The Effect of Font Characteristics on Large Format Display Legibility

By Philip M. Garvey, Wei-Yin Eie, and M. Jennifer Klena

The Thomas D. Larson Pennsylvania Transportation Institute
at The Pennsylvania State University, University Park, PA, USA

Abstract

Objective: To assess the legibility of a large set of existing large format display fonts.
Background: The enormous selection of fonts allows for creative design; however, while there has been a lot of research on print and computer font legibility, only a limited number of large format display font studies have been conducted.
Method: Sixty-four subjects from 19-87 years of age viewed 64 displays using 33 fonts shown on a computer monitor. Viewing began at a very small size, which grew larger to simulate a driver or pedestrian approaching a sign. Subjects attempted to read the displays at the smallest possible size. Threshold legibility was determined for each font.
Results and Conclusions: Font selection can make a very big difference in the distance at which a display can be read; however, there are many fonts that have equivalent legibility. Case can sometimes, but not always, have a large impact on display legibility, with uppercase often performing significantly better than lowercase. The choice of serif versus sans-serif alone does not have an important effect on display legibility. Age impaired sign reading ability, but not until the participants were over sixty. Finally, fonts that share a family name (e.g., Times Bold versus Times New Roman) can have dramatically different legibility distances.
Application: The results of this research can immediately and directly aid letter manufacturers, display designers, and display owners, as they now know how far away a large number of fonts can be read, and the impact of choosing one font style over another.
Keywords: Vision, driver, legibility threshold, font style, display, letter height

Background and Objectives

Many elements, such as internal contrast, letter height, and letter width, contribute to the readability of large visual displays (e.g., highway and on-premise signs, billboards, banners, posters, etc.); however, one of the main factors is letter style or font (Garvey and Kuhn, 2011). While there have been many studies on print and computer font legibility (e.g., Yager et al., 1998; Legge and Bigelow, 2011), most evaluations of font legibility for large displays has been conducted in the highway, airport, and railway environments and, therefore, have been restricted to simple and unembellished fonts (Garvey et al., 1995). The font choice for large displays is limited only by the imagination of designers. While the enormous available selection of fonts (and limitless potential for future fonts) allows for creative design, it also creates difficulties for letter manufacturers, sign shops, designers, and display owners, as they have no way of assessing these fonts' relative legibility distances. This is because only a limited number of studies have been conducted in this field, assessing only only a small number of fonts each (e.g., Kuhn et al., 1998; Garvey et al., 2001, Zineddin et al., 2003; Garvey, 2007). The present study is the first to address the visibility of a substantial set of existing large display fonts.

Laboratory Experiment to Evaluate Large Visual Display Font Legibility

Overview

The study was conducted in a laboratory setting where many fonts could be evaluated in a short period of time using high-resolution, computer-generated graphics.

Method

Fonts

A set of 33 fonts was selected for evaluation (Table 1 lists the fonts with their exact names; abbreviated versions of these names are used throughout the paper). They represented the most popular fonts used in the commercial signing industry and a selection of additional fonts that are asked for by designers, but that have questionable legibility according to sign industry representatives. Thirty-one of the fonts were tested in both all uppercase and lowercase (initial capital letter followed by lowercase letters). The lowercase of two of these 31 fonts (i.e., Copperplate Gothic and Trajan Bold) consisted of a larger capital letter followed by smaller uppercase letters. Two of the 33 fonts are only available in all uppercase (i.e., Country Gothic and Ribbon). This resulted in a total of 64 unique conditions being tested. The fonts were displayed as scale-sized, one-word displays on a high-resolution computer monitor (for example, Figure 1). Each of the fonts was tested using all of the words in Table 2. The subjects viewed the displays under a simulated daytime lighting environment.

Table 1. Fonts evaluated, in alphabetical order.

Adobe Garamond Pro	Garamond Bold *(Monotype version, bundled with windows)*	Mistral
Architectural GT	Georgia	Myriad Pro
Arial Bold	Gill Sans MT	Old English Text MT
Arial	Gotham Medium	Optima Bold
Avant Garde Medium BT	Goudy Old Style Bold BT	Optima Regular
Avenir LT Std 65 Medium	Helvetica Bold	Palatino Linotype
Brush Script MT Italic	Helvetica	Papyrus
Copperplate Gothic Bold	Helvetica Neue LT Std 45 Light	Ribbon GT
Country Gothic GT	Helvetica Neue LT Std 67 Medium Condensed	Times Bold
Frutiger LT Std 55 Roman	Kabel Ultra BT	Times New Roman
Futura Bk BT Book	Minion Pro	Trajan Pro Bold

