White Paper 2012-12

This is an Author's Original Manuscript of an article that has been updated and published by Taylor & Francis Group in Ergonomics on [date of publication], available online: http://dx.doi.org/10.1080/00140139.2014.940000. We suggest that the updated version be referenced in all citations as the archival version of this work.

An Evaluation of Typeface Design in a Text-Rich Automotive User Interface

Bryan Reimer, Bruce Mehler, Joseph F. Coughlin

September 23, 2012

This paper reports on the results of a project examining the impact of typeface design on glance behavior away from the roadway when a driver interacts with a multi-line menu display designed to model a text-rich automotive human machine interface (HMI). Data from two studies are considered. Across the two studies, usable data was collected from 82 participants ranging from 36 to 75 years of age in a driving simulation experiment in which participants were asked to respond to a series of address, restaurant identification, and content search menus that were implemented using two different typeface designs. The second study served as a replication of the first with the sole exception that the brightness of the display screen was changed. Across the two studies, among men, a "humanist" typeface resulted in a 10.6% lower visual demand as measured by total glance time as compared to the "square grotesque" typeface. Total response time and number of glances required to complete a response showed similar patterns. Interestingly, the impact of different typeface style was either more modest or not apparent for women on these variables. Error rates for both males and females were 3.1% less for the humanist typeface. This research suggests that optimizing typeface characteristics may be viewed as a simple and effective method of providing a significant reduction in interface demand and associated distractions. Future work will need to assess if other typeface characteristics can be tuned to provide further reductions in demand.

1. Introduction

The importance of providing a driver with a visual user interface in which controls can be rapidly identified and information content easily read appears self-evident. If text or numeric characters are hard to read, user satisfaction is negatively impacted and the risk of accident may increase due to both increased time of the eyes being directed away from the roadway and from cognitive distraction.

Until relatively recently, the total amount of text presented as part of the user interface in automobiles was fairly limited and largely associated with stationary dials, buttons and knobs. However, the advent of nomadic navigation systems, followed by the emergence of in-dash integrated infotainment display screens, has dramatically increased the amount of text-based information that can be presented to the driver. Moreover, these displays are dynamic in nature so that content cannot be deduced on the basis of a memorized location. As a result, legibility, the degree to which individual characters are understandable or recognizable, is of increasing significance as a fundamental consideration in human machine interface (HMI) design in automobiles. In other areas of the automotive operating environment, considerable investment has already been placed on legibility. For instance, the Clearview typeface was developed and tested to specifically enhance legibility of roadway signage (Funkhouser, Chrysler, Nelson, & Park, 2008; Holick, Chrysler, Park, & Carlson, 2006).

1.1 Background

From the perspective of typographic designers, factors that influence legibility can be grouped as *extrinsic* and *intrinsic* (Bigelow & Matteson, 2011). Extrinsic factors are physical considerations such as size, illumination, contrast, polarity, and color. These factors have received respectable attention within the automotive design community and are covered in various standards documents (e.g. ISO 15008, 2009). Text size is known to have a significant effect on reading speed and this has been confirmed in automotive oriented research (Cai & Green, 2005; Fujikake, Hasegawa, Omori, Takada, & Miyano, 2007; O'Day & Tijerina, 2011).

Intrinsic factors involve the actual shape of characters and include features such as case, width, weight, stroke modulation, form groups, serifs (projecting features at the end of strokes), and slant. The effect of shape-based factors on legibility has not been studied as extensively as extrinsic factors. Nonetheless, Bigelow and Matteson (2011) note that the "relative dearth of rigorous studies of design features and legibility has not, however, prevented cultural and aesthetic preferences from giving rise to anecdotal claims of superior (or inferior) legibility for various typeface designs and design categories"; they go on to suggest areas for further investigation to establish empirical data to support design choices.

Because the reading of displays by the driver in an automobile is limited to brief glances, reading in this environment is substantially different from continuous or immersive reading considered in typical legibility studies. Some typographers suggest that "humanist" (Frutiger®) sans-serif typefaces with strongly differentiated form groups may be more legible in the context of brief glances than the widely used geometric sans-serif (Century Gothic™), "grotesque" sans-serifs (Helvetica®) and "square grotesque" sans-serifs (Eurostile®) typefaces. Additional study is required not only on intrinsic factors of typefaces, but also of the arrangement of text into short text groupings or segments as might be used in an automotive display.

1.2 Typeface Considerations

The present study represents a collaboration undertaken between typographic specialists from Monotype Imaging Inc. and human factors researchers in the New England University Transportation Center at MIT to examine whether typeface design characteristics can impact legibility in an automotive display context in a manner that can be objectively measured. While it would be relatively easy to select from the universe of existing typefaces examples with clear differences in legibility, a basic design consideration was to select as a reference point a typeface representative of a form that is currently in use in the automotive industry and compare it against a form that expert opinion suggests might offer advantages. In other words, the comparison would be made between a typeface design that has a recognized level of acceptability in automotive applications and evaluate whether that level of legibility can be improved upon. As a starting point for this work, two commercially available typeface genres were selected for comparison purposes. These were a square grotesque typeface, Eurostile, which is known to be used in current automotive applications and a humanist style typeface, Frutiger, which has a number of features that Monotype typographers believed should improve legibility on in-vehicle display screens.

As illustrated in Figure 1, humanist genre typefaces are considered to be more legible because of the following:

- open space inside the letterforms to prevent from blurring their shapes
- ample space between the letterforms to prevent them from clashing or blurring together
- highly distinguishable shapes to prevent 'at a glance' ambiguity
- varied horizontal proportions to add distinguishing characteristics

In contrast, grotesque and square grotesque typefaces are considered less legible due to the following:

- nearly closed letterforms (long terminal features) blur their form
- highly assimilated letterforms increase ambiguity
- highly assimilated horizontal proportions increase ambiguity
- typically tight letter spacing causes letterforms to blur together

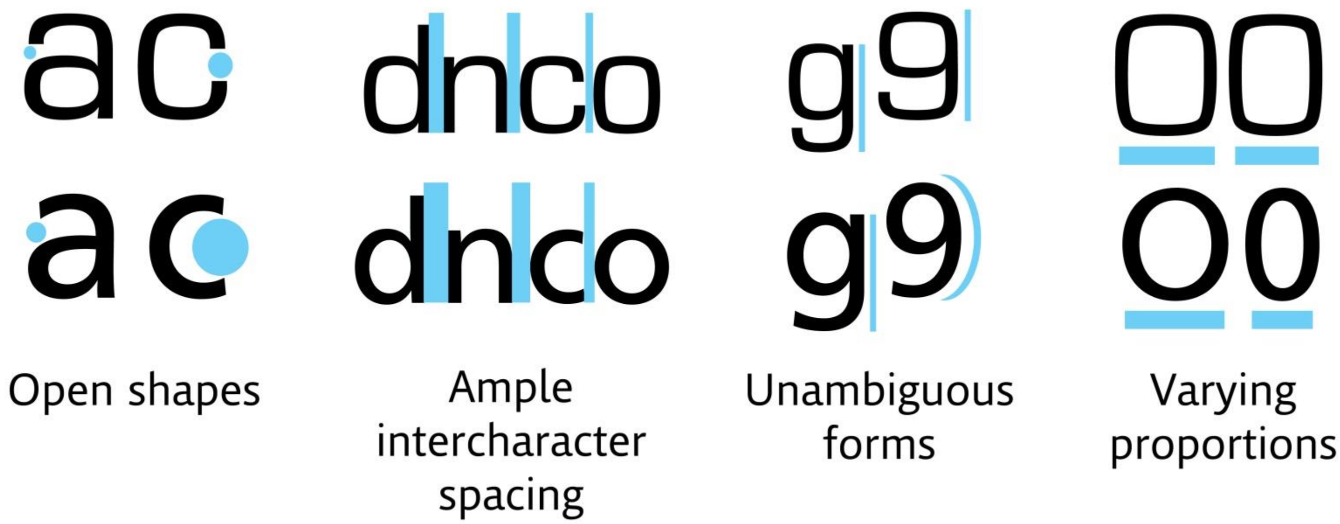


Figure 1. The top line of characters are a square grotesque design (Eurostile) and the bottom line a humanist design (Frutiger) highlighting various characteristics thought to improve legibility. (Graphic courtesy of Steve Matteson of Monotype Imaging.)

The most important feature in the recognition of Latin letterforms is the terminations (Fiset et al., 2008). The open space design of the humanist typefaces supports distinctive and highly visible forms and the distance between the terminations works to avoid the meshing together of forms and keeps these features easily identifiable (Pelli et al., 2009). A sampling of the range of openness of aperture in popular commercial type designs is illustrated in Figure 2 by the terminations of strokes in the letter "c" starting from a square grotesque typeface (Eurostile) and continuing through a humanist typestyle (Frutiger). The letters shown below are all displayed at 100 point – no adjustments have been made to regularize their height.

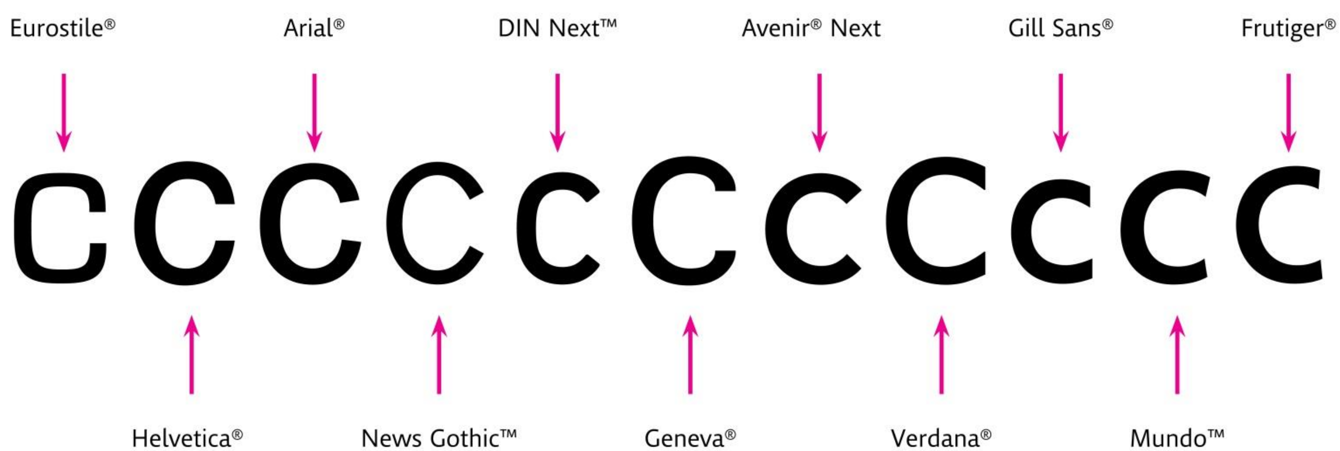


Figure 2. This illustration begins on the left with a very closed aperture of a square grotesque design and progresses to the right with more open apertures found in the humanist sans serif genre. (Graphic courtesy of Steve Matteson of Monotype Imaging.)

The ample space between letterforms protects from too much crowding (Pelli et al., 2007), therefore increasing the visual span and resulting in better legibility. The third and fourth attributes are particularly relevant for ease of rapid identification. Letter identification is facilitated when there is a lower number of shapes that can be confused with one another (Attneave & Arnoult, 1956), which should be the case for humanist vs. square grotesque typefaces. The square grotesque shapes adhere to a rectangular form that is repeated in a large number of characters. This results in similarly shaped letterforms. On the other hand, the humanist letters are differentiated from one another through their structural make-up and subtle stroke modulations. The resulting forms are easier to distinguish from one another.

1.3 Character Height, Width & Stoke Width

As noted previously, character size is a significant variable underlying legibility (Cai & Green, 2005; Fujikake, et al., 2007; O'Day & Tijerina, 2011). While some typeface designers have argued for using the height of the lowercase "x" to characterize the physical size of typefaces (x-height) (Bigelow & Matteson, 2011), current international standards for automotive displays (ISO 15008, 2009) specify that character height for a particular font is to be measured as the distance between the base line and the cap line height of the font, using the capital "H" as the reference. With this in mind, Monotype typographers constructed scaled versions of a humanist typeface (Frutiger) and square grotesque typeface (Eurostile) in which the capital letter heights were equivalent across the two fonts to assess the significance of the intrinsic shape characteristics of the two type styles while conforming to automotive industry standards (see Figure 3).

Figure 3. The fonts were constructed to have equivalent letter heights based on the capital letter "H" in line with ISO 15008 standards for defining automotive font sizes. The square grotesque typeface (Eurostile) is on the left and humanist typeface (Frutiger) is on the right.

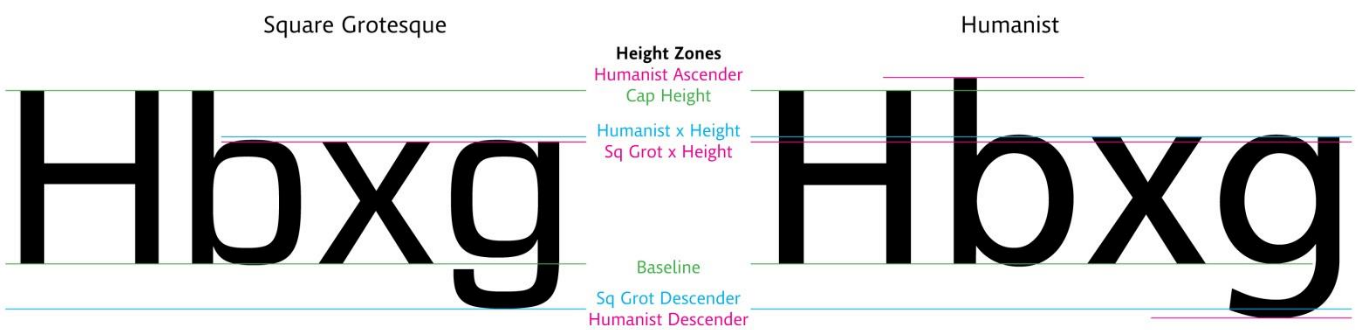


Figure 4. Subtle differences in the heights of other characters may be present when fonts are normalized around the height of the capital "H" reference standard. The square grotesque typeface (Eurostile) is on the left and humanist typeface (Frutiger is on the right. (Graphic courtesy of Steve Matteson of Monotype Imaging.)

Figure 4 highlights some of the subtle differences that may appear across typeface designs in terms of a seemingly simple variable such as character height. When the capital letter "H" is used as a reference, a comparison of the square grotesque font (Eurostile) with the humanist font (Frutiger), shows that the lower case letters in the humanist typeface are slightly larger. This can be seen in the height of the ascender in the character "b" which extends above the top of the capital "H", in the "x-height, and in the descender of the character "g" which drops lower below the reference line than is the case for the square grotesque typeface. At the same time, the character size in the square grotesque design is slightly wider and has a rather squarish proportion, while the humanist has an upright, rectangular proportion (see Figure 4). In effect, the humanist design has a taller x-height, while the square grotesque characters are wider. The end result for the capital "H" height normalized versions of the two typefaces was that the overall areas of the counters, or insides of the letters, were very close in size and the two fonts were similar in optical size. The magnitude of these size differences is very small compared to the difference in openness seen in a comparison of the characters “c” and “g” between the two typefaces as can be observed in the figures above. These differences would be challenging for the untrained observer to consciously detect at the sizes typically used in-vehicle information display systems. Nonetheless, these factors may combine with the more overt features of openness of shapes, character spacing, varied proportions, and other shape distinguishing features that impact overall legibility.

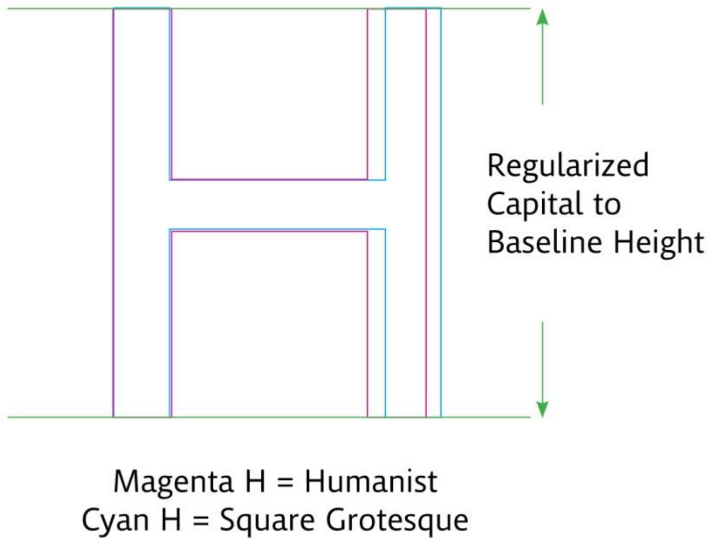


Figure 5. The graphic above compares the relative difference between the two typefaces studied in stroke width (difference between the cyan and magenta lines on the left side of the “H”) and character width (difference on the right side of the “H”). (Graphic courtesy of Steve Matteson of Monotype Imaging.)

