS contrast as a function of the proportions of text pixels.

Table 1 shows the original text contrast values based on the average background luminance and the adjusted text contrast values for 20% text pixels. The absolute values of the adjusted text contrast are always lower than

Table 2. Original and Adjusted Background RMS Values

Conditions	Text contrast	Background RMS contrast		
	Original	Background	Original	Adjusted
Additive transparency	0.45	Culture	0.1533	0.217
	0.30			0.184
	0.45	Wave	0.2731	0.300
	0.30			0.284
	0.45	Plain	0.0	0.165
	0.30			0.113
Multiplicative transparency	–0.45	Culture	0.1533	0.252
	–0.30			0.201
	–0.45	Wave	0.2731	0.342
	–0.30			0.304
	–0.45	Plain	0.0	0.198
	–0.30			0.128
Opaque text	–0.95	Culture	0.1533	0.499
	–0.35			0.211
	–0.95	Plain	0.0	0.469
	–0.35			0.151
	–0.1			0.062
Low –contrast opaque text	–0.30	Plain	0.0	0.128
	–0.25			0.105
	–0.20			0.083
	–0.15			0.062
	–0.10			0.041

Negative numbers indicate that the text was darker than the background. Unlike the adjusted text contrast values, the adjusted background root mean square (RMS) values are affected by both the background pattern and the text contrast.

the absolute value of the text contrast. But for the additive cases where the text was more luminous than the background, this lowering is greater than for the other cases where the text is darker than the background. Table 2 shows the effect of including the text on background RMS contrast. Because the absolute text contrasts are smaller in the additive than the multiplicative conditions, the RMS contrasts will also be smaller. When combined with text contrast, according to Equation 5, these smaller RMS values work against the effects of text contrast, but text contrast prevails. The result is that the adjusted global masking index does predict worse performance for the additive condition. The adjusted global masking index is still an equivalent text contrast, but now it represents the contrast of an equally detectable letter or word with no background or text around it.

Table 3. Spearman Correlation Values

Data set \ metric	Global masking index	Adjusted index
Transparent text	0.83	0.87
Opaque text	0.90	0.90
Low-contrast text	0.90	0.90
Combined data	0.89	0.91

Shown are search times for all experiments separately and the combined data for both the original and the adjusted global masking index using $p_T = 0.20$. The adjusted global masking index leads to the best predictability of readability.

When using the adjusted values to calculate the global masking index, we found an improvement in predictability for the transparent text data alone. There was no change for the opaque text data or the low-contrast data rank correlations. There was also an improvement in the predictability for the combined data ($r = 0.89 \rightarrow r = 0.91$). Table 3 shows the correlation values for the reported conditions.

Figure 5 illustrates the relationship between the adjusted index and search times. The figure shows that although the adjusted measure predicts the direction of the difference between the multiplicative and the additive conditions by assigning lower effective contrast values to the additive conditions, it does not predict enough of an effect. The multiplicative culture conditions line up with those on the plain backgrounds, but the wave conditions appear to be shifted to the left (the wave background masks less than the metric predicts).

Discussion of Predictability

When there was background texture/variation, the global masking index led to good predictability. The

