Introduction.

Fonts are ubiquitous in modern life. Nearly everywhere you look, there are screens and printed words that show text in a variety of fonts. Some fonts have an intended purpose for which they are meant to outperform other fonts. This project, in continuation of a project conducted two years ago, is to create a survey tool to test among fonts designed to be used on highway and street signs, which is the easiest to read under blurred conditions. The type of font and the degree to which the text has been blurred will be the independent variables, and the dependent variable will be how accurately the participant can read the text. These findings will help governments and other authoritative organizations decide which fonts to use on signage, and possibly in other scenarios as well, such as forms or websites.

Six total fonts were used in this experiment. Two fonts are based on Highway Gothic (formally FHWA Series), the official font that the United States uses on the Interstate Highway System and that many other countries use as well. These fonts are Blue Highway, designed to closely resemble Highway Gothic, and Overpass, an open source font inspired by and meant to improve upon the original design. The third font is Transport, designed for use on the United Kingdom's highways. (Capps, 2016) Many other highway systems, particularly in Europe, use fonts derived from Transport. The fourth font is DIN 1451, which Germany uses on many of their roads. (Herrmann, "DIN 1451 (Germany)" 2008) The fifth is RutaCL, designed for use on road signs in Chile, and the sixth is Helvetica, one of the most popular humanist fonts ever created and used on many transportation systems around the world.

There are a lot of people who struggle with blurred/fuzzy vision or related eye problems. Blurred vision affects over 3.6 million people age 40 and older in the United States alone, and over 22 million people of the same age group are affected by cataracts (Friedman, 2008). With this many people who can be at risk, it is important to find the right font for its purpose.

Many studies have shown that certain characteristics of fonts change how well they can be read under certain conditions. For example, there have been many studies that show how sans serif fonts often perform better than serif fonts. One study showed that Swiss, a font very similar to Helvetica, is easier to read in low-light environments than Dutch, a relatively popular serif font comparable to Times New Roman. (Silver & Braun, 2002) In a different study, Helvetica was perceived to be more readable than Times or Goudy, both common serif fonts. (Dean, Aquilante, & Plass, 1998)

Courier, a monospaced font known to be very "blocky" in style, has been proven to outperform other fonts in certain conditions. In one study, Courier was found to be easier to read and have a higher acuity than Times for people with low vision, but Times was slightly better for speed reading by people with normal vision. (Mansfield, Legge, & Bane, 1996) Another study focused on what fonts would ease data retrieval from webpages found little to no difference between fonts. (Ling & Schaik, 2005) In the previous iteration of this project, it was concluded that Courier was easier to read than Times New Roman and Helvetica when under blurred conditions. The current iteration of this project considers whether highway fonts that are more spaced out are easier to read under blurred conditions.

Procedures.

The survey tool was created in the form of a website. The survey was designed to be as accessible as possible and to take only minutes for each participant to complete. The survey posed minimal risk to the participants. English-speaking participants five years old and older were recruited by word-of-mouth and on online forums to take the survey.

The survey tool is secure, operating on a password-protected computer hosted by a company that adheres to strict security protocols. (Amazon.com, 2016) The survey tool stored all results in a secure database on the same server as the survey tool, both of which were only accessible by the researcher. The survey tool anonymously collected the IP addresses of the visitors to the site, but stored the addresses in a separate database from the survey answers, never linking any of the survey data to the IP addresses.

Before beginning the survey, the survey tool presented each participant with an agreement that the participant would willingly take the test and answer the questions honestly, and that the participant would not take the survey more than once, among a few other requirements. The participants were also reminded that they were not required to complete the survey, and that the information collected would not be linked to any personally identifiable information. The participants were then asked to fill out a short questionnaire about themselves, including their age, gender, information regarding their eyesight, and if they drive.

The survey tool presented each participant with a series of five images with a series of four words in each image. Each series was a random sequence of common vocabulary words between four and seven characters long, selected from the top 500 entries in the Corpus of Contemporary American English (COCA). Each series was displayed in one of six common fonts used on highway and street signs. The images were blurred to varying degrees using a Gaussian blur algorithm. The participants were asked to type into a text box their best guesses of what the images read.

Results.

There were 1,357 participants who completed their surveys, providing data on a total of exactly 6,785 images. Since the average level of blur was relatively light, over 80% of the images were correctly read by the participants. Despite this, plenty of data was still available to find correlations and draw conclusions.

A major challenge faced when analyzing the data was how to find a reliable way to compare the original text from which the image was generated to the text submitted by the participants. The metric used is called the Levenshtein distance. The Levenshtein distance between two strings is a value that represents how many characters would have to be changed in order to change one of the strings into the other. Since a Levenshtein distance of zero represents equal strings, a lower Levenshtein distance indicates more similarity between the two strings.

About 53% of participants identified as female, with all but 17 of the other 47% identifying as male. No correlation was found between gender and average Levenshtein distance.

ANOVA sidenote.

