

Assessing the impact of typeface design in a text-rich automotive user interface

Bryan Reimer^{a*}, Bruce Mehler^a, Jonathan Dobres^a, Joseph F. Coughlin^a, Steve Matteson^b, David Gould^b,
Nadine Chahine^b and Vladimir Levantovsky^b

^aAgeLab, Massachusetts Institute of Technology, Cambridge, MA, USA; ^bMonotype Imaging Inc., Woburn, MA, USA

(Received 20 September 2013; accepted 26 June 2014)

Text-rich driver–vehicle interfaces are increasingly common in new vehicles, yet the effects of different typeface characteristics on task performance in this brief off-road based glance context remains sparsely examined. Subjects completed menu selection tasks while in a driving simulator. Menu text was set either in a ‘humanist’ or ‘square grotesque’ typeface. Among men, use of the humanist typeface resulted in a 10.6% reduction in total glance time as compared to the square grotesque typeface. Total response time and number of glances showed similar reductions. The impact of typeface was either more modest or not apparent for women. Error rates for both males and females were 3.1% lower for the humanist typeface. This research suggests that optimised typefaces may mitigate some interface demands. Future work will need to assess whether other typeface characteristics can be optimised to further reduce demand, improve legibility, increase usability and help meet new governmental distraction guidelines.

Practitioner Summary: Text-rich in-vehicle interfaces are increasingly common, but the effects of typeface on task performance remain sparsely studied. We show that among male drivers, menu selection tasks are completed with 10.6% less visual glance time when text is displayed in a ‘humanist’ typeface, as compared to a ‘square grotesque’.

Keywords: typography; driving; distraction; display; legibility

1. Introduction

It is important to provide a driver with a visual user interface in which controls can be rapidly identified and information content easily read. 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 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 navigation systems, followed by the emergence of in-dash embedded 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 the content cannot be deduced on the basis of a memorised location. As a result, legibility, the degree to which individual characters are understandable or recognisable, is of increasing significance as a fundamental consideration in human–machine interface (HMI) design in automobiles. Considerable investment has already been placed on legibility in other areas of the automotive operating environment. For instance, the Clearview typeface was developed and tested to specifically enhance legibility of positive contrast roadway signage (Chrysler, Carlson, and Hawkins 2003; Garvey, Pietrucha, and Meeker 1997, 1998; Holick et al. 2006). The legibility benefits of Clearview were shown for night-time legibility with a high-brightness reflective material, but were not present in a study of negative contrast signs in uppercase (Chrysler, Carlson, and Hawkins 2003; Holick et al. 2006). However, the effect of typeface legibility on in-vehicle task performance appears largely unexamined in the published literature.

1.1 Background

From the perspective of typographic designers, factors that influence legibility can be thought of as *extrinsic* and *intrinsic* (Bigelow and Matteson 2011). Extrinsic factors are psychophysical considerations such as size, illumination, contrast, polarity and colour. 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

An earlier presentation of this work appears as a white paper (Reimer, Mehler, and Coughlin 2012).

*Corresponding author. Email: reimer@mit.edu

speed (Legge and Bigelow 2011; Tinker 1963), and this has been confirmed in automotive-oriented research (Cai and Green 2005; Fujikake et al. 2007; O'Day and 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' sans-serif typefaces with strongly differentiated forms and contours (such as Frutiger®) may be more legible in the context of brief glances than typefaces that employ tighter spacing and more uniform letter shapes, including geometric sans-serifs (such as Century Gothic™), 'grotesque' sans-serifs (such as Helvetica®) and 'square grotesque' sans-serifs (such as Eurostile®) (Schulz-Anker 1970).

1.2 Typeface considerations

The present study aims 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 exemplar typefaces with clear differences in legibility, for the purposes of this investigation, only typefaces that are representative of ones currently in use in the automotive industry for vehicle displays were considered. A comparison was targeted to assess empirical advantages of a typeface design that, based upon expert opinion, has a recognised level of enhanced legibility in automotive applications. 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 production vehicles by one manufacturer, and a humanist style typeface, Frutiger, which is highly consistent with a typeface used by another. Typographers on the research team believed that the Frutiger typeface contained a number of intrinsic characteristics that would be expected to improve legibility on in-vehicle display screens.

As illustrated in Figure 1, 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.

In contrast, humanist genre typefaces are considered to be more legible because of the following:

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

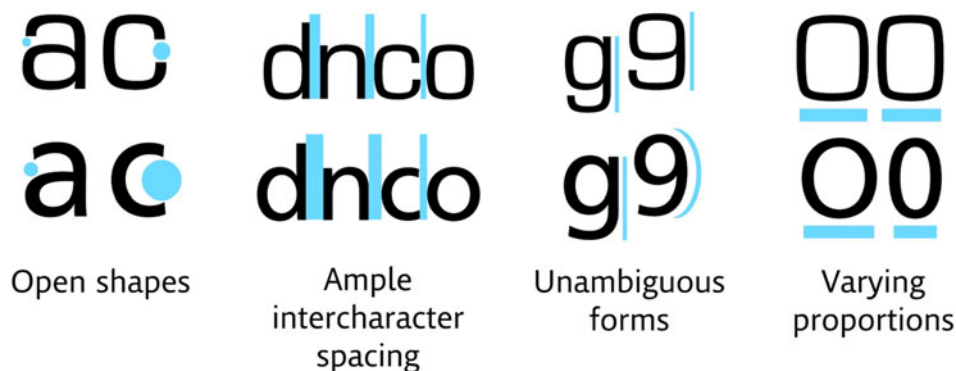


Figure 1. The top-line characters are a square grotesque design (Eurostile) and the bottom line a humanist design (Frutiger), highlighting various characteristics thought to improve legibility.

