The effects of natural scene statistics on text readability in additive displays

Daryn R. Blanc-Goldhammer, Facebook Reality Labs, Redmond, WA
Kevin J. MacKenzie, Facebook Reality Labs, Redmond, WA

The minimum contrast needed for optimal text readability with additive displays (e.g. AR devices) will depend on the spatial structure of the background and text. Natural scenes and text follow similar spectral patterns. Therefore, natural scenes can mask low contrast text – making it difficult to read. In a set of experiments, we determine the minimum viable contrast for readability on an additive display. Reading performance was assessed with an RSVP task. Text was additively overlaid on algorithmically generated images that mimicked natural scene statistics. When the text to background contrast ratio was below 1.6:1, participants' reading rates decreased rapidly. At low contrast ratios (<1.4:1), this effect was slightly mitigated for out of focus backgrounds, which reduced power from masking frequencies. Above 1.6:1, reading rate did not significantly increase but subjective quality ratings did. These results help inform the possible requirement for the development of additive display systems for consumer products.

The increased use of additive displays (for example in augmented reality head mounted displays, advertisement window displays, in car heads up displays) has led us to re-evaluate current standards for viable text contrast on video display terminals (VDTs). In an additive display text light is added to the background, so the image at the retina is the sum of light from the background and the display as a function of spatial location. Thus, the image at the retina fluctuates with the background. The current study examines issues related to presenting text over a potentially interfering background with the goals being to 1) develop an efficient method to assess reading rate on additive displays, and 2) determine the minimum viable contrast for text readability on additive displays.

BACKGROUND

Additive displays present a new challenge to display design because information is presented over real-world backgrounds, which are often textured and complex. Unlike the plain backgrounds typically used for information display on traditional VDTs, the complex background of the environment may interfere with our ability to retrieve information from the display. There is evidence that bandpass backgrounds made up of spatial frequencies that are important for letter identification (about 3 cycle/letter; Majaj, Pelli, Kurshan, & Palomares, 2002) result in the masking of letters (Solomon & Pelli, 1994) and words (Liang, 2002). To understand how a real-world scene could interfere with text presented on an additive display, we should consider the spatial structure of the scene. In general, the spatial structure of natural scenes follows a 1/f spectrum, meaning that contrast energy decreases with spatial frequency (Field & Brady, 1997). Text also follows this same spatial structure, and having environmental content with the same contrast energy across spatial channels contributes to this masking effect.

One of the most straightforward ways to increase text readability is to increase the contrast between the text and the background – either through increasing display brightness or reducing the background light transmission. Indeed, most standards for video display of text focus on the minimum contrast for text readability (HFES, 2007; ISO, 2011). Current ANSI standards (HFES, 2007) are based on qualitative judgments from standard VDTs, and do not encompass the specialized case of additive displays. ANSI suggests a minimum 3:1 contrast ratio for a typical observer for text larger than human visual acuity limits. Furthermore, this contrast ratio should be increased for differences in expected users. For example, to ensure an equivalent reading experience for an elderly population, a 7:1 minimum ratio allow those whose visual acuity is approaching 20/80 to read normal size text ("Understanding WCAG 2.0", 2016).

These existing standards are possibly insufficient for understanding additive text in a real-world environment because they ignore the potential masking that can occur. Furthermore, simply applying these standards to an additive display system may result in displays that are not power efficient. Thus, we need to re-evaluate these standards for cases in which information will be presented in under additive display conditions, using both metrics of text readability and text quality.

While the ideal contrast for quality text is quite high, there is ample evidence that information can be extracted from text at contrasts lower than current standards. For example, if we just consider the minimum contrast to support readability or word identification, the literature suggests that contrast can be as low as 1.15:1 on a plain background (Legge, Rubin, & Luebker, 1987). This value jumps to 1.5:1 for a textured background (Scharff & Ahumada, 2002).

