Applied Ergonomics 60 (2017) 68–73

Contents lists available at ScienceDirect

Applied Ergonomics

journal homepage: www.elsevier.com/locate/apergo

Effects of ambient illumination, contrast polarity, and letter size on text legibility under glance-like reading

Jonathan Dobres [a, *], Nadine Chahine [b], Bryan Reimer [a]

[a] *Massachusetts Institute of Technology AgeLab and New England University Transportation Center, 77 Massachusetts Avenue, E40-209, Cambridge, MA 02139, United States*

[b] *Monotype Imaging, Inc., 600 Unicorn Park Drive, Woburn, MA 01801, United States*

ARTICLE INFO

Article history:
Received 17 April 2016
Received in revised form
27 September 2016
Accepted 4 November 2016

Keywords:
Legibility
Illumination
Visual perception

ABSTRACT

Recent research on the legibility of digital displays has demonstrated a "positive polarity advantage", in which black-on-white text configurations are more legible than their negative polarity, white-on-black counterparts. Existing research in this area suggests that the positive polarity advantage stems from the brighter illumination emitted by positive polarity displays, as opposed to the darker backgrounds of negative polarity displays. In the present study, legibility thresholds were measured under glance-like reading conditions using a lexical decision paradigm, testing two type sizes, display polarities, and ambient illuminations (near-dark and daylight-like). Results indicate that legibility thresholds, quantified as the amount of time needed to read a word accurately, were highest for the negative polarity configurations under dark ambient illumination, indicated worse performance. Conversely, the positive polarity conditions under dark ambient illumination and all conditions under bright illumination demonstrated significantly reduced thresholds, indicating greater legibility. These results are consistent with the hypothesis that the "positive polarity advantage" arises because brighter illumination produces pupillary contraction that reduces optical aberrations as light enters the eye. These results have implications for the design of automotive interfaces and other scenarios in which an interface must be optimized for glance-like reading under variations in ambient lighting conditions.

1. Introduction

Digital displays have made it easy to display text in arbitrary color and contrast combinations. In combination with advanced sensing and computing capabilities, the format of the display can be rapidly shifted based upon the intrinsic characteristics of the content, ambient conditions, or even perceived characteristics of a reader (Burke, 2006). Negative or "reverse" polarity displays—so named because they utilize light text on a dark background, as opposed to black-on-white positive polarity displays—have been in common use since the days of microfiche reading devices (Cushman, 1986) and have more recently become popular in mobile and automotive interfaces. In the automotive sector, such displays are preferred because the darker background of the negative polarity display hides wear and tear on the screen, blends in with the interior of the car, and reduces ambient illumination in the cabin during nighttime driving (i. e., positive polarity displays may emit more light in the cabin and increase glare). In some production applications, changes in the polarity of the display are made in response to ambient conditions, while other systems use a negative polarity display at all times. In the mobile device sector, negative polarity designs are less dominant, and their use appears to be more aesthetically motivated, or are used in response to the perceived optimization of the display for ambient illumination. For example, guidelines for development on the Apple Watch platform strongly encourage the use of negative polarity displays because the dark background blends in with the hardware's dark bezel. More generally, negative polarity designs popularly connote a more "high tech" aesthetic.

The relative legibility tradeoffs of negative versus positive polarity displays have garnered considerable attention in recent years. Recent research has shown that positive polarity text has superior legibility compared to negative polarity (Buchner and Baumgartner, 2007; Mayr and Buchner, 2010; Piepenbrock et al., 2013a, 2014; Piepenbrock et al., 2013b; Taptagaporn and Saito, 1990, 1993;

* Corresponding author.
E-mail addresses: jdobres@mit.edu (J. Dobres), nadine.chahine@monotype.com (N. Chahine), reimer@mit.edu (B. Reimer).