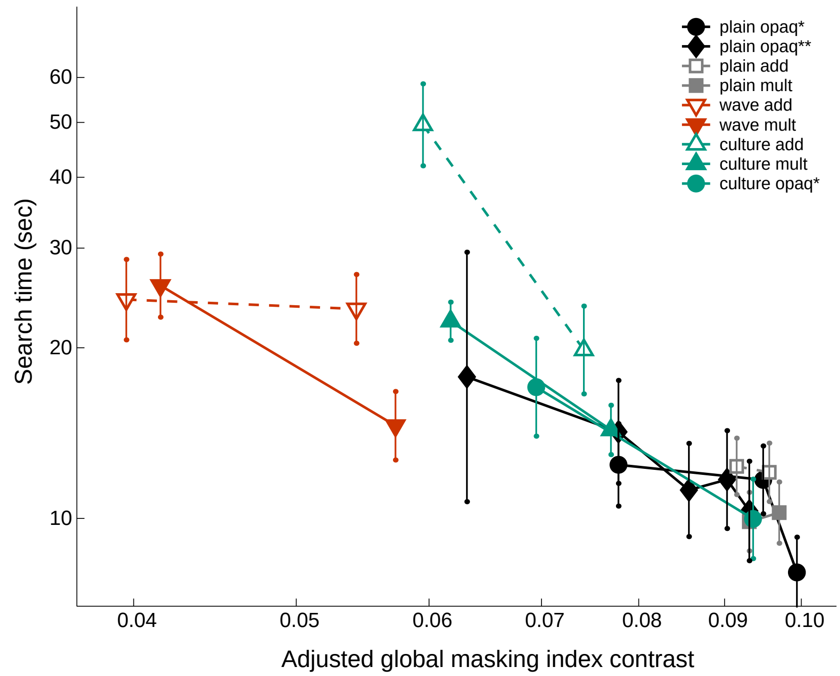

Figure 5. The same search times shown in Figures 3 and 4 plotted as a function of the adjusted global masking index for the transparent text data of Experiment I, the opaque text data of Experiment II (*), and the low-contrast data of Experiment III (**). As compared with Figures 3 and 4, where the additive text search times (open symbols) were above the multiplicative times (corresponding closed symbols), the additive times are now shifted to the left, but not enough to form a monotonic curve with their corresponding multiplicative text search times. The dark text on the culture background conditions (solid green symbols) falls into line with the plain background conditions (black symbols), but the index gives the wave background conditions (red symbols) an equivalent contrast that is too low.

adjusted index further improved predictability, but only when there was transparent text. The small improvement seen when using the adjusted global masking index for the transparent text data occurred because type of transparency (one had brighter text and one had dimmer text) as well as text contrast influenced the adjusted text contrast and background RMS contrast terms. Also, including the text in the background RMS contrast calculation decreased the influence of the background variance. Because the wave pattern had more background variation, but led to faster search times overall, the adjusted index slightly reduced this discrepancy.

An important implication that arises from the adjustment procedure results is how stimulus contrast should be calculated for text stimuli. A frequently used measure of stimulus contrast is the Mickelson contrast

$$(L_{MAX} - L_{MIN})/(L_{MAX} + L_{MIN}), \qquad (6)$$

where L_{MAX} is the peak luminance and L_{MIN} is the minimum luminance. For a stimulus, such as a grating, where the midrange equals the average luminance, this is the same as

$$(L_{MAX} - L_{AVE})/L_{AVE}\,, \qquad (7)$$

where L_{AVE} is the average luminance. For uniform text on a plain background, the contrast is commonly specified as

$$(L_T - L_B)/L_B, \qquad (8)$$

where L_T is the text luminance and L_B is the background luminance. However, all three equations can be represented by the more general equation,

$$(L_T - L_0)/L_0\,, \qquad (9)$$

where

$$L_0 = p\,L_T + (1 - p)\,L_B\,. \qquad (10)$$

When $p = .5$, Equation 9 becomes the Mickelson contrast (Equation 6). When $p = 0$, then Equation 9 becomes the commonly used text contrast (Equation 8). Finally, by setting $p = p_{TEXT}$, the proportion of text pixels, Equation 9 becomes the average luminance-based contrast (Equation 7).

In our unadjusted contrast calculations, we computed text contrast using Equation 8, which essentially assumes that the proportion of text pixels is negligible. Our adjusted contrast calculations using Equation 7 and setting the visually effective proportion of text pixels to 20% more accurately predicted the readability results. Using a 50% proportion of text pixels, the Mickelson contrast, actually generated slightly better predictions. This might mean that the mid range is better than the average as a predictor of zero contrast. It also might be that a maladjustment helps correct for other problems with the index.

General Discussion

Our results show that the effects on search times of specific background textures (wave vs. culture) are not simply predicted by background RMS contrast, although both were significantly slower than the plain background. The effects of the opaque and multiplicative text on the culture pattern were predicted by the adjusted index. The culture pattern contained less RMS contrast than the wave pattern, but the condition with the slowest search times was the low-contrast culture pattern with additive transparency. In the other contrast and transparency conditions, there was not a significant difference between the two patterns. As recommended by Ward, Parks, and Crone (1995), placing the transparent text information over less textured areas should increase readability. However, when this is not possible, background RMS contrast may not be the best predictor for readability.

Type of transparency also influenced readability. There was no evidence that the multiplicative text was either better or worse than the corresponding opaque text. However, in general, the additive text led to slower search times even taking into account the lowering of the contrast from the text luminance. Unfortunately, we do not have corresponding data on light, nontransparent text, so we do not know whether transparency or lightness is the problem.