Another factor that is known to influence legibility is stroke width (O’Day & Tijerina, 2001). For a given character height, very thin characters are going to be relatively difficult to read at a glance, increasing the thickness will improve legibility up to a point, and then further thickening will begin to obscure legibility. O’Day and Tijerina examined stroke widths of 7%, 9%, 20%, 28%, and 30% of character height. For the combinations of character height, character width, and stoke width that they considered, the combinations with thinner stroke widths were associated with fewer errors and faster reading time. Individual typefaces within the same typeface genre or across different genres may differ slightly in stroke widths, even though their assigned weight category (e.g. light, regular, medium or semi-bold) may be the same. The humanist font selected for this study is approximately 4.74% heavier in weight than the square grotesque. Specifically, the stroke width for the humanist font is 14% of character height and the value for the square grotesque is 13.6% of character height. This is a subtle difference that is not likely to be easily detectable except when the fonts are enlarged as in the Figures 1 - 5. Figure 5 highlights how subtle this stroke width difference is. The very fine difference between the cyan and magenta lines on the left side of the “H” indicates the relative difference in stroke width between the two typefaces selected. The difference in the cyan and the magenta lines on the right side of the “H” indicate the relatively larger difference in the fonts in terms of character width, with the square grotesque font being noticeably wider. For the conditions of this study taking the physical dimensions and resolution of the target display into account, the rendering of the square grotesque font was calculated as requiring 2.87 pixels versus 3.0 pixels, so in physical pixel count this represents a 0.13 pixel difference in vertical stroke width between the two fonts. Modifying the fonts to match the stroke widths would change an intrinsic characteristic of the typeface design. A decision was thus made to leave the stroke widths of the two fonts as they are normally proportioned for these families.

1.4 Research Intent

This paper reports on the results of two experiments designed to assess the extent to which typeface design impacts how a driver interacts with a multi-line menu display designed to model a text-rich automotive HMI. The first study aimed to assess the hypothesis that menu selection tasks performed while reading a humanist style typeface will require less visual demand then tasks completed while reading a square grotesque style typeface. The second study assessed the extent to which a modification in contrast between the text and screen would impact glance behavior as well as whether the results obtained in Study I would replicate to determine whether our initial findings on the impact of typeface design on glance behavior were robust and not a chance finding.

2. Methods

2.1 Participants

The recruitment procedure and research protocol were approved by MIT's institutional review board. Recruitment was directed at drivers 35 - 75 years old since visual acuity tends to become more of an issue as individuals approach middle age. Participants were required to be active, experienced drivers, based on having held a valid driving license for 3+ years and self-reported average driving frequency of 3 or more times a week. Additional requirements consisted of being in self-reported reasonably good health for one's age, being fully comfortable speaking and reading English, and having no major illness resulting in hospitalization in the past 6 months. A diagnosis of Parkinson's or other neurological problems was also an exclusion criterion due to possible impact on fine motor control. Compensation of $30 was provided for participation.

2.2 Apparatus

Data collection was carried out in the MIT AgeLab driving simulator which is built around a fixed base, full cab 2001 Volkswagen New Beetle. An 8' by 6' (2.44m by 1.83m) projection screen was positioned 76" (1.93m) in front of the mid-point of the windshield and provided approximately a 40 degree view of the virtual world at a resolution of 1024 x 768 pixels. Graphical updates were generated at a minimum frame rate of 20 Hz using STISIM Drive version 2.08.02 (Systems Technology, Inc., Hawthorne, CA) based upon a driver's interaction with the steering wheel, brake and accelerator. Force feedback was provided through the steering wheel and auditory feedback consisting of engine noise, cornering, and braking sounds was provided through the vehicle's sound system. Instructions and audio tasks were pre-recorded and also presented through the vehicle sound system. Driving performance data was captured at 10 Hz. A FaceLAB® 5.0.5 eye tracking system (Seeing Machines, Canberra, Australia) recorded data at up to 60 Hz. Two video cameras, one mounted in front of and one behind and to the side of the driver, captured images of the participant's face and hands to monitor general behavior and interaction with a 7" LCD touch screen interface (model CTF400L; cartft.com, Reutlingen, Germany). A MEDAC System/3 physiological monitoring unit (NeuroDyne Medical, Cambridge MA) was sampled at a rate of 250 Hz. to obtain heart rate (modified lead II EKG configuration) and electrodermal activity (skin conductance). Previous validation work has established a high correspondence in the allocation of visual attention in relation to interaction with visual manipulative human machine interfaces HMIs (Wang et al., 2010) and physiological reactivity to cognitive demands (Reimer & Mehler, 2011) between this simulator configuration and on-road behavior.

The CTF-400-L 7" display was selected as being relatively representative of touch screen interfaces being installed in current generation automobiles; it has an aspect ratio of 16:9 with a native resolution of 800 x 480 pixels. The touch screen was mounted on top of the center console which placed it approximately 700 mm distant from the center point between the eyes of the average participant (see Figure 6). As noted earlier, ISO standard 15008 (ISO 15008, 2009) calls for characterizing font character height in terms of the height of the capital letter "H". At the touch screen face, the height of the *H* character for both typefaces was 4mm. The effective size of the character depends on the distance of the driver's eyes from the screen. The standard for representing this feature is to represent the value as the subtended angle from the rearmost point of the cyclopean eyellipse. Represented in arc minutes, this corresponds to a value of approximately 19.6 arc minutes for a representative driver in the simulator. Iso standard 15008 rates the suitability level of effective character size as follows: ≥ 20 = recommended, ≥ 16 = acceptable, and ≥ 12 = minimum (for situations where requirements for accuracy and speed of reading are modest). This would place the font size used at the very top end of the acceptable range.

Figure 6. Touch screen mounted in simulator. Note also one of the two eye tracking cameras, an IR illumination pod, and the face video camera mounted on the dash.

The simulation scenario consisted of a divided highway with two lanes in each direction plus a 2 foot (0.61 m) shoulder on each side of the roadway. Lane width was 15 feet (3.62 m) and posted speed limit was 65 mph (104.6 km/h). Typical traffic events on the virtual highway included passing vehicles, lane changes, and slow downs. The average traffic density in the virtual scenario was set at 23 vehicles/mile (14.3/km). Average traffic speed for vehicles in the left lane was set equal to the posted speed limit of 65 mph (104.6 km/h) and 5 mph slower (96.5 km/h) for the right lane.

2.3 Stimulus Material

A touch screen style menu / list selection display template was developed drawing on elements commonly employed across various automotive HMI display screens without specifically modeling a particular commercial implementation. The key element in this study was a 5 line "Destination Selection" list (see Figure 7). Entries in the list changed while the remaining elements were held constant except for font; the font type of the other elements always matched the font used in the selection list.

As described above, the two fonts compared using this display were specialized TrueType versions of humanist (Frutiger®) and square grotesque (Eurostile®) (see Figures 7 & 8). Monotype typographers adjusted the implementations so that capital letter heights were equivalent across the two fonts to conform to automotive standards for character size measurement (ISO 15008).

For the simulated display, high resolution (3334 x 2000 pixel; 300 dpi) screen images were first created in Adobe Illustrator at a point size of 27. The files were subsequently converted to bitmap (.bmp) format using the *Type Optimized (Hinted)* anti-aliasing and 32-bit *depth* settings. These images were then reduced to 1280 x 768 pixel resolution, 96 dpi, bitmap files. The CTF-400-L display hardware downscaled these images to the screen's native 800 x 480 pixel resolution. When displayed on the touch screen, the resulting characters had a capital letter height of 4 mm measured at the screen face. As noted earlier, this corresponded in this set-up to an effective visual measure of 19.6 arc minutes.

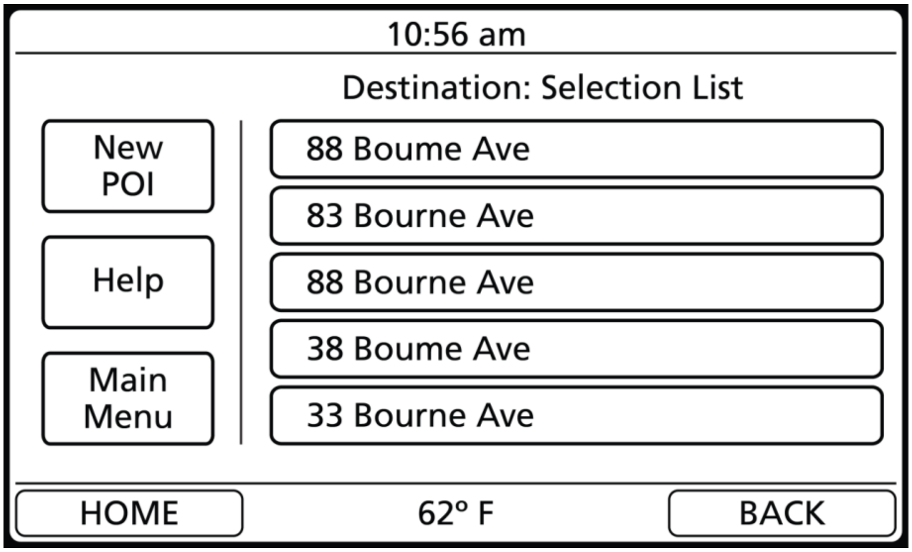


Figure 7. Menu screen in a humanist font

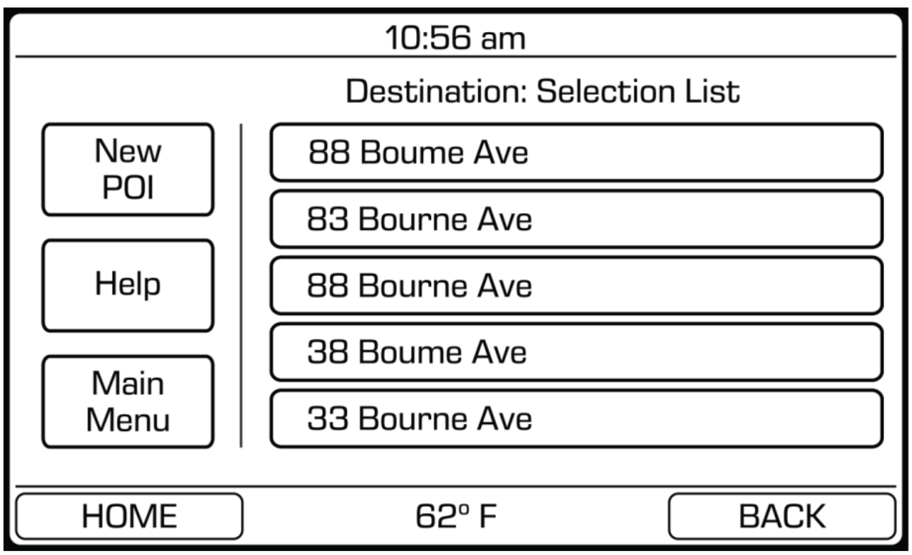


Figure 8. Menu screen in a square grotesque font

Three types of menu lists were presented: addresses, restaurant names, and content searches. Addresses all consisted of leading 2 digit numbers, a name, and a descriptor such as "Street" or "Ave" (see Figures 7 & 8). Restaurant names were all 2 to 3 words in length. The address and restaurant menus deliberately employed characters and name combinations that were visually similar, making accurate visual differentiation of characters important for correct target identification (e.g. "88" vs. "83"; "Boume" vs. "Bourne"). Content search lists contained selection lines ranging from 2 to 4 words in length and did not deliberately employ visually challenging character combinations as was the case in the address and restaurant names. For example, one content search task requested locating a financial services company out of a list of business names. The full set of menu stimuli are reproduced in both typefaces in Appendix A and the target items are listed in Appendix B.

Five menu lists, each with unique content, were created for each task type (5 x 3 = 15 menus). The menus were then produced in two font types (humanist and square grotesque), resulting in a total of 30 menu screens to be presented to each participant. Targets were selected such that each line position was used only once per list type for a given font. Two forms of the target location assignments were created (A&B) such that a given item and location combination that was presented in the humanist font in form A, was presented in the square grotesque font in form B. Participant assignment was balanced so that approximately half the final sample was presented with form A and half with form B and so that the distribution across genders was also balanced.

Figure 9. Prompt screen presented using a Times New Roman font

A prompt screen was used to cue participants as to what item they were to search for on the menu. Each prompt screen consisted of the heading "Please Select:" with a target underneath and the image of a touch screen button labeled "START" below (see Figure 9). A Times New Roman font was employed and the target was presented in capital letters to minimize shape carry-over between the prompt screen and the font employed on the menu display.

2.4 Procedure

Participants read and signed an informed consent, eligibility was verified by interview, and a questionnaire covering demographic variables, driving history, technology experience, and current state (degree of drowsiness, stress level) was completed. Corrected vision was assessed using the Snellen eye chart. Physiological sensors were attached (see Mehler, Reimer, & Coughlin, 2012 for details). Participants then moved to the simulator and adjusted the driver's seat and steering wheel so that they were comfortable and their eyes and mouth nominally visible for the recording and eye tracking cameras. An eye tracking head model was then created.

Recorded audio instructions described the simulator and provided the following guidance and incentive: "During the study, you will receive a monetary award for performing the tasks while you continue driving the simulator. While performance on the tasks is important, you should balance driving safety while you attempt to complete the tasks, just as you would when driving a real car. Since in the real world you cannot disregard the traffic code, you may be penalized $2 for every ticket you receive and $5 for any collision." These instructions are frequently used in our simulation protocols and are intended to encourage a realistic balance between secondary task engagement and driving safety. They reinforced text presented in the informed consent form where it was specified that the monetary award for performing the secondary tasks could be up to $10. In actuality, all participants received equivalent compensation regardless of performance.

A brief drive of 2.65 miles (approximately 4 minutes) followed to provide a degree of familiarization with the simulator environment. Participants were then instructed to pull over to the side of the virtual highway and stop the car. Participants were informed that they were taking part in a study of drivers' interactions with menus that are presented on touch screen displays. The instructions continued: "At numerous points during the drive, a chime will sound and a prompt will appear on the display screen. The prompt will indicate the selection we would like you to locate on the menu screen that will be displayed next. Please carefully read the prompt so that you know exactly what to look for on the menu screen. When you have carefully read the prompt, press the START button on the screen and the menu will be displayed. Locate the correct selection on the menu screen and touch it. The screen will then go blank. In another 20 to 40 seconds, a chime will sound indicating that another prompt is now being displayed." The chime was employed to cue the participant that a new stimulus was ready and the START button allowed the participant to self-pace when they were ready to engage with the menu.

A research associate (RA) then manually triggered presentation of a series of practice trials and provided further explanation of the task as needed. A minimum of 3 examples (1 each of an *address*, *restaurant selection*, and *content search* task) were presented to each participant and the RA had the option to present up to 2 additional examples to ensure that participants understood the tasks.

An audio recording provided the following guidance before driving resumed: "During the drive you will need to balance the demands of driving safely with the demands of the task, just as you would if you were actually driving on a real highway. You will have the opportunity to earn a small monetary bonus by engaging in each of the tasks. Both speed and accuracy are important, so you will want to take enough time to carefully read each prompt to ensure that you make a correct menu selection. At the same time, you will want to get back to paying attention to the roadway quickly enough so that your driving performance and safety are not adversely affected. While we want you to do your best to complete each task to the best of your ability, you should always give priority to safe driving." As stated, these instructions were intended to encourage a balance between attending to the task and an awareness that it was important to attend to safe driving as would be the case under actual driving conditions. Participants were then prompted to resume driving. Shortly after highway speed was regained, automated presentation of stimuli was initiated using a program that randomized the presentation order of the 30 tasks. As noted previously, the presentation intervals between the end of one task and the prompt that another task was ready, varied randomly between 20 and 40 seconds.

A post-experimental questionnaire reassessed current state and assessed symptoms of negative experiences in the simulator using the Simulator Sickness Questionnaire (SSQ) (Kennedy, Lane, Berbaum, & Lilienthal, 1993).

2.5 Data Reduction & Analysis

Eye data was processed following ISO standards (ISO 15007-1, 2002; ISO 15007-2, 2001) and the time spent focused on the touch screen, number of inspections of the touch screen and counts of glances greater than 1.5 seconds and 2.0 seconds. The 1.5 second value corresponds to the maximum occlusion time proposed in the NHTSA distraction guidelines (National Highway Traffic Safety Administration, 2012). The 2.0 second value corresponds to guidelines suggested by the Alliance of Automobile Manufacturers (2006), and currently maintained in the proposed NHTSA distraction guidelines, as the maximum duration for single glances. Total response time was recorded from the point when a participant pressed the start button on the prompt screen to the participant's final selection in the menu list. Trials of the same type, i.e. addresses, restaurant names, or content search, were averaged within each participant to compute average response per font and menu type. A 2 x (2 x 3) design resulted with gender treated as a between subject variable and font type and content type treated as within subject variables.