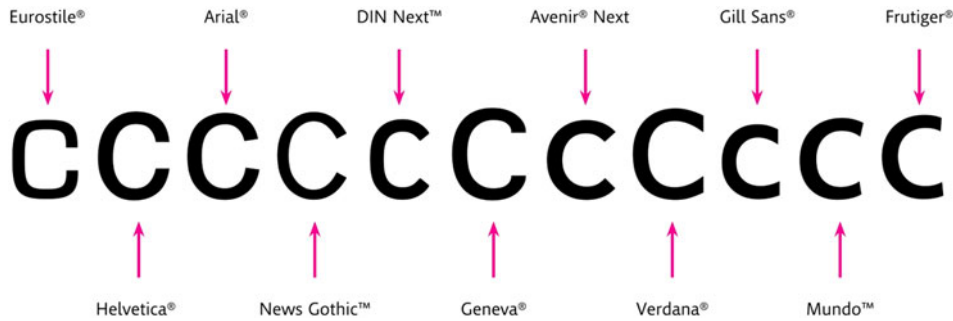


Figure 2. This illustration begins on the left with a very closed aperture of a square grotesque design and progresses to the right with more open apertures found in the humanist sans-serif genre.

Note: The letters are all displayed at 100 point – no adjustments have been made to regularise their height.

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; the distance between the terminations works to avoid the meshing together of forms and keeps these features easily identifiable (Bouma 1970, 1973; 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 fine-tuning of space between letterforms protects against crowding (Bouma 1970, 1973; Pelli et al. 2007) and the merging of characters into one another, therefore possibly increasing the perceptual span and resulting in better legibility. Although the space between individual letters may seem to be an extrinsic legibility factor, it is important to keep in mind that modern typefaces may contain detailed letter-pair specific positioning adjustments (‘kerning’, in typographic parlance), and thus letter spacing is a prominent intrinsic component of type design. Letter identification is facilitated when there are a lower number of shapes that can be confused with one another (Attneave and Arnoult 1956), which should be the case for humanist versus 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.

As noted previously, character size is a significant variable underlying legibility (Cai and Green 2005; Fujikake et al. 2007; O’Day and Tijerina 2011). While some typeface designers have argued for using the height of the lowercase ‘x’ to characterise the physical size of typefaces (x-height) (Bigelow and Matteson 2011; Legge and Bigelow 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, scaled versions of a humanist typeface (Frutiger) and square grotesque typeface (Eurostile) were constructed 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, therefore conforming to automotive industry standards (see Figure 3).

Figure 4 highlights some of the subtle differences that may appear across typeface designs, even with 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 lowercase 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 (see Figure 5) and has rather square proportions, while the humanist has upright, rectangular proportions. 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 normalised versions of the two typefaces was that the overall areas of the counters, or insides of the letters, were very close



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.

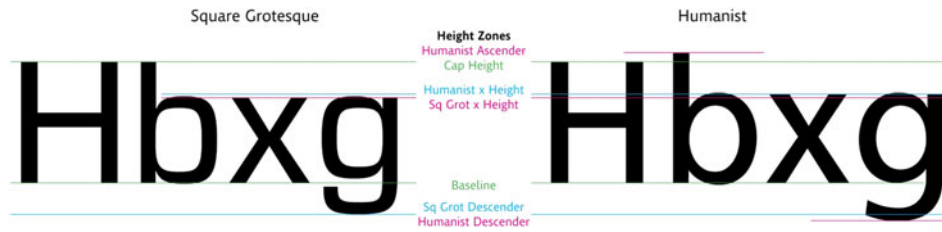


Figure 4. Subtle differences in the heights of other characters may be present when fonts are normalised around the height of the capital 'H' reference standard. The square grotesque typeface (Eurostile) is on the left and humanist typeface (Frutiger) is on the right.

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 (see Figures 1–5). These differences would be challenging for the untrained observer to consciously detect at the sizes typically used in 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.

Stroke width is also known to impact legibility (O'Day and Tijerina 2011). For any character of a given height, very thin strokes will be relatively difficult to read at a glance; increasing their thickness will improve legibility up to a point, and then further thickening will begin to obscure legibility. O'Day and Tijerina (2011) examined stroke widths of 7%, 9%, 20%, 28% and 30% of character height. For the combinations of character height, character width and stroke 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 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' indicates the relatively larger difference in the fonts in terms of character width, with the square grotesque font being noticeably wider. Taking the physical dimensions and resolution of the target display used in this study into account, the square grotesque font has an average stroke width of 2.87 pixels, and the humanist font has an average stroke width of 3.0 pixels. This represents a 0.13 pixel difference in vertical stroke width between the two fonts. These fonts were used to render the text components of a predetermined set of anti-aliased stimulus images. Although anti-aliasing smoothes fractional pixel values by introducing subtle shades of grey into character strokes, the resulting difference in luminosity between the two fonts is negligible. Moreover, attempts to rigorously equalise these minute smoothing effects would necessitate structural modifications to the fonts, in essence creating customised versions that no longer represent their intrinsic qualities as designed. Anti-aliasing effects may alter a stroke's perceived width if a stroke happens to fall across a

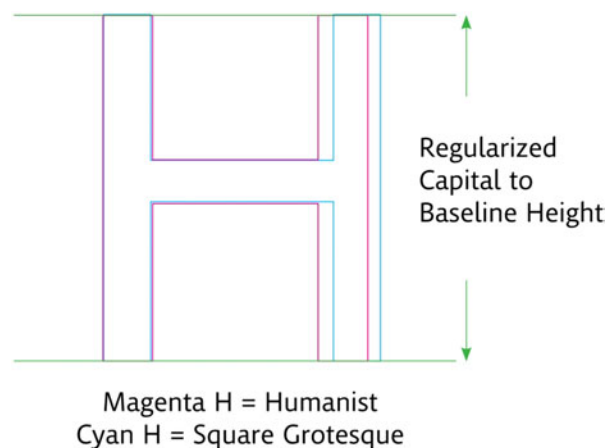


Figure 5. The graphic compares the relative difference between the two typefaces studied in stroke width (difference between the cyan and magenta lines on the left side of the 'H') and character width (difference on the right side of the 'H').

pixel boundary, regardless of any special effort to fine-tune the glyphs themselves. Therefore, a decision was made to leave the stroke widths of the two fonts as they are normally proportioned for these families.