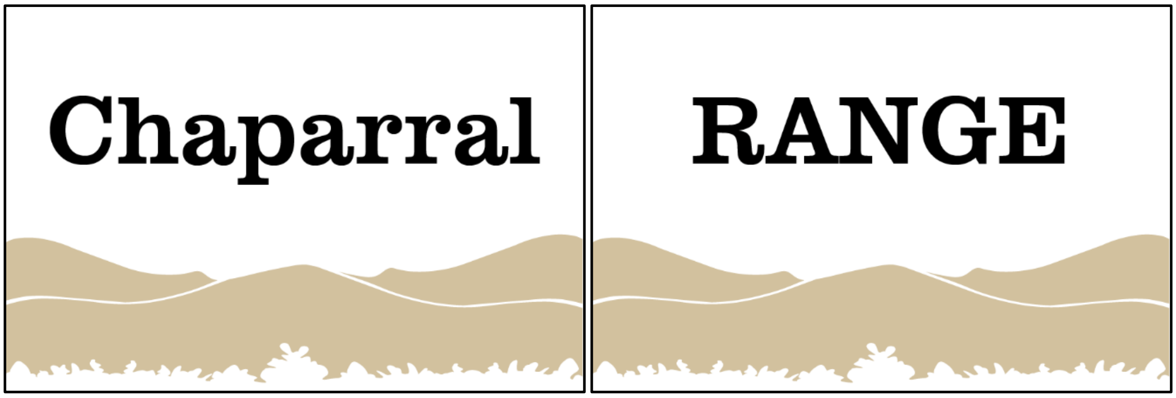

Figure 1. Example displays in the Clarendon font using lowercase and all uppercase.

Table 2. Words used.

Sunday	Gardens	Wadsworth	Perennials
Radiant	Trailhead	Riverside	Rutherford
Appleton	Harding's	Cardinal	Crawfordsville
Benneton	Pershing's	Frederick	Agency
Hangar	Cafeteria	Stables	Forestry
Traffic	Gelateria	Fountain	Smith
Solarized	Concourse	Marketplace	Thomas
Pharaoh	Fairway	Maplewood	Siracusa
Magnolia	Cloverton	Arboretum	Aurum
Flowers	Wilmington	Moonbrook	Goldberg
Appalachian	Campus	Planetarium	Savannah
Groundwater	Acorns	Byzantine	Freeway
Canals	Galleria	Emporium	Gloucester
Waterways	Trenton	Pavilion	Metro
Junction	Station	Davenport	Turnpike
Distribute	Terminal	Umbria	Trails

Subjects

A total of 64 subjects participated in the research project (27 males and 37 females). The subjects ranged from 19 to 87 years of age. All subjects were tested for binocular static visual acuity using a GOOD-LITE Co. light box and Sloan Letters at 10 feet (for results see Table 3).

Table 3. Age groups and visual acuity.

Subject Age Group	Number of Subjects			Mean Acuity
	Total	M	F	
Younger (19-34)	20	14	6	20/17.82
Middle Age (35-59)	23	6	17	20/17.41
Older (60-87)	21	7	14	20/20.84
Total	64	27	37	20/18.66

Test Site and Apparatus

The study was conducted at the Thomas D. Larson Pennsylvania Transportation Institute on The Pennsylvania State University's University Park campus (Figure 2). To display the fonts and record the subjects' performance, the apparatus consisted of a Sony 48-Inch 1080p 60Hz Smart LED TV and associated Dell OptiPlex 7020 Mini Tower desktop computer. A program was written using MATLAB to display the stimuli and collect the legibility data.

