Predicting the Readability of Transparent Text

Lauren F. V. Scharff
Psychology, Stephen F. Austin State University, Nacogdoches, TX, USA
lscharff@sfasu.edu

Albert J. Ahumada, Jr.
NASA Ames Research Center, Moffett Field, CA, USA
aahumada@mail.arc.nasa.gov

Text readability was measured for two types of transparent text (additive and multiplicative) at two contrast levels (0.3 and 0.45) on three background textures (culture, wave, plain), and it was measured for five levels of low text contrast (0.1, 0.15, 0.2, 0.25, 0.3) on plain backgrounds. For the transparent text, reading search times were longer for additive transparency, the low contrast, and the culture then the wave and then the plain background. For the low contrast experiment the 0.1 contrast level led to significantly slower search times when compared to all other contrast levels. When there were background textures a masking index that combined text contrast and background RMS contrast predicted search times much better than either measure alone. 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.

Introduction

Text readability is influenced by a large number of factors, many of which have been previously studied, e.g. luminance and/or chromatic contrast (Legge, Rubin, and Luebker, 1987; Knoblauch, Arditi, and Szlyk, 1990), color (Legge, and Rubin, 1986; Pastoor, 1990), blur (Legge, Pelli, Rubin, and Schleske, 1985; Farrell and Fitzhugh, 1990), the addition of noise (Parish and Sperling, 1991; Solomon and Pelli, 1994; Regan and Hong, 1994), case (Kember and Varley, 1987), polarity (Legge, et al., 1985; Parker and Scharff, 1997), and the use of textured backgrounds (Hill and Scharff, 1999; Scharff, Ahumada, and Hill, 1999; Scharff, Hill, and Ahumada, 2000). However, the large number of possible combinations of even this non-comprehensive list of factors means that, if a display designer desired to maximize readability, the particular combination of factors of interest probably would not have previously been examined for readability. Thus, a metric to predict readability would be quite useful.

Scharff, Ahumada, and Hill (1999), and subsequently Scharff, Hill, and Ahumada (2000) investigated the ability of two image measures (text contrast and background 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. Their indices better predicted readability than the image measures alone. 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 metrics generalize to text displays incorporating additional factors? One such factor, which may influence life-or-death decisions, is transparent text as is used in head-up displays (HUDs) in some airplanes and automobiles. Of additional interest is the use of very low contrast text. Because it is obviously detrimental to readability, most (but not all) display designers know to avoid low contrast text. However, in HUDs, there may be regions of a display that result in very low text contrast, simply because the background may show large variations in luminance or because high text contrast would mask critical features of the image.

There has been much previous research on HUDs, especially with respect to accommodation issues (e.g. Edgar and Reeves, 1997; Iavecchia, Iavecchia, and Roscoe, 1988; Leitner and Haines, 1981). Of more relevance to the current work are HUD studies of legibility as a function of the background (Ward, Parks, and Crone, 1995), and text contrast (Weintraub and Ensing, 1992, as cited in Ververs and 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 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, there may not be an easy way to avoid the influence of background textures. Overall, their work supports the merit of investigating the influence of background textures and the development of a metric to predict readability of transparent text.

With respect to the best contrast for HUDs, Weintraub and Ensing (1992, as cited in Ververs and Wickens, 1996) concluded that, for moderate ambient illumination conditions, at least a 1.5/1 luminance-contrast ratio is the most ideal. If the contrast is too high, it can be distracting and obscure items in the background, and if it is too low, it can be difficult to read. 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. If our metric predicts readability for low as well as higher text contrasts, it may be useful in determining appropriate high and low contrasts for a dual-contrast-level HUD.

The purpose of this current work is first, to measure readability (search times) for two types of transparent text (additive and multiplicative) at two contrast levels on three background textures. Second, readability was measured for six levels of low text contrast on plain backgrounds. Both experiments used the same basic procedure to measure readability as was previously used by 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.

Following these readability measures, we investigated how well the Global Masking Index, in comparison to the simple image measures of text contrast and background RMS contrast, would predict the readability of such text displays. Although this index again better predicted readability than the image measures, we performed a second series of calculations using an adjusted Global Masking Index. With this adjusted index we moved away from the assumption from signal detection that the signal would have no effect on masking and adaptation. When both the text and the background were used to calculate image contrast, predictability of readability was improved.

Methods

Two experiments measuring readability

Macintosh Power PC 7200/120 computers were used to create and run both 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) and resulted in a viewing angle of 0.04 deg for each pixel.

Experiment I: Measuring the effects 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 included a plain (uniform) background and two periodic textures taken from a webpage 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). 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, leading to a 15.5 cm square texture (18.36 deg /side). The plain background was matched in size. See Figure 1 for examples of single 75 by 75 pixel tiles of the three backgrounds.