However, the methods used to derive the 1.5:1 contrast ratio for text on a textured background are not sufficient to apply the ratio to additive displays. Scharff and Ahumada (2002) recorded the time it took for participants to identify a target word embedded in a paragraph of text presented on a background. The text varied in contrast and type (opaque, multiplicative transparency or additive transparency), although not all text types were presented at all contrasts. Additive transparent text was only presented at 1.3:1 and 1.45:1 text to background contrast ratio. The background was either plain or one of two textures. By combining all conditions (including the two additive text contrast conditions and the conditions using other text type and contrast), the authors were able to show that there was a monotonic increase in reading performance up to a contrast ratio of 2:1. They suggest that below 1.5 the readability of text on textured backgrounds diminishes rapidly. Because this minimum value was generated outside of the conditions that they tested for an additive display, it requires validation prior to being considered a minimum requirement for additive transparent text. It is also important that we extend this work beyond simple textured backgrounds to investigate the implications that natural statistics may have in masking the extraction of information for text as a function of overall contrast.

The minimum relative contrast of text will likely be dependent on the complex nature and use cases of additive displays. However, it will also be dependent on the interaction of device design decisions and content placement. For example, if content is displayed at a different focal depth than a display plane, focal blur may reduce the impact of the background masking as a result of the low pass filtering from increased blur. Scharff and Ahumada (2002) did not explore reducing contrast energy in background spatial frequency components as would occur through focal blur. A blur-induced reduction in contrast energy at the masking spatial frequencies can mitigate this background interference. Importantly, this suggests that there could be potential to exploit defocus differences between content presentation and background in an additive display. Second, there is evidence that interference from surrounding stimuli is reduced when the interfering stimulus is presented with a different binocular disparity than the target (i.e. appearing to be in a different depth plane; Lehmkuhle & Fox, 1980; Wardle, Cass, Brooks, & Alais, 2010).

Here, we present a set of four experiments that determine the minimum contrast required for a user to extract text information from an additive display. In Experiment 1 we identify the minimum viable contrast required to prevent a complex background from masking text. In Experiments 2 and 3 we determine the extent to which the masking effect is mitigated by background defocus and disparity. Then we extend these findings in Experiment 4 to determine how subjective quality of text is impacted by contrast.

GENERAL METHODS

We presented additive text on top of algorithmically generated images using a custom built haploscope with plate beam splitters (70/30). Two 32-inch (1920 x 1080) Cambridge Display ++ monitors with a refresh rate of 120 Hz were positioned 94 cm from the eye. After calibration, the maximum luminance for the display system was 444 cd/m^2. Stimuli presented on the monitors were calibrated to converge at eye level for an interpupillary distance (IPD) of 60 mm (higher IPDs were adjusted for by physically separating the screen). The experiment was conducted in a black curtained area with the lights off.

Reading performance task

To determine the reading performance of the text (Experiments 1-3), we assessed reading rate with a two-alternative forced choice rapid serial presentation (RSVP) task programmed in PsychoPy (Peirce, 2009).

Each trial started with the presentation of a red target word on a grey screen. The target word appeared until the participant advanced the trial by pressing the space bar. Then there was a 500ms mask of white x's on a plain background. This was followed by the set of words sequentially overlaid in random order on a background image. The target word was randomly determined to be present in the word set half of the time. After the presentation of the words, the screen remained blank until the participants responded with one of two keys to indicate if the target word was present or absent in the word set. Their response was recorded and the presentation speed of the next trial from that condition was adjusted using a basic adaptive staircase procedure. The staircases converged to 80% correct using a 3-up-1-down stepping method. There were two staircases for each contrast condition, with starting speeds well above and well below our target 80% point. Our dependent variable, reading rate, was the speed at which the participant got 80% correct for a given condition.

This is an effective method for determining the cases in which participants are able to reach their maximum reading rate. Although the absolute reading rate in this type of task may be higher than it would be in a natural reading task, the range of best performance would be the same (Legge & Bigelow, 2011).