http://dx.doi.org/10.1016/j.apergo.2016.11.001

Tsang et al., 2012). This "positive polarity advantage", as some have termed it, has been shown to increase as text size decreases (Piepenbrock et al., 2013a), and is more pronounced for younger observers (Piepenbrock et al., 2013b). Several competing theories have been put forth to explain the positive polarity advantage, which include simple familiarity effects (Hall and Hanna, 2004), a "luminance asymmetry effect", in which luminance decrements against the background are perceived as creating a greater change in luminance than increments of equal magnitude (Lu and Sperling, 2012), and the influence of spherical aberrations of the eye on visual input (Lombardo and Lombardo, 2010) Among these theories, a converging stream of evidence strongly suggests that the positive polarity advantage arises from the differing levels of illumination produced by the two display configurations (Buchner et al., 2009; Piepenbrock et al., 2014; Taptagaporn and Saito, 1990). Positive polarity displays feature a bright background and cause the pupil to contract, which in turn reduces distortions of visual input due to the aberrations of the eye. Conversely, darker negative polarity displays produce pupillary dilation, making it more likely that visual input will be affected by spherical aberrations. At least one study has demonstrated that when display illumination is held constant across polarity conditions, the positive polarity advantage is eliminated, and only the overall illumination of the display itself affects reading accuracy (Buchner et al., 2009). It should be noted, however, that this study manipulated on-screen brightness projected directly at the observer, rather than ambient illumination *per se*. While one study has shown that text polarity affects reading accuracy regardless of the available ambient illumination, this study employed a relatively narrow range of illuminations, from a near darkness of 5 lx to standard office lighting of 550 lx (Buchner and Baumgartner, 2007).

Historically, reading was performed in long stretches, as with a book or newspaper (Cushman, 1986; Judisch, 1969; Seppala, 1975). Opportunities for reading at a glance were relatively limited, and primarily involved glances to roadway signage (Ells and Dewar, 1979; Jacobs et al., 1976; Sivak et al., 1981). As a result, the bulk of legibility studies, such as those outlined above, quantify legibility using long-form reading tasks and metrics, such as proofreading and words read per minute, all of which rely on self-paced paradigms. It remains to be seen whether findings from long-form reading studies are consistent under glance-like reading scenarios, in which the observer has a limited amount of time to encode the available visual and lexical information. The increasing prominence of the smartphone and the availability of information at a glance make this a key research question in contemporary studies of legibility. Such scenarios are especially relevant in environments where information may only be available in short glances, as when using an in-vehicle interface while driving, glancing at a smartphone notification, or viewing a rapidly moving advertisement. In addition, it is unclear whether a positive or negative polarity display would "wash out" under high ambient illumination, potentially creating a pattern of results different from those observed under the relatively dim illuminations used in previous studies.

Recent research has been conducted to explicitly investigate the relative legibility of a variety of typographic factors under glance or glance-like reading conditions. A study conducted in a full cab driving simulator, in which a menu system was set in one of two possible typefaces, showed that the choice of typeface significantly impacted drivers' task completion time and number of glances to the display (Reimer et al., 2014). Later work extended these findings by showing that the same pattern of results regarding typeface could be demonstrated using a simpler desktop-based method (Dobres et al., 2016a; 2016b). While these studies show that glance legibility can be probed using empirical methods, they were all conducted under relatively dim illumination (a simulator approximating evening illumination or a dimly lit room).

Here we present a study in which legibility thresholds are measured under a glance-like reading paradigm for two contrast polarities, type sizes, and ambient lighting conditions. Legibility thresholds are operationalized as the amount of on-screen display time needed to read the stimuli with approximately 80% accuracy. This work extends earlier research by addressing limitations in the generalizability of results across variations in ambient lighting conditions. In addition, it extends the methodological underpinnings of the approach (Dobres et al., 2016b) from English to Italian. Based on previous research, we expect that legibility thresholds will be lower in the bright ambient light condition, while under the dark ambient condition, the negative polarity displays should show significantly elevated legibility thresholds compared to positive polarity displays. We also expect that legibility thresholds will be elevated at the smaller of the two text sizes. Lastly, we expect the positive polarity advantage to be more pronounced at the smaller text size.

2. Methods

2.1. Participants

The participant sample was recruited from within the Fiat Chrysler Automobiles (FCA) Italian headquarters in Torino, Italy. Participants were required to be between the ages of 20 and 65, to be in self-reported good health for their age, to drive a motor vehicle at least once per week, to have normal or corrected-to-normal vision, and to speak and read Italian as a first language. All participants provided an informed verbal consent consistent with the United States Department of Health and Human Services' "Common Rule", developed with the approval of the Massachusetts Institute of Technology's Committee on the Use of Humans as Experimental Subjects.

A total of 50 participants meeting these criteria were recruited. Of these, 1 participant withdrew due to discomfort, 5 were excluded because at least one of their estimated thresholds (see below) were in excess of 300 ms, 6 were excluded because their mean response times were greater than 1000 ms, 3 were excluded due to a probable failure to reach a stable threshold in at least one condition (defined as an absence of staircase reversals during the last 20 trials of a condition block, see below), and 1 was excluded because he/she was unable to attend all data collection sessions. This left a total of 34 participants in the analysis sample, including 13 women (mean age 36.2 years, SD 8.1) and 21 men (mean age 39.0 years, SD 9.9). There was no significant difference in age between the genders ($t(29.3) = 0.91$, $p = 0.369$).

