

Legibility: Back to the Basics

2011-01-0597

Published
04/12/2011Shannon O'Day
Ford Motor CompanyLouis Tijerina
Ford Motor Company

Copyright © 2011 SAE International

doi:[10.4271/2011-01-0597](https://doi.org/10.4271/2011-01-0597)

ABSTRACT

The Objective for this study was to revisit some of the known factors that affect legibility including font characteristics, as well as, contrast polarity, luminance contrast, and color contrast under high ambient conditions as specified in SAE J1757. The study focused on older drivers due to their increased visual needs and limitations.

The study was conducted in 2 phases: 1) a study of font characteristics; character height, character width, and stroke width using a central composite design. Subjects read a group of letters and numerals displayed on a laptop display using occlusion goggles. The reading time (Total Shutter Open Time or TSOT), reading errors, and a subjective Readability Rating (using a 4 point scale “Very Easy”, “Easy”, “Difficult”, “Very Difficult”) were recorded. Licensed drivers in three age groups, 25 to 44 yrs, 45 to 59 yrs, and 61 to 91 yrs participated. The response surfaces were generated and compared to the character sizes recommended in ISO 15008. Results showed that a wide range of characters heights can be legible as long as character width and stroke width were carefully chosen.

The second study investigated legibility under SAE J1757 defined daylight conditions, Overcast and Direct sunlight. The study was run in the daylight simulation lab at Ford Motor Co. capable of simulating near full daylight illuminance levels, both diffuse and direct sunlight. 32 combinations of character and background colors were tested by varying color contrast, luminance contrast, and contrast polarity. 15 younger (< 60yrs) and 15 older (≥60yrs) drivers participated. Participants read a group of letters and numerals

from each color and luminance combination under each ambient condition while reading time and reading errors were recorded along with a subjective readability rating (Unacceptable, Minimally acceptable, Preferred). The subjective data indicated that the contrast ratios participants considered minimally acceptable were close to those listed as minimally acceptable in ISO15008. The ISO15008 minimum acceptable contrast ratios resulted in reading accuracies of 98% (Overcast) and 94% (Direct Sunlight). Color difference was found to play a minor role in legibility under daylight ambient conditions.

INTRODUCTION

Legible is defined as “capable of being read or deciphered” [8]. The first known use of the word was the 14th century [8]. In modern times Legibility has been studied in many contexts starting in the 1930's, mainly with respect to character sizes and proportions for road signage and military applications [15]. With the development of flat panel displays and the unique issues resulting from raster scan and pixilated screens, further studies were conducted [9, 13, 14, 15] and many standards and guidelines were written [5]. In 1988 Ford contracted R & R Research to develop a computer model of legibility based on the empirical research of Rockwell [10] and H.R. Blackwell [2]. This model determined the contrast ratios for a range of ages necessary for legibility based on character size (height, width, stroke width) and luminance. Although a very useful model, it was limited to less than full daytime ambient conditions. All of these studies have resulted in ranges of character sizes, character proportions, and contrast ratios deemed legible. However, feedback from

Ford customers indicated that these ranges may need closer definition and constraint.

The relative vast number of studies over so many years is a testament to the importance of legibility. This is certainly true in the automobile cockpit where legible controls and displays help to minimize the amount of time the driver's eyes are diverted from the road scene. Fonts used to label switches and buttons often are limited size due to the size of the button, the number of controls, and the limited amount of instrument panel real estate available. The same goes for text within displays. Displays are asked to convey information from so many sources and functions that font size often gets squeezed to fit all the desired text. Increasing the size of the display may not alleviate the situation since a larger display invites an increase in content to justify the display taking up so much instrument panel real estate. Some features like media RDS and navigation menus require a lot of text to be displayed making a large font problematic. There is no escape, as more technology goes into the cockpit, the font size used on switches, buttons, controls and displays will increasingly feel the pressure to shrink in size and proportion. How much can the font shrink before legibility is compromised? What are the trade-offs between the various font characteristics? ISO 15008 [6] lists acceptable character sizes. Do drivers, especially older drivers, agree with those character sizes?