Figure 4 shows that while equivalent contrast could be said to generate significantly lower reading performance when it is below a critical value of 0.15 (Hill, 2001), the figure is also consistent with no critical value and a continuous improvement of performance up to a contrast of 1. The possible discrepancy between this and other results strongly indicates that a critical contrast (Legge et al., 1987; and Pastoor, 1990) may be dependent on the task and the individuals performing it. For example, using a different task and two participants, Legge et al. concluded that the critical contrast for opaque text on a plain background was 0.30, whereas the results from Scharff et al. (2000) and Hill (2001) suggest that the cutoff is lower. The large variance for the 0.10 contrast level from Hill (2001) suggests that there will be individual differences (perhaps due to age in this case) with respect to such a cut off and the slow increase in performance at high contrasts may result from individual differences.

Much HUD research is concerned with accommodation issues (Edgar & Reeves, 1997; Iavecchia, Iavecchia, & Roscoe, 1988; Leitner & Haines, 1981). Rarer are HUD studies of legibility as a function of the background (Ward et al., 1995) and text contrast (Weintraub & Ensing, 1992, as cited in Ververs & Wickens, 1996). Ward et al. (1995) investigated participants' ability to identify targets and speedometer changes in simulated automobile HUDs as a function of high-, medium-, or low-background complexity (subjectively defined) and position of the HUD within the visual field. Not surprisingly, performance was better with less complex backgrounds, and better when the HUD was placed over the roadway rather than in the areas of the visual scene that contained more background variation. Unfortunately, in automobiles, there may be heavy traffic obscuring the roadway, and in airplanes, there is no analogy to a roadway; although, in general, the sky shows less variation than does a ground scene. Thus, unlike Web pages, there may not be an easy way to avoid the influence of background textures.

For text displays, such as Web pages, it is easy to recommend the use of plain backgrounds with moderate-to-high-contrast text, and very high text contrasts if patterned backgrounds are used. This recommendation is not useful for HUDs or see-through LCD displays; they will inevitably contain textured backgrounds, and while very high-contrast text may aid readability of the information, it will decrease discriminability in the background. Weintraub and Ensing (1992) concluded that, for moderate ambient illumination HUD conditions, at least a 1.5/1 luminance-contrast ratio (0.5 contrast) is the most ideal. Our results suggest that such a contrast would still lead to occasional conditions where readability would be significantly reduced. Ververs and Wickens (1996) investigated the use of different levels of contrast for different information items in the HUD. When less relevant information was presented with lower contrast, performance was better than when all information was presented with the higher contrast. However, they did not specify the contrast levels used, nor did they systematically manipulate contrast in order to determine the best values for the low- versus the high-contrast items. Our results suggest that, for plain backgrounds, the low-contrast level could be 0.30 and still be equally readable while offering the dual contrast advantage. However, for textured backgrounds, if 0.50 is used as the high-contrast level, reducing contrast much below that for the lower contrast level could easily lead to conditions where readability would be significantly hampered.

Conclusions

All of the text display variables in our experiments (transparency type, text contrast, background texture pattern) influenced readability in at least an interactive manner. Display designers would have a difficult time determining these influences when creating their displays; therefore, a metric that outputs a prediction of readability would be useful.

While the global masking index predicted readability well, the adjusted global masking index resulted in somewhat better predictability. Therefore, although the global masking index does not include display variables other than text contrast and background RMS contrast, it has successfully predicted readability search times for displays manipulating a variety of variables. The simple adjustment of the contrast calculations makes it more accurate, and its simplicity makes it appealing for use as an application metric for text display designers.

Appendix A: ANOVA Summary Tables

Table A1. Search Times Analysis Summary or Transparent Text Data (Log sec)

Effect	MS	df	MSE	df	F value	*p* level
Transparency	1.636	1	0.0352	27	46.509	0.0000003
Background	3.611	2	0.0174	54	206.983	0
Contrast	1.676	1	0.016	27	123.443	0
T x B	0.228	2	0.0156	54	14.664	0.000008
T x C	0.000002	1	0.0161	27	0.0001	0.99
B x C	0.620	2	0.0156	54	39.870	0
T x B x C	0.337	2	0.0221	54	15.249	0.000006

MS=mean squares; df=degrees of freedom; MSE=mean squares error.