1.3 Research intent

This report presents 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 than 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 behaviour 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 behaviour were robust and not a chance finding. The two studies utilised separate but demographically similar subject samples.

2. Study 1: 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. We elected to begin this work considering participants 35 and older since vision typically begins to degrade (acuity, contrast and luminosity thresholds, noise, etc.) as individuals approach middle age and it was taken as a working hypothesis that subtle differences in legibility might have more impact in such a sample and thus be more observable. Participants were required to be active, experienced drivers, based on having held a valid driving licence for 3 + years and self-reported average driving frequency of three or more times a week. Additional requirements consisted of being in self-reported good health for one's age, being fully comfortable speaking and reading English, and having no major illness resulting in hospitalisation in the past 6 months. A diagnosis of Parkinson's disease or other neurological problems was also an exclusion criterion, as these disorders can impact 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' (2.44 m by 1.83 m) projection screen was positioned 76" (1.93 m) in front of the mid-point of the windshield and provided approximately a 40° view of the virtual world at a resolution of 1024 × 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 were 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 behaviour and interaction with a 7" LCD touch screen interface (model CTF400L; cartft.com, Reutlingen, Germany). Previous validation work has established a high correspondence in the allocation of visual attention in relation to interaction with visual manipulative HMIs (Wang et al. 2010) between this simulator configuration and on-road behaviour.

The CTF-400-L 7" display is 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 × 480 pixels. The touch screen was mounted on top of the centre console, placing it at a distance of approximately 700 mm from the eyes of the average participant (see Figure 6). The background illumination of the touch screen display was set to its *bright* mode to ensure a high contrast between the background and text.

The simulation scenario consisted of a divided highway with two lanes in each direction plus a 2' (0.61 m) shoulder on each side of the roadway. Lane width was 15' (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 slowdowns. 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 menu/list selection display template was developed from elements commonly employed across various automotive HMI display screens without specifically modelling a particular commercial implementation. The key element



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.

in this study was a five-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 selection list always matched the font used for the other elements.

The two fonts compared using this display were specialised TrueType versions of Frutiger® (humanist) and Eurostile® (square grotesque). The native versions of these fonts were adjusted so that capital letter heights were equivalent across the two fonts to conform to ISO 15008, i.e. defined using the capital letter ‘H’ as a reference. The height of the *H* character for both typefaces on the touch screen was 4 mm.

The effective size of the character depends on the distance of the driver’s eyes from the screen, and is typically represented as the angle subtended by the character on the retina. 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 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 top end of the acceptable range.

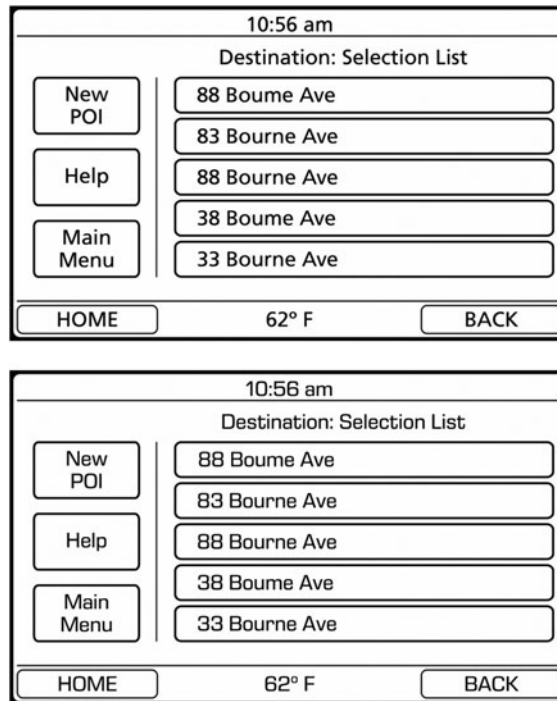


Figure 7. Menu screen in a (a) humanist font and (b) square grotesque font.

For the simulated display, high-resolution (3334 × 2000 pixel; 300 dpi) screen images with font size of 27 pt were first created in Adobe Illustrator. The files were subsequently converted to bitmap (.bmp) format using the *Type Optimised (Hinted)* anti-aliasing and 32-bit *depth* settings. These images were then reduced to 1280 × 768 pixel resolution, 96 dpi, bitmap files. The CTF-400-L display hardware downscaled these images to the screen's native 800 × 480 pixel resolution. A set of stimuli in each typeface appear in [Figure 7](#).

Three types of menu lists were presented: addresses, restaurant names and content searches. Addresses all consisted of leading two-digit numbers, a name and a descriptor such as 'Street' or 'Ave' (see [Figure 7](#)). Restaurant names were all two to three 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 two to four words in length and did not deliberately employ visually challenging character combinations. For example, one content search task requested locating a financial services company out of a list of business names.

Five menu lists, each with unique content, were created for each task type (5 × 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. The full set of menu stimuli are reproduced in both typefaces and the target items appear in the appendixes of Reimer, Mehler, and Coughlin (2012). Targets were selected such that each line position was used only once per list type for a given font. Two variants of the target location assignments were created (A and B) such that a given item and location combination presented in the humanist font in form A was presented in the square grotesque font in form B. Form variant (A or B) and gender were counterbalanced across the sample.

A prompt screen cued participants to search for a specific item on the subsequently presented menu (see [Figure 8](#)). A Times New Roman font was employed and the target was presented in capital letters to minimise 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. 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.

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 penalised \$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 (Mehler et al. 2009; Reimer et al. 2006). They reinforced text presented in the informed consent form where it was stated that the monetary award for performing the secondary tasks could be up to \$10; in reality, all participants received equal compensation regardless of performance.