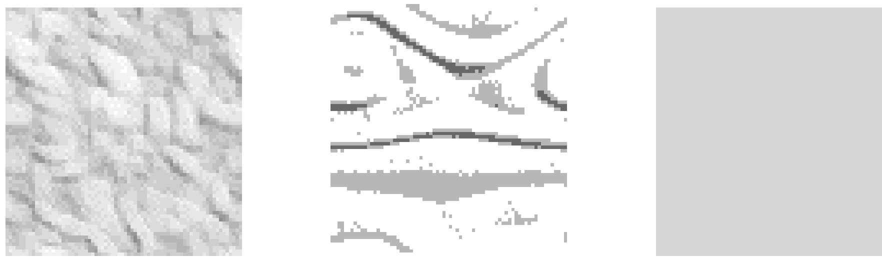
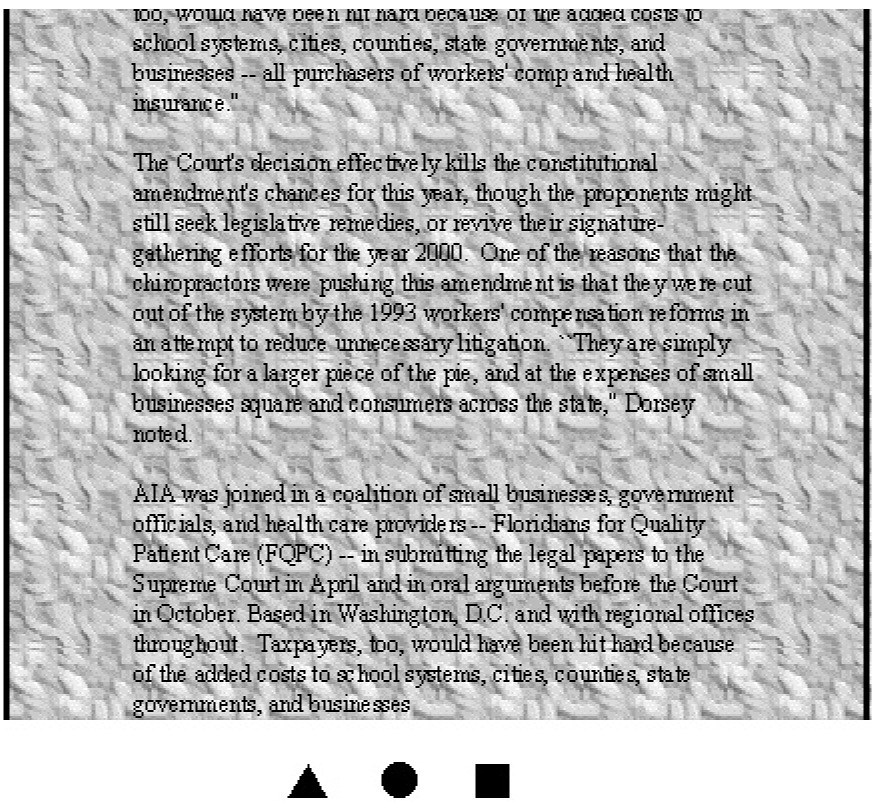


Figure 1. From left to right, tiling elements of the three backgrounds used in the transparent text experiment: culture, wave, plain.

Seven newspaper excerpts presented in 12 point (6 vertical pixels per letter x, 0.25° at our viewing distance) Times New Roman font were used to create the text arrays. These 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) each contained 99-101 words. A target word ("triangle", "circle", or "square") was placed in a counterbalanced manner within each text block, so that there were 12 of each text excerpt (one for each of the 12 conditions). Figure 2 shows an example text display on the culture background together with the response choices.



too, would have been hit hard because of the added costs to school systems, cities, counties, state governments, and businesses -- all purchasers of workers' comp and health insurance."

The Court's decision effectively kills the constitutional amendment's chances for this year, though the proponents might still seek legislative remedies, or revive their signature-gathering efforts for the year 2000. One of the reasons that the chiropractors were pushing this amendment is that they were cut out of the system by the 1993 workers' compensation reforms in an attempt to reduce unnecessary litigation. ``They are simply looking for a larger piece of the pie, and at the expenses of small businesses square and consumers across the state," Dorsey noted.

AIA was joined in a coalition of small businesses, government officials, and health care providers -- Floridians for Quality Patient Care (FQPC) -- in submitting the legal papers to the Supreme Court in April and in oral arguments before the Court in October. Based in Washington, D.C. and with regional offices throughout. Taxpayers, too, would have been hit hard because of the added costs to school systems, cities, counties, state governments, and businesses

Figure 2. An example text display with a multiplicative contrast of 1.0 on the culture background. The correct response is to click on the square.

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). 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.)