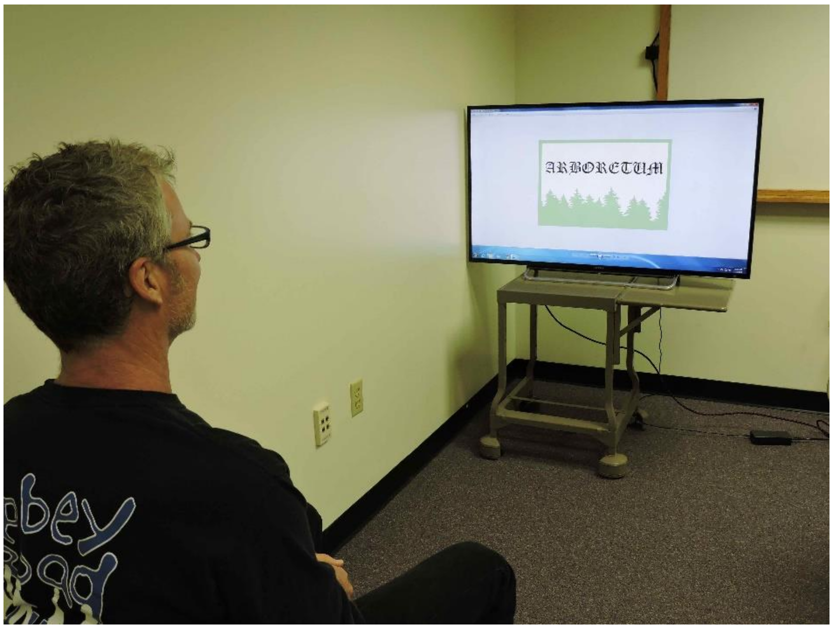

Figure 2. Test Site: "Arboretum" in Old English uppercase.

Variables

The main independent variables were Age Group (Young [19-34], Middle [35-59], and Old [60-87]) and Font (the 64 levels described above). In addition, the following variables were evaluated: Case (uppercase vs. lowercase), Style (serif vs. sans serif), Font Weight (e.g., bold vs. condensed), Word Superiority (rank ordering of the 64 words), and Art/Word combination (some displays were shown with graphics that matched the test word).

The dependent variable was threshold legibility size (the smallest size at which the participant could read the word). To standardize the readability of the fonts to larger displays used in the built environment, threshold legibility size (in millimeters) was converted to Legibility Index (LI). LI is the standard used in the transportation field to express the legibility of a display as a function of the number of feet of legibility distance that can be expected for each inch of letter height. For example, if a font had an LI of 35, a display with 10-inch letters would be readable 350 feet away (35 x 10 = 350), and 500 feet away if the LI was 50 (50 x 10 = 500).

Procedure

The 64 subjects each viewed all 64 fonts, for a total of 4,096 individual observations. The fonts were shown randomly beginning at a very small capital letter height (5 mm) and growing larger (up to a maximum of 85 mm) to simulate a driver or pedestrian approaching a display. The subjects were seated 21.34 feet from the screen that displayed the fonts. The subjects attempted to read the displays at the smallest possible size; however, they were instructed not to read it aloud until they were sure what it said. The threshold LI was determined for each font for each subject. Statistical analyses were conducted to evaluate the effect of the independent variables on LI.

Analyses and Results

Age Group Effect

The subjects were divided into three age groups. The youngest age group (19-34) had a mean LI of 35.99, the middle age group (35-59) had a mean LI of 35.61, and the oldest age group (60-87) had a mean LI of 31.31. To determine whether the differences among age groups were statistically significant, the effect of age group on display legibility was evaluated using a one-way analysis of variance (ANOVA). The p-value was set at 0.05.

The ANOVA showed that there were indeed statistically significant differences among the age group mean LIs ($F(2, 61) = 4.76$, $p = 0.01$). To determine which of the groups differed significantly from the others, a Scheffé post-hoc test was conducted. The Scheffé was used because the group sizes were different. There was no significant difference between the young and middle age groups ($p = 0.98$); however, both the young group and the middle age group had significantly higher LIs than the old group ($p = 0.03$ and $p = 0.04$, respectively). The mean LI of the young group was 4.68 points higher than that of the old group, and the mean LI of the middle group was 4.30 points higher than the mean LI of the old group.

While it is possible, of course, for small differences in mean LI to result in statistical significance with large sample sizes like those used in this study, a difference in mean LI of about 5.0 has been operationally defined to be a minimum important or practical difference (see Mace et al., 1994). A 5-ft/in of letter height difference in LI would, for example, result in 50 feet more legibility distance for a display with 10-inch letter heights. A practical implication is that at 25 miles per hour, this would give a driver an additional 1.36 seconds to read the sign. At 4.3 and 4.7, the differences among the age groups, while being statistically significant, only just approach practical significance.

Font Effect

The primary objective of the research, the effect of font on large visual display legibility, was evaluated. As discussed above, there were 64 conditions tested in this research study. In this analysis each will be considered a unique "font," even though, as discussed above, many are simply upper and lowercase versions of the same fonts. A separate Case Effect analysis is included below to tease out any differences due to the case variable.