Having the right character height, width and stroke width is, of course, only the start. When a vehicle is in the showroom, the display may be clearly legible with no annoying glare or reflections. If only this were the most extreme ambient condition a vehicle display was subjected to! Unfortunately, when vehicles are driven in daylight, the display is exposed to sunlight that can overpower the brightness of the LCD panel and effectively washout the display. If sunlight can reflect in a specular manner off the display's first-surface and into the driver's eyes, it is doubtful that any display would be legible. However, even when the display is free of specular sunlight reflections it can still be a challenge to achieve legibility. Many display characteristics have been studied by the authors that affect legibility under high ambient conditions, such as, the type of polarizer, level of AR (antireflective) and AG (antiglare) coatings, touchpanel material, display luminance, etc. This may be the subject of a future paper. However, the goal of these display characteristics is to enable the display to achieve a certain contrast under daylight conditions. What level of contrast is acceptable? What contrast is preferred? What effect does color contrast have on legibility? How does legibility change when ambient conditions change?

This paper describes two studies that went back to the basics of legibility to better understand older drivers' needs under challenging daytime ambient conditions. The first study investigated the relationship of Font characteristics to

legibility. The second study investigated the relationship of luminance contrast, color contrast, and contrast polarity to legibility under daytime ambient conditions as defined in SAE J1757 [11].

FONT CHARACTERISTICS

The first study was a within subjects study of three font characteristics and their effect of legibility. The font characteristics were 1) character height, 2) character width, and 3) stroke width. For each font characteristic 5 levels were studied (see Table 1). With 3 factors and 5 levels, the total number of fonts tested was 15. Each participant read from all 15 fonts with 2 replications. The reading task involved reading aloud an unbroken string of 10 characters, all in the same font, that were displayed on a laptop computer screen directly in front of the participant (see figure 1). Participants wore occlusion goggles to simulate typical eye glance behavior. Shutter open time was 1.5 sec. and the shutter closed time was 2 sec. The characteristics of each tested font are shown in Table 2. Figure 2 shows an example of each of the tested fonts. Legibility was measured by the Total Shutter Open Time (TSOT) to read each string of 10 characters, the number of reading errors, and a subjective Readability Rating that used the following scale: Very Easy, Easy, Difficult, and Very Difficult.

The study used a Central Composite design [3] that allowed for efficient reconstruction of the complete response surface for each Legibility measure. The Central Composite design α was 1.216, so that all 1st and 2nd order regression β weights for the coded model were orthogonal to one another. This simplified the calculation of the regression and ANOVA results.

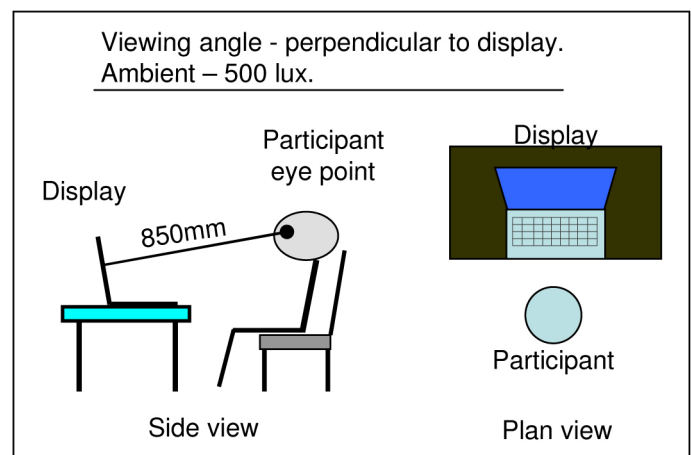


Figure 1. Font Study layout. Display was directly in front of the participant under general office lighting ~500 lux.