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

$$T_A = B + C_T L_B T \quad , \qquad (1)$$

where T is the text array with text pixels having a value of one and non-text 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 computed to have given text contrasts with respect to the average background luminance. Their combination rule was

$$T_M = B * (1 + C_T T) \,, \qquad (2)$$

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

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 of the text excerpts was used for the four practice trials, while the remaining six were each presented once for each of the 12 conditions. (Links to actual stimuli can be found at http://hubel.sfasu.edu/research/tt_stim/extransstim.html.)

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 one hour, 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 great majority of undergraduate students are between the ages of 18 and 21.

Participants were instructed to scan the text and find a target shape word (“triangle”, “square”, or “circle”). Once 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 minutes.

Experiment II: Measuring the effects of very low contrast

Design and Stimuli

This experiment (summarized from Hill, 2001) originally employed a 3 (background luminance levels: 70, 80, 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 current 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.

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 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 as described above, except that there were three blocks of trials rather than two.

Results of the two experiments

For both experiments, the search time data were sorted by each condition for each participant and the median of the correct trials for each condition was calculated as long as the participant got three or more correct in that condition. (With a three-alternative task and six trials per condition, at least three correct was needed to perform above chance.)

Following these criteria, for the transparent text experiment, there were twenty-eight participants who had complete search time data sets to be analyzed. A second analysis including all participants was performed using the error rate data.

For the low contrast 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.

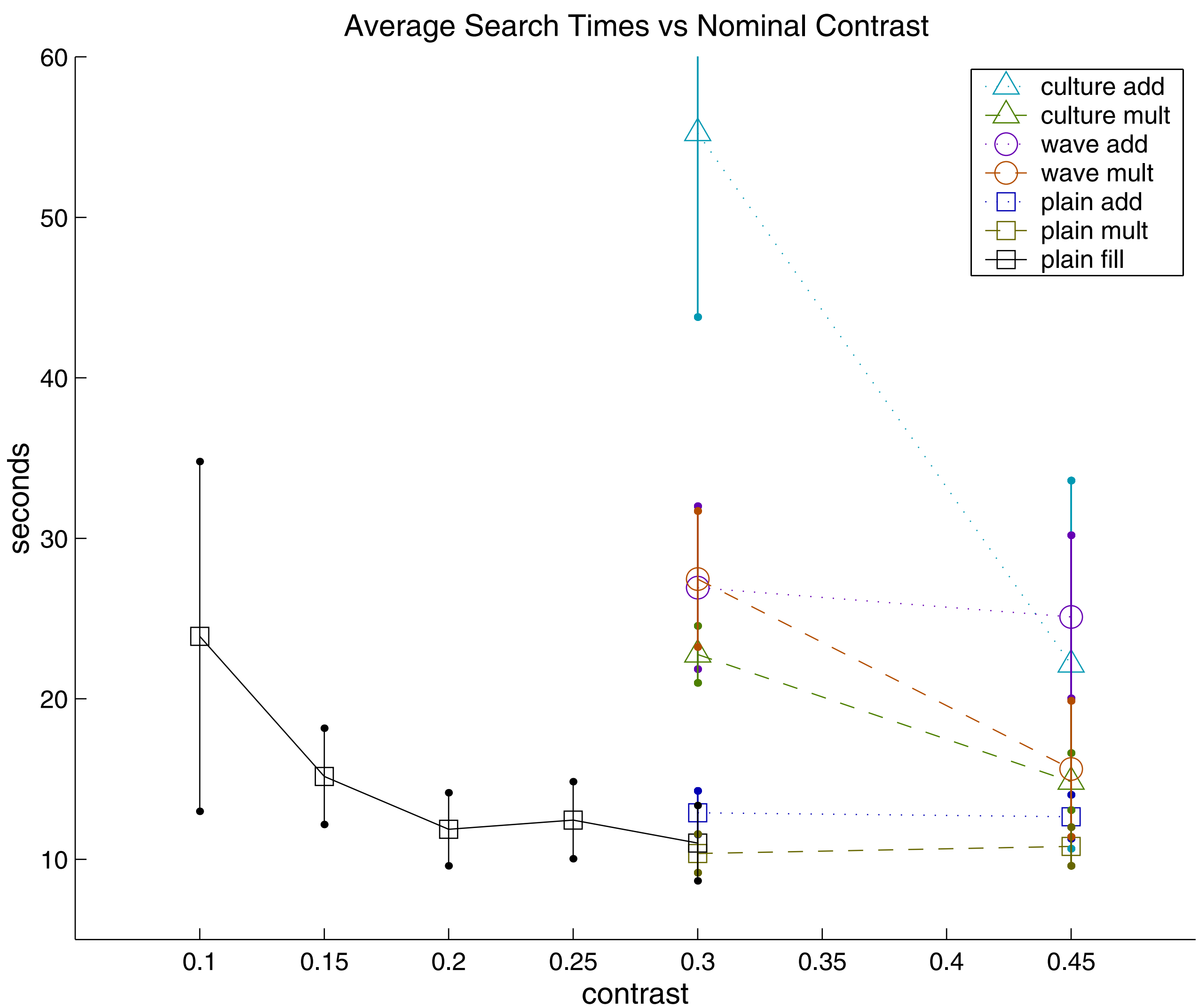


Figure 3. Search times and standard error bars for the transparent text and the low contrast experiments.