Background stimuli

Images of natural scenes follow a $1/f^a$ amplitude spectrum (Field & Brady, 1997). We generated 11.5 x 11.5 degree background stimuli that follow this spectral pattern using an inverse fast Fourier transformation in Matlab (ifft). We set a to -2.2. We amplified the power at the horizontal and vertical orientations to conform to the documented bias for these orientations in natural and man-made scenes (Coppola, Purves, McCoy, & Purves, 1998). Even after this amplification, the image statistics still had a 1/f spectrum.

We scaled images to have a mean luminance in the central region around the words to be ~=222 cd/m^2. This luminance is within the range of typical indoor lighting (Panasonic, 2012).

The background for each trial was randomly selected from a set of our generated images, which all had the same image statistics (mean luminance and a).

Text stimuli

We added text over the background such that the text contrast (C) was the proportion of the text luminance (L_T) to the background luminance (L_B) minus one.

$$C = L_T/L_B - 1 \qquad (1)$$

Given that the maximum luminance of the display (and therefore the maximum luminance of the text) was 444 cd/m^2 and the background luminance was 222 cd/m^2, text contrast could range between 0 and 1. Contrast conditions were set for each experiment.

The luminance of each pixel of the text was determined by multiplying the pixel luminance of the background by the text contrast and adding it to the background luminance.

$$L_T = L_B + C * L_B \qquad (2)$$

This mimics the expected outcome of an additive display where the light hitting the eye is a sum of the light from the background and the light emitted from some display. Unlike multiplicative LCDs, which can present black text, the information presented on an additive display will always appear lighter than the background.

Words were selected by transcribing the text from the book *A Wasted Day* (Davis, 2006). To ensure that task difficulty was maintained throughout the task, we removed any words with fewer than three letters. The average word length was five letters. Text in the word set was 0.25 degree (15 arc min) high Arial font. This size is above the acuity limit (20/20 Snellen letters are 5 arc min), so participants could read the text.

EXPERIMENT 1: THE EFFECT OF TEXT CONTRAST ON READING RATE

In our first experiment, we determined the minimum required contrast for additive text presented over a textured scene.

Methods

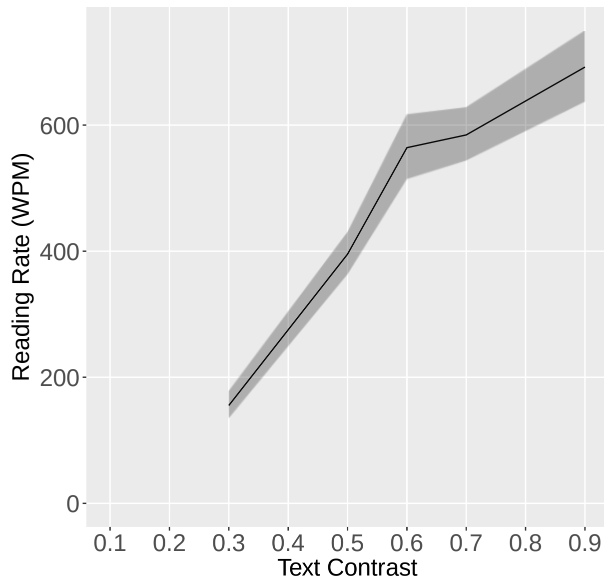
Twenty-five participants volunteered in exchange for a $50 Amazon gift card. Participants included employees as well as externally recruited volunteers. We verified that each participant had normal or corrected-to-normal binocular vision and stereo depth perception.

In this experiment, participants were presented with a target word, followed by the rapid serial presentation of a sequence of five words. All words in the set were presented at one of five contrast conditions (0.3, 0.5, 0.6, 0.7, 0.9). Participants were asked to respond if the target word was present or absent in the word set. Reading rate was the speed at which the participant got 80% correct for a given contrast condition.

Results

We found that reading rate increased with text contrast using a repeated measures ANOVA ($F(4,96) = 60.47$, $p < 0.01$). A post hoc test revealed that each contrast condition was significantly different from all other conditions ($p < 0.01$) except for comparisons between conditions 0.6, 0.7 and 0.9. This indicates that above a contrast of 0.6 there is no statistically reliable increase, and below 0.6 reading rate drops off rapidly (see Figure 1). Participants could still read the text, with difficulty, as low as 0.3 – 0.5 contrast. However, many participants could not get their accuracy up to 80% in these conditions, even with very long presentation times.