2.2. Apparatus

The experiment utilized custom software developed by the Massachusetts Institute of Technology AgeLab, built on the PsychoPy platform (Peirce, 2008). The experiment was run on a 1.4 GHz Mac Mini under Mac OS 10.10.1 ("Yosemite"). Stimuli were displayed on a 17″ Dell 1707FPT LCD monitor with a resolution of 1280×1024 pixels and a refresh rate of 60 Hz. Participants were seated such that their eyes were approximately 0.7 m from the display. While head restraints were not used, participants were encouraged to maintain a consistent posture throughout the experiment. Participants were instructed to wear their preferred optical correction (if any) for that reading distance, and to do so throughout the experiment.

2.3. Stimuli & task

Following the methodology utilized in (Dobres et al., 2016b), participants performed a 1-interval forced choice lexical decision task (Meyer and Schvaneveldt, 1971). A schematic of the task is presented in Fig. 1. Each trial begins with a 400 ms display of a fixation rectangle (200px by 100px, or 5.29° by 2.65° at a 0.7 m viewing distance), centered on the screen, indicating the general area where stimuli will appear (all stimuli and masks are displayed at the screen's center). The fixation rectangle is followed by a 200 ms mask composed of randomized non-letter characters. Then a single word (or pseudoword) stimulus is displayed for a variable presentation time, as determined by an adaptive staircase procedure (see "Adaptive Staircase Procedure", below). This is immediately followed by another 200 ms mask. Finally, the participant is prompted to decide whether the stimulus was a word or pseudoword. Participants are given a maximum of 5000 ms to respond by pressing either the '1' or '3' key of the numeric keypad (the keys corresponded to "word" and "pseudoword", respectively, and were marked with either green or red tape for clarity). Participants were not provided with feedback regarding the accuracy of their responses, other than during the practice section described below. Each mask was unique, constructed by randomly selecting eight characters from a small pool of non-letter characters. The sandwiching of the stimulus between the two masks minimizes the stimulus's visible persistence in iconic memory, ensuring that it will only be perceptually accessible for the intended presentation time (Coltheart, 1980).

The experiment began with a series of ten practice trials, with stimulus duration set to 1000 ms. After five consecutive correct answers, participants were permitted to move on to the main experiment. If the participant reached the end of the ten trials without making five consecutive correct responses, he/she was allowed to repeat the practice block. If the participant was still unable to complete the practice at this minimum performance criterion, he/she would be excluded from the experiment (no participants were excluded for this reason).

The primary stimuli of this experiment were Italian words and pseudowords. Words of equal frequency in the Italian lexicon were sampled from CoLFIS (Bertinetto et al., 2005). Only words without diacritical marks (accents) were used to control for the visual complexity and associated reading times of the words (Abdelhadi et al., 2011). Pseudowords were generated from the word stimuli using Wuggy (Keuleers and Brysbaert, 2010). From these sets of words and pseudowords, 410 words and 410 pseudowords were sampled at random for use in the experiment. Each list was split in half for use across the two data collection sessions, such that no words or pseudowords were repeated for a participant.

"Neue Helvetica 55 Roman" was used to display all stimulus text. Neue Helvetica is a widely-used sans-serif typeface, and an experienced typographer determined that the 55 Roman variant had the clearest on-screen rendering under PsychoPy. To limit exposure and familiarity effects outside of the experimental conditions, a serif typeface that looked substantially different from the one used during main data collection, "Georgia", was used to display practice trial stimuli and all prompt text. Prompt text set in Georgia was also displayed at approximately double the size of the largest word and pseudoword stimuli.

Primary data collection (400 trials per session) began after the practice block. After every 50 trials (approximately every 4–5 min), participants were allowed to take a short rest of up to 30 s (the participant could terminate the rest periods early if so desired). Break periods occurring after the 100th, 200th, and 300th trial were mandatory and could not be skipped. These periods coincided with a transition to a new typographic condition.