For the transparent text experiment, results of the three-way ANOVA for search times showed significant main effects for all variables, and all interactions were significant. In general, type of text transparency significantly affected search times, F(1, 27) = 30.79, p<.01, with additive transparency slower than multiplicative transparency. Background influenced search times, F(2, 54) = 76.19, p<.01, with the plain background leading to significantly faster search times than the wave background, and the wave background leading to significantly faster search times than the culture background. Contrast also significantly influenced search times, F(1,27) = 58.82, p<.01, with the higher contrast leading to faster search times. The two-way interaction between transparency and contrast was significant at the 0.05 level, and all other interactions were significant at the 0.01 level. Figure 3 shows the three-way interaction from the transparent text experiment, plus the low contrast data. For the plain background, contrast showed no effect. Contrast showed significant effects for all textured backgrounds except the wave pattern with the additive transparency. The condition that most strongly affected search times was that using the low contrast culture pattern when the transparent text was additive.

Results of the three-way ANOVA for error rate also showed significant main effects for all variables (p<0.01), and all interactions were significant (at least p<0.05). 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 there were no 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.

For the low contrast experiment there was a significant effect of contrast (p<0.25). The 0.1 contrast level led to significantly slower search times when compared to all other contrast levels. There were no other significant differences between the remaining contrast levels.

Discussion of the two experiments

These two experiments show that multiple factors interact to influence readability. As was seen for non-transparent text (Scharff, et al. 2000), background variation (texture) effects are more apparent for the low contrast stimuli. At the 0.45 contrast level, especially for the multiplicative transparency, there was little difference in readability for the textured backgrounds versus the plain background, while there were large differences at the 0.30 contrast level. Both our contrasts were below the 1.5/1 ratio recommended by Weintraub and Ensing (1992). If their ratio is considered the ideal for important HUD items, and less important items are displayed at a lower contrast (as suggested by Ververs and Wickens, 1996), then our results suggest that this lower contrast be 0.45 rather than 0.30.

The results from Hill (2001) show that, even for plain backgrounds and nontransparent text, very low contrasts can affect readability. Other researchers have also shown this effect of contrast (e.g. Legge, Rubin, and Luebker, 1986, and Pastoor, 1990), however, the critical contrast 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 was 0.30, while Hill found the critical drop-off to occur between 0.15 and 0.10.

With respect to background complexity, the amount of background RMS contrast did not correlate with search times for the two patterns, although both were significantly slower than the plain background. 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, et al. (1995), placing the HUD 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; in general, the multiplicative transparency conditions were less detrimental to readability. However, as with the higher contrasts, type of transparency did not influence the readability of the plain backgrounds.

For text displays such as web pages, it is easy to recommend the use of only plain backgrounds with moderate to high contrasts, and very high text contrasts if patterned backgrounds are used. It is not possible to make this simple recommendation for HUDs; they will inevitably contain textured backgrounds, and while very high contrasts may aid readability of the HUD information, they will decrease the discriminability of items in the background. Further, most HUDs use additive transparency, which may have more detrimental effects on readability of the HUD. It is important to note, however, that the stimulus displays (especially those with nontransparent text) and readability task used in our experiments are very different from those that would be used in a HUD or HUD-related task.

Predicting Readability

Of interest here is how well our previously developed metric (based on the Global Masking Model, Scharff et al., 2000) will predict readability when using transparent text and very low contrast text. As we have done in the past, the success of the metric is compared to that of the two image measures of text contrast and background RMS contrast.

Following this test of the Global Masking Index, we modified our approach so that the text as well as the background was used to compute text contrast and masking RMS contrast. The Global Masking Index was based on signal detection models, where the effect of the signal on masking and adaptation can be ignored. Because the text comprises a relatively large part of the stimulus (~20%), we felt that readability would be better predicted if the text was also included in the contrast calculations.

Image Measure Regressions

As described in Scharff et al. (2000), the specific image measures used in the analyses included text contrast and background RMS contrast. The text contrast was defined as

$$C_T = (L_T - L_B)/L_B = L_T/L_B - 1, \qquad (3)$$

where L_B is the average background luminance and L_T is the average text luminance. The background RMS contrast was defined as

$$C_{RMS} = E[(L_i - L_B)^2]^{0.5} / L_B = ((\Sigma(L_i - L_B)^2)/n)^{0.5} / L_B, \qquad (4)$$

where E[.] is the averaging operator, the summation (Σ) is over all background image pixels, L_i is the luminance of the i*th* pixel, and n is the number of pixels.