Table A2. Mean Log Search Times for Each Condition in All Three Experiments

Conditions		Text contrast	Background RMS	Search times
	Background	Original	Original	(log ms)
Additive transparency	Culture	0.45	0.1533	4.30
		0.30		4.69
	Wave	0.45	0.2731	4.37
		0.30		4.39
	Plain	0.45	0.0	4.08
		0.30		4.09
Multiplicative transparency	Culture	-0.45	0.1533	4.16
		-0.30		4.35
	Wave	-0.45	0.2731	4.07
		-0.30		4.05
	Plain	-0.45	0.0	4.01
		-0.30		3.99
Opaque text	Culture	-0.95	0.1533	4.00
		-0.35		4.23
	Plain	-0.95	0.0	3.90
		-0.35		4.07
		-0.1		4.09
Low-contrast opaque text	Plain	-0.30	0.0	4.01
		-0.25		4.07
		-0.20		4.05
		-0.15		4.15
		-0.10		4.25

RMS=root mean square.

Table A3. Error Analysis Summary for Transparent Text Data

Effect	MS	df	MSE	df	F value	*p* level
Transparency	34.261	1	0.607	53	56.469	0
Background	37.113	2	0.872	106	42.556	0
Contrast	43.0386	1	0.718	53	59.957	0
T x B	13.455	2	0.466	106	28.858	0
T x C	6.520	1	0.470	53	13.880	0.000474
B x C	14.789	2	0.359	106	41.155	0
T x B x C	2.761	2	0.394	106	6.999	0.0014

MS=mean squares; df=degrees of freedom; MSE=mean squares error.

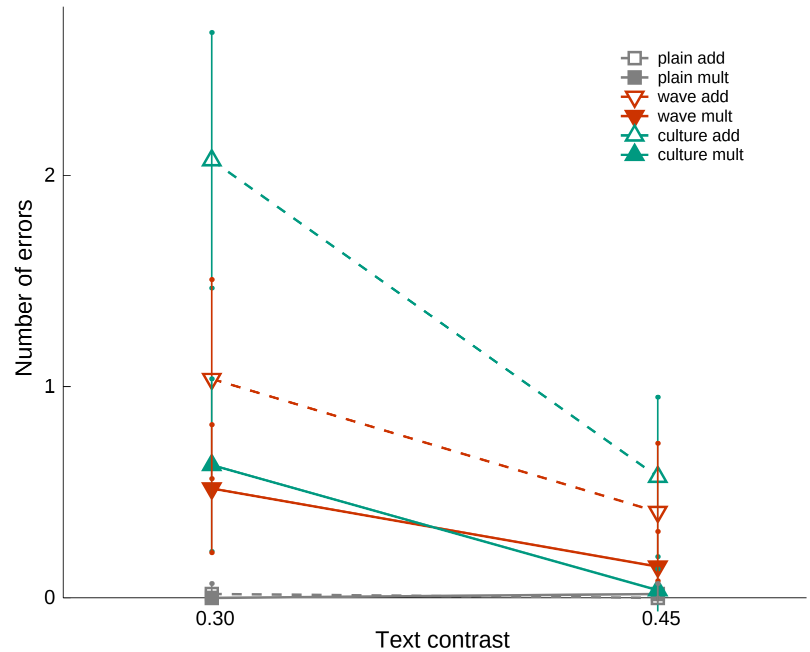

Figure A1. Average number of errors in the transparency data set.

Table A4. Opaque Text ANOVA Summary for Log Search Time Data (log sec)

Effect	MS	df	MSE	df	F value	*p* level
Conditions	0.686	4	0.0388	184	17.7	0

Conditions had five levels (0.15, 0.35, and 0.95 text contrast on a plain background, and 0.35 and 0.95 contrast on the culture background). MS=mean squares; df=degrees of freedom; MSE=mean squares error.

Table A5. Low-Contrast Text ANOVA Summary for Search Time Data (sec).

Effect	MS	df	MSE	df	F value	*p* level
Text contrast	389	4	61.2	52	6.363	0.0003

MS=mean squares; df=degrees of freedom; MSE=mean squares error.

Appendix B: Derivation of the Adjusted Global Masking Index

As stated above in "The Global Masking Index with Adjusted Contrast," to improve our index, we removed our assumption from signal detection models that the signal would have no effect on masking and adaptation. This appendix shows the derivation of the formulas used to adjust our text contrast and background RMS values shown in Tables 1 and 2.

The luminance of the text and background stimulus L_{TB} is given by

$$L_{TB} = p_T L_T + q_T L_B, \qquad (B1)$$

where L_T is the average text luminance, L_B is the average background luminance (Equation 2), p_T is the proportion of text pixels, and $q_T = 1 - p_T$. Using Equation B1, we adjusted our calculations of text contrast so that both the text and the background were used in the contrast calculations.