Table 1. The Font study factors and factor level values.

	Factor Level Values				
Character Height (mm)	2.6	3.0	5.0	7.0	7.4
Character Width (% of height)	45%	52%	85%	118%	125%
Stroke Width (% of height)	7%	9%	20%	28%	30%

Table 2. List of the tested font characteristics.

Font	Character Height (mm)	Character Width (% of height)	Stroke Width (% of height)
1	7.0	118%	28%
2	7.0	52%	28%
3	7.0	118%	9%
4	7.0	52%	9%
5	3.0	118%	28%
6	3.0	52%	28%
7	3.0	118%	9%
8	3.0	52%	9%
9	7.4	85%	20%
10	2.6	85%	20%
11	5.0	125%	20%
12	5.0	45%	20%
13	5.0	85%	30%
14	5.0	85%	7%
15	5.0	85%	20%



Figure 2. Examples of the fonts used in the study. Actual text strings used in the study were randomized.

Response Surface Generation

There were a total of 37 participants who ranged in age from 25 years to 91 years collected into 3 age groups (see Table 3). Reading accuracy, or proportion correct, for each trial was defined as the number of characters read correctly divided by 10 (total number of character presented per trial). Figure 3 shows the percent of characters misread (1 - Reading Accuracy), across all fonts, as a function of age group. The two younger age groups had virtually the same error rate whereas the 60 yrs and older age group had an error rate approximately 4 times greater. This trend can also be seen in Figure 4, where the percent of characters misread is shown for each of the 15 fonts and each age group. Although the younger age groups had far fewer errors overall, they did have more errors for some fonts, such as fonts 2, 5, 6, 7, 8, 10, and 12. These were the same fonts that the older age group had most difficulty reading and their reading errors were far higher than the younger groups. For the other fonts (fonts 1, 3, 4, 9, 11, 13, 14, 15), the reading errors were about the same across all age groups. So, the error rates for the older age group mimic the error rates of the younger age groups, but they magnified error rates of the more difficult fonts. The older age group therefore provided a larger range of responses than either of the younger age groups. Based on these results it was decided to focus on the older age group for detailed analysis.

Table 3. The percent of total (37) participants in each age group.

Age Group	Females	Males	Total
1. < 45 yrs	8%	19%	27%
2. 45 to 59	8%	22%	30%
3. ≥ 60 yrs	32%	11%	43%
Total	49%	51%	100%

The 60 yrs and older age group had 15 participants, 12 females and 3 males. Each participant completed 2 blocks (repetitions) of the 15 fonts, for a total of $15 \times 2 \times 15 = 450$ data points. There was no missing Reading Accuracy data. If a participant was not able to read a trial, then zero characters read correctly was recorded. Due to high positive skew in the data (toward high accuracy), an arcsine transformation (equation 1) was applied to the Proportion Correct data.

$$p' = \sin^{-1}(\sqrt{\text{proportion_correct}}) \quad (1)$$

The residuals of the transformed data were improved though still non-normal in the tails. However, the F-test is robust to non-normal data. The Central Composite design of the study allowed for construction of the complete 2nd order polynomial response surface. The Central Composite regression for proportion correct (equation 2) was significant with R² of .485. All the terms were significant except the squared stroke width term. The character height, width, height*width, height², and stroke width terms were, in order, the most important contributors to the model.

$$\begin{aligned} \text{Sin}^{-1}(\text{Proportion Correct}^{1/2}) = & b_0 + b_1 (\text{height}_c) + b_2 (\text{width}_c) \\ & + b_3 (\text{stroke}_c) + b_4 (\text{height}_c * \text{width}_c) + b_5 (\text{height}_c * \text{stroke}_c) \\ & + b_6 (\text{width}_c * \text{stroke}_c) + b_7 (\text{height}_c^2) + b_8 (\text{width}_c^2) \\ & + b_9 (\text{stroke}_c^2) + \text{error} \end{aligned}$$