Figure 1. The relationship between reading rate and text contrast for additive text on a textured background.



EXPERIMENT 2: THE EFFECT OF BACKGROUND DEFOCUS ON READING RATE

For Experiment 2, we examined if simulated defocus blur of the retinal image would reduce the masking effect that we show in Experiment 1.

Methods

We set the overall text contrast to 0.4. This contrast level was selected because it is at the edge of readability based on the results of Experiment 1 and we wanted to avoid ceiling effects on reading rate.

We simulated defocus blur by creating a blur kernel, based on a disk. The disc size was estimated by computing the circle of confusion (CoC):

$$CoC = A*(abs(d_T - d_B)/ d_T)*(f/(d_B - f)) \qquad (3)$$

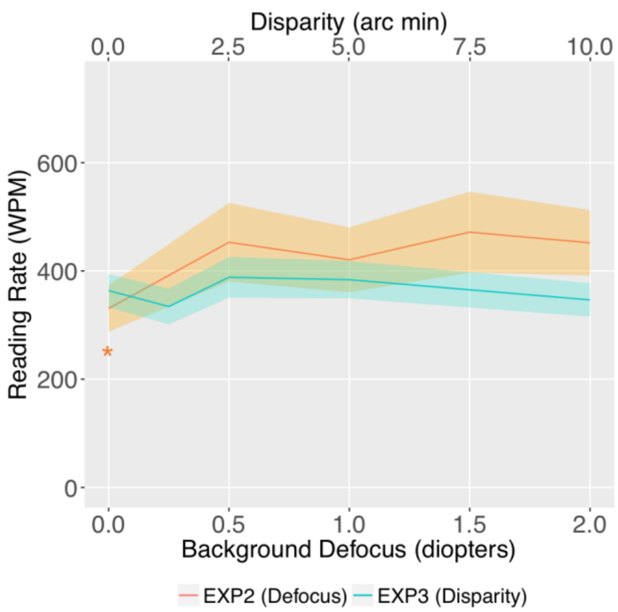
with the pupil (A) set to 5 mm, and the focal length of the eye (f) estimated to be 1.7 cm. Although we did not actually separate the background and text in space, the calculation uses distance to the text (d_T) and distance to the background (d_B). d_B was set to 94 cm and d_T was calculated based on the defocus amount in diopters (D). We convolved each background with the blur kernel for defocus amounts of 0.5, 1.0, 1.5 and 2.0 D.

Eleven participants volunteered in exchange for a gift card.

Results

We found that reading rate was lowest with zero defocus (i.e., without low-pass filtering the background image). By adding defocus (i.e., by low-pass filtering the background image), the reading rate improved by about 120 words per minute (see Figure 2; $F(4,40) = 3.87$, $p < .01$). A post hoc Tukey test revealed that reading rate was lower for zero defocus compared to 0.5, 1.5 or 2.0 diopters of defocus. All other comparisons were not significantly different. We have no theoretical reason to predict that there would be a difference between zero defocus and all other conditions, but not 1.0 diopters defocus. The fact that there was no significant difference between zero defocus and 1.0 diopters of defocus is therefore likely due to random variability.

Figure 2. The relationships between reading rate and disparity and background blur for low contrast additive text on a textured background.



Although this is a significant improvement, it does not completely recover the reading rate to the level that it is at higher contrasts. This suggests that small, low contrast additive text can be read when presented over a complex background, as long as that background is out of focus by at least 0.5 D. However, preliminary data from a pilot version of this experiment, using 0.5 contrast, suggests that this effect disappears for contrasts higher than 0.4 (i.e. defocus does not improve reading rate when the contrast ratio is higher than 1.4:1). Therefore, if we end up achieving the suggested ideal contrast ratio (3:1), defocus would likely not improve reading rate.