2.4. Conditions tested

A total of 8 conditions were measured. Text was displayed in two possible contrast polarities and at two possible sizes (capital letter height of 3 mm or 4 mm, approximately 14.7 arcmin and 19.6 arcmin, respectively; sizes chosen based on International Standards Organization guidance for in-vehicle devices). Text size was measured according to the on-screen height of the typeface's capital letter 'H' (International Standards Organization, 2009), though all stimuli were presented in lowercase letters. Text contrast polarity was either positive (black text on a white background) or negative (white text on a black background). White elements used RGB values of 255, 255, 255, while black elements used RGB values of 0, 0, 0. Pixels in various shades of gray were also present, depending on PsychoPy's text smoothing algorithm (based on Python's Pyglet rendering engine, which utilizes grayscale

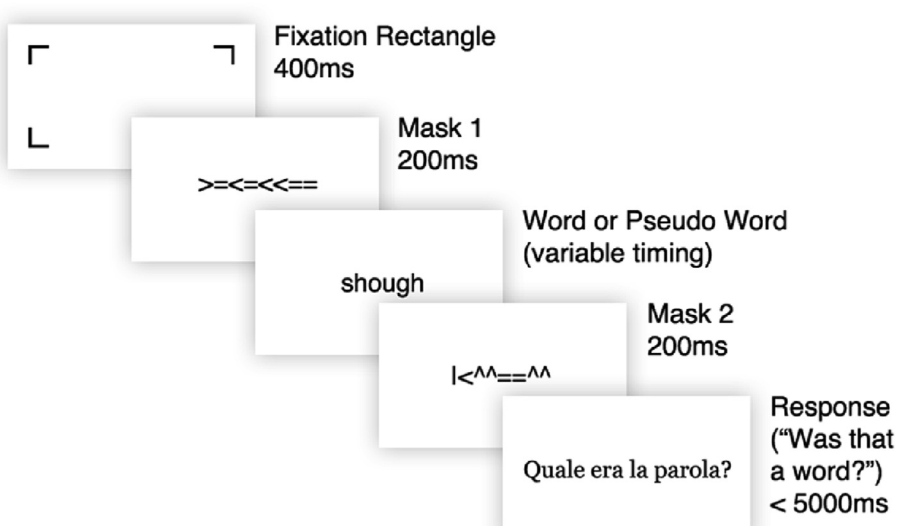

Fig. 1. The structure of an individual trial of the experiment. See Methods for details.

smoothing).

These 4 contrast/size conditions were viewed either under dark ambient lighting (near 0 lux) or bright ambient lighting (4750 lux, equivalent to slightly overcast daylight), in compliance with established guidelines for studying the effects of ambient illumination on device use (International Standards Organization, 2009). Under bright illumination, monitor brightness was set to 100% (estimated 300 cd/m^2 when displaying a white screen), and under dark illumination, 75% (estimated 225 cd/m^2). All participants completed two sessions of the experiment; the bright ambient conditions were presented on the participant's first testing day, while the dark ambient conditions were presented on the participant's second day. All data was collected within a 13-day period, with all bright ambient data collected within the first week, and all dark ambient data in the second week. Owing to necessary setup time and facility limitations, the order of the ambient lighting conditions could not be counterbalanced.

Each typographic condition was presented in a separate block for 100 trials per condition. Condition order was randomized per participant and session, which effectively prevented presentation order from influencing threshold estimates (bright condition X2(3) = 0.42, p = 0.935; dark condition X2(3) = 2.36, p = 0.500; Friedman's test on condition order during each session).

2.5. Adaptive staircase procedure

During the main data collection blocks, task difficulty was controlled via an adaptive staircase procedure (Leek, 2001; Levitt, 1971). This technique changes the difficulty of the task based on a participant's pattern of correct and incorrect responses. Using a “3-down, 1-up” rule, the task is made more difficult (stimulus display time is decreased) after three consecutive correct responses, and made easier (stimulus display time is increased) after one incorrect response. Following this rule, stimulus display time will converge on a difficulty that produces 79.4% accuracy (Leek, 2001).

We modified the staircase algorithm to accommodate the experiment's workflow in the following ways. First, stimulus duration was initially decremented in a controlled manner to allow the participant to adapt to the expected task difficulty. At the start of each condition, stimulus duration was set at 800 ms. Three trials were performed at this setting, regardless of the participant's responses. Stimulus duration was then decremented to 600 ms for the next 3 trials, 400 ms for 3 trials after that, and finally, 200 ms for another 3 trials. Staircase control of stimulus duration was initiated on the 13th trial of the condition.

The staircase's step size (the increment by which stimulus duration was adjusted) was gradually decreased throughout each condition, allowing the staircase to make finer adjustments as the condition progressed. Over the course of 100 trials per condition, step size reached a minimum of 1 frame. Third, stimulus duration was constrained to be at least 16.7 ms and at most 1000 ms.

Staircase parameters were reset at the start of each condition, allowing for the calculation of separate stimulus duration thresholds for each of the 4 conditions. Each condition is calibrated to the same hypothetical accuracy level. Therefore, a less legible typeface should require a longer presentation time (and thus a higher threshold) to reach the same accuracy level as a more legible typeface.