Primary comparisons were computed using a repeated measures general linear model (GLM). Where significant main effects appeared, post hoc comparisons were computed using paired t-tests. All statistical computations were conducted using SPSS V.20. Where percentage differences between the two typefaces are presented in the results and discussion, the values are based on the following calculation: (value for square grotesque – value for humanist) / value for humanist.

3. Study I Results

3.1 Sample Characteristics

Fifty-one participants were recruited and 48 completed the simulation. All three of the participants who failed to complete the simulation were male. Reasons for these losses were simulator sickness, a protocol error, and a hardware configuration error. Six of the participants (1 male) who completed the simulation were excluded from the analysis. Three cases (1 male) were participants who reported not needing to wear glasses to drive but who chose to use reading glasses during a portion of the experiment to see the touch screen. This resulted in a behavior where they observed the simulated roadway by looking above the lenses of their glasses and looked through the lenses to observe the touch screen. Since this may or may not reflect behavior they might exhibit under actual driving conditions, these cases were excluded. Three other cases (all female) with average response times of 13.3, 15.0, and 19.5 seconds were excluded as outliers. When compared to the response times of the remainder of the sample (see Figure 10), these long response delays are clearly disproportionate as all other cases had an average response time of 8.3 seconds or less. The final analysis sample consisted of 42 subjects, split evenly between males and females. The age range for the male participants was between 36 and 75 with a mean of 55.1 (SD=11.3). Female participants ranged from 37 to 74 years of age with a mean of 56.0 (SD=12.1).The ages of male and female participants did not differ statistically $F(1,40)=.05$, $p=.82$. Male and female participants did not statistically differ in total or subscales of the SSQ (p-values >.05).

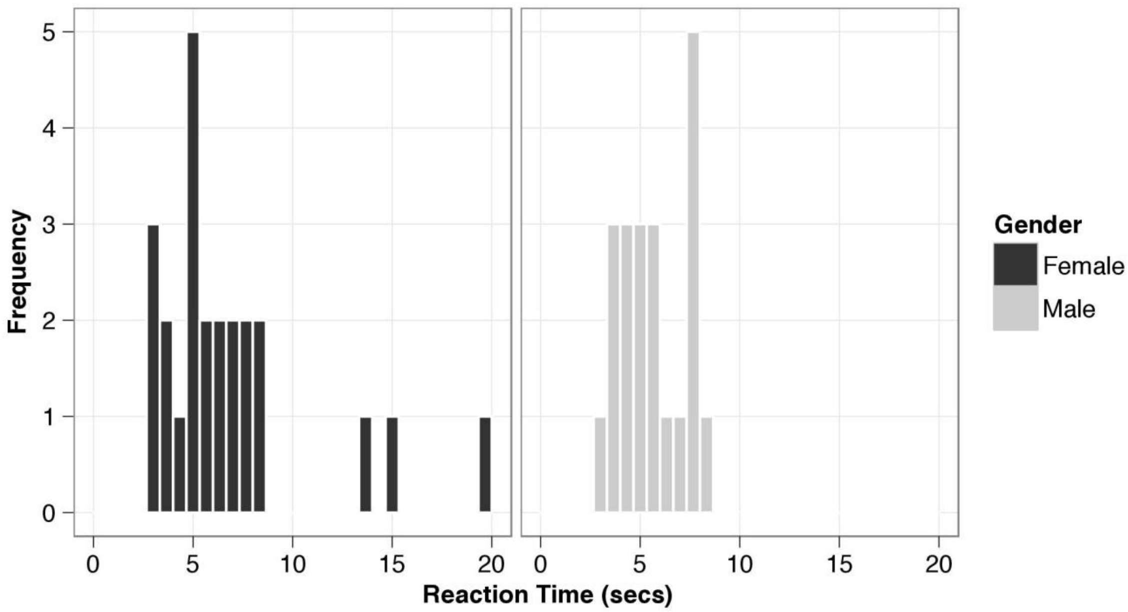


Figure 10. Histogram of average reaction times across typeface design and menu type for participants in Study I. The three right most female cases were classified as outliers.

Corrected visual acuity measured using the Snellen Eye Chart did not differ between male and female participants ($F(1,40)=1.05$, $p=.31$). Males ranged from 20/15 to 20/50 (between line 9 and 4 on the Snellen Eye Chart) while averaging 6.7 (SD=1.5), i.e. just under 20/25. Females ranged from 20/15 to 20/40 (between line 9 and 5) and averaged 7.1 (SD=0.89).

3.2 Task Response Behavior

Task response times by gender, typeface design and menu type appear in Table 1. Response time was significantly impacted by menu type ($F(2, 80)=43.95$, $p<.001$) with content search tasks taking significantly longer than the address ($t(41)=5.66$, $p<.001$) or restaurant name identification tasks ($t(41)=5.96$, $p<.001$). Response time did not differ between the address and restaurant conditions ($t(41)=.67$, $p=.504$). Across the two font conditions, drivers took 5.07 (SD=1.60), 5.18 (SD=1.84), and 6.37 (SD=2.11) seconds to respond to address, restaurant and content selection tasks, respectively.

Table 1: Task response time in Study I (seconds)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	4.64 (1.62)	4.95 (1.73)	5.97 (2.41)	5.19 (1.64)
	SG	5.25 (1.94)	5.49 (2.22)	7.27 (2.71)	6.00 (1.96)
Female	Hum	4.89 (1.32)	5.49 (2.37)	6.19 (2.41)	5.53 (1.81)
	SG	5.48 (2.09)	4.81 (1.72)	6.06 (1.93)	5.45 (1.78)
All	Hum	4.77 (1.47)	5.22 (2.07)	6.08 (2.38)	5.36 (1.71)
	SG	5.37 (1.99)	5.15 (1.99)	6.67 (2.40)	5.73 (1.87)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

A main effect of typeface design on response time appears in the model (F(1,40)=5.39, p=.025) with responses for the sample as whole being faster for the humanist font. The effect of typeface is best considered in combination with a significant interaction with gender (F(1, 40)=7.94, p=.007). Decomposing the interaction effect, separate models assessing the effect of typeface design were developed for male and female participants. A main effect of typeface design appears for male (F(1,20)=12.40, p=.002) participants. Men responded to menus in the humanist typeface in an average of 5.19 (SD=1.64) seconds. Responses to menus with the square grotesque typeface took 6.00 (SD=1.95) seconds or 15.7% longer than the humanist typeface. On the other hand, female participants response times were not significantly different across the two typefaces (F(1,20)=.13, p=.72), with menus in humanist typeface requiring an average of 5.53 (SD=1.81) seconds per response and menus with the square grotesque typeface taking an average of 5.45 (SD=1.78) seconds per response. The impact of menu type on response time remained significant (p values <.01) when the genders were assessed independently. This suggests that the observed difference in response time to the three different menu types is fairly robust.

Table 2: Error rates in Study I (percentages)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	20.5 (18.0)	18.6 (21.5)	12.6 (17.4)	17.2 (14.7)
	SG	26.4 (22.5)	20.2 (22.8)	17.6 (20.0)	21.4 (15.8)
Female	Hum	21.9 (18.6)	18.3 (22.8)	16.2 (16.3)	18.8 (11.1)
	SG	27.9 (25.2)	20.7 (27.3)	18.1 (22.4)	22.2 (20.9)
All	Hum	21.2 (18.1)	18.5 (21.9)	14.4 (16.8)	18.0 (12.9)
	SG	27.1 (23.6)	20.5 (24.9)	17.9 (21.0)	21.8 (18.3)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

The number of errors by gender, typeface design and menu type appear in Table 2. Errors differed statistically across the three menu types (F(2,80)=4.52, p=.014). On average, incorrect selections were made 24.2% (SD=17.1) of the time for address entries, 19.5% (SD=18.7) for restaurant names, and 16.1% (SD=13.7) for content searches. Errors to address menu tasks were marginally larger than restaurant menu tasks (t(41)=1.96, p=.056) and significantly larger than content search menu tasks (t(41)=3.08, p=.004). Restaurant menu tasks and content search tasks, however, did not differ (t(41)=1.14, p=.262). While not statistically significant (F(1,40)=2.04, p=.161), a nominal differences in error rates appeared between menus with the humanist typeface (M=18.0%, SD=12.9) and menus drawn with the square grotesque typeface (M=21.8%, SD=18.3). Participants' gender did not appear to be a predictor of error rates. In addition, across all content types and two typeface designs, error rates and response time were not significantly correlated r(42)=-.257, p=.100.

3.3 Glance Behavior

Total glance time to the display (Table 3) was impacted significantly by menu type (F(2,80)=12.60, p<.001) and typeface design (F(1,40)=7.63, p=.009). An interaction between typeface design and gender (F(1,40)=7.03, p=.011) also influences this model. The total glance time for content search (M=4.23, SD=1.44) was significantly longer than the time required to identify an address (M=3.57, SD=.94) or restaurant name (M=3.62, SD=1.21), (t(41)=3.55, p=.001) and (t(41)=4.79, p<.001) respectively. No difference in off-road glance time appeared between the address and restaurant conditions (t(41)=.408, p=.686).

Table 3: Total glance time to the display in Study I (seconds)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	3.54 (0.98)	3.75 (1.34)	4.28 (1.62)	3.86 (1.08)
	SG	3.84 (1.20)	4.05 (1.60)	5.10 (2.01)	4.33 (1.36)
Female	Hum	3.37 (0.94)	3.50 (1.16)	3.66 (1.19)	3.51 (0.93)
	SG	3.52 (1.02)	3.17 (1.04)	3.87 (1.03)	3.52 (0.93)
All	Hum	3.46 (0.95)	3.63 (1.25)	3.97 (1.44)	3.68 (1.01)
	SG	3.68 (1.11)	3.61 (1.41)	4.48 (1.70)	3.93 (1.22)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

The main effect of typeface design is best considered in relation to the significant interaction between typeface design and gender. As illustrated in Figure 11, the main effect of typeface design appears to be driven by the male participants. Statistically, this is assessed by separate GLMs constructed for the male and female participants. For male participants there was a main effect of typeface design (F(1,40)=10.78, p=.004). This corresponds to a .47 second increase in total glance time to the touch screen with the square grotesque typeface as opposed to the humanist typeface, a 12.2% difference. No effect of typeface on total glace time appears for the female participants (F(1,40)=.010, p=.92). In both the models for men and women, the relationship in glance demands between the three menu types remains consistent with main effect significant for the men (F(2,40)=8.45, p=.001) and women (F(2,40)=5.01, p=.011).

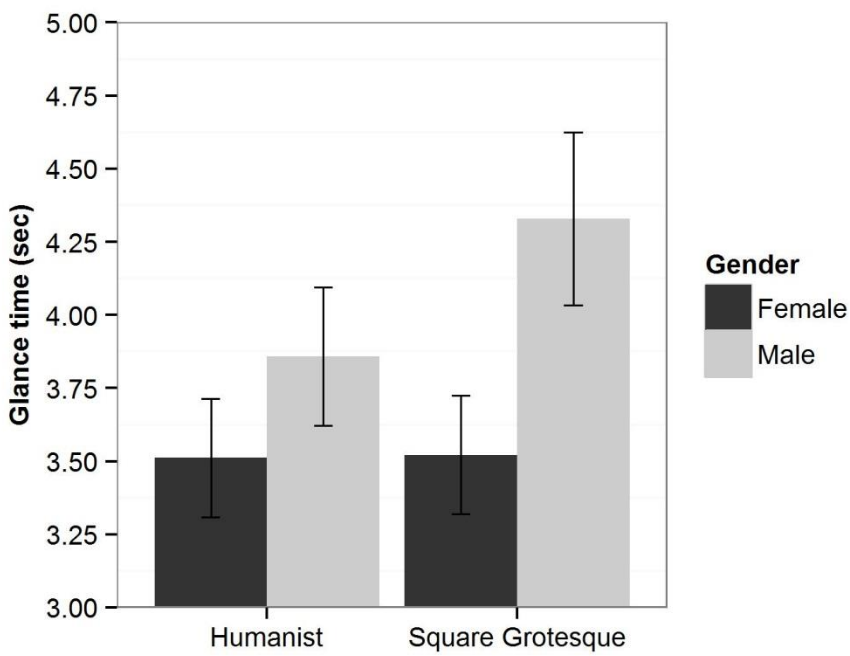


Figure 11. Total glance time to the display screen in Study I across all three menu types.

Consistent with the total allocation of visual attention to the display, the average number of glances to the display (Table 4) is impacted by menu type (F(2,80)=35.16, p<.001) and typeface design (F(1,40)=10.46, p=.002). In addition, typeface and gender appear as a significant interaction effect in the model (F(1,40)=7.87, p=.008). Consistent with the relationship observed for glance time, the average number of glances to the display required for each entry was greater for the content search task (M=3.45, SD=1.18) than address menus (t(41)=6.26, p<.001) or restaurant menus (t(41)=6.72, p<.001). In comparison, address identification required on average 2.71 (SD=1.02) glances per response and restaurant name identification 2.71 (SD=1.01) glances.

Table 4: Glance frequency to the display in Study I (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	2.49 (1.13)	2.57 (1.04)	3.26 (1.30)	2.77 (1.01)
	SG	2.69 (1.15)	2.94 (1.36)	3.86 (1.32)	3.16 (1.13)
Female	Hum	2.73 (0.96)	2.79 (1.05)	3.29 (1.27)	2.94 (0.98)
	SG	2.94 (1.16)	2.55 (0.93)	3.39 (1.17)	2.96 (1.01)
All	Hum	2.61 (1.04)	2.68 (1.04)	3.28 (1.27)	2.85 (0.99)
	SG	2.82 (1.15)	2.75 (1.17)	3.63 (1.25)	3.06 (1.06)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

To decompose the significant typeface gender interaction, separate GLMs constructed for the male and female participants show that the average number of glances increase significantly by typeface design for the male participants (F(1, 40)=26.96, p<.001), but not for female participants (F(1, 40)=.070, p=.795) (see Figure 5). Among men, menus with the humanist typeface required on average 2.77 (SD=1.01) glances per response, while square grotesque menus required on average 3.16 (SD=1.13). This corresponds to a .39 glances per response increase with the square grotesque typeface. Alternatively, this can be viewed as the square grotesque typeface requiring a 14% greater glance demand than humanist typeface. In contrast, among women there was virtually no difference in the number of glances between the two typefaces (see Figure 12). In the models for both males and females, a main effect of menu type remains (p values <.001).

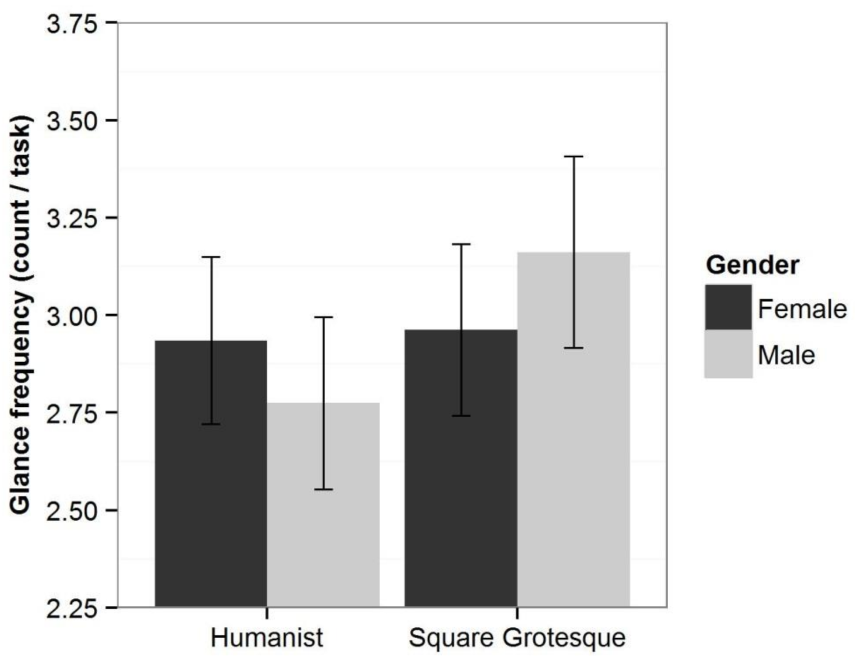


Figure 12. Glance frequency to the display screen in Study I across all three menu types.

The National Highway Traffic Safety Administration (NHTSA) proposed Visual-Manual Driver Distraction Guidelines (National Highway Traffic Safety Administration, 2012) suggest occlusion testing with a shutter open time of 1.5 seconds as one of the alternative interface evaluation methods. Following this construct, an exploratory analysis of the number of glances in excess of 1.5 seconds was computed for each task (see Table 5; Figure 13). The average number of glances per response greater than 1.5 is impacted by gender ($F(1,40)=5.92$, $p=.020$) and a trend appears for typeface ($F(1,40)=4.06$, $p=.051$). A significant interaction effect does not appear between gender and typeface ($F(1,40)=1.79$, $p=.188$). Across typeface, males exhibit .34 more glances greater than 1.5 seconds per response then females. A .07 glance per response increase in the number of glances in excess of 1.5 seconds was observed with square grotesque typeface as opposed to the humanist typeface. While there was no significant gender interaction, the effect appears to be modestly driven by men where there was a .12 (11.9%) increase in the number of glances in excess of 1.5 seconds between the humanist and square grotesque typeface. In comparison females show a .02 (3.4%) increase.