A brief drive of 2.65 miles (approximately 4 minutes) provided a degree of familiarisation with the simulator environment. Participants were then instructed to pull over to the side of the virtual highway and stop the car. Participants

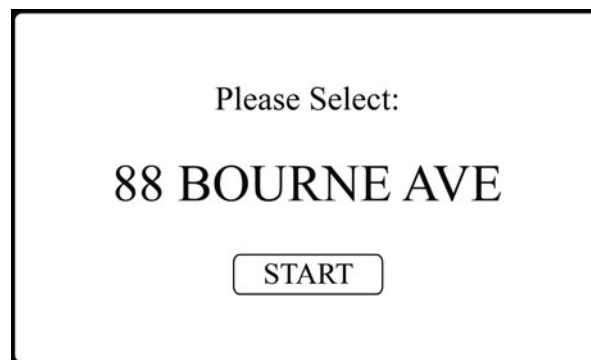


Figure 8. Prompt screen presented using a Times New Roman font.

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 three examples (one each of an *address*, *restaurant selection* and *content search* task) was presented to each participant, and the RA could present up to two additional examples to ensure that participants understood the tasks.

At the conclusion of the practice trials, an audio recording reiterated the instruction to balance accurate completion of the secondary tasks with the demands of safe driving. Participants were then prompted to resume driving. Shortly after highway speed was regained, the main menu selection tasks began. An automated programme presented the 30 task stimuli in a randomised order. The presentation intervals between the end of one task and the prompt for the next varied randomly between 20 and 40 s.

A post-experimental questionnaire assessed symptoms of negative experiences in the simulator using the Simulator Sickness Questionnaire (SSQ) (Kennedy et al. 1993).

2.5 Data reduction and analysis

Eye data were processed following ISO standards (ISO 15007-1 2002; ISO 15007-2 2001) to determine the time spent looking at the touch screen, number of inspections of the touch screen and counts of glances greater than 2.0 s. The 2.0-s value corresponds to guidelines suggested by the Alliance of Automobile Manufacturers (Driver Focus-Telematics Working Group 2006) and the Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices (National Highway Traffic Safety Administration 2013), and has been identified as a relevant risk threshold (Klauer et al. 2006). Total response time was computed as the total time between the press of the 'Start' button on the prompt screen and 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. This resulted in a $2 \times (2 \times 3)$ design with gender treated as a between-subject variable and font type and content type treated as within-subject variables.

Primary comparisons were computed using repeated-measures ANOVA (general linear model). Owing to non-normality of the response time data, ANOVA models were calculated based on the grand ranks of the observations (Conover and Iman 1981; Iman 1974). Post-hoc comparisons were computed using paired Wilcoxon signed-rank tests. All statistical computations were conducted using R (R Core Team 2012). 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. Note that although we include menu type as a factor in our model, in the interest of clarity we choose not to report its effects here as the present paper deals with the effect of typeface on secondary task completion while driving a motor vehicle. A breakdown of the data by menu type is available in a technical report (Reimer, Mehler, and Coughlin 2012). Although a main effect of menu type was frequently apparent, this factor did not interact with typeface.

3. Study 1: results

3.1 Sample characteristics

A total of 51 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 (one male) who completed the simulation were excluded from the analysis. Three cases (one 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 behaviour 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 behaviour 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 s were excluded as outliers (response times in the remainder of the sample all averaged less than or equal to 8.3 s). 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 $F(1,40) = 0.05, p = 0.82$. Male and female participants did not differ in total or subscales of the SSQ (p -values > 0.05).

Corrected visual acuity measured using the Snellen eye chart did not differ between male and female participants ($F(1,40) = 1.05, p = 0.31$). 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 behaviour

Measures of interest summarised by gender and typeface design appear in Table 1. A main effect of typeface design on total response time appears in the model ($F(1,40) = 9.61, p = 0.035$) 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 ($F(1,40) = 7.91, p = 0.008$). Decomposing the interaction effect, separate tests assessing the effect of typeface design were developed for male and female participants. A main effect of typeface design appears for male participants ($V = 4.04, p < 0.001$, Wilcoxon signed-rank test). Men responded to menus in the humanist typeface at an average of 5.19 s (SD = 1.64). Responses to menus with the square grotesque typeface took 6.00 s (SD = 1.95) or 15.7% longer than the humanist typeface. On the other hand, female participants response times were not significantly different across the two typefaces ($V = 1051, p = 0.769$), with menus set in the humanist typeface requiring an average of 5.53 s (SD = 1.81) per response, compared to menus set in the square grotesque typeface taking an average of 5.45 s (SD = 1.78) per response.

While not statistically significant ($F(1,40) = 1.05, p = 0.311$), a nominal difference 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 was not a predictor of error rates.

3.3 Glance behaviour

Typeface significantly impacts total glance time to the display ($F(1,40) = 9.70, p = 0.003$). An interaction between typeface design and gender ($F(1,40) = 4.22, p = 0.046$) is also evident. The main effect of typeface design is best considered in relation to the significant interaction between typeface design and gender. As illustrated in Figure 9, the main effect of typeface design appears to be driven by the male participants. Statistically, this is assessed by separate tests for the male and female participants. For male participants, there was a main effect of typeface design ($V = 569.5, p = 0.003$). This corresponds to a 0.47 s 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 glance time appears for the female participants ($V = 934.5, p = 0.617$).

Consistent with the total allocation of visual attention to the display, typeface significantly impacts the average number of glances to the display ($F(1,40) = 34.96, p < 0.001$). In addition, typeface and gender interact significantly ($F(1,40) = 8.15, p = 0.007$). Paired Wilcoxon tests show that typeface design significantly impacted the total number of glances made by male participants ($V = 395, p < 0.001$), but not female participants ($V = 839.5, p = 0.904$) (see Figure 10). 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 0.39 more glances per response with the square grotesque typeface. In other words, the square grotesque typeface required 14% more glances than the humanist typeface. In contrast, among women there was virtually no difference in the number of glances between the two typefaces.

Table 1. Measures of interest for Study I summarised by gender and typeface.