2.6. Data analysis

Thresholds were obtained for each condition by calculating the median stimulus duration of each condition's final 20 trials. Lower stimulus display time thresholds are taken as indications of superior legibility. Response accuracy was calculated as the mean accuracy during the final 20 trials of each condition, when threshold estimates were likely to have stabilized. Stimulus presentation time thresholds and performance accuracy data were analyzed in a (2 × 2 × 2) repeated-measures design (lighting × typeface × polarity). All statistics were computed and visualized using R (R Core Team., 2016).

3. Results

3.1. Performance accuracy

Summary statistics for response accuracy are shown in Table 1. The staircase procedure used to estimate glance legibility thresholds in this experiment should cause response accuracy to converge on approximately 79.4% accuracy, and therefore response accuracy should not differ between conditions. Mean response accuracy across all participations was 79.1%. Response accuracy did not differ significantly from 79.4% in any of the 8 conditions tested (all $p > 0.27$, one-sample *t* tests), suggesting that the staircase procedure successfully converged on an accurate threshold estimate in most cases.

3.2. Display time thresholds

Display time thresholds are presented in Table 2 and Fig. 2. A statistical model including factors of ambient lighting, contrast polarity, and type size as main effects indicates significant effects for all three factors (lighting $F(1, 33) = 7.61$, $p = 0.009$; polarity $F(1, 33) = 24.40$, $p < 0.001$; size $F(1, 33) = 4.21$, $p = 0.048$), as well as a significant interaction between ambient lighting and contrast polarity ($F(1, 33) = 14.18$, <0.001). Considering the bright ambient lighting condition on its own, only a significant effect of size is apparent ($F(1, 33) = 5.49$, $p = 0.025$), with the 3 mm display size resulting in a 3.0% increase in legibility thresholds (84.8 ms at 4 mm vs. 87.4 ms at 3 mm). Polarity did not produce significant differences in reading time thresholds ($F(1, 33) = 0.19$, $p = 0.665$), nor did polarity and size interact under bright ambient lighting ($F(1, 33) = 1.46$, $p = 0.236$). Conversely, in the dark ambient lighting condition, there is no significant effect of size (likely owing to the very similar thresholds for positive polarity text), but the effect of polarity is highly significant ($F(1, 33) = 49.60$, $p < 0.001$). Further posthoc testing within the dark ambient lighting conditions shows that there is no effect of size between the positive polarity conditions ($t(33) = 0.13$, $p = 0.898$), but there is a significant effect of size between the negative polarity conditions ($t(33) = 2.31$, $p = 0.027$).

Taken together, these results suggest that positive polarity text retains some legibility advantage regardless of ambient lighting. Positive polarity thresholds are nominally lower than negative polarity text under bright ambient lighting, and significantly lower under dark ambient lighting. More importantly, these results are consistent with the hypothesis that the “positive polarity advantage” results from the tighter contraction of the pupil due to bright illumination. Thresholds are relatively low for all conditions under bright ambient illumination, while thresholds elevate dramatically in the relatively dark conditions of negative polarity text under dim

Table 1
Means (and standard deviations) of percentage of correct responses for each condition.

Ambient lighting	Contrast polarity	3 mm	4 mm
Simulated Day-Time	Negative	78.4 (6.2)	80.1 (6.0)
	Positive	79.3 (7.2)	78.7 (5.4)
Simulated Night-Time	Negative	77.6 (9.0)	79.0 (7.4)
	Positive	80.1 (8.4)	80.1 (6.9)

Table 2
Means (and standard deviations) of presentation time thresholds for each condition.

Ambient lighting	Contrast polarity	3 mm	4 mm
Simulated Day-Time	Negative	88.7 (50.0)	92.2 (61.7)
	Positive	86.0 (34.7)	77.5 (41.9)
Simulated Night-Time	Negative	122.3 (56.8)	106.9 (50.9)
	Positive	84.1 (43.7)	83.3 (53.5)

ambient lighting.

4. Discussion

4.1. General discussion

In the present paper, we have shown that legibility thresholds, quantified as the amount of on-screen display time needed to identify a word with approximately 80% accuracy, are reduced globally under conditions of high ambient illumination, suggestive of an increase in legibility under high ambient illumination. Under dark ambient illumination, thresholds remain comparably low for positive polarity text, but are significantly elevated for negative polarity text. Ignoring other factors such as eye accommodation time to and from glances to non-task related activities (for example, in the context of an automotive environment), and considering this result in terms of pure legibility, the combination of dark ambient conditions and negative polarity displays proves to be significantly less legible than positive polarity under the same lighting conditions, and either polarity under bright ambient conditions. The difference between polarities under dark ambient lighting is especially pronounced at the smaller 3 mm text size, further supporting earlier observations of legibility decrements in suboptimal designs (Dobres et al., 2016b; Reimer et al., 2014).