Substantial mean LI differences were found among the 64 fonts, ranging from Gill Sans uppercase with a high of almost 50 ft/in of letter height, to Mistral lowercase, with a low LI of 15.5 (Table 4). The statistical analyses (one-way ANOVA) revealed a statistically significant effect of font on LI ($F(63,4032) = 41.16$, $p < 0.01$). Because there were 64 levels of the variable, a post-hoc test was used to determine which of the fonts were statistically significantly different from the others. As multiple comparisons were made, a post-hoc test that reduces the chance of Type I errors (which could lead to incorrectly stating that a paired comparison was significant when it in fact was not) was used. The Fisher's LSD method was selected for this study. While Fisher's LSD is often considered to be overly liberal (allowing more Type-I errors), the common alternative of using the Tukey HSD (honestly significant difference) is often considered to be too conservative. The consequences of incorrectly concluding that one font is statistically significantly more legible than another are not particularly serious, so the Fisher's LSD method was selected. To further address this liberality issue, mere statistical significance was not the only criterion used for font recommendation, but rather the combination of statistical significance and practical importance described above.

For all cases where one font was at least 5-ft/in of letter height larger, the findings of the Fisher's LSD analysis were that they were statistically significantly more legible. As a result of this, simply choosing a font that has at least a 5-ft/in of letter height larger mean in LI in Table 4 will ensure the selection of a font that is both statistically and practically more legible. For example: Goudy Old Style Bold UC is more legible than Helvetica Light UC.

Table 4. Ranking of font effect: LI from high to low.

Rank	Font	Example	Mean LI (ft/in)
1	Gill Sans UC	MAGNOLIA	49.64
2	Avenir Medium UC	MAGNOLIA	46.37
3	Copperplate Gothic UC	MAGNOLIA	46.29
4	Helvetica UC	MAGNOLIA	44.86
5	Kabel Ultra UC	MAGNOLIA	44.14
6	Times Bold UC	MAGNOLIA	43.22
7	Futura Medium UC	MAGNOLIA	42.83
8	Garamond Bold UC	MAGNOLIA	41.98
9	Optima Bold UC	MAGNOLIA	41.62
10	Architectural UC	MAGNOLIA	40.18
11	Goudy Old Style Bold UC	MAGNOLIA	40.12
12	Helvetica Bold UC	MAGNOLIA	39.88
13	Arial UC	MAGNOLIA	39.32
14	Avenir Medium LC	Magnolia	39.28
15	Kabel Ultra LC	Magnolia	39.04
16	Futura Medium LC	Magnolia	38.83
17	Georgia UC	MAGNOLIA	38.81
18	Copperplate Gothic SC	MAGNOLIA	38.58
19	Ribbon UC	MAGNOLIA	38.11
20	Optima Bold LC	Magnolia	37.90
21	Times Bold LC	Magnolia	37.80
22	Papyrus UC	MAGNOLIA	37.74
23	Helvetica Medium Condensed UC	MAGNOLIA	37.53
24	Architectural LC	Magnolia	37.36
25	Trajan Bold UC	MAGNOLIA	37.07
26	Gotham Medium UC	MAGNOLIA	36.72
27	GillSans LC	Magnolia	36.32
28	Palatino UC	MAGNOLIA	36.19
29	Garamond Bold LC	Magnolia	36.14
30	Arial Bold UC	MAGNOLIA	36.03
31	Frutiger UC	MAGNOLIA	35.74
32	Avant Garde Medium UC	MAGNOLIA	35.73

Table 4. Ranking of font effect: LI from high to low (continued).

33	Minion UC	MAGNOLIA	35.42
34	Helvetica Light UC	MAGNOLIA	35.14
35	Country Gothic UC	MAGNOLIA	33.89
36	Helvetica Medium Condensed LC	Magnolia	33.62
37	Goudy Old Style Bold LC	Magnolia	33.57
38	Arial Bold LC	Magnolia	33.52
39	Arial LC	Magnolia	33.40
40	Palatino LC	Magnolia	32.96
41	Trajan Bold SC	MAGNOLIA	32.88
42	Frutiger LC	Magnolia	32.85
43	Avant Garde Medium LC	Magnolia	32.56
44	Optima UC	MAGNOLIA	31.98
45	Adobe Garamond UC	MAGNOLIA	31.49
46	Georgia LC	Magnolia	31.39
47	Minion LC	Magnolia	31.22
48	Helvetica Bold LC	Magnolia	31.22
49	Gotham Medium LC	Magnolia	30.73
50	Times New Roman UC	MAGNOLIA	30.67
51	Helvetica Light LC	Magnolia	30.30
52	Helvetica LC	Magnolia	30.08
53	Optima LC	Magnolia	29.61
54	Myriad UC	MAGNOLIA	27.42
55	Adobe Garamond LC	Magnolia	25.89
56	Times New Roman LC	Magnolia	25.79
57	Myriad LC	Magnolia	25.27
58	Brush Script UC	MAGNOLIA	24.20
59	Papyrus LC	Magnolia	21.95
60	Old English UC	MAGNOLIA	21.92
61	Mistral UC	MAGNOLIA	19.29
62	Old English LC	Magnolia	18.42
63	Brush Script LC	Magnolia	15.49
64	Mistral LC	Magnolia	14.52