EXPERIMENT 3: THE EFFECT OF BINOCULAR DISPARITY ON READING RATE

We also wanted to test the potential mitigating effect of binocular disparity, the difference in the image between the eyes that allows us to perceive objects in depth. There is evidence that interference from surrounding stimuli is reduced when the interfering stimulus is presented with a different binocular disparity than the target (i.e. appearing to be in a different depth plane; Lehmkuhle & Fox, 1980).

Methods

We varied the difference in binocular disparity between the background and the text so that the text appeared to be popping out in front of the background. The apparent depth between the text and background was between 0 and 16 JNDs, which was calculated using values derived by De Sivla et al. (2011). We presented 21 participants with low contrast (0.4) additive text over textured backgrounds.

Results

There were no significant effects of disparity on reading rate ($F(4,80) = 2.04$, $p = 0.1$; see Figure 2). This seems to indicate that disparity alone does not mitigate the text masking effects of a natural scene. However, it is possible that in a more ecologically valid situation there could be an interaction between depth cues, such optical blur and disparity that would lead to an improvement in reading rate. This should be the focus of future work.

EXPERIMENT 4: THE EFFECT OF TEXT CONTRAST ON RATINGS OF SUBJECTIVE QUALITY

The previous experiments helped to determine a minimum contrast for text legibility in an additive display. Our next step was to understand the quality of the reading experience and the potential benefit of increasing contrast beyond the minimum value. We ran Experiment 4 to see how contrast affects subjective quality ratings of additive text on textured backgrounds.

Methods

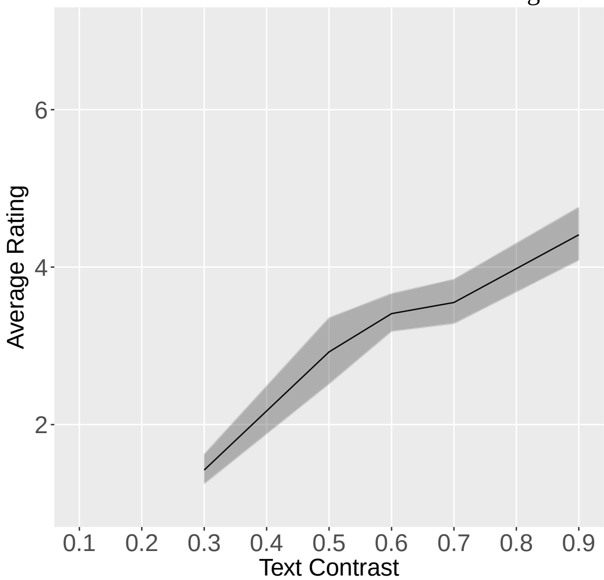
We asked 13 participants to read a paragraph of text and rate it on overall visual quality, ease and difficulty using a 1-7 Likert scale. We decided to use a full paragraph with unlimited reading time rather than a single rapidly presented word to provide participants with more information to make their quality judgment.

Each paragraph was a list of 200 words from *A Wasted Day* (Davis, R.H., 2006) in the order that they appeared in the story. A new paragraph was presented on each trial.

Results

The ratings of the three quality questions were all correlated so we combined them into an average rating. We ran a general linear model with subject as a random effect and found that the average rating increases linearly with contrast (b=4.79, SE = 0.48, p<0.001; see Figure 3). Unlike reading rate, subjective quality ratings do not plateau. This is consistent with our expectations if the minimum contrast for quality text were greater than 2:1. We will be expanding our future experiments in this direction to determine if there is a point at which quality ratings plateau.

Figure 3. The relationship between subjective quality of text and text contrast for additive text on a textured background.