These results are strikingly consistent with an emerging body of research on the effects of contrast polarity on legibility. Several studies have found that positive polarity stimuli are more easily perceived than their negative counterparts, an effect dubbed the "positive polarity advantage" (Buchner and Baumgartner, 2007; Lu and Sperling, 2012; Piepenbrock et al., 2014; Piepenbrock et al., 2013a). Subsequent work has shown that this effect is likely the result of pupillary dilation caused by the difference in illumination between positive and negative polarity displays, or more precisely, the luminance of the large background areas in these conditions. As the pupil dilates over the imperfect surface of the eye, incoming light and its associated optics are distorted, thus impairing legibility (Buchner et al., 2009; Piepenbrock et al., 2014; Tsang et al., 2012).

The results of the present study agree with research from Piepenbrock et al. (2013a) showing that the "positive polarity advantage" exists and is stronger at smaller type sizes. However, the pattern of results observed here argues for what might be termed "negative polarity disadvantage". In the bright ambient conditions (approximately 4750 lux), typographic thresholds were globally reduced and modestly, though significantly, affected by type size for both positive and negative displays. Under dark (near 0 lux) ambient illumination, the changes in illumination emanating from the positive and negative polarity backgrounds appear to have significantly impacted legibility. Negative polarity thresholds were significantly elevated compared to all others, and this effect was especially pronounced at the smaller 3 mm size. Conversely, the positive polarity displays produced thresholds similar to those observed under bright ambient illumination. Consistent with the idea of a negative polarity disadvantage, Piepenbrock et al., (2013a) data demonstrate that the mean number of words read increases with type size under negative polarity conditions, but is unaffected by size under positive polarity conditions (see that paper's Fig. 2).

Previous research has shown that the positive polarity advantage persists regardless of ambient illumination (Buchner and Baumgartner, 2007). However, that study employed a dark ambient condition (5 lux) and "typical office lighting" (550 lux). In contrast, the present study employs a bright ambient condition of 4750 lux, comparable to mild outdoor lighting and an order of magnitude brighter than the condition studied in Buchner and Baumgartner (2007). We argue that this wider illumination range exposes a key limit of the positive polarity advantage. While the positive polarity displays still had nominally lower thresholds under bright ambient illumination, the difference was non-significant.

Notably, the present work employs a lexical decision paradigm that essentially enforces glance-like reading conditions, and is thus relevant to automotive interface designs, mobile computing applications, and others. Much of the previous work in this area employed long-form reading tasks and relatively high-level metrics such as proofreading and comprehension, while the lexical decision tasks abstracts away much of these high-level complications and allows for a more direct and rapid measurement of legibility.

4.2. Limitations

This study drew its sample from a relatively constrained group

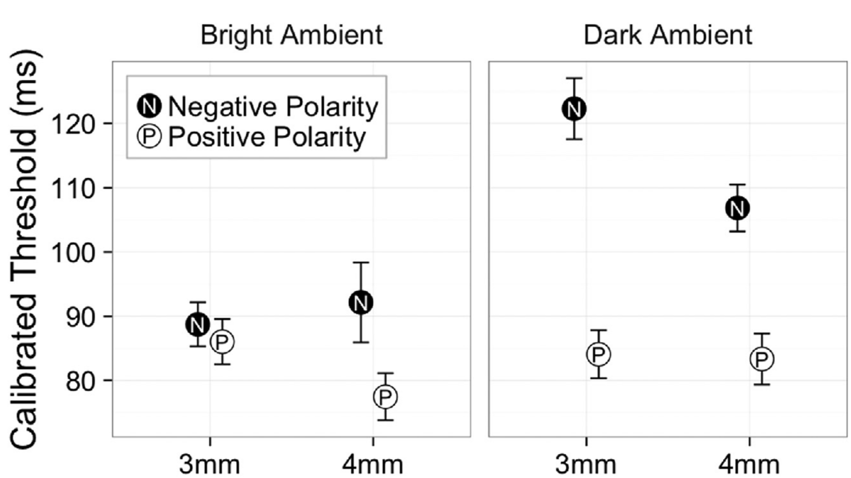

Fig. 2. Mean on-screen display time thresholds for all conditions under study (lower numbers indicate greater legibility under the conditions studied). Error bars represent ±1 within-subject standard error.

of full-time working professionals. While participants in the study were as old as 59, it is likely that all were high functioning individuals with minimal or no visual/cognitive impairments, as evidenced by the lack of age effects observed. It is difficult to say whether a less high-functioning sample would reveal a different pattern of results than those seen here.