Analogous to Equation 2, the adjusted contrast (C_A) is defined to be

$$C_A = L_T/L_{TB} - 1. \qquad (B2)$$

To convert our previous calculations of text contrast to the adjusted version, we substituted Equation B1 into Equation B2

$$C_A = L_T/(p_T L_T + q_T L_B) - 1 ,$$

divided the top and bottom of the fraction by L_B to get

$$C_A = (L_T/L_B)/(p_T (L_T/L_B) + q_T) - 1 ,$$

used the definition of the unadjusted text contrast C_T in Equation 2 to obtain

$$C_A = (C_T + 1)/(p_T (C_T + 1) + q_T) - 1 ,$$

which simplifies to

$$C_A = q_T C_T/(p_T C_T + 1). \quad \text{(B3)}$$

A similar approach adjusts the background RMS contrast so that it includes both the text and the background. Analogous to Equation 1, the standard deviation S_{TB} of the combined contrast image is given by

$$S_{TB}^2 = E[(L_i - L_{TB})^2]/L_{TB}^2, \quad \text{(B4)}$$

where the expectation operator E[.] again takes the average over all individual pixels, indexed by i. Letting $E_T[.]$ and $E_B[.]$ be operators that average over only the text and background pixels, respectively, we can expand this as

$$S_{TB}^2 = p_T E_T [(L_i - L_{TB})^2]/L_{TB}^2 + q_T E_B [(L_i - L_{TB})^2]/L_{TB}^2.$$

This can be rewritten as

$$S_{TB}^2 = (L_B/L_{TB})^2 (p_T E_T [(L_i - L_{TB})^2]/L_B^2$$

$$+ q_T E_B [(L_i - L_{TB})^2]/L_B^2).$$

Expanding the variances about the local means, we obtain

$$S_{TB}^2 = (L_B/L_{TB})^2 (p_T (E_T [(L_i - L_T)^2] + (L_T - L_{TB})^2)/L_B^2$$

$$+ q_T (E_B [(L_i - L_B)^2] + (L_B - L_{TB})^2)/L_B^2).$$

Let S_T and S_B be the contrast standard deviations in the text and the background, respectively, with respect to the average luminance in the text or background, so that

$$S_T^2 = E_T [(L_i - L_T)^2]/L_B^2. \quad \text{(B5)}$$

and

$$S_B^2 = E_B[(L_i - L_B)^2]/L_B^2. \quad \text{(B6)}$$

Substituting the expressions B5 and B6 we obtain

$$S_{TB}^2 = (L_B/L_{TB})^2 (p_T S_T^2 + p_T (L_T - L_{TB})^2/L_B^2$$

$$+ q_T S_B^2 + q_T (L_B - L_{TB})^2/L_B^2).$$

Substituting the Equation B1 and Equation 2 and simplifying, we obtain

$$S_{TB}^2 = (1/(p_T C_T + 1)^2) \times$$

$$[p_T S_T^2 + q_T S_B^2 + p_T (C_T - p_T C_T)^2 + q_T (- p_T C_T)^2],$$

which simplifies to

$$S_{TB}^2 = [p_T S_T^2 + q_T S_B^2 + p_T q_T C_T^2]/(p_T C_T + 1)^2. \quad \text{(B7)}$$

For additive transparency text, the variation in the text is the same as in the background, so

$$S_T^2 = S_B^2. \quad \text{(B8a)}$$

and for multiplicative transparency text, the text variation is scaled by $1 + C_T$, so

$$S_T^2 = ((1 + C_T) S_B)^2. \quad \text{(B8b)}$$

For opaque text (Experiment II),

$$S_T = 0. \quad \text{(B8c)}$$

Finally, for text on a *plain* background (all low-contrast conditions),

$$S_T = S_B = 0. \quad \text{(B8d)}$$

The final adjusted masking index is obtained by substituting C_A for C_T and S_{TB} for C_{RMS} in Equation 5 giving

$$C_{AM} = C_A/(1+ (S_{TB}/C_2)^2)^{0.5}. \quad \text{(B9)}$$

Acknowledgments

We would like to thank Lori Shird for her assistance in the data collection process. Funding for this work was provided by the Airspace Operations Systems (AOS) Project of NASA's Airspace Systems Program and supported by NASA RTOP 711-51-12 through the Cooperative Agreement NCC 2-1095 with the San Jose State University Foundation, San Jose, CA. Commercial Relationships: None.