		Response time (s)	Error rate (%)	Glance time (s)	Glance frequency	Glances > 2.0 s
Female	Hum	5.53 (2.13)	18.8 (19.2)	3.51 (1.09)	2.94 (1.11)	0.366 (0.436)
	SG	5.45 (1.96)	22.2 (25.0)	3.52 (1.05)	2.96 (1.13)	0.327 (0.405)
Male	Hum	5.19 (2.00)*	17.2 (19.0)	3.86 (1.36)*	2.77 (1.19)*	0.495 (0.433)
	SG	6.00 (2.45)*	21.4 (21.8)	4.33 (1.71)*	3.16 (1.36)*	0.545 (0.484)

Note: Hum = humanist and SG = square grotesque; standard deviations are in parenthesis. *Significant difference between paired comparisons with a gender group.

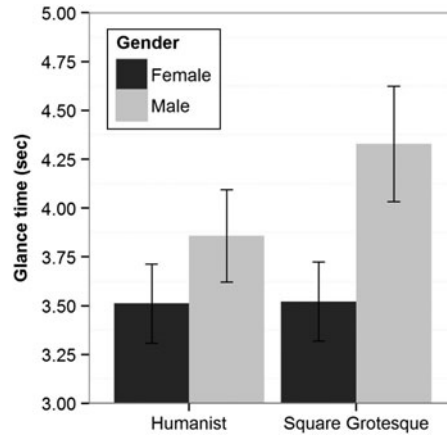


Figure 9. Total glance time to the display screen in Study I by gender and typeface style.

The number of glances greater than 2.0 s did not differ significantly between typeface ($F(1, 40) = 0.002, p = 0.958$) or gender ($F(1, 40) = 3.03, p = 0.089$), and these factors did not interact ($F(1, 40) = 1.27, p = 0.267$).

4. Study II: methods

Study II was conducted and analysed using the driving simulator and methodologies described for Study I except for one change. The simulator is located in a dimly lit room, and a participant's main field of view is defined by the virtual roadway screen. Compared to typical outdoor daylight driving conditions, the driving simulator environment offers significantly reduced levels of ambient 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 behaviour away from a roadway (toward the HMI), and because the tasks were presented using a separate display mounted on top of the centre console within a driver's main field of view, we wished to evaluate whether a somewhat less bright display illumination level closer to the ISO 15008 standard would impact glance performance. Therefore, the setting on the CF-400-L touch screen interface was changed from the *bright* setting used in Study I to the lower contrast *normal* setting for Study II.

In addition to the primary analysis of data in Study II, a statistical comparison is provided across Studies I and II. This extended analysis was conducted to provide an assessment of how changes made to the contrast of the display impacted drivers' behaviour. The extended $2 \times 2 \times (2 \times 3)$ design considers contrast and gender as between subject variables and font type and menu type as within subject variables. Once again, statistical effects due to menu type do not interact with typeface and therefore are not reported here in the interest of clarity.

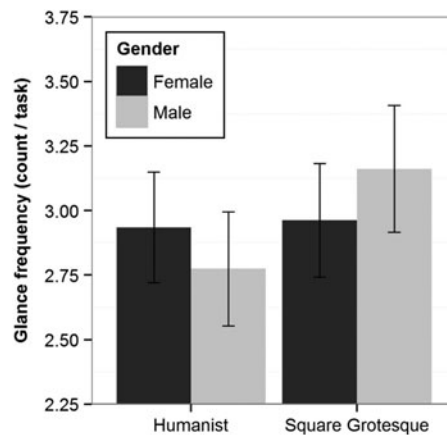


Figure 10. Glance frequency to the display screen in Study I by gender and typeface style.

5. Study II: results

5.1 Sample characteristics

A total of 46 participants, who did not participate in Study I, were recruited for 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 (one 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 were 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. 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 ($F(1,38) = 0.13, p = 0.72$). Male and female participants did not statistically differ in total or subscales of the SSQ (p -values > 0.05).

Corrected visual acuity measured using the Snellen eye chart in Study II did not differ between male and female participants ($F(1,38) = 0.70, p = 0.41$). 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).

5.2 Task response behaviour

Measures of interest summarised by gender and typeface design appear in Table 2. Consistent with Study I, a main effect of typeface design on response time appears ($F(1,38) = 8.18, p = 0.007$) in which there was an 8.7% faster 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 ($F(1,38) = 1.12, p = 0.296$).

Post-hoc tests show that typeface significantly impacted response time for the male ($V = 613, p = 0.026$) but not the female ($V = 669.5, p = 0.071$) participants. Men responded 0.66 s (10.6%) faster to menus in the humanist typeface. Comparing the two studies, response times to the higher contrast condition in Study I ($M = 5.54, SD = 1.70$) are nominally 1 s (17.9%) faster than the lower contrast condition in Study II ($M = 6.53, SD = 2.70$), though the effect is not statistically significant ($F(1,78) = 1.41, p = 0.239$). A main effect of typeface design ($F(1,78) = 13.01, p < 0.001$), as well as an interaction effect between gender and typeface ($F(1,78) = 6.60, p = 0.012$), appears. As previously reported, the effect of typeface is best considered in relation to the interaction with gender. Across the sample, male participants responded to humanist typeface ($M = 5.71, SD = 2.00$) 0.74 s (13.0%) faster than square grotesque typeface ($M = 6.45, SD = 2.38$), while women's response time only differed by 0.16 seconds (2.7%), $M = 5.89, SD = 2.36$ for humanist and $M = 6.05, SD = 2.69$ for square grotesque. Taken together, these results provide strong evidence for an effect of typeface on reaction time among males.

Along the lines of the results of Study I, the effect of typeface on error rate trended toward significance ($F(1,38) = 3.33, p = 0.076$), with menus drawn in the humanist typeface appearing to have lower error rates ($M = 15.88\%, SD = 11.22$) than 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 in Study I. Participants' gender was not a predictor of error rates ($F(1, 38) = 0.72, p = 0.402$).

Considering the data across studies, error rates were not significantly affected by contrast ($F(1,78) = 0.359, p = 0.551$), 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 typeface appears ($F(1,78) = 4.71, p = 0.033$). 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 ($M = 21.1\%, SD = 16.0$).