In addition, the complications of experiment setup and breakdown made counterbalancing the bright and dark ambient conditions impossible, and the bright conditions were always conducted as part of the first experiment session. We speculate that a practice effect could have reduced all thresholds measured during the second session (Karni and Sagi, 1993), which may in part explain why no difference was observed between 4 mm and 3 mm thresholds for the positive polarity condition during that session.

5. Conclusion

This study demonstrates that the “positive polarity advantage” is readily apparent under conditions of dark ambient illumination, but is only nominally in evidence under bright ambient illumination. Legibility thresholds, as measured via a lexical decision task, are globally reduced under bright ambient illumination, and elevate under dark illumination only for negative polarity text configurations. This would suggest that the “positive polarity advantage” is perhaps better termed “negative polarity disadvantage.” These results have implications for general interface design guidelines, particularly in the automotive and mobile device sector, where concerns over nighttime driving and the potential for visual distraction necessitate a balance between interface brightness and the readability of its text. Moreover, the results described here suggest that larger text sizes should be employed whenever possible, particularly on negative polarity displays. Future work will need to assess the degree to which these results can be extended to a real operating context where slow and rapid changes in the environmental conditions (e.g. ambient illumination, glare, etc.) impact legibility as an operator shifts attention between the operating context and an in-vehicle display.

Acknowledgements

Data collection was facilitated by Monotype and Fiat Automotive leveraging software and methodologies developed by the MIT AgeLab. MIT authors performed all data analysis and drafted the initial manuscript. Support for MIT's contribution to this work was provided by the US Department of Transportation's Region I New England University Transportation Center at MIT and the Toyota Class Action Settlement Safety Research and Education Program. The views and conclusions being expressed are those of the authors, and have not been sponsored, approved, or endorsed by Toyota or plaintiffs' class counsel.

References

Abdelhadi, S., Ibrahim, R., Eviatar, Z., 2011. Perceptual load in the reading of Arabic: effects of orthographic visual complexity on detection. Writ. Syst. Res. 3 (2), 117–127. http://doi.org/10.1093/wsr/wsr014.

Bertinetto, P.M., Burani, C., Laudanna, A., Marconi, L., Ratti, D., Rolando, C., Thornton, A.M., 2005. Corpus e Lessico di Frequenza dell'Italiano Scritto (CoLFIS). Retrieved from. http://linguistica.sns.it/CoLFIS/Home.htm.

Buchner, A., Baumgartner, N., 2007. Text - background polarity affects performance irrespective of ambient illumination and colour contrast. Ergonomics 50 (7), 1036–1063. http://doi.org/10.1080/00140130701306413.

Buchner, A., Mayr, S., Brandt, M., 2009. The advantage of positive text-background polarity is due to high display luminance. Ergonomics 52 (7), 882–886. http://doi.org/10.1080/00140130802641635.

Burke, R., 2006. The third wave of marketing intelligence. In: Krafft, M., Mantrala, M. (Eds.), Retailing in the 21st Century, pp. 113–125.

Coltheart, M., 1980. Iconic memory and visible persistence. Percept. Psychophys. 27 (3), 183–228.

Cushman, W.H., 1986. Reading from microfiche, a VDT, and the printed page: subjective fatigue and performance. Hum. Factors J. Hum. Factors Ergonomics Soc. 28 (1), 63–73. http://doi.org/10.1177/001872088602800107.

Dobres, J., Chahine, N., Reimer, B., Gould, D., Zhao, N., 2016a. The effects of Chinese typeface design, stroke weight, and contrast polarity on glance based legibility. Displays 41 (C), 42–49. http://doi.org/10.1016/j.displa.2015.12.001.

Dobres, J., Chahine, N., Reimer, B., Gould, D., Mehler, B., Coughlin, J.F., 2016b. Utilising psychophysical techniques to investigate the effects of age, typeface design, size and display polarity on glance legibility. Ergonomics 1–15. http://doi.org/10.1080/00140139.2015.1137637.

Ells, J.G., Dewar, R.E., 1979. Rapid comprehension of verbal and symbolic traffic sign messages. Hum. Factors J. Hum. Factors Ergonomics Soc. 21 (2), 161–168. http://doi.org/10.1177/001872087902100203.

Hall, R.H., Hanna, P., 2004. The impact of web page text-background colour combinations on readability, retention, aesthetics and behavioural intention. Behav. Inf. Technol. 23 (3), 183–195. http://doi.org/10.1080/01449290410001669932.