References

D'Zmura, M., Colantoni, P., Knoblauch, K., & Laget, B. (1997). Color transparency. *Perception, 26*, 471-492. [PubMed]

Edgar, G., & Reeves, C. (1997). Visual accommodation and virtual images: Do attentional factors mediate the interacting effects of perceived distance, mental workload, and stimulus presentation modality? *Human Factors, 39*, 374-382. [PubMed]

Farrell, J., & Fitzhugh, A. (1990). Discriminability metric based on human contrast sensitivity. *Journal of the Optical Society of America A, 7*, 1976-1984. [PubMed].

Hill, A. (2001). Readability of screen displays as a function of color contrast and luminance contrast. Unpublished master's thesis, Stephen F. Austin State University, Nacogdoches, Texas.

Hill, A., & Scharff, L. (1999). Readability of computer displays as a function of colour, saturation, and background texture. In D. Harris (Ed.), *Engineering psychology and cognitive ergonomics: Vol 4.* Aldershot, UK: Ashgate.

Hill, A., & Scharff, L. V. (1997). Readability of screen displays with various foreground/background color combinations, font styles, and font types. *Proceedings of the Eleventh National Conference on Undergraduate Research, II*, 742-746.

Iavecchia, J., Iavecchia, H., & Roscoe, I. (1988). Eye accommodation to head-up virtual images. *Human Factors, 30*, 689-702. [PubMed]

Kember, P., & Varley, D. (1987). The legibility and readability of a visual display unit at threshold. *Ergonomics, 30*, 925-931. [PubMed]

Knoblauch, K., Arditi A., & Szlyk, J. (1991). Effects of chromatic and luminance contrast on reading. *Journal of the Optical Society of America A, 8*, 428-439. [PubMed]

Legge, G., Pelli, D., Rubin, G., & Schleske, M. (1985). Psychophysics of reading. I. Normal vision. *Vision Research, 25*, 239-252. [PubMed]

Legge, G., & Rubin, G. (1986). Psychophysics of reading. IV. Wavelength effects in normal and low vision. *Journal of the Optical Society of America A, 3*, 40-51.

Legge, G., Rubin, G., & Luebker A. (1987). Psychophysics of reading. V. The role of contrast in normal vision. *Vision Research, 27*, 1165-1177. [PubMed]

Leitner, E., & Haines, R. (1981). Magnitude of visual accommodation to a head-up display. *NASA Technical Paper* 1796, National Technical Information Service, Springfield, VA.

Mulligan, J. (1993). Nonlinear combination rules and the perception of motion transparency. *Vision Research, 33*, 2021-2030. [PubMed]

Parish, D. H., & Sperling, G. (1991). Object spatial frequencies, retinal spatial frequencies, noise, and the efficiency of letter discrimination. *Vision Research, 31*, 1399-1415. [PubMed]

Parker, B., & Scharff, L. V. (1997). *Influences of contrast sensitivity on text readability in the context of a GUI*. Retrieved November 22, 2002 from http://hubel.sfasu.edu/research/agecontrast.html

Pastoor, S. (1990). Legibility and subjective preference for color combinations in text. *Human Factors, 32*, 157-171. [PubMed]

Regan, D., & Hong, X. H. (1994). Recognition and detection of texture-defined letters. *Vision Research, 34*, 2403-2407. [PubMed]

Scharff, L., Ahumada, A. J., & Hill, A. (1999). Discriminability measures for predicting readability, human vision and electronic imaging IV. *Proceedings of SPIE, 3644*, 270-277.

Scharff, L. V., Hill, A. L., & Ahumada, A. J. (2000). Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express, 6*, 81-91. [Article]

Schorno, G. (1996). Texture tiles. Retrieved September 9, 1996 from http://mars.ark.com/~gschorno/tiles/

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. *Nature, 369*, 395-397. [PubMed]

Ververs, P., & Wickens, C. (1996). The effect of clutter and lowlighting symbology on pilot performance with head-up displays. *Proceedings of the Human Factors and Ergonomics Society, 40*, 62-66.

Ward, N., Parks, A., & Crone, P. (1995). Effects of background scene complexity and field dependence on the legibility of head-up displays for automotive applications. *Human Factors, 37*, 735-745. [PubMed]

Weintraub, D., & Ensing, M. (1992). Human factors issues in head-up display design: The Book of HUD: State-of-the-art report. Wright-Patterson AFB, Dayton, OH: Crew System Ergonomics Information Analysis Center.