Table 5: Glances greater than 1.5 seconds to the display in Study I (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	1.02 (0.31)	1.08 (0.46)	0.95 (0.64)	1.02 (0.39)
	SG	1.11 (0.39)	1.05 (0.56)	1.25 (0.93)	1.14 (0.53)
Female	Hum	0.75 (0.49)	0.77 (0.48)	0.66 (0.65)	0.73 (0.48)
	SG	0.84 (0.47)	0.76 (0.59)	0.65 (0.52)	0.75 (0.46)
All	Hum	0.88 (0.42)	0.93 (0.49)	0.81 (0.65)	0.87 (0.46)
	SG	0.98 (0.45)	0.90 (0.59)	0.95 (0.81)	0.94 (0.53)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

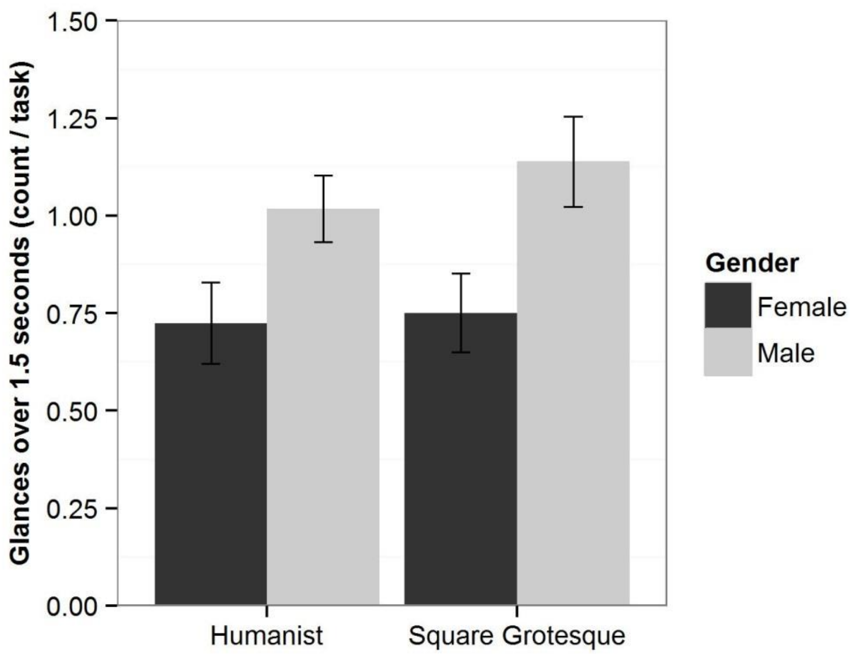


Figure 13. Glances greater than 1.5 seconds to the display in Study I across all three menu types.

The number of glances to the display per response greater than 2 seconds are summarized by gender, menu type and typeface design in Table 6. No significant or substantive differences appear between the two typefaces ($F(1,40)=.033$, $p=.858$), among the different menu types ($F(2,80)=1.62$, $p=.204$) or by gender ($F(1,40)=2.11$, $p=.154$).

Table 6: Glances greater than 2 seconds to the display in Study I (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	0.49 (0.38)	0.56 (0.40)	0.44 (0.52)	0.50 (0.38)
	SG	0.54 (0.35)	0.54 (0.42)	0.56 (0.65)	0.55 (0.43)
Female	Hum	0.38 (0.42)	0.41 (0.48)	0.31 (0.42)	0.37 (0.40)
	SG	0.30 (0.35)	0.39 (0.45)	0.29 (0.42)	0.33 (0.37)
All	Hum	0.43 (0.40)	0.48 (0.44)	0.37 (0.47)	0.43 (0.49)
	SG	0.42 (0.37)	0.46 (0.44)	0.43 (0.56)	0.44 (0.41)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

4. Study I - Summary Review and Discussion

The three different types of menus were included in the study since it was possible that particular features of each content type might differentially impact legibility in the HMI context. In particular, the address menus emphasized the use of numbers that might be easier to differentiate in a humanist typeface. The restaurant names did not include numbers and only focused on letter character form issues. The content search menus did not deliberately attempt to use particular characters or numbers that typographic experts have identified as being less legible in the square grotesque typeface. Instead, the content search items contained what might be considered a more typical distribution of text content. While the results show, for example, that task time was fairly similar for both addresses and restaurant names and that task time for content searches was notably longer, there were no marked interactions between menu type and font type. This indicates that legibility differences between the two font types were fairly broadly distributed across the content studied. The essential question then has to do with the impact of typeface design on each of the dependent variables (response time, glance time, number of glances, etc.) independent of menu type.

In brief, when considering males in the sample, there was a clear and highly statistically significant impact of typeface design on the primary dependent measures. Total glance time was almost a half second faster for the humanist font which represented a 12.2% difference. Presentations in the humanist typeface resulted in a 14% better performance based on the glance frequency metric and total time to complete tasks was 15.7% faster. Men also had nominally fewer moderate (>1.5 second) and long duration (>2.0 second) glances to the touch screen when interacting with the humanist typeface. A complete lack of a difference by typeface in women for these variables was an unexpected finding. In contrast, both men and women showed lower error rates with the humanist vs. the square grotesque typeface. A second study was then conducted to determine the extent to which this overall pattern of results was replicable or represented a chance finding. In addition, the possible impact of contrast on the results was investigated.

5. Study II Methods

Study II was again conducted using the driving simulator described in section 2.2. The simulator is located in a dimly lit room, and a participant's main field of view is defined by the graphic images projected on the 8' by 6' (2.44m by 1.83m) virtual roadway screen. Compared to typical outdoor daylight driving conditions, the driving simulator environment offers significantly reduced levels of ambient lighting with limited dynamic range between various sources of lighting with the projected display of the virtual roadway typically being the brightest light source in the driver's field of view. In Study I, the illumination of the touch screen display was set to its *bright* mode, which results in the HMI display standing out quite clearly in relation to the vehicle interior and the forward roadway scene. Since the main focus of the overall project was to evaluate the impact of typeface design on timing and glance behavior away from a roadway (toward the HMI), and because the tasks were presented using a separate display mounted on top of the center console within a driver's main field of view – we believe that it was important to evaluate whether the overall dynamic range between various lighting sources (projection display, internal display, etc.) and the resulting eye adaptation levels might have influenced the basic findings of Study I.

In order to assess the effect of eye adaptation on glance behavior, the brightness of the internal display, compared to the projection display and overall ambient lighting condition, was reduced in Study II to be within a 3 exposure value (EV) range of illumination. (A single step in EV corresponds to a change of illumination level where amount of light entering an eye, or a camera lens, doubles. See Appendix C for additional background.) This corresponded to changing the CF-400-L touch screen interface from its *bright* setting used in Study I to the *normal* setting in study II.) The light intensity levels were confirmed using a digital SLR camera to make sure that the difference in exposure values between the projection screen and the touch screen display did not exceed 3EV range, and that overall driver's field of view fell within a total scene dynamic range of under 4EV.

Following the procedures outlined in Study I, data was collected, reduced and analyzed. Consistent with Study I, a 2 x (2 x 3) experimental design was initially developed with gender as a between subject variable and font type and menu type as within subject variables. In addition, a statistical comparison is provided for each key measure in Study II with the results from Study I. This extended analysis was conducted to provide an assessment of how changes made to the contrast of the display impacted drivers' behavior and how influences of the contrast change may have impacted behaviors in higher order interactions. The extended 2 x 2 x (2 x 3) design considers contrast and gender as between subject variables and font type and menu type as within subject variables.

6. Study II Results

6.1 Sample Characteristics

Forty-six participants took part in Study II. Of these, two female participants failed to complete the simulation due to simulator sickness. Eye data from an additional male participant could not be coded due to video and eye tracking equipment problems. Two of the participants who completed the simulation (1 male) were excluded from the analysis for using reading glasses during a portion of the experiment to see the touch screen. Finally, data from another male participant was dropped to balance the number of participants in each gender group. In contrast to Study I, no overall reaction time outliers appear in the dataset (see Figure 14). The final analysis sample consisted of 40 subjects, split evenly between males and females. The age range for the male participants was between 36 and 74 with a mean of 55.0 (SD=11.8). Female participants ranged from 37 to 74 years of age with a mean of 53.8 (SD=9.4).The ages of male and female participants did not differ statistically $F(1,38)=.13$, $p=.72$. Male and female participants did not statistically differ in total or subscales of the SSQ (p-values >.05).

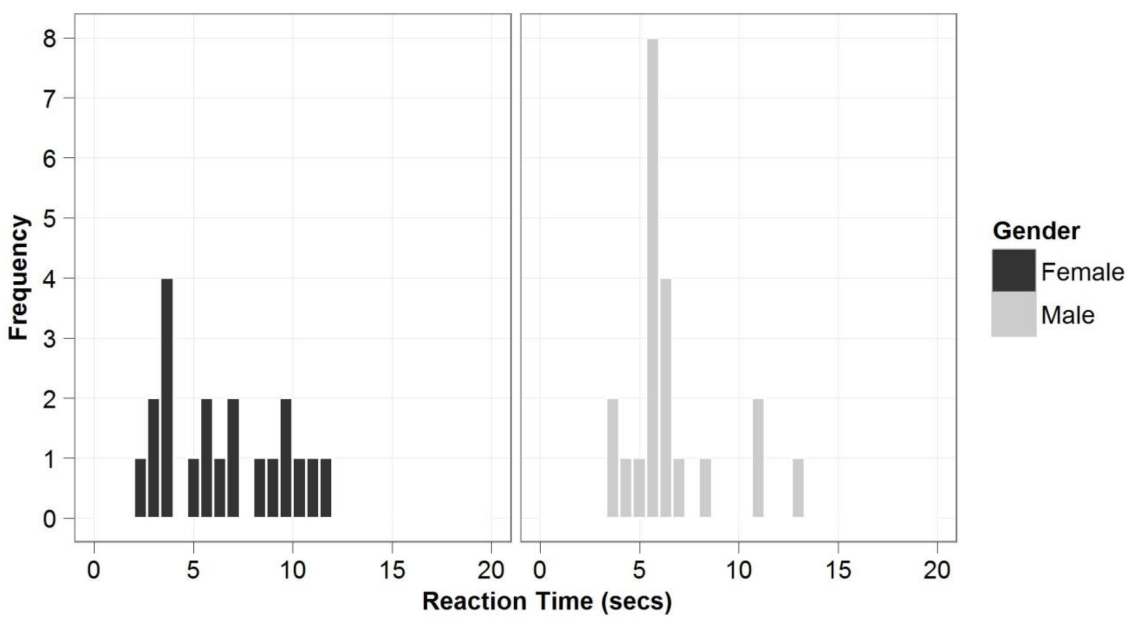


Figure 14. Histogram of average reaction times across typeface and menu type for participants in Study II.

Corrected visual acuity measured using the Snellen Eye Chart in Study II did not differ between male and female participants (F(1,38)=.70, p=.41). Males ranged from 20/15 to 20/50 (between line 10 and 4 on the Snellen Eye Chart) while averaging 6.70 (SD=1.42), i.e. between just under 20/25 and 20/30. Females ranged from 20/20 to 20/50 (between line 8 and 4) and averaged 6.35 (SD=1.23).

6.2 Task Response Behavior

Task response times by gender, typeface design and menu type appear in Table 7. In contrast to Study I, response time was not significantly impacted by content type (F(2, 76)=1.96, p=.148). Consistent with Study I, a main effect of typeface design on response time appears (F(1, 38)=7.41, p=.010) suggesting that across the sample there was an 8.7% improvement in response time with the humanist typeface as compared to the square grotesque typeface. Unlike Study I, the interaction with gender fails to reach statistical significance (F(1,38)=.344, p =.561). However, a significant three way interaction between content type, typeface style, gender (F(2,76)=.3.20, p =.046) does appear.

Table 7: Task response times in Study II (seconds)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	6.13 (2.43)	5.88 (2.46)	6.74 (2.41)	6.25 (2.23)
	SG	7.00 (3.05)	6.70 (2.79)	7.04 (3.07)	6.91 (2.74)
Female	Hum	6.06 (2.89)	6.54 (3.75)	6.18 (2.66)	6.26 (2.83)
	SG	6.96 (3.73)	5.98 (2.69)	7.12 (4.03)	6.69 (3.33)
All	Hum	6.10 (2.63)	6.21 (3.15)	6.46 (2.52)	6.26 (2.51)
	SG	6.98 (3.36)	6.34 (2.73)	7.08 (3.54)	6.80 (3.01)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

The three way interaction effect was decomposed into separate models for male and female participants. A main effect of typeface design appears for the male (F(1,19)=13.20, p=.019) participants. Men responded .66 seconds (10.6%) faster to menus in the humanist typeface. Differences in female participants' response times between the two typefaces did not reach statistical significance (F(1,19)=1.96, p=.178). A significant interaction effect between typeface and menu type, however, did appear in the model (F(2,38)=4.13, p =.024). The interaction effect suggests that females respond more slowly to restaurant menus that were presented in the humanist typeface than the square grotesque typeface. This result may be somewhat influenced by two cases where restaurant menu responses for the humanist typeface were in excess of 3 seconds greater than the remaining samples for both typefaces in men and women. In contrast to the direction of the effect observed for restaurant menus, females responded .91 seconds (15.0%) faster to addresses and .94 seconds (15.2%) faster to content search tasks in the humanist typeface as compared to the square grotesque typeface. The later result appears consistent with effects observed in men across both studies.

Looking statistically across the two studies, a marginal effect of contrast appears (F(1,78)=3.86, p=.053) suggesting that response times to the higher contrast condition in Study I (M=5.54, SD=1.70) are 1 second (17.9%) faster than the lower contrast condition in Study II (M=6.53,SD=2.70). A main effect of condition (F(2, 156)=17.15, p<.001) and typeface design (F(1,78)=12.91, p=.001) appear along with an interaction effect between gender and typeface design (F(1,78)=4.93, p =0.29) and an interaction effect between menu type and contrast (F(2,156)=5.91, p =.003). Consistent with Study I alone and versus Study II alone, in the combined sample content search tasks took significantly longer than the address (t(84)=4.35, p<.001) or restaurant name identification tasks (t(81)=5.10, p<.001). Response time did not differ between the address and restaurant conditions (t(81)=.47, p=.643). Across the studies, drivers took 5.78 (SD=2.42), 5.72 (SD=2.40), and 6.57 (SD=2.51) seconds to respond to address, restaurant and content selection tasks, respectively. The two-way interaction effect with contrast is described by the significant effect of menu type observed in the assessment of high contrast (Study I) and non-significant effect observed with a reduced contrast (Study II). As previously reported, the effect of typeface is best considered in relation to the interaction with gender where a main effect of font only appears in a model of the male participants (F(1,39)=18.20, p<.001). Across the sample, male participants responded to humanist typefaces (M=5.71, SD=2.00) .74 seconds (13.0%) faster than square grotesque typefaces (M=6.45, SD=2.38), while women's response time only differed by .16 seconds (2.7%), (M=5.89, SD=2.36), (M=6.05, SD=2.69) for humanist and square grotesque respectively. Taken together these effects further reinforce the strength of observation of the independent studies for a clear effect of font type in reaction time among males.

Table 8: Error rates in Study II (percentages)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	17.5 (17.7)	17.8 (18.9)	7.8 (15.1)	14.3 (10.0)
	SG	31.0 (26.6)	14.0 (19.6)	12.5 (15.4)	19.2 (13.7)
Female	Hum	15.8 (18.0)	16.8 (17.6)	19.8 (26.4)	17.4 (12.4)
	SG	27.5 (23.1)	21.0 (21.0)	16.5 (17.6)	21.7 (13.2)
All	Hum	16.6 (17.6)	17.3 (18.0)	13.8 (22.1)	15.9 (11.2)
	SG	29.3 (24.7)	17.5 (20.4)	14.5 (16.4)	20.4 (13.3)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

Consistent with Study I, errors in responses to menus (Table 8) differed statistically across the three menu types (F(2,76)=3.83, p=.026). On average, 22.9% (SD=22.2) of address entries, 17.4% (SD=19.1) of restaurant names and 14.1% (SD=19.3) of content searches ended with incorrect responses. Post-hoc tests show that the error rate on address entries is significantly larger than content selection (t(39)=2.27, p=.029) and marginally larger than restaurant name selections (t(39)=1.28, p=.207). Restaurant name selections and content searches do not markedly differ (t(39)=1.76, p=.086). A significant difference (F(1,38)=4.87, p=.033) in error rates appeared between menus with the humanist typeface (M=15.88%, SD=11.22) and menus drawn with the square grotesque typeface (M=20.42%, SD=13.31). This 4.5% difference in error rates between the typefaces observed in this study appears modestly larger than the 3.8% difference observed as a statistical trend in Study I. Participants' gender did not appear to be a predictor of error rates. As in Study I, error rates and response times across all content types and typeface designs were not significantly correlated r(40)=.200, p=.215.