Case Effect

For over 60 years, research has shown that using lowercase words can improve *recognition* distance over all-uppercase words (Forbes et al., 1950). The current study, however, used a *legibility* paradigm, which has not been shown to benefit from the use of lowercase letters (Forbes et al., 1950; Mace et al., 1994; and Garvey et al., 1997). The difference is that in recognition tasks, the reader knows what word he or she is looking for and merely has to match a mental image of that word with the word on the display; this is helped with the use of lowercase because the ascenders and descenders create a unique overall word shape or footprint. In a legibility task, the reader does not know what the display will say and therefore has to read all or most of the individual letters to build the word.

The effect of case (uppercase vs. lowercase) on font legibility was evaluated using separate ANOVAs. For all the 31 fonts that had upper and lowercase conditions, the uppercase words had higher mean LIs than the lower. In 22 of the cases, that difference was statistically significant (Table 5). The comparisons that were not statistically significant are shaded in red, those that were statistically significant, but not practically important are shaded in yellow, and those that were both statistically significant and practically important are shaded in green. The statistically significant differences in LI as a function of case, ranged in magnitude from 3.91 for Helvetica Medium Condensed to 15.79 for Papyrus.

Table 5. Case effect on LI.

Font	Case	Mean LI	LI Difference	F-value	p-value
Adobe Garamond	LC	25.89	5.60	18.21	<0.01
	UC	31.49			
Architectural	LC	37.36	2.82	2.46	0.12
	UC	40.18			
Arial Bold	LC	33.52	2.51	2.16	0.14
	UC	36.03			
Arial	LC	33.40	5.92	14.59	<0.01
	UC	39.32			
Avant Garde Medium	LC	32.56	3.17	3.86	0.05
	UC	35.73			
Avenir Medium	LC	39.72	6.65	12.69	<0.01
	UC	46.37			
Brush Script	LC	15.49	8.71	40.71	<0.01
	UC	24.20			
Copperplate Gothic	SC	38.58	7.71	18.37	<0.01
	UC	46.29			
Frutiger	LC	32.85	2.89	3.22	0.08
	UC	35.74			
Futura Medium	LC	38.83	4.00	4.32	0.04
	UC	42.83			
Garamond Bold	LC	36.14	5.84	9.65	0.00

	UC	41.98			
Georgia	LC	31.39	7.42	25.39	<0.01
	UC	38.81			
Gill Sans	LC	36.32	13.32	56.86	<0.01
	UC	49.64			
Gotham Medium	LC	30.73	5.99	14.78	<0.01
	UC	36.72			
Goudy Old Style Bold	LC	33.57	6.55	14.98	<0.01
	UC	40.12			
Helvetica Bold	LC	31.22	8.66	28.42	<0.01
	UC	39.88			
Helvetica	LC	30.08	14.78	23.86	<0.01
	UC	44.86			
Helvetica Light	LC	30.10	5.04	10.61	0.01
	UC	35.14			
Helvetica Medium Condensed	LC	33.62	3.91	5.34	0.02
	UC	37.53			
Kabel Ultra	LC	39.04	5.10	7.06	0.01
	UC	44.14			
Minion	LC	31.22	4.20	7.35	0.01
	UC	35.42			
Mistral	LC	14.52	4.77	35.72	<0.01
	UC	19.29			
Myriad	LC	25.27	2.15	3.76	0.05
	UC	27.42			
Old English	LC	18.42	3.50	3.92	0.05
	UC	21.92			
Optima Bold	LC	37.90	3.72	3.71	0.06
	UC	41.62			
Optima	LC	29.61	2.37	3.15	0.08
	UC	31.98			
Palatino	LC	32.96	3.23	3.88	0.05
	UC	36.19			
Papyrus	LC	21.95	15.79	122.98	<0.01
	UC	37.74			
Times Bold	LC	37.81	5.41	10.45	0.02
	UC	43.22			
Times New Roman	LC	25.79	4.88	17.14	<0.01
	UC	30.67			
Trajan Bold	SC	32.88	4.19	7.42	0.01
	UC	37.07			