Table 2. Measures of interest for Study II summarised by gender and typeface.

		Response time (s)	Error rate (%)	Glance time (s)	Glance frequency	Glances > 2.0 s
Female	Hum	6.26 (2.83)	0.174 (0.124)	3.61 (1.22)	3.18 (1.24)	0.301 (0.351)
	SG	6.69 (3.33)	0.217 (0.132)	3.73 (1.31)	3.29 (1.31)	0.313 (0.292)
Male	Hum	6.25 (2.23)*	0.143 (0.100)	4.27 (1.14)*	3.19 (1.15)	0.451 (0.365)
	SG	6.91 (2.74)*	0.192 (0.137)	4.66 (1.41)*	3.41 (1.49)	0.511 (0.416)

Note: Hum = humanist and SG = square grotesque; standard deviations are in parenthesis. *Significant difference between paired comparisons with a gender group.

5.3 Glance behaviour

Typeface design significantly impacts total glance time to the display ($F(1,38) = 4.42, p = 0.042$), with overall response times to menus in the humanist typeface ($M = 3.94, SD = 1.21$) being 0.26 s (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 failed to reach statistical significance ($F(1,38) = 1.21, p = 0.278$); however, a main effect of gender does appear ($F(1,38) = 4.52, p = 0.040$). As illustrated in Figure 11, 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 9), in Study II female participants' glance times tend to be slightly lower (3.3%) with the humanist typeface compared to the square grotesque typeface. Conversely, in Study I the mean glance times for women were essentially the same across typefaces. The 9.1% higher visual demand for the square grotesque typeface observed among the men in this study is consistent with the result from Study I.

Between the two studies, contrast appears to have a modest but non-significant impact on glance time ($F(1,78) = 1.08, p = 0.302$). Main effects of typeface ($F(1,78) = 12.33, p < 0.001$) and gender ($F(1, 78) = 6.79, p = 0.010$) are evident. A significant interaction between gender and typeface ($F(1,78) = 5.32, p = 0.024$) appears, 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 0.43 s (10.6%) longer than menus in the humanist typeface ($M = 4.06, SD = 1.12$). Female participants glance time showed a more modest 0.06-s difference (1.7%) between the square grotesque ($M = 3.62, SD = 1.12$) and humanist ($M = 3.56, SD = 1.07$) typefaces.

Figure 12 shows the average frequency of glances to the display screen by gender and typeface. As compared to Study I, the effect of typeface on the average number of glances to the display is more modest ($F(1,38) = 2.90, p = 0.097$). As illustrated in Figure 12, gender does not influence the pattern of response as markedly as in Study I ($F(1, 38) = 0.01, p = 0.908$).

Overall glance frequency did not differ significantly between the two studies ($F(1,78) = 1.15, p = 0.286$). Across the two studies, a main effect of typeface ($F(1,78) = 11.65, p = 0.001$) and an interaction between typeface and gender appear ($F(1,78) = 4.95, p = 0.029$). Breaking the effect of typeface down across gender, male participants glanced to the menus in the square grotesque typeface ($M = 3.28, SD = 1.30$) 0.31 times more per task (10.1%) than menus in the humanist typeface ($M = 2.98, SD = 1.09$). Female participants' glances to menus were more consistent across typeface, with square grotesque ($M = 3.12, SD = 1.16$) and humanist ($M = 3.05, SD = 1.11$) typefaces differing by only 0.07 glances per task (2.3%). Although the observed effects are stronger in the combined sample, observations between the two studies are consistent.

Consistent with Study I, the average number of glances greater than 2 s per response in Study II was not significantly affected by typeface ($F(1,38) = 3.07, p = 0.088$) or gender ($F(1,38) = 2.32, p = 0.136$). Across the two studies, a main effect of gender ($F(1,78) = 5.41, p = 0.023$) is associated with men ($M = 0.50, SD = 0.39$) exhibiting 0.17 (51.5%) more glances greater than 2 s per response than women ($M = 0.33, SD = 0.34$). The effect of typeface was not significant ($F(1,78) = 1.71, p = 0.195$), nor did typeface interact with gender in this case ($F(1,78) = 1.10, p = 0.297$).

6. 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

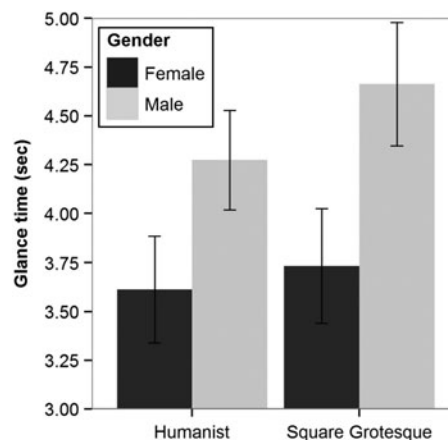


Figure 11. Glance time to the display screen in Study II across all three task types by typeface design for male and females.

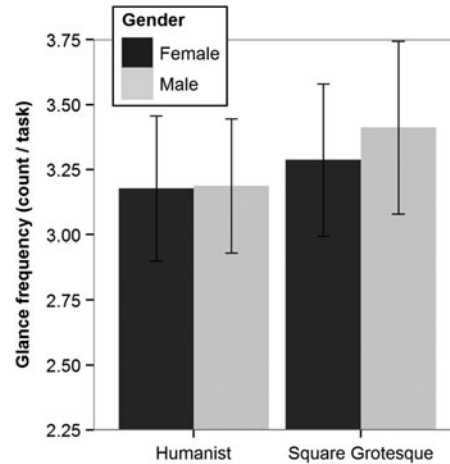


Figure 12. Glance frequency to the display screen in Study II across all three menu types.

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. Optimised 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 2006), and recent events have resulted in the development of *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* (National Highway Traffic Safety Administration 2013). 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 optimising 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 demand characteristics.