Considering the data across studies, error rates were not significantly affected by contrast (F(1,78)=.438, p=.510), with mean values of 20.0% (SD=13.3) and 18.1% (SD=10.5) for Study I and Study II respectively. Following the results outlined for Study I, in Study II, a main effect of menu type (F(2,156)=8.26, p<.001) and typeface appear (F(1,78)=6.06, p=.016). More error occurred during address entries (M=23.6%, SD=17.3) than restaurant names (M=18.5%, SD=16.5; t(81)=2.62, p=.011) and content searches (M=15.2%, SD=14.6; t(81)=3.66, p<.001). Restaurant name selections and content searches only marginally differ (t(81)=1.73, p=.092). Across the sample, 17.0% (SD=12.1) of the responses to menus in the humanist typeface were incorrect. This was 3.1% less than the percentage of incorrect responses to menus in the square grotesque typeface (M=21.1%,SD=16.0).

6.3 Glance Behavior

Total glance time to the display (Table 9) was impacted significantly by menu type (F(2,76)=4.44, p=.015). The total glance time for address menus, restaurant menus and content search tasks was (M=4.28, SD=1.52), (M=3.86, SD=1.28), and (M=4.08, SD=1.34) respectively. A significant difference in off-road glance time appeared between the address and restaurant menus (t(39)=3.58, p =.001) but not the address and content search (t(39)=1.19, p=.240) or restaurant and content search (t(39)=1.65, p=.108). The pattern observed between menu types was slightly different than what was observed in Study I where glance times for the content search task were longer than the address or restaurant tasks.

Table 9: Total glance time to the display in Study II (seconds)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	4.45 (1.53)	3.91 (1.18)	4.46 (1.34)	4.27 (1.14)
	SG	4.92 (1.66)	4.43 (1.54)	4.64 (1.71)	4.66 (1.41)
Female	Hum	3.67 (1.41)	3.57 (1.42)	3.60 (1.16)	3.61 (1.22)
	SG	4.07 (1.76)	3.51 (1.18)	3.62 (1.34)	3.73 (1.31)
All	Hum	4.06 (1.50)	3.74 (1.30)	4.03 (1.31)	3.94 (1.21)
	SG	4.49 (1.74)	3.97 (1.43)	4.13 (1.60)	4.20 (1.42)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

Typeface design significantly impacts total glance time (F(1,38)=6.83, p=.013), with overall response to menus in the humanist typeface (M=3.94, SD=1.21) appearing .26 seconds (6.6%) faster than the square grotesque typeface (M=4.20, SD=1.42). In contrast to Study I, the interaction between typeface design and gender (F(1,38)=1.93, p=.173) failed to reach statistical significance, however, a main effect of gender does appear (F(1,38)=4.14, p =.049). As illustrated in Figure 15, the effect of typeface design on glance time appears stronger for the male participants. While this appears quite consistent with the glance times observed in Study I (Figure 10), what differs statistically is that in this Study female participants glance times tends to decrease slightly (3.3%) with the humanist typeface compared to the square grotesque typeface as opposed to in study I where the mean glance time for women was essentially the same across typefaces. The 9.1% increase in visual demand observed among the men in this study is consistent with the result from Study I.

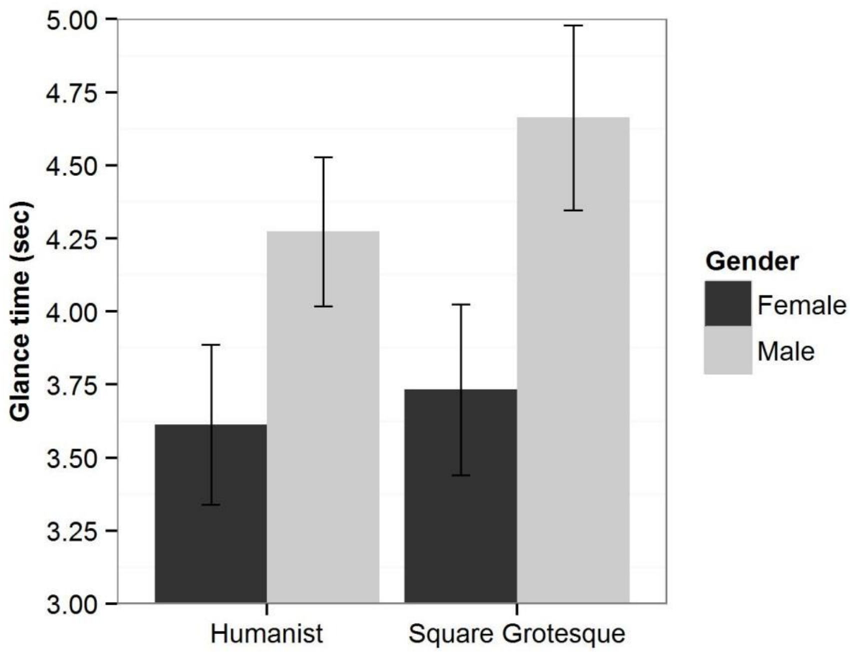


Figure 15. Glance time to the display screen in Study II across all three task types by typeface design for male and females.

Looking across the two studies, contrast appears to have a modest but non-significant (F(1,78)=1.11, p=.296) impact on glance time between the high contrast in Study I (M=3.81, SD=1.08) and lower contrast in Study II (M=4.07, SD=1.29). A main effect of menu type (F(2,156)=8.44, p<.001) and interaction between menu type and contrast (F(2,156)=8.84, p<.001) appear in which across the overall sample glance times for the content search task (M=4.16, SD=1.39) are significantly longer than in restaurant menu tasks (M=3.73 SD=1.24; t(81)=4.43, p<.001) and marginally longer than address menu tasks (M=3.91, SD=1.30; t(81)=1.85, p=.069). Across the studies, the glance time to address menus appears marginally longer than restaurant menus (t(81)=1.98, p=.051). A main effect of typeface (F(1,78)=14.42, p<.001) and interaction between gender and typeface (F(1,78)=7.89, p=.006) appear in line with results presented earlier. Across the studies, male participants glanced at menus in the square grotesque typeface (M=4.49, SD=1.38) for .43 seconds (10.6%) longer than menus in the humanist typeface (M=4.06, SD=1.12). Female participants glance time showed a more modest .06 second difference (1.7%) between the square grotesque (M=3.62, SD=1.12) and humanist (M=3.56, SD=1.07) typefaces.

Table 10: Glance frequency to the display in Study II (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	3.11 (1.37)	2.88 (1.19)	3.58 (1.22)	3.19 (1.15)
	SG	3.36 (1.67)	3.20 (1.46)	3.68 (1.69)	3.41 (1.49)
Female	Hum	3.03 (1.25)	3.06 (1.40)	3.44 (1.46)	3.18 (1.24)
	SG	3.42 (1.51)	3.07 (1.17)	3.37 (1.48)	3.29 (1.31)
All	Hum	3.07 (1.30)	2.97 (1.29)	3.51 (1.33)	3.18 (1.18)
	SG	3.39 (1.57)	3.14 (1.31)	3.52 (1.57)	3.35 (1.38)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

Table 10 displays the average frequency of glances to the display screen by gender, typeface design and menu type. Consistent with the total allocation of visual attention to the touch screen, the average number of glances to the touch screen is impacted by content type (F(2,76)=8.88, p<.001). Across typeface designs the average number of glances to the display for address menus (M=3.23, SD=1.38), restaurant menus (M=3.05, SD=1.24), and content search menus (M=3.52, SD=1.34) are all significantly different (address vs. restaurant t(39)=2.04, p=.048; address vs. content search t(39)=2.41, p=.021; restaurant vs. content search t(39)=3.78, p=.001).

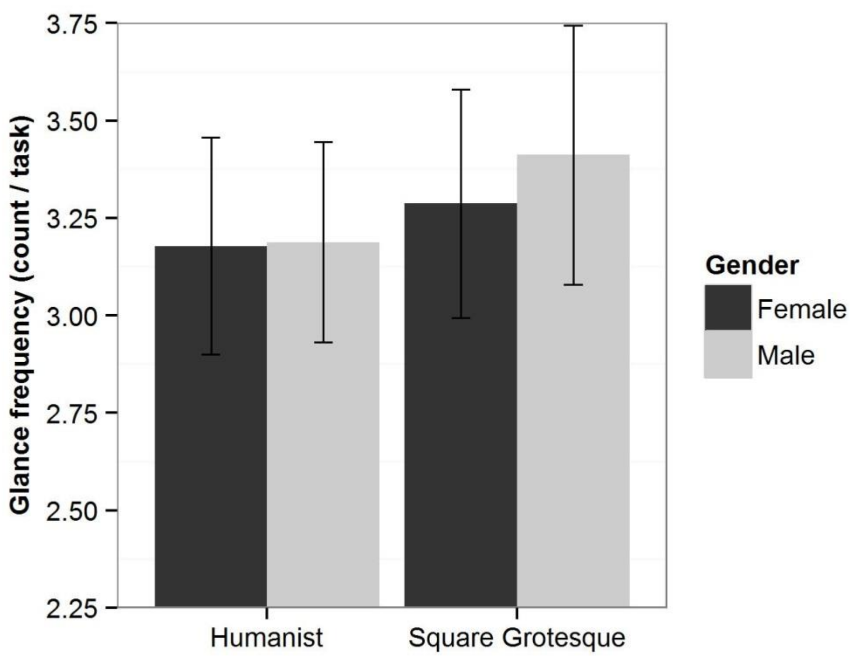


Figure 16. Glance frequency to the display screen in Study II across all three menu types

A more modest trend than observed in Study I appears on the effect of typeface design on the average number of glances to the display (F(1,38)=3.82, p=.058). As illustrated in Figure 16, gender does not influence the pattern of response as markedly as in Study I. Overall glance frequency did not differ significantly between studies (F(1,78)=1.47, p=.228).

Across the two studies a main effect of menu type appears (F(2,156)=37.26, p<.001) in which glance times to the content search menus (M=3.48, SD=1.25) is larger than address menus (M=2.97, SD=1.23; t(81)=5.97, p<.001) and restaurant menus (M=2.88, SD=1.13; t(81)=7.27, p<.001). A significant interaction between menu type and the contrast (F(2,156)=4.68, p=.011) is best considered in terms of the results presented above. In Study I, differences in the frequency of glances to the display were greater for the content search task than the address or restaurant tasks. While in Study II, the glance frequency between all three of the different menu types significantly differ. Furthermore, a main effect of typeface F(1,78)=12.47, p=.001) and interaction between typeface and gender appear (F(1,78)=5.03, p=.028). Breaking the effect of typeface down across gender, male participants glanced to the menus in the square grotesque typeface (M=3.28, SD=1.30) .31 times more per task (10.1%) than menus in the humanist typeface (M=2.98, SD=1.09). Female participants glances to menus in the different typefaces was more equivalent, with square grotesque (M=3.12, SD=1.16) and humanist (M=3.05, SD=1.11) typefaces differing by only .07 glances per task (2.3%). Although the observed effects are stronger in the combined sample, they are consistent with observations in the two independent studies.

Following the construct outlined in Study I, the average number of glances greater than 1.5 second to the display per menu interaction appears in Table 11 by gender, typeface design and menu type. In contrast to study I, a main effect of menu type appears on the average number of glances per interaction greater than 1.5 seconds (F(2,76)=6.71, p=.002). Responses to address menus (M=1.02, SD=.59) involved significantly more glances greater than 1.5 seconds than responses to restaurant menus (M=.86, SD=.56) and content search menus (M=.77, SD=.50), (t(39)=2.84, p=.007) and (t(39)=3.05. p=.004) respectively. The average number of glances greater than 1.5 seconds in responses to restaurant menus and content search menus did not differ statistically (t(39)=1.34, p=.188).

Table 11: Glances greater than 1.5 seconds to the display in Study II (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	1.14 (0.60)	0.99 (0.58)	0.91 (0.58)	1.01 (0.45)
	SG	1.33 (0.67)	1.06 (0.56)	0.92 (0.55)	1.10 (0.48)
Female	Hum	0.76 (0.55)	0.71 (0.61)	0.60 (0.39)	0.69 (0.46)
	SG	0.85 (0.67)	0.67 (0.57)	0.66 (0.52)	0.73 (0.52)
All	Hum	0.95 (0.60)	0.85 (0.61)	0.75 (0.51)	0.85 (0.48)
	SG	1.09 (0.71)	0.87 (0.59)	0.79 (0.55)	0.91 (0.53)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

The average number of glances per task greater than 1.5 was significantly impacted by gender (F(1,38)=5.69, p=.022) and a statistical trend appears with typeface design (F(1,38)=3.07, p=.088). The effect of gender, marginal effect of typeface and non-significant interaction between gender and typeface (F(1,38)=.47, p=.496) are consistent with results from Study I. Figure 17, displays the mean differences in the number of glances over 1.5 seconds by gender and typeface design. Across typeface, male participants exhibited on average .35 more glances greater than 1.5 seconds per response then females. This finding is highly consistent with the .34 difference observed in Study I. Equivalent to Study I, a .07 glance per response increase in the number of glances in excess of 1.5 seconds was observed with square grotesque typeface as opposed to the humanist typeface here. While there was no significant gender interaction, the effect appears to be modestly less driven by men than in Study I. In this study, among the male participants, there was a .09 (9.0%) increase in the number of glances in excess of 1.5 seconds between the humanist and square grotesque typeface as compared to the .12 (11.9%) observed in Study I. In comparison, females in study II show a .04 (5.8%) increase as compared to .02 (3.4%) in Study I.

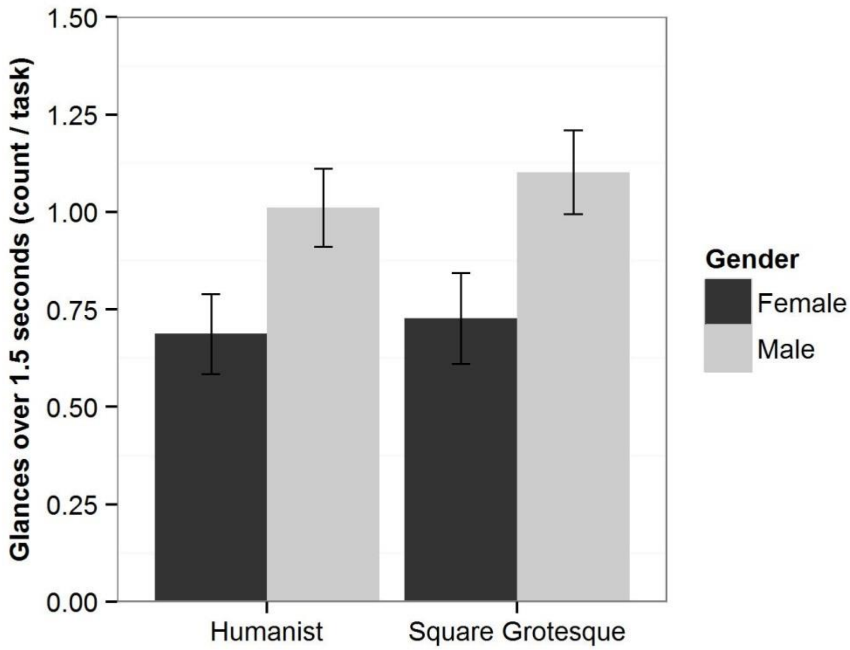


Figure 17. Glances greater than 1.5 seconds to the display screen in study II across all three task types by typeface design for male and females.

Considering the data from the two studies together, the average number of glances per task greater than 1.5 seconds was significantly impacted by gender (F(1,78)=11.61, p<.001), menu type (F(2,156)=4.77, p=.010), and typeface design (F(1,78)=7.08, p=.009), but not contrast level (i.e. no main effect of study) (F(1,78)=.067, p=.796). No significant interaction effects appear in the model. The effect of gender was highly consistent across the two studies with men in Study I exhibiting .34 more glances greater than 1.5 seconds per response then females. In Study II, the difference was .35 more glances. While menu type did not appear significant in either study, the effect appears significant with the combined power of the larger sample. Responses to address menus (M=.97, SD=.50) require .07 more glances greater than 1.5 seconds then restaurant menus (M=.89, SD=.53; t(81)=2.13, p=.036) and .14 more glances then content search menus (M=.83, SD=.60; t(81)=2.55, p=.013). Restaurant menu responses and content search responses did not differ (t(81)=1.31, p=.196). In each of the two studies, a statistical trend suggests that typeface influences the number of glances greater than 1.5 seconds per response. With the combined power of both samples the effect is statistically significant as reported. Across the two samples there were .07 less glances (8.1%) greater than 1.5 seconds per response observed with the humanist typeface (M=.86, SD=.46) as compared to the square grotesque typeface (M=.93, SD=.53). As in the independent studies, while no interaction effect appears between typeface design and gender, the effect appears to be mostly driven by men. Among the male participants, .11 (10.9%) fewer glances in excess of 1.5 seconds per response are observed with the humanist typeface as compared to the square grotesque typeface, while women showed a .03 (4.2%) difference.

