Applied Ergonomics 70 (2018) 240–246

Contents lists available at ScienceDirect

Applied Ergonomics

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

The effects of visual crowding, text size, and positional uncertainty on text legibility at a glance

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

[a] MIT AgeLab and New England University Transportation Center, 77 Massachusetts Avenue, E40-275, Cambridge, MA 02139, USA
[b] Monotype Imaging, Inc., 600 Unicorn Park Drive, Woburn, MA 01801, USA

ARTICLE INFO

Keywords:
Typography
Legibility
Visual crowding

ABSTRACT

Reading at a glance, once a relatively infrequent mode of reading, is becoming common. Mobile interaction paradigms increasingly dominate the way in which users obtain information about the world, which often requires reading at a glance, whether from a smartphone, wearable device, or in-vehicle interface. Recent research in these areas has shown that a number of factors can affect text legibility when words are briefly presented in isolation. Here we expand upon this work by examining how legibility is affected by more crowded presentations. Word arrays were combined with a lexical decision task, in which the size of the text elements and the inter-line spacing (leading) between individual items were manipulated to gauge their relative impacts on text legibility. In addition, a single-word presentation condition that randomized the location of presentation was compared with previous work that held position constant. Results show that larger text was more legible than smaller text. Wider leading significantly enhanced legibility as well, but contrary to expectations, wider leading did not fully counteract decrements in legibility at smaller text sizes. Single-word stimuli presented with random positioning were more difficult to read than stationary counterparts from earlier studies. Finally, crowded displays required much greater processing time compared to single-word displays. These results have implications for modern interface design, which often present interactions in the form of scrollable and/or selectable lists. The present findings are of practical interest to the wide community of graphic designers and interface engineers responsible for developing our interfaces of daily use.

1. Introduction

Over the last decade, the advent of mobile computing has fundamentally impacted the ways in which users interact with their devices and the information accessible from them. This rapid evolution has led to changes in common use cases for human-computer interaction. Interfaces that were once relatively stable and simplistic, such as the infotainment dashboards of motor vehicles, have now become digital front-ends full of dynamically changing content. Where once computers were reliably anchored to desktops, now they can be pulled from pockets or read from the wrist. As the use cases have evolved, so too have the user behaviors associated with them. Perhaps most prominently, users are now accustomed to reading pieces of text in brief glances, a behavior previously limited to the intake of information from signage or in-vehicle gauges.

Although the factors affecting legibility have been of interest to researchers for over a century, most of this body of work has focused on long-form paragraph reading or threshold acuity assessments. Both of these are unpressured perception paradigms, in which the observer reads and responds at his or her own pace. In contrast, more modern mobile interaction paradigms, and especially in-vehicle interfaces, often force users to multitask, placing constraints on the amount of time the user has available to perceive the interface and process secondary information. Recent laboratory-based research examining legibility at a glance has shown that legibility can be affected by the typeface family used and the size at which it is set (Dobres et al., 2016a, b), the boldness or weight of the font (Dobres et al., 2016b), and the amount of ambient illumination available in the interface's environment (Dobres et al., 2017a). Furthermore, these laboratory assessments are corroborated by driving simulator research showing that the typeface used for a menu interface can affect the amount of time spent glancing off the road to the device screen, the total time needed to complete secondary tasks, and the number of errors made (Reimer et al., 2014).

While the aforementioned body of work represents a first step in the investigation of glance-based legibility, the work also highlights the large number of interacting factors that affect complex reading behavior. Those studies presented their target stimuli in isolation, without distracting elements, and thus neglected the impact of visually crowded

* Corresponding author.
E-mail address: jdobres@mit.edu (J. Dobres).

https://doi.org/10.1016/j.apergo.2018.03.007
Received 17 July 2017; Received in revised form 22 December 2017; Accepted 12 March 2018
0003-6870/

displays on legibility. Crowded displays are common in modern HCI designs, as users are often required to scroll through lists of information or differentiate between columns of a design layout, as when reading a website. Note that by "crowding", we are referring generally to displays which feature distractor elements alongside some target of interest. This is not to be confused with the strict classical definition of visual crowding, which specifically refers to a loss of visual discriminability experienced when a target presented in the visual periphery is surrounded by flanking elements ((Bouma, 1970, 1973); for extensive reviews of this phenomenon, see (Pelli and Tillman, 2008; Pelli et al., 2007; Whitney and Levi, 2011)). While these types of phenomena grant insight into the workings of visual perception and processing, they are of less direct use to practicing engineers and designers. Instead, in the present paper we rely on a looser definition of visual crowding, which refers simply to the presence and density of task-irrelevant visual information in a given display.

Even so, there is reason to believe that classical crowding affects legibility even in central (foveal) vision (Chung et al., 2007; J.-Y. Zhang et al., 2009). Over a century ago, Roethlein found that fonts which featured more negative space (i.e., were less visually cramped) were associated with greater legibility (Roethlein, 1912), a finding supported at least in part by more recent research showing that lighter-weight fonts, which have more negative space by definition, may be more legible than thicker fonts (Dobres et al., 2016b). Along the same lines, increasing the spacing between individual characters generally aids legibility, even when words are read foveally under normal conditions (Montani et al., 2014; Perea and Gomez, 2012; Perea et al., 2011).