Serif vs. Sans Serif

Arditi and Cho (2005) studied the effect of serif on font legibility and found very little effect on either reading speed or threshold letter size. In their study, they held all aspects of the font constant except for the serif variable. Unlike those researchers, the current study allowed all other aspects of the fonts (e.g., x-height, stroke width, letter width:height) to vary naturally, and simply combined the results of all of the serif fonts and compared that with the results of all of the sans-serif fonts. Although the method differed, the results were similar to those of Arditi and Cho.

Of the 33 fonts tested, 11 had serifs and 18 did not (Table 6). Four fonts were not used in this analysis because their unusual character did not lend itself to this distinction; these were Brush Script, Old English, Country Gothic, and Mistral. Separate analyses were conducted for the fonts in uppercase and lowercase with the data from all the observations combined. With mean LIs of 32.99 and 33.13 respectively, there was no statistical difference between the serif and the san-serif fonts in the mixed case analysis ($F(1, 1726) = 0.08$, $p = 0.77$). A statistically significant effect was found in the lowercase analysis ($F(1,1726) = 5.35$, $p = 0.02$); however, with mean LIs of 37.91 for the serif and 39.12 for the sans-serif fonts, the difference (i.e., 1.21 ft/in of letter height) is not practically significant.

Table 6. Fonts used in the serif/sans serif analysis.

Serif	Sans Serif
Adobe Garamond, Garamond Bold	Arial, Arial Bold
Architectural	Avant Garde Medium
Copperplate Gothic	Avenir Medium
Georgia	Frutiger
Gaudy Old Style Bold	Futura Medium
Minion	Gill Sans
Palatino	Gotham Medium
Times Bold, Times New Roman	Helvetica, Helvetica Light, Helvetica Medium Condensed, Helvetica Bold
Trajan Bold	Kabel Ultra
	Myriad
	Optima, Optima Bold
	Papyrus
	Ribbon

Font Family

Five of the fonts tested in the study had more than one "weight," such as bold or condensed (Table 7). ANOVAs were conducted on these "font families" to determine if this had an effect on legibility distance. Separate one-way ANOVAs (and a post-hoc test for Helvetica, as it had four levels) were conducted for the fonts in both uppercase and lowercase.

Table 7. Fonts used in the font family analysis.

Font	Weight			
Times	Times New Roman	Times Bold		
Arial	Arial	Arial Bold		
Optima	Optima	Optima Bold		
Garamond	Adobe Garamond	Garamond Bold		
Helvetica	Helvetica	Helvetica Light	Helvetica Medium Condensed	Helvetica Bold

Uppercase

- **Times Bold**, with a mean LI of 43.22, was significantly more legible than Times New Roman, with a mean LI of 30.67 ($F(1, 126) = 69.64$, $p < 0.01$).
- **Optima Bold**, with a mean LI of 41.62, was significantly more legible than Optima, with a mean LI of 31.98 ($F(1, 126) = 33.35$, $p < 0.01$).

- **Garamond Bold**, with a mean LI of 41.98, was significantly more legible than Adobe Garamond, with a mean LI of 31.49 ($F(1, 126) = 37.00$, $p < 0.01$).
- **Helvetica's** ANOVA revealed a statistically significant effect ($F(3, 252) = 11.36$, $p < 0.01$). As discussed with earlier analyses, because there were more than two levels of this variable, a post-hoc test was necessary to determine which of the Helvetica weights were significantly different than the others. The post-hoc test known as the Bonferroni was used, showing that Helvetica (LI = 44.86) was significantly more legible than Helvetica Bold (LI = 39.88), Helvetica Light (LI = 35.14), and Helvetica Medium Condensed (LI = 37.53), with p values of 0.03, <0.01, and <0.01, respectively. Also, Helvetica Bold was statistically more legible than Helvetica Light ($p = 0.04$).