For the transparent text experiment, both text contrast and background RMS contrast showed fairly equivalent Spearman rank correlations with search times. The text contrast had a negative correlation with search time ($r = -0.34$) and background RMS contrast had a positive correlation ($r = 0.77$).

For the low contrast experiment, the text contrast also had a negative correlation with search time, although it was much higher ($r = -0.90$). Not surprisingly, since all the backgrounds were plain / without texture, background RMS contrast showed a zero correlation with search times.

When the transparent text and low contrast text data were combined, the text contrast image measure was not effective in predicting the search times ($r = -0.08$), while the combined background RMS contrast measure was slightly less effective than it was for the transparent text alone ($r = 0.72$).

The Global Masking Index

As described in Scharff et al. (1999, 2000), this index, modified from a global masking model of signal detection, allows us to investigate the combined influence of both text contrast and background RMS contrast without estimating additional parameters. Although our reading search task (which in every trial a target was always present) was different from the typical detectability task (which uses a target present/absent decision), we felt that a measure of text detectablity on the background might predict search times. As in our previous work, our calculations assume a flat contrast sensitivity function, since the readers sat close enough to the display that the frequencies relevant to reading were in the optimal visual range (about 6cpd) or lower. Additionally, the global masking model assumes that the masking contrast energy is uniform over the target background and is similar to the target in spatial frequency.

We define the readability index as the text contrast that would give the same discriminability on a uniform background. As derived in Scharff et al. (1999) for binary text, the readability index is independent of the size of the text target and the contrast sensitivity, giving the equivalent luminance contrast C_M of the masked text as

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

For the transparent text data, the global masking index leads to much better predictability of search times than either of the image measures alone ($r = 0.83$). However, this index predicts no effect of transparency type, and it predicts more masking by the wave pattern, since it has a larger background RMS contrast. For the low contrast data, the global masking index leads to no improvement in predictability when compared to the text contrast image measure ($r = 0.90$).

When the two data sets are combined, the global masking index results in a correlation value of r = 0.87, which is greater than either image measure alone.