The average number of glances greater than 2 seconds per response in study II (Table 12) differed by menu type (F(2,76)=4.71, p=.012), but not typeface design (F(1,38)=2.02, p=.164) or gender (F(1,38)=2.48, p=.124). The effect suggest that the mean number of glances greater than 2 seconds in responses to the address menus (M=.47, SD=.44) tended to be larger than the restaurant menus (M=.38, SD=.33), and was significantly larger than the content search menus (M=.33, SD=.41), (t(39)=1.99, p=.054) and (t(39)=2.63, p=.012) respectively. Glances greater than 2 seconds per response during the restaurant menu and content search menu tasks were not statistically different (t(39)=1.43, p=.160).

Table 12: Glances greater than 2 seconds to the display in Study II (count per task)

		Addresses	Restaurant Names	Content Searches	Average
Male	Hum	0.51 (0.44)	0.41 (0.37)	0.43 (0.54)	0.45 (0.37)
	SG	0.65 (0.52)	0.48 (0.40)	0.40 (0.48)	0.51 (0.42)
Female	Hum	0.35 (0.45)	0.32 (0.35)	0.24 (0.38)	0.30 (0.35)
	SG	0.37 (0.41)	0.34 (0.31)	0.23 (0.25)	0.31 (0.29)
All	Hum	0.43 (0.45)	0.36 (0.36)	0.34 (0.47)	0.38 (0.36)
	SG	0.51 (0.48)	0.41 (0.36)	0.31 (0.39)	0.41 (0.37)

Note: Standard deviations are in parenthesis; Hum = Humanist and SG = Square Grotesque.

Across the two studies a main effect of menu type (F(2,156)=4.22, p=.016) appears along with a marginal interaction between menu type and contrast (study) (F(2,156)=2.55, p=.082). The interaction effect is a result of the non-significant influence of menu type observed in Study I that contrasts with the significant effect observed in Study II. Averaging across the two studies, the number of glances greater than 2 seconds in response to content search menus (M=.36, SD=.45) is less than the number observed for address menus (M=.45, SD=.40; t(81)=2.40, p=.019) and the number observed for restaurant menus (M=.43, SD=.37; t981)=2.24, p=.028). No statistical difference in the number of glances greater than 2 seconds exists between the address and restaurant menus (t(81)=.59, p=.557). A main effect of gender (F(1,78)=4.53, p=.036) is associated with men (M=.50, SD=.39) exhibiting .17 (51.5%) more glances greater than 2 seconds per response then women (M=.33, SD=.34). While the effect of typeface was not significant (F(1,78)=1.09, p=.299), a marginal interaction between typeface and gender appears (F(1,78)=2.97, p=.089). This is associated with .06 (12.8%) fewer glances in excess of 2 seconds being observed among the men with the humanist typeface (M=.47, SD=.37) as compared to the square grotesque typeface (M=.53, SD=.42). In the case of women, however, the humanist (M=.33, SD=.38) and the square grotesque typefaces (M=.32, SD=.33) produced essentially the same number of glances per response.

7. Study II – Summary Review

Reducing the contrast of the HMI display screen resulted in a nominal increase in task completion time, glance time, and glance frequency, although only task completion time approached statistical significance. No impact on error rates was observed. This indicates that the relative difference between the brighter touch screen and the illumination level of the roadway scene did not trigger an adaptation adjustment in the eye that produced any negative impact on processing time or error rates. If anything, lowering the contrast in Study II resulted in drivers taking slightly longer to complete the task.

Interestingly, lowering the contrast to levels that might be more typical of much of normal day time driving conditions, the magnitude of the gender differences seen in Study I decreased somewhat in Study II. Specifically, woman began to show some advantages in responding to the humanist over the square grotesque typeface more in line with the pattern seen in males. Thus, while woman showed no effective difference in glance time between the two typefaces in Study I, they did show a 3.3% reduction in glance time with the humanist typeface in Study II. Similar advantages were seen in women in Study II in total task time and glance frequency. As in Study I, however, the advantages of the humanist font were much more pronounced in males. Combining the data across studies, providing information to male participants in the humanist font resulted in a 13% improvement in overall response time, 10.6% in glance time, and 10.1% in glance frequency.

8. Overall Discussion

As pointed out in the international standards document (ISO 15008, 2009), information and control systems are expected to be designed in a manner that enhances performance and comfort and does not negatively influence workload. The design specialist wants to provide the customer with a visually appealing display and the human factors engineer is responsible for seeing that interface characteristics support efficient and safe operation. Optimized font design should ideally support all of these goals.

Consumer demand for in-vehicle telematics systems supporting navigation, infotainment and communication has resulted in increasingly complex and information dense in-vehicle interfaces. The Alliance of Automobile Manufacturers has agreed upon a *Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information Systems (Driver Focus-Telematics Working Group, 2002)* and recent events have resulted in the initial drafting of voluntary governmental guidance (National Highway Traffic Safety Administration, 2012). Both sets of guidelines provide vehicle manufacturers with a variety of criteria for evaluating driver focused electronics systems regarding the reduction of distraction. While manufacturers have placed considerable effort into optimizing the driver vehicle interface to meet or exceed these guidelines, one area not fully developed is an understanding of how differences in typestyle usage in electronic interfaces may contribute to reduced demand.

This exploratory work demonstrates that the adjustment of typeface design resulted in a reduction of 10.6% in visual demand measured as total glance time across two studies in a menu selection task in male participants. Males also clearly benefited from the humanist typeface in terms of total task time and number of glances. Both males and females showed a 3.1% lower error rate when presented with the humanist font. There was no meaningful impact of font type in women for the task response time, glance frequency or glance duration in Study I under the high contrast condition. In Study II, where the brightness of the display screen relative to the outside driving scene was reduced, women showed a modest, but more similar pattern as the men in which response time and glance frequency were improved with the humanist typeface. The apparent gender differences observed in this sample are, to the best of our knowledge, novel and were unexpected. While males and females did not differ in visual acuity as measured by the Snellen Eye Chart, this does raise the question as to whether there might be other visual acuity or perceptual differences associated with gender that might account for these interesting findings.

The use of eye tracking provides a sensitive method for assessing the allocation of gaze that other design assessment tools such as occlusion or the recording of total task time may not fully capture. The choice of typestyles compared here was not random. Square grotesque represents a typeface design style used by a number of vehicle manufacturers. Humanist, on the other hand, offers a number of attributes that expert typographers believe offer distinct advantages in legibility in the context of limited glance time applications. Humanist style fonts are used by other vehicle manufacturers and in popular mobile computing user interfaces. Some manufacturers have been observed to use a mix of humanist and square grotesque typefaces. The present studies provide objective data supporting the position that the intrinsic font characteristics evaluated can have a positive impact on reducing the glance time demands of a text-rich, multi-line menu interface.

The termination of task trials without the participant encountering negative consequences of an incorrect response (i.e. frustration of not obtaining the desired selection and having to start over) is an artificial requirement of the experimental design employed here. It is worth noting that this may result in a conservative / under representation of the magnitude of the benefits of a more legible font. In actual driving conditions, task engagement would likely continue until a correct menu item is selected, resulting in additional time with the eyes off the roadway. Future work will need to establish whether observed differences in error rates by typeface represent an underestimation of the magnitude of underlying differences and thus provide further illustration of the advantages of a humanist typeface.

This paper focuses on an exploration of the impact of typeface design on the level of visual demand experienced by a driver when interacting with a text-rich HMI. Nonetheless, it is recognized that other characteristics of a typeface, such as character size, capitalization, shadowing, rendering, and foreground/background color combinations (e.g. white on black or black on white), may be adjusted to further reduce demand. In addition, the optimization of other aspects of a display layout, such as white space and design elements is needed. In summary, this research suggests that optimizing typeface characteristics may be viewed as an effective method of providing a meaningful reduction in interface demand and associated distractions. Future work will need to assess if other font characteristics or user customization can be tuned to provide further reductions in demand.

9. Limitations

These were exploratory studies and a number of limitations should be noted. Many variables interact in making-up the characteristic features of a particular typeface and this presents a challenge to systematically evaluating what specific features contribute in what degree to an overall difference in legibility between different typefaces. Character height has been established as a significant variable and this was explicitly controlled in this study by setting the absolute height of the capital letter "H" to be the same for both fonts in line with the ISO 15008 standard. While the reasoning behind the selection of the humanist and square grotesque typefaces in this study had to do with features such as the openness of shapes, inter-character spacing, and ambiguity of forms, another attribute of the humanist typeface used in this study is a slightly wider stroke width than the square grotesque. While the magnitude of this difference is difficult to discern at the display sizes used in this study, to what extent this attribute contributed to the overall difference observed is unknown and would require additional testing to assess. The same could be said of the slight difference in x-heights between the two fonts. These variables highlight the fact that there are many subtle features that contribute to a given typeface design. As previously mentioned, the experimental design did not attempt to control for incorrect responses which may have underestimated the overall impact the modestly higher error rates for the square grotesque type face might have had on driver behavior. Several cases were excluded based upon an unexpected behavior pattern, i.e. the attempt to drive and read a display at the same time using reading glasses. However, it can be argued that these cases illustrate the extreme lengths some drivers need to use to interact with new generation in-vehicle interfaces. This may argue for the desirability of being able to customize some aspects of the display to tune it to match the visual capacities of individual drivers. No measures of near or intermediate visual acuity were collected. While differences in near or intermediate visual acuity could be predictors of performance, it can be argued that the results presented here are in line with the expected behavior of actual drivers. In essence, drivers need to complete HMI related activities while maintaining a high level of acuity in the far reaches of the visual field.

10. Acknowledgments

This collaborative project was underwritten in part by Monotype Imaging Inc. through funding provided to MIT and in contribution of staff time. In particular, Steve Matteson, Vladimir Levantovsky, David Gould, Nadine Chahine, and Geoff Greve of Monotype Imaging provided significant background on factors believed to influence the legibility of font forms and Vikki Quick also provided useful comments. The authors would also like to acknowledge the US Department of Transportation's Region I New England University Transportation Center at MIT for additional support. We would also like to acknowledge the contribution of Steve Proulx in the creation of the stimulus screens and Kirsten Olson for programming. Alea Mehler, Hale McAnulty, and Erin Mckissick contributed significantly to item development, recruitment of participants and the actual running of the simulation. Alexander Chiclana and Nicole Cazares also assisted in conducting study sessions. Ying Wang oversaw eye tracking data reduction and quality review.

11. References

Attneave, F., & Arnoult, M. D. (1956). The quantitative study of shape and pattern perception. *Psychological Bulletin, 53*(6), 452-471.

Bigelow, C., & Matteson, S. (2011). *Font improvements in cockpit displays and their relevance to automotive safety*. Paper presented at the Society of Information Displays 2011 Vehicle Displays and Interfaces Symposium, University of Michigan-Dearborn.

Cai, H., & Green, P. (2005). Range of character heights for vehicle displays as predicted by 22 equations. *Proceedings of the SID Vehicle Display Symposium,* Dearborn, MI.

Driver Focus-Telematics Working Group. (2002). Statement of Principles, Criteria and Verification Procedures on Driver Interactions with Advanced In-Vehicle Information and Communication Systems, Version 2.0: Alliance of Automotive Manufacturers.

Fiset, D., Blais, C., Ethier-Majcher, C., Arguin, M., Bub, D., & Gosselin, F. (2008). Features for identification of uppercase and lowercase letters. *Psychological Science, 19*(11), 1161-1168.

Fujikake, K., Hasegawa, S., Omori, M., Takada, H., & Miyano, M. (2007). Readability of character size for car navigation systems. Human Interface and the Management of Information. Interacting in Information Environments. *Lecture Notes in Computer Science, 4558/2007*, 503-509. doi: DOI: 10.1007/978-3-540-73354-6_55

Funkhouser, D., Chrysler, S., Nelson, A., & Park, E. S. (2008). Traffic Sign Legibility for Different Sign Background Colors: Results of an Open Road Study at Freeway Speeds. *Proceedings of the Human Factors and Ergonomics Society 52nd Annual Meeting*, New York.

Holick, A. J., Chrysler, S. T., Park, E. S., & Carlson, P. J. (2006). Evaluation of the clearview font for negative contrast traffic signs. Austin, Texas: Texas Department of Transportation.

ISO 15007-1. (2002). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 1: Definitions and parameters. Geneva, Switzerland: International Standards Organization.

ISO 15007-2. (2001). Road vehicles - Measurement of driver visual behaviour with respect to transport information and control systems - Part 2: Equipment and procedures. Geneva, Switzerland: International Standards Organization.

ISO 15008. (2009). Ergonomic aspects of transport information and control systems – Specification and test procedures for in-vehicle visual presentation. Geneva, Switzerland: International Standards Organization.

Kennedy, R. S., Lane, N. E., Berbaum, K. S., & Lilienthal, M. G. (1993). Simulator sickness questionnaire: an enhanced method for quantifying simulator sickness. *The Interantional Journal of Aviation Psychology, 3*(3), 203-220.

Mehler, B., Reimer, B., & Coughlin, J. F. (2012). Sensitivity of physiological measures for detecting systematic variations in cognitive demand from a working memory task: an on-road study across three age groups. *Human Factors, 54*(3), 396-412. doi: 10.1177/0018720812442086

National Highway Traffic Safety Administration. (2012). Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices. Washington, DC: National Highway Traffic Safety Administration (NHTSA), Department of Transportation (DOT).

O'Day, S., & Tijerina, L. (2011). Legibility: Back to the Basics. SAE. *International Journal of Passenger Cars - Mechanical Systems, 4*(1), 591-604.

Pelli, D. G., Majaj, N. J., Raizman, N., Christian, C. J., Kim, E., & Palomares, M. C. (2009). Grouping in object recognition: the role of a Gestalt law in letter identification. *Cognitive Neuropsychology, 26*(1), 36-49.

Pelli, D. G., Tillman, K. A., Freeman, J., Su, M., Berger, T. D., & Majaj, N. J. (2007). Crowding and eccentricity determine reading rate. *Journal of Vision, 7*(2), 1.

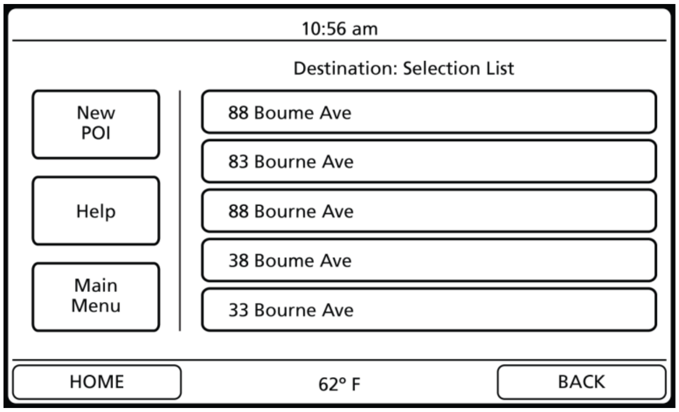
Reimer, B., & Mehler, B. (2011). The impact of cognitive workload on physiological arousal in young adult drivers: a field study and simulation validation. *Ergonomics, 54*(10), 932-942.

Wang, Y., Reimer, B., Mehler, B., Lammers, V., D'Ambrosio, L. A., & Coughlin, J. F. (2010). The validity of driving simulation for assessing differences between in-vehicle informational interfaces: a comparison with field testing. *Ergonomics 53*(3), 404-420.

Appendix A – Stimulus Items

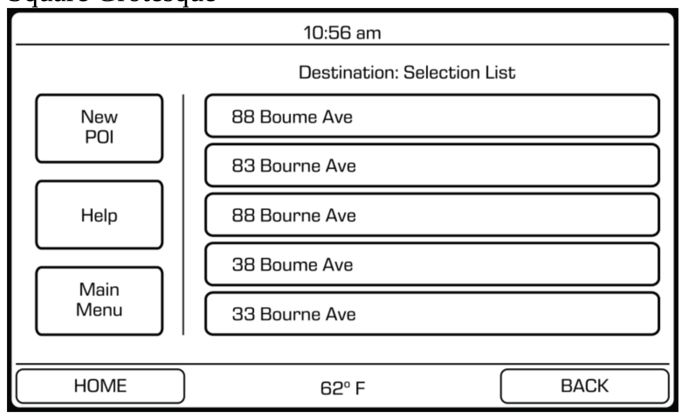
Each of 15 menus was presented once in the humanist font and once in the square grotesque font. The presentation order of the menu screens was randomized across the sample. See Methods section for details. Note – the images below are smaller than those used in the study and clarity is distorted as a result. Please see Figures 6 and 7 in the introductory section of this paper for a more exact representation of the stimulus items.