However, such intra-character crowding effects are relatively subtle compared to effects arising from the overall density of text on the page or display. One factor governing the density of text is the leading, or vertical space that separates lines of text. Research on this issue, though sparse, has consistently shown increased leading to be markedly beneficial for legibility compared to "set solid" or maximally dense typesetting (at least up to extreme leadings greater than twice the line height), and that readers subjectively prefer text set with some amount of leading (Bentley, 1921; Paterson and Tinker, 1944, 1947; Poulton, 1972; Tinker, 1963; Wilkins and Nimmo-Smith, 1987). This line of work also highlights the fact that the effect of increased leading may be amplified in variation with other factors, such as text size, horizontal line width, or the typeface used (Becker et al., 1970; Paterson and Tinker, 1944; Tinker, 1963). These same findings are also apparent in investigations of digital typography (Holleran and Bauersfeld, 1993; van Nes, 1986), albeit in paradigms that require unpressured responses from the reader.

In the present study, we expand this line of research by examining the effect of crowded displays on glance-based legibility, more consistent with modern, mobile-oriented reading behaviors. Legibility is assessed using a lexical decision task, which requires readers to classify a briefly presented stimulus as either a word or a nonsense pseudoword. The display time of these stimuli is adjusted via a staircase procedure in accordance with participant performance, to arrive at a reading time threshold per condition studied. This allows for the legibility of different typographic configurations to be compared. Legibility is assessed at two text sizes and with two different degrees of leading. In addition, a condition that presents the lexical decision task in isolation, without crowding distractors, is also compared. We hypothesize that 1) crowded displays will require more time for accurate reading compared to isolated displays, 2) text set at a larger size will be read more easily than text at a smaller size, 3) text with more generous leading (more vertical spacing) will be read more easily than more crowded text, 4) wider leading should ameliorate decrements in legibility arising from smaller text sizes, 5) the effect of crowding will be more pronounced at the smaller text size, and 6) older readers will experience greater increases in reading time for the more difficult conditions.

2. Methods

2.1. Participants

37 participants (between the ages of 35 and 75) were recruited from the Massachusetts Institute of Technology AgeLab's participant pool. Prior to participating in the study, all participants provided written informed consent in accordance with the MIT Institutional Review Board as required by the Declaration of Helsinki. Participants were required to be in "reasonably good health" as reported to experimenters. Participants were excluded from the study if they had experienced a major medical illness or had been hospitalized in the previous six months, or if they had medical conditions that impair vision (beyond those which can be treated with corrective lenses). Participants were excluded if they reported a history or diagnosis of epilepsy, Parkinson's disease, Alzheimer's disease, dementia or mild cognitive impairment, or other neurological problems. All participants were native English speakers, and had normal or corrected-to-normal vision (e.g., with glasses or contacts worn for the experiment). In addition, all participants were assessed for near acuity using the Federal Aviation Administration's test for near acuity (Form 8500-1), and for far acuity using a Snellen eye chart. No participants were excluded due to excessively low acuity (summary statistics are provided in Table 1). As the table indicates, one participant had unusually low distance acuity but adequate near acuity. The opposite is true for one participant with unusually low near acuity. Excluding these participants from analysis did not appreciably change the reported results. In the interest of providing a more varied and robust sample, these participants are retained. Participants were permitted to choose whether they wore corrective lenses during the experiment, and were asked to abide by their choice throughout data collection.

Of the 37 total participants, 7 were excluded, leaving a final pool of 30: 5 failed to reach a stable stimulus duration threshold, resulting in unreliable measurements, and 2 were excluded due to a failure to comply with the experiment protocol. Failure to reach a stable threshold was defined as a calculated threshold value of greater than 600 ms, or if a participant's staircase showed no reversals in the last twenty trials, indicating that the participant had failed to achieve a stable level of performance.

This left a total of 30 participants (mean age = 53.0 years). Age distribution did not differ significantly between genders ($t(28) = 0.32$, $p = 0.754$, *t*-test). Descriptive statistics for the final participant pool are provided in Table 2.

2.2. Task, apparatus, & stimuli

2.2.1. Task

Participants performed a 1-interval forced choice lexical decision task (Meyer and Schvaneveldt, 1971), modified to accommodate an array of distractor words. The stimulus sequence and timings are shown Fig. 1. The word arrays contained a single target word or pseudoword embedded within an array (three columns and five rows) of distractor words. The target word/pseudoword was always presented in the array's center column, and never appeared in the top or bottom row, ensuring that the target was always crowded on all sides. Each lexical decision trial begins with a 1000 ms screen cueing the participant to

Table 1
Summary statistics for binocular acuity measures (all measures taken with optical correction worn).

Acuity	Mean	SD	Min	**25th percentile**	**50th percentile**	**75th percentile**	Max
Near	32.17	8.68	25	25	30	30	60
Distance	25.87	11.01	13	20	25	28.75	70