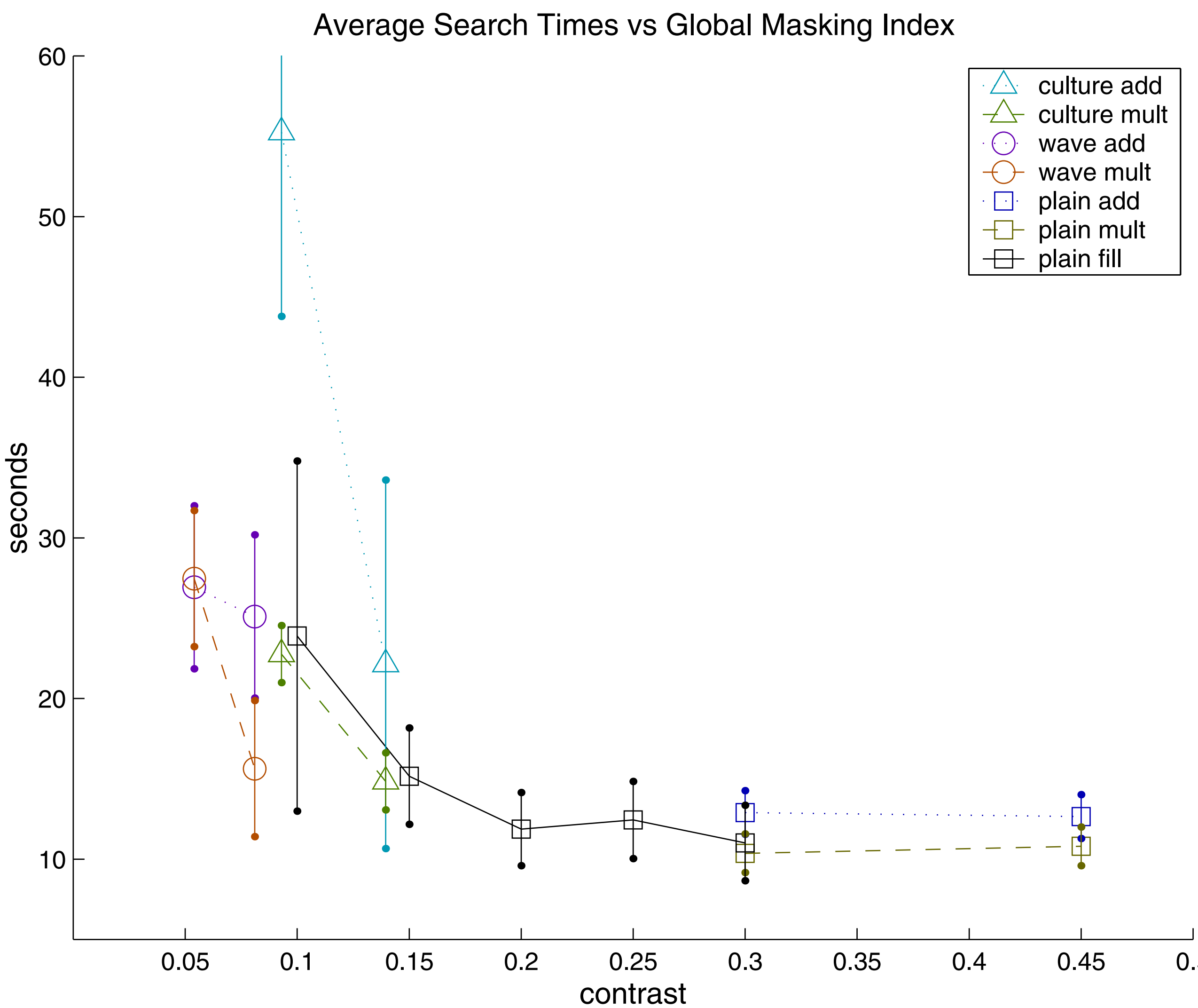


Figure 4 shows the relationship between the search times and the global masking index for the transparent text experiment and the low contrast experiment combined. As compared with Figure 3, where the latencies are plotted as a function of nominal text contrast, the index assigns low equivalent contrasts to the textured background conditions.

The Global Masking Index with Adjusted Contrast

In an effort to improve our index, we re-examined our assumption from signal detection models that the signal would have no effect on masking and adaptation. In our stimuli, the proportion of text pixels in the stimulus is $p_T = 0.17$. The luminance of the text and background stimulus L_{TB} is thus given by

$$L_{TB} = p_T L_T + (1 - p_T) L_B \quad , \tag{6}$$

where L_T is the average text luminance, and L_B is the average background luminance as above. Using Eq. 6 we adjusted our calculations of text contrast so that both the text and the background were used in the contrast calculations.

The adjusted contrast (C_A) is defined to be

$$C_A = L_T / L_{TB} - 1 \, . \tag{7}$$

In order to convert our previous calculations of text contrast to the adjusted version, we substituted Eq. 6 into Eq. 7 and used the definition of the unadjusted text contrast C_T in Eq. 3 to obtain

$$C_A = C_T (1 - p_T) / (p_T C_T + 1) \, . \tag{8}$$

When solving for the two transparency cases, text contrast (C_T) had two values, 0.30 and 0.45. For the additive case, the text was brighter than the background, and the two C_T values 0.30 and 0.45 convert to C_A values of 0.235 and 0.345. However, for the multiplicative case, the text was dimmer than the background, so the C_T values -0.30 and -0.45 convert to the C_A values -0.261 and -0.403. In both cases, the absolute contrast is reduced, but the reduction is smaller for the multiplicative text. In the low contrast experiment, the text was dimmer than the background, so C_T values -0.1, -0.15, -0.2, -0.25, -0.3 converted to C_A values -0.084, -0.127, -0.171, -0.216, -0.261.

We took a similar approach to adjust the background RMS contrast so that it included both the text and the background. The standard deviation S_{TB} of the combined contrast image is given by

$$S_{TB}^2 = E[(L_i - L_{TB})^2] / L_{TB}^2, \quad (9)$$

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 write this as

$$S_{TB}^2 = p_T E_T[(L_i - L_{TB})^2] / L_{TB}^2 + (1 - p_T) E_B[(L_i - L_{TB})^2] / L_{TB}^2. \quad (10)$$

When we express Eq. 10 in terms of our earlier contrast variables, we find

$$S_{TB}^2 = p_T ((C_T + 1)^2 S_T^2 + C_T^2) + (1 - p_T) ((C_B + 1)^2 S_B^2 + C_T^2) \quad (11)$$

where S_T and S_B are the contrast standard deviations in the text and the background, respectively, and C_B is the contrast of the average background with respect to the combined stimulus,