DISCUSSION

We found that the luminance and depth position of text is important when text is presented on an interfering background. The minimum contrast required for reading text on an additive display with an interfering background is 1.6:1. This is only slightly higher than the value (1.5:1) suggested by Scharff and Ahumada (2002). For contrast levels below 1.6:1, we observed that reading rates decline rapidly. This effect can be slightly mitigated by background defocus blur, but only when the text contrast is near detection threshold. For contrast levels below 1.4:1, we observed that participants were unable to reliably extract information from the display. For contrast levels above 1.6:1, reading rate did not significantly increase but quality ratings did. In future studies, it will be useful to find the point at which the quality ratings plateau to determine the ideal contrast for high quality text on an additive display. Unfortunately, with our current system we were not able to get text contrast higher than 2:1 so such a manipulation would require a different system. Another option would be to lower the background image luminance so there would be a wider range of text contrasts possible. However, reducing the background luminance would make the contrast energy and luminance of the scene much lower than what would be expected in a real-world environment.

Our findings have implications for the development of augmented reality displays. When creating content for minimal background interference for readability/legibility, developers should carefully consider the spatial structure and luminance of the locations that the devices will be used. We have demonstrated that these environmental factors will have an impact on determining where and how brightly to present text on an additive display.

REFERENCES

Coppola, D. M., Purves, H. R., McCoy, A. N., & Purves, D. (1998). The distribution of oriented contours in the real world. *Proceedings of the National Academy of Sciences*, *95*(7), 4002-4006.

Davis, R.H. (released 2006). A Wasted Day. Retrieved from https://www.gutenberg.org/files/1820/1820-h/1820-h.htm

De Silva, V., Fernando, A., Worrall, S., Arachchi, H. K., & Kondoz, A. (2011). Sensitivity analysis of the human visual system for depth cues in stereoscopic 3-D displays. *IEEE Transactions on Multimedia*, *13*(3), 498-506.

Field, D. J., & Brady, N. (1997). Visual sensitivity, blur and the sources of variability in the amplitude spectra of natural scenes. *Vision research*, *37*(23), 3367-3383.

HFES. (2007). American National Standard for Human Factors Engineering of Computer Workstations (ANSI/HFES Standard No. 100-2007). Human Factors & Ergonomics Society, Santa Monica, California.

ISO. (2011). Ergonomics of human-system interaction – Part 303: Requirements for electronic visual displays. ISO, Switzerland.

Legge, G. E., & Bigelow, C. A. (2011). Does print size matter for reading? A review of findings from vision science and typography. *Journal of vision*, *11*(5), 8-8.

Legge, G. E., Rubin, G. S., & Luebker, A. (1987). Psychophysics of reading—V. The role of contrast in normal vision. *Vision research*, *27*(7), 1165-1177.

Lehmkuhle, S., & Fox, R. (1980). Effect of depth separation on metacontrast masking. *Journal of Experimental Psychology: Human Perception and Performance*, *6*(4), 605.

Liang, Y. X. (2002). Readers beware! Effects of visual noise on the channel for reading. Intel Science Talent Search. Semifinalist. http://psych.nyu.edu/pelli/#intel

Majaj, N. J., Pelli, D. G., Kurshan, P., & Palomares, M. (2002). The role of spatial frequency channels in letter identification. Vision research, 42(9), 1165-1184.

Panasonic. (2012). The light required for each room. Retrieved from https://panasonic.net/ecosolutions/lighting/technology/knowledge/03/

Peirce, J.W. (2009). Generating stimuli for neuroscience using PsychoPy. Front. Neuroinform. 2:10. doi:10.3389/neuro.11.010.2008

Scharff, L. F., & Ahumada, A. J. (2002). Predicting the readability of transparent text. Journal of vision, 2(9), 7-7.

Solomon, J. A., & Pelli, D. G. (1994). The visual filter mediating letter identification. Nature, 369(6479), 395-397.

Understanding WCAG 2.0: A guide to understanding and implementing Web Content Accessibility Guidelines 2.0. (2016, October 7). Retrieved from https://www.w3.org/TR/UNDERSTANDING-WCAG20/visual-audio-contrast-contrast.html

Wardle, S. G., Cass, J., Brooks, K. R., & Alais, D. (2010). Breaking camouflage: Binocular disparity reduces contrast masking in natural images. *Journal of vision*, *10*(14), 38-38.