(2)

where $b_0=1.572$, $b_1=.1780$, $b_2=.1403$, $b_3=-.07289$, $b_4=-.1323$, $b_5=.0683$, $b_6=.06165$, $b_7=-.1245$, $b_8=-.102$, $b_9=-.01317$
height_c, width_c, stroke_c coded factor levels
coded factors based on $\alpha = 1.216$

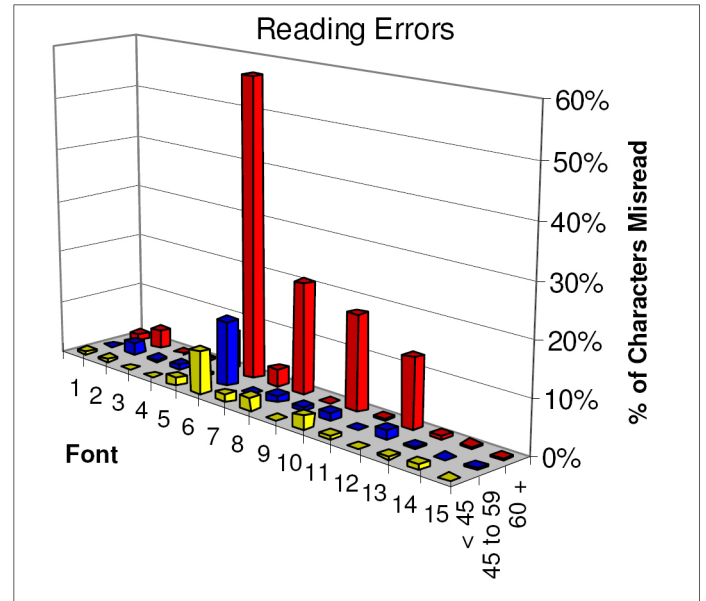


Figure 4. Reading errors for each font and each age group. The older age group (60+) magnify the errors of the younger age groups.

Reading time data was acquired via the occlusion goggle software. The Total Shutter Open Time (TSOT) has been validated in road and track testing to be a highly correlated surrogate measure for Total Glance Time away from the road scene to complete a visual-manual task while driving [1]. There was missing data due to the inadvertent loss of one repetition for one participant and the inability to complete certain trials by other participants (“unreadable” trials were not included in the TSOT model). This left only 416 data points available for analysis. Preliminary model fitting revealed that TSOT produced highly non-normal residuals with strong positive skew. In such cases an inverse transformation is recommended [4], and this indeed did produce normally distributed residuals. The resulting Central Composite regression is shown in equation 3.

$$\begin{aligned} \text{TSOT}^{-1} = & a_0 + a_1 (\text{height}_c) + a_2 (\text{width}_c) + a_3 (\text{stroke}_c) \\ & + a_4 (\text{height}_c * \text{width}_c) + a_5 (\text{height}_c * \text{stroke}_c) \\ & + a_6 (\text{width}_c * \text{stroke}_c) + a_7 (\text{height}_c^2) + a_8 (\text{width}_c^2) \\ & + a_9 (\text{stroke}_c^2) + \text{error} \end{aligned}$$

(3)

where $a_0=.2507$, $a_1=.03273$, $a_2=.01998$, $a_3=-.01914$,
 $a_4=-.008510$, $a_5=.009941$,
 $a_6=.006008$, $a_7=-.01191$, $a_8=-.005945$, $a_9=-.008187$
height_c, width_c, stroke_c coded factor levels
coded factors based on $\alpha = 1.216$

The ANOVA (using Type III sum-of-squares) showed that all variables were significant except the squared character width term. Since there was missing data it was not advisable to calculate the proportion of variance accounted for by the

model. However, comparison of the magnitude of the sum-of-squares indicated that the character height, character width and stroke width were the most important contributors to the model by an order of magnitude compared to the squared terms and the two-way interactions.