Humanist

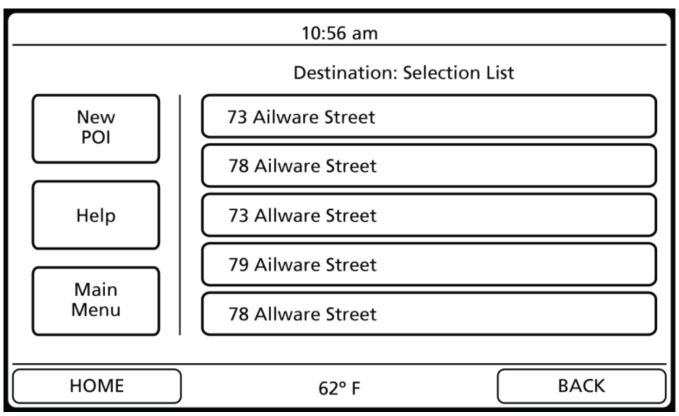


SET A-1

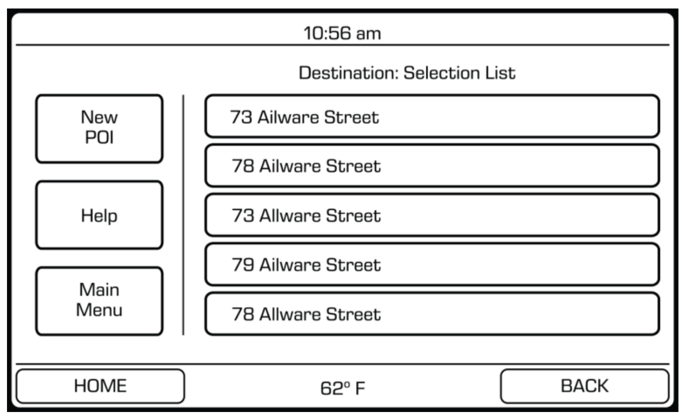
Square Grotesque



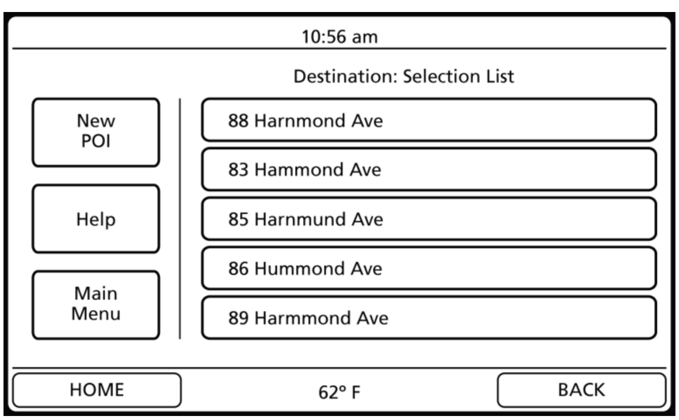
SET A-16



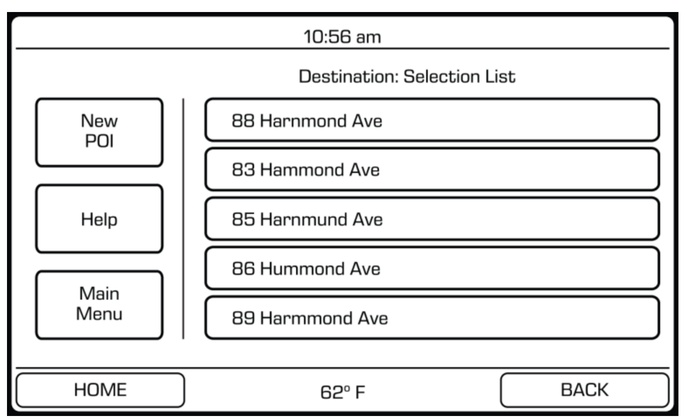
SET A-2



SET A-17

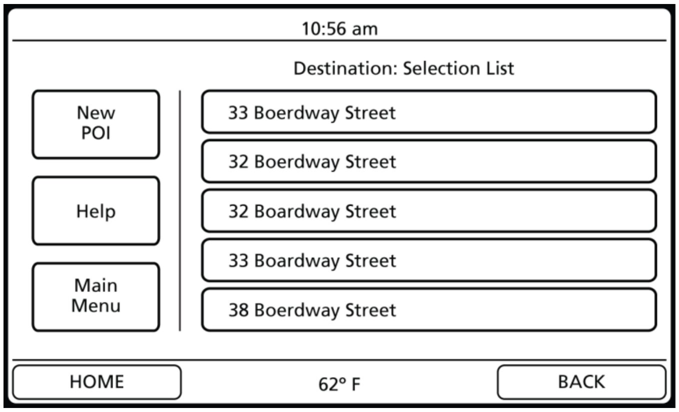


SET A-3

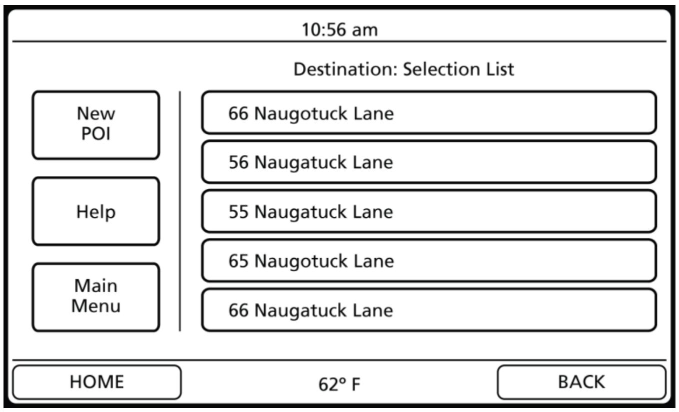


SET A-18

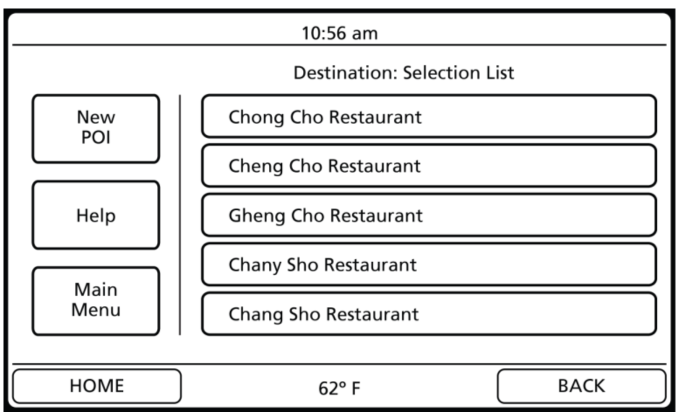
Humanist



SET A-4

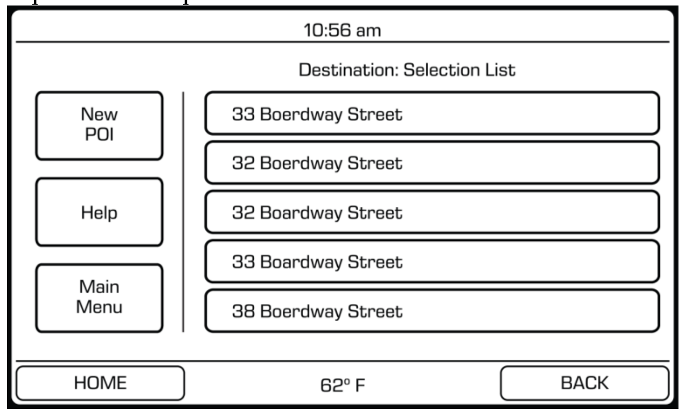


SET A-5

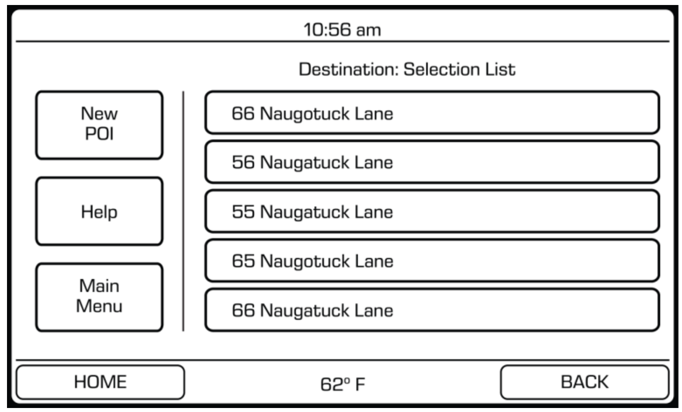


SET A-6

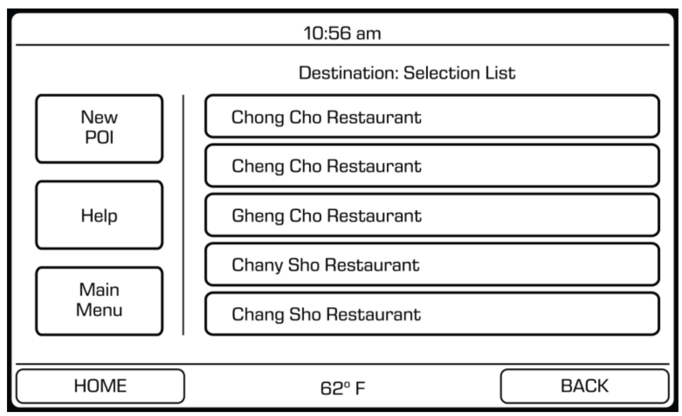
Square Grotesque



SET A-19

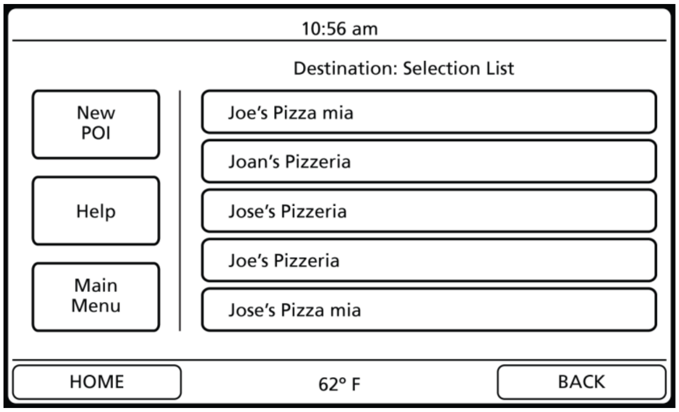


SET A-20

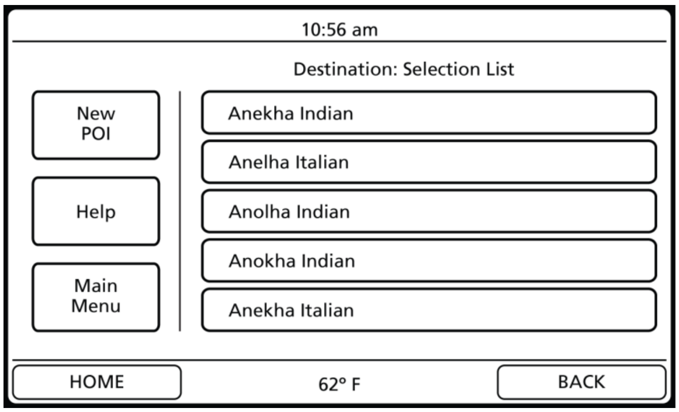


SET A-21

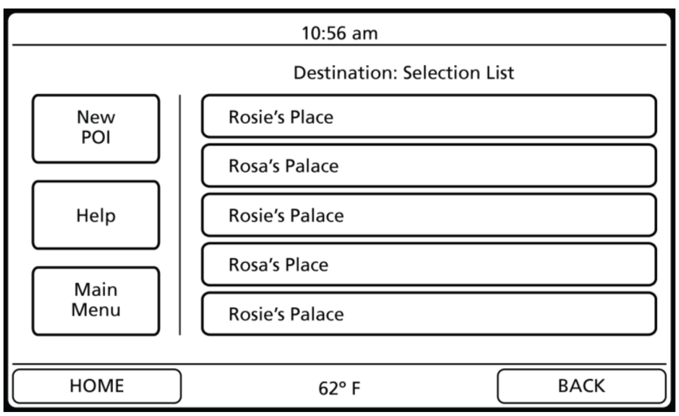
Humanist



SET A-7

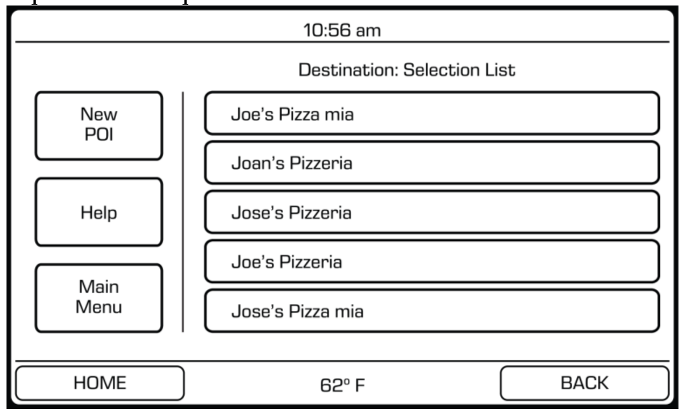


SET A-8

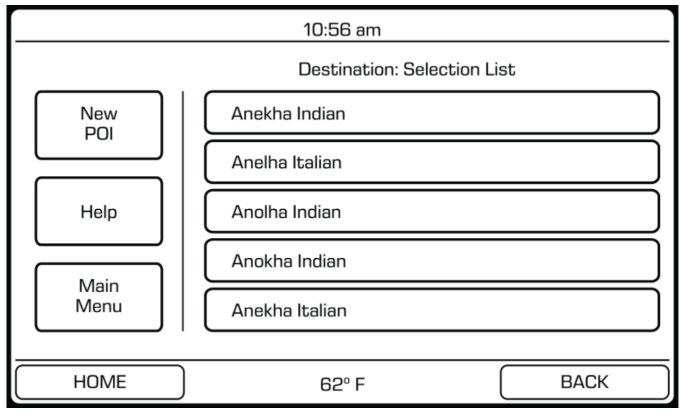


SET A-9

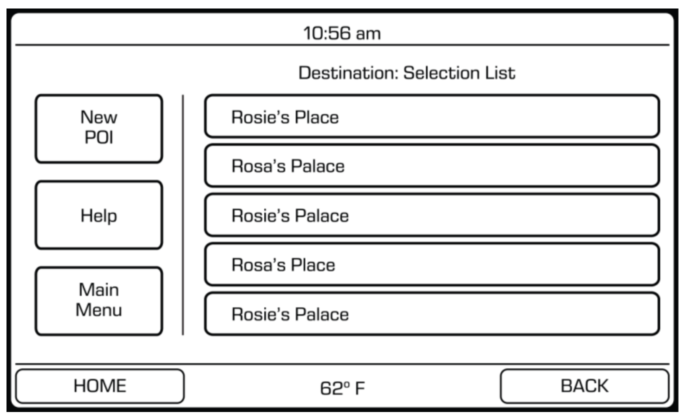
Square Grotesque



SET A-22

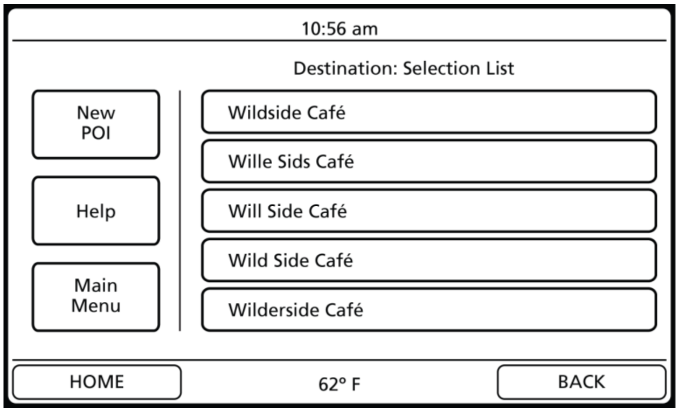


SET A-23

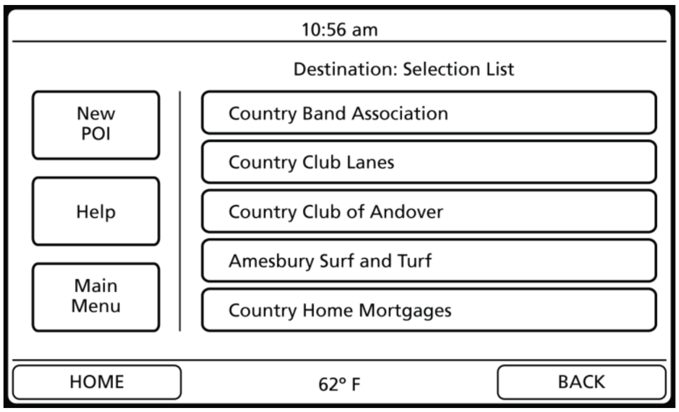


SET A-24

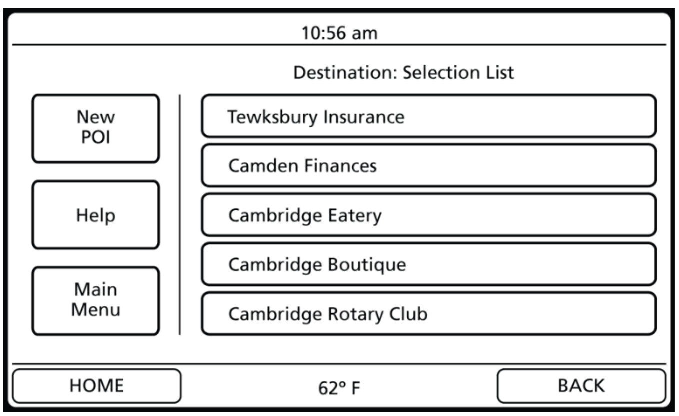
Humanist



SET A-10

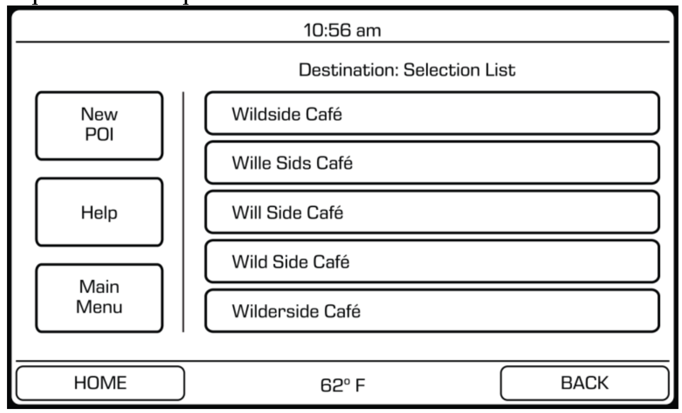


SET A-11

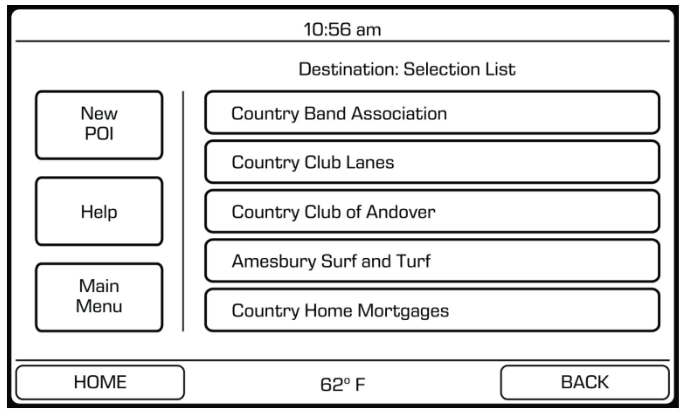


SET A-12

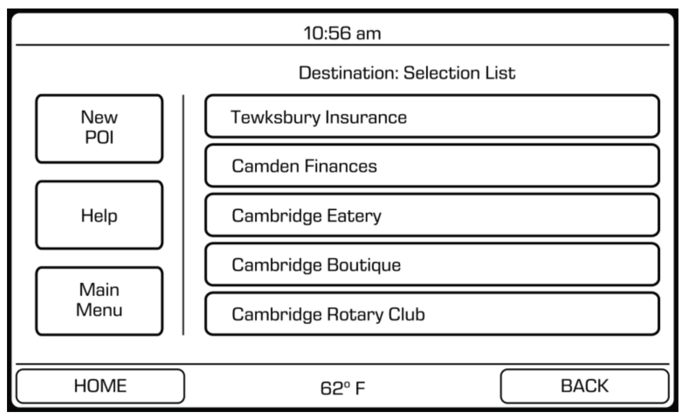
Square Grotesque



SET A-25

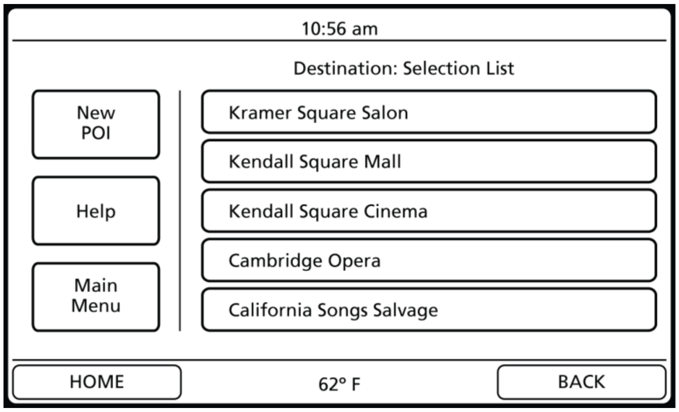


SET A-26

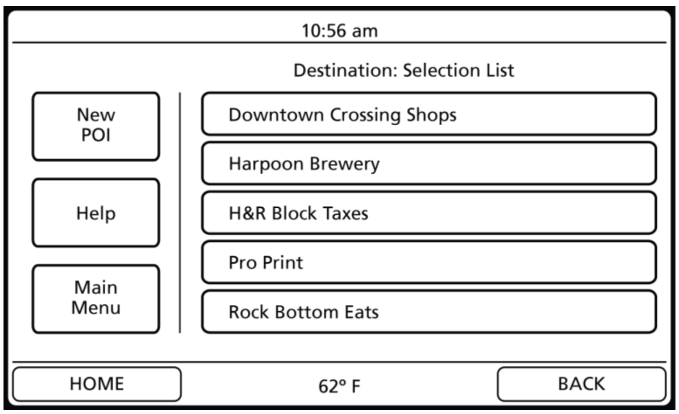


SET A-27

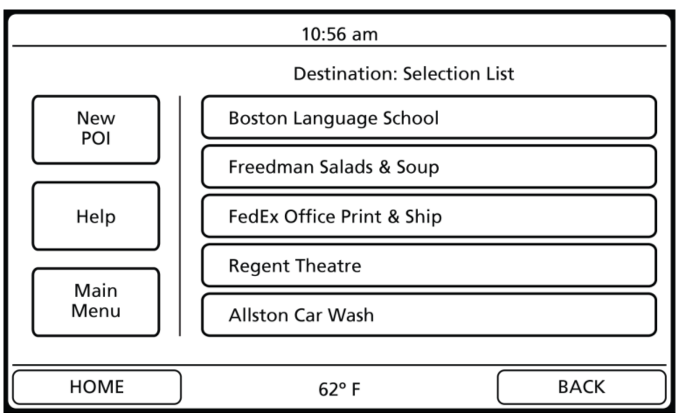
Humanist



SET A-13

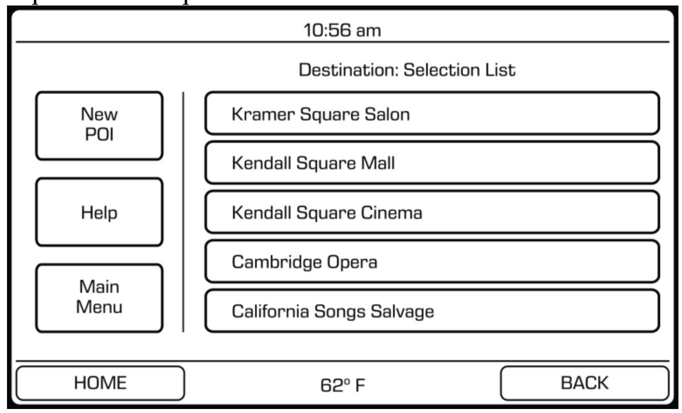


SET A-14

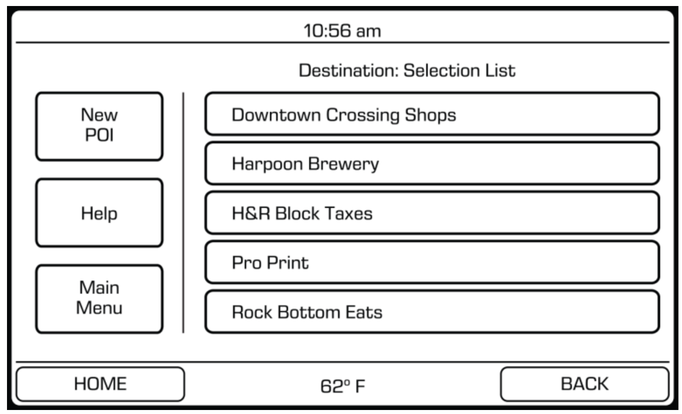


SET A-15

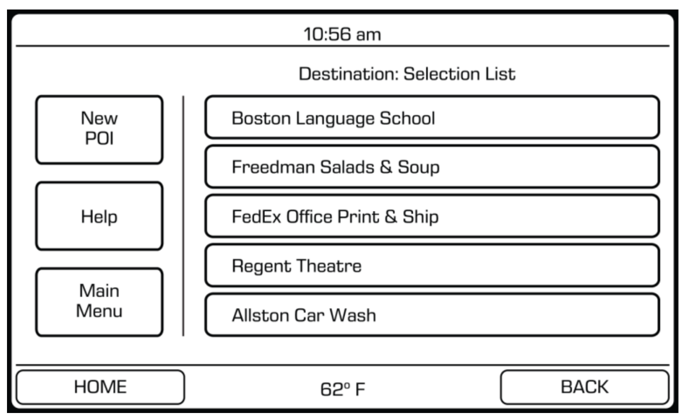
Square Grotesque



SET A-28



SET A-29



SET A-30

Appendix B – Target Items

A total of 30 tasks (target items) were presented to each participant. For any given target item, half the participants were presented with the item in the humanist (H) typeface and half viewed the item in the square grotesque (SG) typeface. The location of targets (lines 1-5 on the menu) was balanced across menu type and typeface (see columns 3 and 6 in the table).

SET A - Study 2012d (April 30)

H	Prompt (target)	I	SG	Prompt (target)	II
(1)	83 Bourne Ave	2	(16)	88 Bourne Ave	3
(2)	78 Allware Street	5	(17)	73 Ailware Street	1
(3)	85 Harnmund Ave	3	(18)	89 Harmmond Ave	5
(4)	33 Boardway Street	4	(19)	32 Boerdway Street	2
(5)	66 Naugotuck Lane	1	(20)	65 Naugotuck Lane	4
(6)	Cheng Cho Restaurant	2	(21)	Chang Sho Restaurant	5
(7)	Jose’s Pizza mia	5	(22)	Joan’s Pizzeria	2
(8)	Anolha Indian	3	(23)	Anokha Indian	4
(9)	Rosa’s Place	4	(24)	Rosie’s Place	1
(10)	Wildside Café	1	(25)	Will Side Café	3
(11)	Music Organization	1	(26)	Restaurant	4
(12)	Restaurant	3	(27)	Financial Service	2
(13)	Shopping Destination	2	(28)	Music Store	5
(14)	Restaurant	5	(29)	Shopping Destination	1
(15)	Movie Theater	4	(30)	Business Service	3

SET B - Study 2012d (April 30)

H	Prompt (target)	II	SG	Prompt (target)	I
(1)	88 Bourne Ave	3	(16)	83 Bourne Ave	2
(2)	73 Ailware Street	1	(17)	78 Allware Street	5
(3)	89 Harmmond Ave	5	(18)	85 Harnmund Ave	3
(4)	32 Boerdway Street	2	(19)	33 Boardway Street	4
(5)	65 Naugotuck Lane	4	(20)	66 Naugotuck Lane	1
(6)	Chang Sho Restaurant	5	(21)	Cheng Cho Restaurant	2
(7)	Joan’s Pizzeria	2	(22)	Jose’s Pizza mia	5
(8)	Anokha Indian	4	(23)	Anolha Indian	3
(9)	Rosie’s Place	1	(24)	Rosa’s Place	4
(10)	Will Side Café	3	(25)	Wildside Café	1
(11)	Restaurant	4	(26)	Music Organization	1
(12)	Financial Service	2	(27)	Restaurant	3
(13)	Music Store	5	(28)	Shopping Destination	2
(14)	Shopping Destination	1	(29)	Restaurant	5
(15)	Business Service	3	(30)	Movie Theater	4

Appendix C – Illumination / Contrast Differences & Eye Adaptation

According to the article published by the American Optometric Association as part of the Aviation Vision studies (http://www.aoa.org/x5352.xml), the eye adaptation mechanism to various lighting conditions include physical, biochemical and neural processes allowing a human visual system (HVS) to successfully function within a wide dynamic range of illumination, with overall changes in brightness levels of as much as 1 billion times covering the dynamic range of almost 30 exposure values (EV). (A single step in exposure value corresponds to a change of illumination level where amount of light entering an eye (or a camera lens) doubles.) In comparison, the majority of today's most advanced digital camera sensors are only capable of recording the dynamic range of up to 14EV. However, the HVS wide dynamic range relies on various eye adaptation mechanisms, where some adaptation processes, such as photochemical regeneration in retina rods and cones, can require significant time – from a few minutes for cones to as much as 30-45 minutes for rods, while other mechanisms such as neural and physical adaptation offer almost instantaneous changes. The ranges of various adaptation mechanism also vary – the changes in neural gain supports light adaptation with approximate factor of 10 (about 3.3EV), while physical adaptation of a pupil size can account for up to 30-fold (~5EV) range in the quantity of light entering the eye.