Lowercase

- **Times Bold**, with a mean LI of 37.80, was significantly more legible than Times New Roman, with a mean LI of 25.79 ($F(1, 126) = 74.41$, $p < 0.01$).
- **Optima Bold**, with a mean LI of 37.90, was significantly more legible than Optima, with a mean LI of 29.61 ($F(1, 126) = 24.98$, $p < 0.01$).
- **Garamond Bold**, with a mean LI of 36.14, was significantly more legible than Adobe Garamond, with a mean LI of 25.89 ($F(1, 126) = 46.15$, $p < 0.01$).
- **Helvetica's** ANOVA revealed a statistically significant effect ($F(3, 252) = 5.53$, $p < 0.01$). The Bonferroni post-hoc test showed that the only significant pairings were Helvetica (LI = 36.08) vs. both Helvetica Bold (with a mean LI of 31.22) and Helvetica Light (LI = 30.30), with p values of 0.01, and <0.01, respectively.

Word Analyses

Word Superiority

Due to various factors (e.g., familiarity and word length), some words are easier to read than others and can be read at smaller sizes or further away. This is why this research design included a complete counterbalancing of words and fonts, where each of the 64 font conditions was tested using each of the 64 words. This avoided the possibility that a font might merely seem more legible because it was tested using easier words. To demonstrate what kind of effect word-selection could have, the words were rank-ordered by LI (Table 8). The most legible word was Sunday, with an LI of 45.62, and the least legible was Crawfordsville (LI = 22.81). The difference between these two words was an LI of almost 23 ft/in of letter height.

Table 8. Word superiority effect: rank ordering of words by LI (high to low).

Rank	Word	Mean LI
1	Sunday	45.62
2	Metro	45.60
3	Station	43.71
4	Smith	43.47
5	Thomas	43.29

6	Traffic	42.91
7	Radiant	41.70
8	Agency	40.85
9	Flowers	40.40
10	Freeway	40.12
11	Hangar	39.82
12	Campus	39.79
13	Trenton	39.17
14	Turnpike	38.93
15	Canals	38.82
16	Gardens	38.75
17	Terminal	38.70
18	Pavilion	38.47
19	Fountain	37.74
20	Trails	37.63
21	Cardinal	37.60
22	Stables	37.16
23	Riverside	36.84
24	Magnolia	36.25
25	Acorns	36.13
26	Galleria	35.68
27	Cafeteria	35.58
28	Frederick	35.57
29	Junction	34.97
30	Aurum	34.68
31	Trailhead	34.56
32	Appleton	33.70
33	Fairway	33.33
34	Emporium	33.14
35	Goldberg	32.79
36	Concourse	32.77
37	Savannah	32.42
38	Maplewood	32.26
39	Pharaoh	32.04
40	Harding’s	31.40
41	Forestry	30.97
42	Umbria	30.87
43	Waterways	30.74
44	Distribute	30.74

45	Wadsworth	30.73
46	Perennials	30.42
47	Appalachian	30.06
48	Solarized	30.01
49	Planetarium	29.60
50	Marketplace	29.56
51	Rutherford	29.49
52	Davenport	29.39
53	Groundwater	29.39
54	Arboretum	29.25
55	Cloverton	29.24
56	Byzantine	29.22
57	Wilmington	29.16
58	Benneton	28.65
59	Moonbrook	27.73
60	Gloucester	26.40
61	Gelateria	26.20
62	Pershing’s	26.11
63	Siracusa	25.39
64	Crawfordsville	22.81

Words and Art

All of the displays tested had a combination of words and a graphic element. In ten instances, the graphic had a relation to the word (e.g., a drawing of a flower and the word “Flowers”). To determine whether this had an effect on LI, a series of one-way ANOVAs were conducted (Table 9). Only two of the ten analyses resulted in statistical and practical significance (these are shown in green shaded cells; as above, the red cells are not statistically significant and the yellow are statistically significant, but not practically important). The large difference in the display with the Coffee Cup graphic is most likely due to the inherent difficulty of the word “Gelateria,” rather than any improvement that the image had on the legibility of the word “Cafeteria.”

Table 9. The effect of matching graphic on word legibility.