The Subjective Readability Rating employed a 4 point scale: “Very Easy”, “Easy”, “Difficult”, “Very Difficult.” [Table 4](#) shows the proportion of the 60+ age group who rated a font “Easy” or “Very Easy” for each of the 15 tested fonts. There was no missing data. The data again showed high positive skew, so the same arcsine transformation ([equation 1](#)) was used. The residuals of the transformed data were improved though still non-normal. Nonetheless, the regression ([equation 4](#)) was significant with an adjusted R² of .875. All terms were significant except for the squared character width term.

$$\begin{aligned} \text{Sin}^{-1}(\text{Proportion rating "Easy" or "Very Easy"}^{1/2}) = \\ c_0 + c_1 (\text{height}) + c_2 (\text{width}) + c_3 (\text{stroke}) \\ + c_4 (\text{height}^2) + c_5 (\text{width}^2) \end{aligned} \quad (4)$$

where $c_0=-2.256$, $c_1=.6960$, $c_2=3.419$, $c_3=-2.864$,
 $c_4=-.04976$, $c_5=-1.508$

Table 4. The proportion of 60 + age group participants rating a font “Easy” or “Very Easy”

Font number	Proportion of Participants rating "Easy" or "Very Easy"
1	93%
2	33%
3	100%
4	100%
5	33%
6	0%
7	73%
8	27%
9	100%
10	13%
11	100%
12	20%
13	60%
14	100%
15	93%

Response Surfaces Analysis

A simple equation transforms TSOT to reading rate ([equation 5](#)). In this equation, 10 is the number of characters presented per trial and 1.5 is the shutter open time per glance. However,

not all of the characters read at this rate were read correctly; some of the fonts had large error rates. The product of Reading Rate ([equation 5](#)) and Reading Accuracy ([equation 2](#)) accounts for both speed and accuracy, called the Accurate Reading Rate ([equation 6](#)).

$$\text{Reading rate (characters/ glance)} = (10 / \text{TSOT}) * 1.5 \quad (5)$$

$$\begin{aligned} \text{Accurate Reading Rate (Characters read correctly / glance)} \\ = \text{Reading rate} * \text{Reading Accuracy} \end{aligned} \quad (6)$$

[Figures 5, 6 and 7](#) show the regression predicted accurate reading rate (in characters read correctly per glance) as a function of character height, character width and stroke width. Accurate reading rate can vary widely for any given character height, depending on the character width and stroke width. In general, larger character heights, wider character widths, and narrower stroke widths each show better performance across the range of font characteristics. What level of accurate reading rate should be considered acceptable was not addressed in the study. However, the range of accurate reading rates for the ISO15008 [\[6\]](#) recommended character heights is listed in [Table 5](#), middle column. For each character height the minimum Accurate Reading Rate occurred at a character width of 45% (of height) and a stroke width of 30% (of height). The maximum occurred at a character width of 125% (of height) and a stroke width of 7% (of height).

Table 5. The range of regression predicted Accurate Reading Rate (characters read correctly/ glance) and predicted percent rating “Easy” or “Very Easy” for the ISO15008 recommended character heights.

Range of Predicted Legibility Responses		
Character Height (arc min)	Accurate Reading Rate	% Rating "Easy" or "Very Easy"
12	0.4 to 3.7	0% to 80%
16	1.1 to 3.9	1% to 99%
20	2.0 to 4.0	12% to 100%

[Figures 8, 9, and 10](#) show the regression predicted proportion of participants giving a font a subjective Readability Rating of “Easy” or “Very Easy” as function of character height, character width and stroke width. These plots show the same trend as the Accurate Reading Rate. In general, the subjective Readability Rating is higher for larger character heights and/ or wider character widths and/ or narrower stroke widths. The effect of stroke width is greatest when the character height is small and the character width is narrow. The effect of character width is greatest when the character height is small