$$C_B = L_B / L_{TB} - 1. \quad (12)$$

For text on a uniform background (all low contrast conditions),

$$S_T = S_B = 0. \quad (13a)$$

For additive transparency text,

$$S_T^2 = S_B^2 / (C_T + 1)^2, \quad (13b)$$

and for multiplicative transparency text,

$$S_T^2 = S_B^2, \quad (13c)$$

where

$$S_B^2 = E_B[(L_i - L_B)^2] / L_B^2. \quad (14)$$

When the adjusted image measures were correlated with search times, the text contrast measures that included transparent text data showed improvements in predictability (r = -0.34 → r = -0.43 for the transparent text data alone, and r = -0.08 → r = -0.19 for the combined data). The low contrast data showed no change. Meanwhile, the adjusted background RMS contrast measure decreased predictability for both data sets that included the transparent text (r = 0.77 → r = 0.59 for the transparent text alone, and r = 0.72 → r = 0.44 for the combined data), but it increased predictability for the low contrast text data (r = 0.0 → r = -0.9). This latter value is the rank correlation of response time with text contrast for the low contrast data. This increase is expected because, for plain backgrounds, the adjusted RMS contrast measure S_{TB} is a function only of the text contrast C_T and the proportion of text pixels p_T. When using the adjusted values to calculate the Global Masking Index, we found an improvement in predictability for the transparent text data (r = 0.83 -> r = 0.92). There was no change for the low contrast data rank correlation (r = 0.9) because in both cases, the index values were just a function of text contrast. However, there was also an improvement in the predictability for the combined data (r = 0.87 -> r = 0.94). Table 1 shows all the correlation values for all of the reported conditions. Figure 5 shows the relationship between the adjusted index and search times. The figure shows that the adjusted measure accounts well for the difference between the multiplicative and the additive conditions by assigning lower effective contrast values to the additive conditions. The high contrast culture condition falls on the plain background curve and the low contrast culture condition appears to be an extension of that curve, but the wave conditions appear to be shifted to the left (the wave background masks less than the metric predicts).

Metric	Transparent Text	Low Contrast Text	Combined Data
Text Contrast	-0.34	-0.9	-0.08
Adjusted Text Contrast	-0.43	-0.9	-0.19
Background RMS Contrast	0.77	0.0	0.72
Adjusted RMS Contrast	0.59	-0.9	0.44
Global Masking Index	0.83	-0.9	0.87
Adjusted Index	0.92	-0.9	0.94

Table 1. Spearman correlation values for the Transparent Text Data, the Low Contrast Text Data, and the combined data for the two images measures and the Global Masking Index, for both the original calculations and the adjusted calculations. The Global Masking Index leads to better predictability for any conditions containing transparent text data, and is no worse for the low contrast text data. The adjusted Global Masking Index leads to the best predictability of readability.