This exploratory work demonstrates that the selection of one typeface design utilised in vehicle systems over another resulted in a 10.6% reduction 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. There was no significant 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 was reduced, women showed modest improvements in response time and glance frequency when using the humanist typeface, similar to the pattern seen in men. The lack of an effect of contrast between the two studies confirms that the experimental design reveals *intrinsic* typeface characteristics, as opposed to *extrinsic* characteristics such as contrast and colour.

The apparent gender differences observed in this sample were unexpected and are, to the best of our knowledge, novel. 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. Brain imaging studies suggest that men and women may process information in substantially different ways at the neurological level. However, the relationship between patterns of brain activation and behavioural performance is complex and somewhat difficult to interpret, particularly as it relates to gender differences (Bell et al. 2006; Speck et al. 2000). We speculate that the gender differences observed in the present study may not arise solely or necessarily from legibility effects *per se*, but from a confluence of multiple factors that are not clearly understood. For example, it has been documented that in many, but not all, situations, females exhibit better reading performance than males (Kang, Wang, and Lin 2009; Rosén 2001). Research also suggests that women are generally more risk averse (Byrnes, Miller, and Schafer 1999) and tend to adopt more cautious or safety-oriented in-vehicle behaviours (Evans and Wasielewski 1983). For instance, when engaging with an in-vehicle device under actual highway driving conditions, men have demonstrated a significantly higher proportion of long glances to the device as compared to women (Reimer et al. 2013). It is possible that a combination of both more effective reading performance and more conservative off-road distribution of attention may underlie the lower glance time among female participants. In this instance, one might suspect that legibility differences would be more pronounced for women, given their presumably shorter window of attentional focus on the menu screen. Alternatively, it is plausible that women may suppress non-critical aspects of the menu text and focus instead on the key

characteristics required to make an effective selection, such as prioritising street numbers as opposed to street names. Such a search strategy would require less glance time to the device, as observed in the data, and might thus be less sensitive to broader differences in legibility. In essence, women may be better at the menu selection *task* utilised in this experiment. Alternatively, there could be functional differences in how males and females perceive the characters in the operating context of the vehicle. Regardless of the reason for the apparent gender effect observed, the findings support the position that the legibility characteristics studied here did have a significant impact on male participants.

The choice of typestyles compared here was not random. Square grotesque typefaces are used by a number of vehicle manufacturers. Humanist typefaces, on the other hand, have 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 several 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 intrinsic font characteristics can have a positive impact on the 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 an under-representation of the magnitude of the benefits of a more legible typeface. In actual driving conditions, task engagement would likely continue until a correct menu item is selected, resulting in additional time spent with the eyes off the roadway. Future work will need to establish the effects of typeface under conditions in which there is an effective penalty for incorrect or sub-optimal task performance.

This paper focuses on an exploration of the impact of typeface design at the level of visual demand experienced by a driver when interacting with a text-rich HMI. Nonetheless, it is recognised that other extrinsic characteristics of a typeface, such as character size, capitalisation, shadowing, rendering and foreground/background colour combinations (e.g. white on black or black on white), may be adjusted to further reduce demand. Future work will need to assess the relative extent to which these factors interact to impact overall demand on the driver. In addition, it will be necessary to examine the optimisation of other aspects of a display layout, such as white space and design elements. In summary, this research suggests that optimising typeface characteristics is an effective method of providing a meaningful reduction in interface demand and associated distractions.

7. Limitations

These were exploratory studies and a number of limitations should be noted. The sample considered adults between 35 and 75. It remains to be determined if the impact of typeface on legibility observed here is more or less apparent in younger individuals. Many variables interact to define the characteristic features of a particular typeface, and this complexity presents challenges to any empirical investigation of the effects of typeface on behaviour. Character height has been established as a significant variable, and this was explicitly controlled in this study by setting an equal absolute height of the capital letter 'H' for both fonts, in line with the ISO 15008 standard. This resulted in slight difference in x-heights between the two fonts. The humanist typeface used in this study also has a slightly wider stroke width than the square grotesque. While the magnitude of this difference is difficult to discern at the display sizes used here, to what extent these attributes contributed to the overall difference observed is unknown and would require additional testing to assess. Finally, the humanist and square grotesque typefaces were used in this study because of features such as the openness of shapes, inter-character spacing and ambiguity of forms. 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 modestly higher error rates for the square grotesque typeface. Several cases were excluded based on an unexpected behaviour 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 display customisation, so that a display can be tuned 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 behaviour 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.

Funding

This collaborative project was underwritten in part by Monotype Imaging Inc. through funding provided to MIT and in contribution of technical support and typographical expertise. 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.