Word	Graphic	Mean LI	F-value	p-value
Sunday	Sun	45.62	2.99	0.09
Radiant		41.70		
Appleton	Apple	33.73	8.08	**0.01**
Benneton		28.65		
Hangar	Airplane	39.82	2.08	0.15
Traffic		42.91		
Magnolia	Flower	36.25	4.64	**0.03**
Flowers		40.40		
Moonbrook	Crescent Moon	27.73	1.72	0.19
Planetarium		29.60		
Cafeteria	Coffee Cup	35.58	27.38	**<0.01**
Gelateria		26.20		
Cloverton	Clover	29.24	0.00	0.96
Wilmington		29.16		
Campus	Squirrel	39.79	3.79	0.05
Acorns		36.13		
Frederick	Horse	35.57	0.53	0.47
Stables		37.16		
Arboretum	Tree	29.24	1.06	0.30
Forestry		30.97		

Summary

The objective of this research was to determine the relative legibility distances of a large set of fonts that are used on large-scale visual displays. This research gives users the ability to compare the legibility distances of these fonts and make an informed decision about which to use on their displays. Several results are clear:

- Font selection can make a very big difference in the legibility distance of large displays; however, there are many fonts that have equivalent legibility (see Table 4).
- Case (upper vs. lowercase) can sometimes, but not always, have a large impact on display legibility, with uppercase often performing significantly and substantially better than lowercase, at least under the conditions of this research study (see Table 5).
- The choice of serif vs. sans serif alone does not have an effect on legibility distance for large format displays (see Table 6).
- Font weight can dramatically impact the distance at which a display can be read. Just because a font shares a family name (e.g., Helvetica) does not mean it will have equivalent legibility (see Table 7).
- Word selection can have a dramatic impact on the legibility distance of displays, with simpler, shorter, more familiar words being read at greater distances, regardless of font (see Table 8).

- The matching of a word to an image or graphic on a display does not, in general, have an impact on legibility (see Table 9).

References

Arditi, A., & Cho, J. (2005) Serifs and font legibility. *Vision Research*. 45(23), 2917-2925.

Forbes, T. W., Moskowitz, K., & Morgan, G. (1950). A comparison of lower case and capital letters for highway signs. *Proceedings, Highway Research Board*, 30, 355-373.

Garvey, P. M. (2007). Urban wayfinding signs: Evaluating exceptions to FHWA's standard alphabets. *Transportation Research Record: Journal of the Transportation Research Board*, (2030), 10-14. National Academy Press, Washington, D.C.

Garvey, P. M., & Kuhn, B. T. (2011). Highway sign visibility. Chapter 11 In M. Kutz (Ed.), *Handbook of Transportation Engineering* (2^{nd} Ed.), McGraw-Hill, New York, New York.

Garvey, P. M., Thompson-Kuhn, B., & Pietrucha, M. T. (1995). Sign visibility literature review. *United States Sign Council (USSC) Research Project, Final Report.*

Garvey, P. M., Pietrucha, M. T., & Meeker, D. (1997). Effects of font and capitalization on legibility of guide signs. *Transportation Research Record: Journal of the Transportation Research Board*, (1605), 73-79. National Academy Press, Washington, D.C.

Garvey, P. M., Zineddin, A. Z., & Pietrucha, M. T. (2001). Letter legibility for signs and other large format applications. *Proceedings of the Human Factors and Ergonomics Society 45th Annual Meeting.* 45(18), 1443-1447. Human Factors and Ergonomics Society, Santa Monica, CA.

Kuhn, B. T., Garvey, P. M., & Pietrucha, M. T. (1998). The impact of color on typical on-premise sign font visibility. In *Transportation Research Board's 14th Biennial Symposium on Visibility*, *Washington, D.C.*

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, 1–22.

Mace, D. M., Garvey, P. M., & Heckard, R. F. (1994). *Relative visibility of increased legend size vs. brighter materials for traffic signs*. Federal Highway Administration publication no. FHWA-RD-94-035, Washington, D.C.

Yager, D., Aquilante, K., & Plass, R. (1998). High and how luminance letters, acuity reserve, and font effects on reading speed. *Vision Research,* 38(17), 2527–2531.

Zineddin, A. Z., Garvey, P. M., Carlson, R. A., & Pietrucha, M. T. (2003). Effects of practice on font legibility. In *Proceedings of the Human Factors and Ergonomics Society's 47th Annual Meeting*. 47(13), 1717 - 1720. Human Factors and Ergonomics Society, Santa Monica, CA.

Biographies

Philip M. Garvey, The Larson Transportation Institute, Penn State; M.S. in Experimental Psychology, 1986, Villanova University

Wei-Yin Eie, PhD Candidate, Department of Civil and Environmental Engineering, Penn State; M.S. in Transportation Technology and Management, 2009, National Chiao Tung University, Taiwan

M. Jennifer Klena, College of Agricultural and Biological Engineering, Penn State; B.S. in Health and Human Development, 2000, Penn State