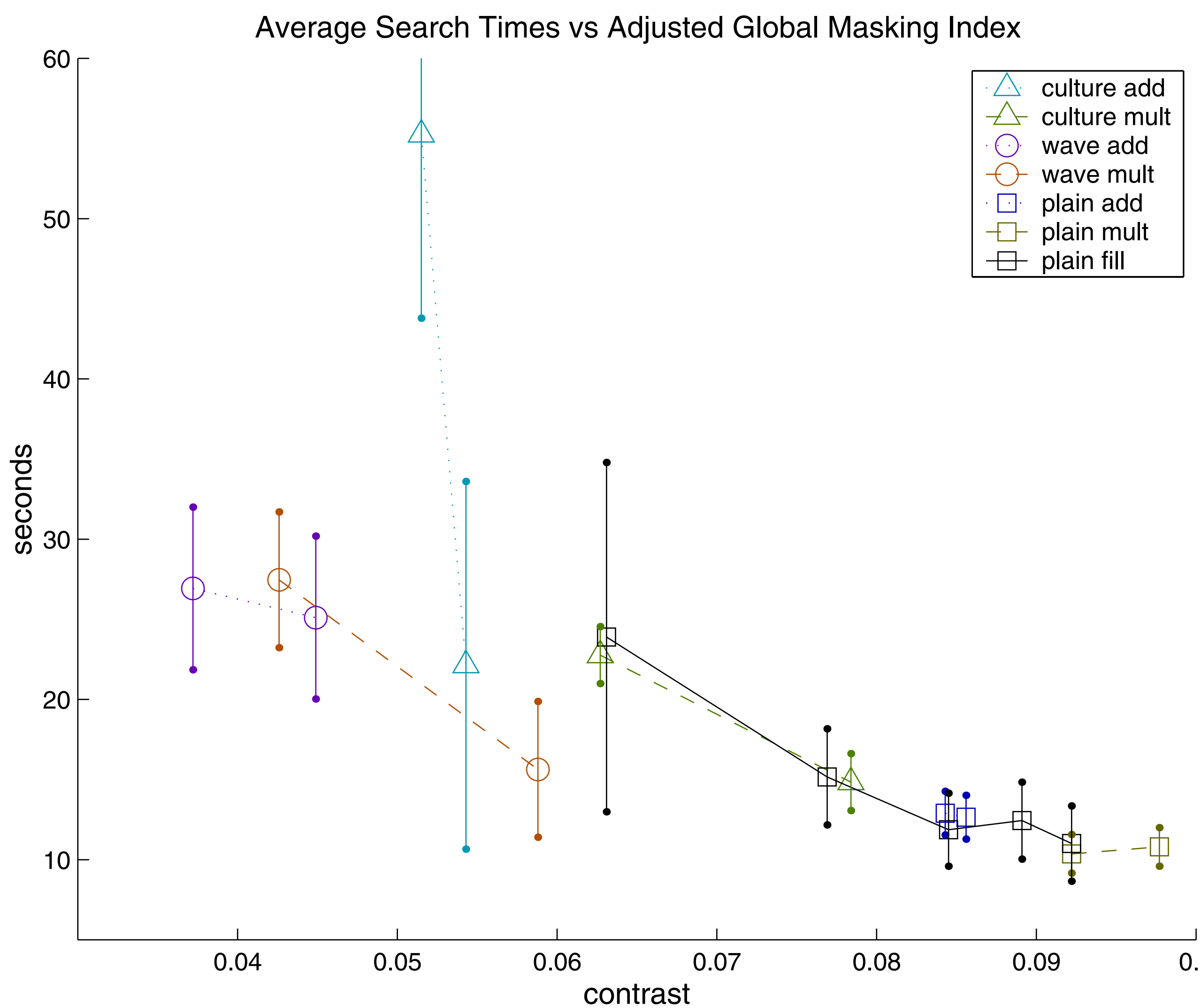


Figure 5 shows the relationship between search times and the adjusted Global Masking Index for the transparent text data and the low contrast data combined. As compared with Figures 3 and 4 where the additive scores were above the multiplicative scores, the additive scores are shifted to the left, so they fall along a curve. Too much masking is predicted for the wave background.

Discussion of Predictability

When there was background texture / variation, the Global Masking Index leads to better predictability of reading search times than either text contrast or background RMS contrast alone. The adjusted index further improved predictability, but again, only when there was background texture. When there was no background texture, however, predictability for either index was not compromised relative to the single image measures, so they were still effective for predicting readability.

The 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. Further, by including the text in the background RMS contrast calculation, the influence of the background variance alone was decreased. As noted above in the experiment discussion, the wave pattern had more background variation, but overall, it led to faster search times. The adjusted index minimized the discrepancy between the predictor and the search times.

Conclusions

All of the text display variables in our two 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 image measure of text contrast alone did a good job of predicting readability for non-transparent text placed on uniform backgrounds, the Global Masking Index predicted readability equally well. Further, when there were background textures (the current transparent text experiment and Scharff et al., 2000), the Global Masking Index resulted in much better predictability than either of the image measures. When the Global Masking Index was adjusted to include the text as well as the background in the calculations of text and background RMS contrast, predictability was further improved for the transparent text displays.

In sum, 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 even more accurate, and its simplicity makes it appealing for use as an application metric for text display designers.

Acknowledgements

We would like to thank Lori Shird, who assisted in the data collection process. This work was supported by NASA RTOP 711-51-12 directly and indirectly through Cooperative Agreement NCC 2-1095 with the San Jose State University Foundation. Commercial Relationships: None.

References

Edgar, G. and 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.

Farrell, J. and Fitzhugh, A. (1990). Discriminability metric based on human contrast sensitivity, *J. Opt. Soc. Am. A*. 7, 1976-1984.

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. and 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.

Hill, A. and Scharff, L. "Readability of computer displays as a function of colour, saturation, and background texture" in *Engineering Psychology and Cognitive Ergonomics Volume Four*, D. Harris, ed. (Ashgate, Aldershot, England, 1999).

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

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

Knoblauch, K., Arditi A., and Szlyk, J. (1990). "Effects of chromatic and luminance contrast on reading, *J. Opt. Soc. Am. A*. **7**, 1976-1984.

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

Legge, G. and Rubin, G. (1986). Psychophysics of reading IV. Wavelength effects in normal and low vision, *J. Opt. Soc. Am. A*. **3**, 40-51.

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

Leitner, E. and Haines, R. (1981). Magnitude of visual accommodation to a head-up display. *NASA Technical Paper* 1796.

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

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

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

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

Scharff, L. , Ahumada, A. J., and Hill, A. (1999). Discriminability Measures for Predicting Readability, Human Vision and Electronic Imaging IV, SPIE Proc. 3644, 270-277.

Scharff, L. V., Hill, A. L., and Ahumada, A. J. (2000). Discriminability measures for predicting readability of text on textured backgrounds. *Optics Express*, 6 (4), pp. 81-91.

Schorno, G. "Texture tiles," http://mars.ark.com/~gschorno/tiles/index.html , 1996.

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

Ververs, P. and 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. and 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.