References

- Attneave, F., and M. D. Arnoult. 1956. "The Quantitative Study of Shape and Pattern Perception." *Psychological Bulletin* 53 (6): 452–471.
- Bell, E. C., M. C. Willson, A. H. Wilman, S. Dave, and P. H. Silverstone. 2006. "Males and Females Differ in Brain Activation during Cognitive Tasks." *NeuroImage* 30 (2): 529–538.
- Bigelow, C., and S. Matteson. 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.
- Bouma, H. 1970. "Interaction Effects in Parafoveal Letter Recognition." *Nature* 226 (5241): 117–178.
- Bouma, H. 1973. "Visual Interference in the Parafoveal Recognition of Initial and Final Letters of Words." *Vision Research* 13 (4): 762–782.
- Byrnes, J. P., D. C. Miller, and W. D. Schafer. 1999. "Gender Differences in Risk Taking: A Meta-Analysis." *Psychological Bulletin* 125 (3): 367–383.
- Cai, H., and P. Green. 2005. "Range of Character Heights for Vehicle Displays as Predicted by 22 Equations." Proceedings of the SID Vehicle Display Symposium, Dearborn, MI.
- Chrysler, S. T., P. J. Carlson, and H. G. Hawkins. 2003. *Headlamp Illumination Provided to Sign Positions by Passenger Vehicles*. Austin: Texas Department of Transportation.
- Conover, W. J., and R. L. Iman. 1981. "Rank Transformations as a Bridge between Parametric and Nonparametric Statistics." *American Statistician* 35 (3): 124–129.
- Driver Focus-Telematics Working Group. 2006. "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.
- Evans, L., and P. Wasielewski. 1983. "Risky Driving Related to Driver and Vehicle Characteristics." *Accident Analysis & Prevention* 15 (2): 121–136.
- Fiset, D., C. Blais, C. Ethier-Majcher, M. Arguin, D. Bub, and F. Gosselin. 2008. "Features for Identification of Uppercase and Lowercase Letters." *Psychological Science* 19 (11): 1161–1168.
- Fujikake, K., S. Hasegawa, M. Omori, H. Takada, and M. Miyano. 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: 503–509.
- Garvey, P. M., M. T. Pietrucha, and D. T. Meeker. 1997. "Effects of Font and Capitalization on Legibility of Guide Signs." *Transportation Research Record* 1605: 73–79.
- Garvey, P. M., M. T. Pietrucha, and D. T. Meeker. 1998. "Clearer Road Signs Ahead." *Ergonomics in Design* 6 (3): 7–11.
- Holick, A. J., S. T. Chrysler, E. S. Park, and P. J. Carlson. 2006. *Evaluation of the Clearview Font for Negative Contrast Traffic Signs*. Austin: Texas Department of Transportation.
- Iman, R. L. 1974. "A Power Study of a Rank Transform for the Two-Way Classification Model When Interactions May Be Present." *Canadian Journal of Statistics* 2 (2): 227–239.
- 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: 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: International Standards Organization.
- ISO 15008. 2009. *Ergonomic Aspects of Transport Information and Control Systems – Specification and Test Procedures for In-Vehicle Visual Presentation*. Geneva: International Standards Organization.
- Kang, Y. Y., M. J. J. Wang, and R. Lin. 2009. "Usability Evaluation of E-Books." *Displays* 30 (2): 49–52.
- Kennedy, R. S., N. E. Lane, K. S. Berbaum, and M. G. Lilienthal. 1993. "Simulator Sickness Questionnaire: An Enhanced Method for Quantifying Simulator Sickness." *The International Journal of Aviation Psychology* 3 (3): 203–220.
- Klauer, S. G., T. A. Dingus, V. L. Neale, J. D. Sudweeks, and D. J. Ramsey. 2006. *The Impact of Driver Inattention on Near-Crash/Crash Risk: An Analysis Using the 100-Car Naturalistic Driving Study Data* (Report No. DOT HS 810 594). Washington, DC: United States Department of Transportation, National Highway Traffic Safety Administration.
- Legge, G. E., and C. A. Bigelow. 2011. "Does Print Size Matter for Reading? A Review of Findings from Vision Science and Typography." *Journal of Vision* 11 (5): 1–22.
- Mehler, B., B. Reimer, J. F. Coughlin, and J. A. Dusek. 2009. "The Impact of Incremental Increases in Cognitive Workload on Physiological Arousal and Performance in Young Adult Drivers." *Transportation Research Record: Journal of the Transportation Research Board* 2138: 6–12.
- National Highway Traffic Safety Administration. 2013. *Visual-Manual NHTSA Driver Distraction Guidelines for In-Vehicle Electronic Devices* (Docket No. NHTSA-2010-0053). Washington, DC: U.S. Department of Transportation National Highway Traffic Safety Administration (NHTSA).
- O'Day, S., and L. Tijerina. 2011. "Legibility: Back to the Basics." *SAE International Journal of Passenger Cars-Mechanical Systems* 4 (1): 591–604.
- Pelli, D. G., N. J. Majaj, N. Raizman, C. J. Christian, E. Kim, and M. C. Palomares. 2009. "Grouping in Object Recognition: The Role of a Gestalt Law in Letter Identification." *Cognitive Neuropsychology* 26 (1): 36–49.
- Pelli, D. G., K. A. Tillman, J. Freeman, M. Su, T. D. Berger, and N. J. Majaj. 2007. "Crowding and Eccentricity Determine Reading Rate." *Journal of Vision* 7 (2): 1.
- R Core Team. 2012. *R: A Language and Environment for Statistical Computing*. Vienna: R Foundation for Statistical Computing. <http://www.R-project.org/>
- Reimer, B., L. A. D'Ambrosio, J. F. Coughlin, M. E. Kafrissen, and J. Biederman. 2006. "Using Self-Reported Data to Assess the Validity of Driving Simulation Data." *Behavior Research Methods* 38 (2): 314–324.

- Reimer, B., B. Mehler, and J. F. Coughlin. 2012. *An Evaluation of Typeface Design in a Text-Rich Automotive User Interface* (MIT AgeLab White Paper 2012-12). Cambridge: Massachusetts Institute of Technology.
- Reimer, B., B. Mehler, J. Dobres, and J. F. Coughlin. 2013. *The Effects of a Production Level 'Voice-Command' Interface on Driver Behavior: Reported Workload, Physiology, Visual Attention, and Driving Performance* (Technical Report 2013-17a). Cambridge, MA: MIT AgeLab.
- Rosén, M. 2001. "Gender Differences in Reading Performance on Documents across Countries." *Reading and Writing* 14 (1–2): 1–38.
- Schulz-Anker, E. 1970. "Syntax-Antiqua, a Sans Serif on a New Basis." *Gebrauchsgraphik* 7: 49–56.
- Speck, O., T. Ernst, J. Braun, C. Koch, E. Miller, and L. Chang. 2000. "Gender Differences in the Functional Organization of the Brain for Working Memory." *Neuroreport* 11 (11): 2581–2585.
- Tinker, M. A. 1963. *Legibility of Print*. Ames: Iowa State University Press.
- Wang, Y., B. Reimer, B. Mehler, V. Lammers, L. A. D'Ambrosio, and J. F. Coughlin. 2010. "The Validity of Driving Simulation for Assessing Differences between In-Vehicle Informational Interfaces: A Comparison with Field Testing." *Ergonomics* 53 (3): 404–420.