A one-way ANOVA test was performed on the data using the font type as the independent variable and the Levenshtein distance as the dependent variable. This test yielded a p-value of 0.00578 and an F value of 7.627. Such a low p-value allowed rejection of the null hypothesis, and the conclusion that the differences between the average Levenshtein distances among the fonts are statistically relevant. Application of the test yet again on images with a Levenshtein score of no more than ten revealed the p-score and F value to be even more supportive, at 0.000148 and 14.41, respectively. This result is shown in a bar chart with the limited Levenshtein distance.

Conclusion.

The purpose of this experiment was to determine which highway font types could be best read under blurry conditions. After performing an analysis of the data, I was successful in identifying which of the fonts used in my experiment can be best read under blurred conditions. After reviewing specific cases in which fonts were misread, patterns began to show themselves. Words that were incorrectly guessed were generally very similar in shape to those in the original image, with specific letters replacing those of a similar shape. For example, the words “effort” and “effect” have two differing letters, but the words are still very similar in spelling and shape. This is a perfect example of the cases in which fonts with more differentiation and more spacing among the characters exceled, and those with uniformly shaped and tightly kerned characters did poorly.

Fonts such as Transport and RutaCL that performed so well in this experiment have very well-differentiated shapes for often similar-looking characters, allowing for them to be identified more easily (e.g. “c” and “o”). Fonts that performed poorly, such as DIN 1451 and Blue Highway, were the opposite: many characters are similar in shape and design, sacrificing legibility for a more uniform look.

This project could be improved in many ways if it were carried out again. I feel that to improve upon this project significantly in the future I would need to reach out to and cooperate with others more. Most of the restrictions that I faced during this process were due to lack of resources, and having the support of one of the government organizations in charge of regulating signage fonts, for example, would enable me to increase the scale not only of my dataset but also of the number of fonts I could test. After presenting this project at the state fair in Virginia, I was put in contact with a few researchers in this field. Their feedback has already been incredibly valuable in helping me to improve upon my experimental design for future experiments, and I have already realized many of these improvements in subsequent versions of my survey tool (see “The Future,” below).

This project is without a doubt the most ambitious I have completed, and most of the challenges that I faced were ones that I did not anticipate when I was planning this experiment. Each step of the way, I found myself problem solving. Whether I was confronting a language barrier when calling the German Ministry of Transport in the middle of the night during their office hours, or figuring out how to perform a one-way ANOVA test, looking back on these challenges makes me very proud of how this project turned out.

The Future.

After proving the potential of this survey tool, I have been working hard to improve upon its design to even better compare highway fonts. Many of the changes I have made are simple refinements to the settings, such as the range of blurriness levels used. I increased the average blurriness level significantly to reduce the number of perfect responses and increase the proportion of relevant data.

Another simple yet important change I have made has been the colors used: in this experiment I only tested participants using images of black text on a white background. However, the majority of road signs in the United States are printed using different color schemes, such as the white-on-green scheme used on most wayfinding signs. The updated survey tool is capable of generating images in the color schemes most-used in the United States.

I also plan to improve the blurring technique used on the images. While I am confident that the Gaussian blur's even distortion of text allows for a fair comparison of font readability, other blur methods could possibly better replicate the perceived distortion of text in highway situations. Most road signs are printed on retroreflective sheeting, which makes text appear to glow outward when light is shined on the sign. This effect is referred to as the overglow effect. I have successfully recreated the overglow effect by lightly blurring an image and then rewriting text onto it, removing any inward blur. By repeating this process many times, I could create a realistic overglow effect. Below are examples of potential survey images created using this technique.

Works Cited.

Capps, Kriston. “Why Every Highway Sign in America Has to Change (Again).” CityLab. N.p., n.d. Web. 19 Dec. 2016.

Davies, M. (2009). The 385+ million word Corpus of Contemporary American English (1990–2008+): Design, architecture, and linguistic insights. International Journal of Corpus Linguistics, 14(2), 159–190. https://doi.org/10.1075/ijcl.14.2.02dav

Herrmann, Ralf. “Traf c Sign Typefaces: DIN 1451 (Germany).” Typography.Guru. N.p., 17 May 2008. Web. 20 Dec. 2016.

Herrmann, Ralf. “Traf c Sign Typefaces: France.” Typography.Guru. N.p., 13 Sept. 2008. Web. 19 Dec. 2016.

Friedman, D. S. (2008, January 1). Vision Problems in the U.S.: Prevalence of Adult Vision Impairment and Age-Related Eye Disease in America. Retrieved October 8, 2014, from http://www.preventblindness.net/site/DocServer/VPUS_2008_ update.pdf

Ling, J., & Schaik, P. (2006). The in uence of font type and line length on visual search and information retrieval in web pages. International Journal of Human-Computer Studies, 64(5), 395-404.

Mans eld, J., Legge, G., & Bane, M. (1996). Psychophysics of reading. XV: Font effects in normal and low vision. Investigative Ophthalmology & Visual Science, 37(8), 1492-1501.

Silver, N., & Braun, C. (1993). Perceived readability of warning labels with varied font sizes and styles. Safety Science, 16(5-6), 615-625.

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

Amazon.com. Amazon Web Services: Overview of Security Processes 2016. Print. Retrieved December 19, 2016, from https://d0.awsstatic.com/whitepapers/Security/AWS_Security_Whitepaper.pdf