International Standards Organization, 2009. Ergonomic Aspects of Transport Information and Control Systems (No. 15008) (Geneva, Switzerland).

Jacobs, R.J., Johnston, A.W., Cole, B.L., 1976. The visibility of alphabetic and symbolic traffic signs. Appl. Ergon. 7 (2), 113. http://doi.org/10.1016/0003-6870(76)90184-8.

Judisch, J.M., 1969. The effect of positive-negative microforms and front-rear projection on reading speed and comprehension. J. Microgr. 2, 58–61.

Karni, A., Sagi, D., 1993. The time course of learning a visual skill. Nature 365 (6443), 250–252. http://doi.org/10.1038/365250a0.

Keuleers, E., Brysbaert, M., 2010. Wuggy: a multilingual pseudoword generator. Behav. Res. Methods 42 (3), 627–633. http://doi.org/10.3758/BRM.42.3.627.

Leek, M.R., 2001. Adaptive procedures in psychophysical research. Percept. Psychophys. 63 (8), 1279–1292.

Levitt, H., 1971. Transformed up-down methods in psychoacoustics. J. Acoust. Soc. Am. 49 (2B), 467–477. http://doi.org/doi:10.1121/1.1912375.

Lombardo, M., Lombardo, G., 2010. Wave aberration of human eyes and new descriptors of image optical quality and visual performance. J. Cartaract Refract. Surg. 36 (2), 313–331. http://doi.org/10.1016/j.jcrs.2009.09.026.

Lu, Z.L., Sperling, G., 2012. Black-white asymmetry in visual perception. J. Vis. 12 (10), 1–8. http://doi.org/10.1167/12.10.8.

Mayr, S., Buchner, A., 2010. After-effects of TFT-LCD display polarity and display colour on the detection of low-contrast objects. Ergonomics 53 (7), 914–925. http://doi.org/10.1080/00140139.2010.484508.

Meyer, D.E., Schvaneveldt, R.W., 1971. Facilitation in recognizing pairs of words: evidence of a dependence between retrieval operations. J. Exp. Psychol. 90 (2), 227–234.

Peirce, J.W., 2008. Generating stimuli for neuroscience using PsychoPy. Front. Neuroinformatics 2, 1–8. http://doi.org/10.3389/neuro.11.010.2008.

Piepenbrock, C., Mayr, S., Buchner, A., 2013a. Positive display polarity is particularly advantageous for small character sizes: implications for display design. Hum. Factors J. Hum. Factors Ergonomics Soc. 56, 942–951. http://doi.org/10.1177/0018720813515509.

Piepenbrock, C., Mayr, S., Mund, I., Buchner, A., 2013b. Positive display polarity is advantageous for both younger and older adults. Ergonomics 56 (7), 1116–1124. http://doi.org/10.1080/00140139.2013.790485.

Piepenbrock, C., Mayr, S., Buchner, A., 2014. Smaller pupil size and better proofreading performance with positive than with negative polarity displays. Ergonomics 57 (11), 1670–1677. http://doi.org/10.1080/00140139.2014.948496.

R Core Team, 2016. R: A Language and Environment for Statistical Computing. Vienna, Austria. Retrieved from. http://www.R-project.org/.

Reimer, B., Mehler, B., Dobres, J., Coughlin, J.F., Matteson, S., Gould, D., et al., 2014. Assessing the impact of typeface design in a text-rich automotive user interface. Ergonomics 57 (11), 1643–1658. http://doi.org/10.1080/00140139.2014.940000.

Seppala, P., 1975. Visual fatigue in reading microforms. Agressologie 16, 147–150.

Sivak, M., Olson, P.L., Pastalan, L.A., 1981. Effect of driver's age on nighttime legibility of highway signs. Hum. Factors J. Hum. Factors Ergonomics Soc. 23 (1), 59–64. http://doi.org/10.1177/001872088102300106.

Taptagaporn, S., Saito, S., 1990. How display polarity and lighting conditions affect the pupil size of vdt operators. Ergonomics 33 (2), 201–208. http://doi.org/10.1080/00140139008927110.

Taptagaporn, S., Saito, S., 1993. Visual comfort in VDT operation: physiological resting states of the eye. Ind. Health 31 (1), 13–28.

Tsang, S.N.H., Chan, A.H.S., Yu, R.F., 2012. Effect of display polarity and luminance contrast on visual lobe shape characteristics. Ergonomics 55 (9), 1028–1042. http://doi.org/10.1080/00140139.2012.688876.