Acknowledgment – The material in this appendix was developed by Vladimir Levantovsky.

ABOUT THE AUTHORS

Bryan Reimer, Ph.D.

Bryan Reimer is a Research Engineer in the Massachusetts Institute of Technology AgeLab and the Associate Director of the New England University Transportation Center. His research seeks to develop new models and methodologies to measure and understand human behavior in dynamic environments utilizing physiological signals, visual behavior monitoring, and overall performance measures. Dr. Reimer leads a multidisciplinary team of researchers and students focused on understanding how drivers respond to the increasing complexity of the operating environment and on finding solutions to the next generation of human factors challenges associated with distracted driving, automation and other in-vehicle technologies. He directs work focused on how drivers across the lifespan are affected by in-vehicle interfaces, safety systems, portable technologies, different types and levels of cognitive load. Dr. Reimer is an author on over 70 peer reviewed journal and conference papers in transportation. Dr. Reimer is a graduate of the University of Rhode Island with a Ph.D. in Industrial and Manufacturing Engineering.

reimer@mit.edu
(617) 452-2177
http://web.mit.edu/reimer/www/

Bruce Mehler, M.A.

Bruce Mehler is a Research Scientist in the Massachusetts Institute of Technology AgeLab and the New England University Transportation Center, and is the former Director of Applications & Development at NeuroDyne Medical Corporation. He has an extensive background in the development and application of non-invasive physiological monitoring technologies and research interests in workload assessment, individual differences in response to cognitive demand and stress in applied environments, and in how individuals adapt to new technologies. Mr. Mehler is an author of numerous peer reviewed journal and conference papers in the biobehavioral and transportation literature. He continues to maintain an interest in health status and behavior from his early work in behavioral medicine. He received an MA in Psychology from Boston University and a BS degree from the University of Washington.

bmehler@mit.edu
(617) 253-3534
http://agelab.mit.edu/bruce-mehler

Joseph F. Coughlin, Ph.D.

Joseph F. Coughlin is founder and Director of the Massachusetts Institute of Technology AgeLab and Director of the US Department of Transportation's Region I New England University Transportation Center. He served as the Chair of the Organization for Economic Cooperation & Developments 21-nation Task Force on Technology and Transportation for Older Persons, is a member of the National Research Council's Transportation Research Board Advisory Committee on the Safe Mobility of Older Persons. He served as a Presidential appointee to the White House Conference on Aging and has consulted or served on technology and design boards for BMW, Daimler, Nissan, and Toyota. Prior to joining MIT, Dr. Coughlin led the transportation technical services consulting practice for EG&G a global Fortune 1000 science and technology firm.

coughlin@mit.edu
(617) 253-4978
http://www.josephcoughlin.com/

About the New England University Transportation Center & MIT Center for Transportation & Logistics

The New England University Transportation Center is a research, education and technology transfer program sponsored by the US Department of Transportation. Together the faculty, researchers and students sponsored by the New England Center conduct work in partnership with industry, state & local governments, foundations and other stakeholders to address the future transportation challenges of aging, new technologies and environmental change on the nation's transportation system. For more information about the New England University Transportation Center, visit utc.mit.edu. For more information about the US Department of Transportation's University Transportation Centers Program, please visit utc.dot.gov. The New England Center is based within MIT's Center for Transportation & Logistics, a world leader in supply chain management education and research. CTL has made significant contributions to transportation and supply chain logistics and helped numerous companies gain competitive advantage from its cutting edge research. For more information on CTL, visit ctl.mit.edu.

About the AgeLab

The Massachusetts Institute of Technology AgeLab conducts research in human behavior and technology to develop new ideas to improve the quality of life of older people. Based within MIT's Engineering Systems Division and Center for Transportation & Logistics, the AgeLab has assembled a multidisciplinary team of researchers, as well as government and industry partners, to develop innovations that will invent how we will live, work and play tomorrow. For more information about AgeLab, visit agelab.mit.edu.