Examining the Legibility of the Number "1" and the "÷" Symbol

by Doug Fox, Barbara S. Chaparro, & Ed Merkle

Summary. This article continues the investigation of the legibility of onscreen typefaces and the influence of individual character features on correct identification. Specific attributes of alphanumeric characters and symbols shown to be the least legible were measured and analyzed using a statistical method called classification tree analysis. Results from this analysis for the number "1" and the "÷" symbol are discussed.

INTRODUCTION

The legibility of individual characters is important in situations that involve codes, such as serial numbers, passwords, email addresses, air traffic controller displays, etc. Computer users are commonly frustrated when they are unable to install a program correctly because they cannot read the serial code, or type in the wrong email address because they are unable to decipher the characters in the address. In more serious situations, unclear characters on computer displays have been blamed for air traffic controllers directing planes into wrong sectors. A report by the Health and Safety Executive suggested one of the biggest safety issues for air traffic displays is the fonts used (newsscientisttech, 2002). For these reasons it is important to study the legibility of individual characters.

In the last issue of Usability News, we evaluated the legibility of two commonly confused characters, the number "0" and the letter "e". Classification tree analysis was used to analyze features which were most influential to zero and "e". From this research, recommendations can be made to typographers on the best way to design characters for optimal legibility.

In this article, two more problematic characters (the number "1" and the "÷" symbol) are examined. These characters were tested in the same conditions as the number "0" and the letter "e" (see Fox, Chaparro, & Merkle, 2007). Based on the frequency of misclassifications, these two characters were less legible than most characters. Using classification tree analysis, the features that influence the legibility of these two characters will be analyzed and discussed.

METHOD

Participants

Ten participants (6 male, 4 female) between the ages of 18-25 volunteered for this study. All participants had at least 20/20 vision. All were compensated $50 for their participation.

Materials

A Dell Core 2 Duo laptop with a TrueLife™ display and ClearType™ font rendering was used. The screen resolution was 1920 x 1080 with 147 dpi. A program was written in C# to display the characters.

Characters were displayed in a 10-point font size, and the laptop was positioned at a distance so that the characters, regardless of typeface, were viewed at a visual angle of .08°. Participants used a chinrest to stabilize their head and maintain a constant distance from the monitor. Twenty-six lowercase letters were used in combination with the digits 0-9 and 11 symbols frequently found in mathematical or scientific documents. The symbols used included: ÷ = + ? % ± $ # @ & !. The chosen symbols had a character height similar to the lowercase letters.

Procedure

Participants were shown a sample of what the characters looked like before each trial started. Each trial started with a "•" to indicate where the characters were going to appear on the monitor and ended with a "•" to indicate that the trial was over. Each participant participated in four trials for each typeface with the first trial being excluded from data collection as practice. Each trial consisted of all 47 characters being exposed briefly (34 ms) one at a time with a blanking time of 1.5 second between each exposure. The order of the characters and the typefaces was randomized. Characters were displayed in black type on a white background. Character identification was read aloud by the participants and accuracy was recorded by the experimenter.

RESULTS

Percent correct was calculated for all characters across the twenty typefaces. The order of percentage correct for the division symbol is shown in Table 1, and the number "1" in Table 2. The division sign was misclassified the most when presented in Bodoni (43.30%), and was misclassified the least when presented in Centaur, Garamond, Georgia, Rockwell, or Verdana (100%). The mean percentage correct for division signs across all 20 typefaces was 88% (SD=15.12). For the number one, the worst typeface was Centaur (36.70%), and the best typeface was Clearview Text (96.70%). The mean percentage correct for the number one was 66% (SD=19.78).

**Table 1. Percentage correct of identifications for the "÷" sign.**

Typeface	Character	Percentage correct
Bodoni	÷	43.30
Gulliver	÷	56.67
Cambria	÷	73.30
Candara	÷	83.30
Constantia	÷	83.30
Corbel	÷	83.33
Baskerville	÷	86.70
Clearview Text	÷	86.70
Calibri	÷	90.00
DIN 1451 Mittleschrift	÷	93.33
Franklin Gothic ITC	÷	93.33
Frutiger	÷	96.67
Futura	÷	96.67
Helvetica	÷	96.67
Consolas	÷	96.70
Centaur	÷	100.00
Garamond	÷	100.00
Georgia	÷	100.00
Rockwell	÷	100.00
Verdana	÷	100.00

Table 2. Percentage of correct identifications for the number "1".

Typeface	Character	Percentage correct
Centaur	I	36.70
Candara	1	43.30
DIN 1451 Mittleschrift	1	43.33
Rockwell	1	43.33
Constantia	1	46.70
Georgia	1	50.00
Garamond	1	53.33
Frutiger	1	56.67
Futura	1	56.67
Corbel	1	60.00
Gulliver	1	66.67
Baskerville	1	66.70
Bodoni	1	73.30
Helvetica	1	80.00
Calibri	1	83.30
Cambria	1	83.30
Consolas	1	93.30
Franklin Gothic ITC	1	93.33
Verdana	1	93.33
Clearview Text	1	96.70

Classification tree analysis was used in this study to determine which features (IVs) of the characters were the most influential in correct identification. Classification tree analysis is a nonparametric statistical procedure that identifies homogenous subgroups (nodes) to accurately predict a dependent variable (DV) chosen by the researcher. The subgroups created share common characteristics that influence the DV. A unique visual structure (see Figure 1) that links the nodes together is the result of the classification tree analysis. The branching that is made is based on the analysis of all the independent variables (IVs). The IV that creates the most differentiating groups is selected as the branch IV. This procedure is repeated creating more nodes until there are no other IVs that influence the DV (Lemon, Roy, Clark, Friedmann, & Rakowski, 2003).

In this study, the DV measure was the number of misclassifications; thus, the more times a character was misclassified the less legible it was. The classification tree analyzed all of the features (apex, area, complexity, contrast, height, midline, perimeter, stroke variation, weight, and width) to determine which feature contributed the most to misclassification.
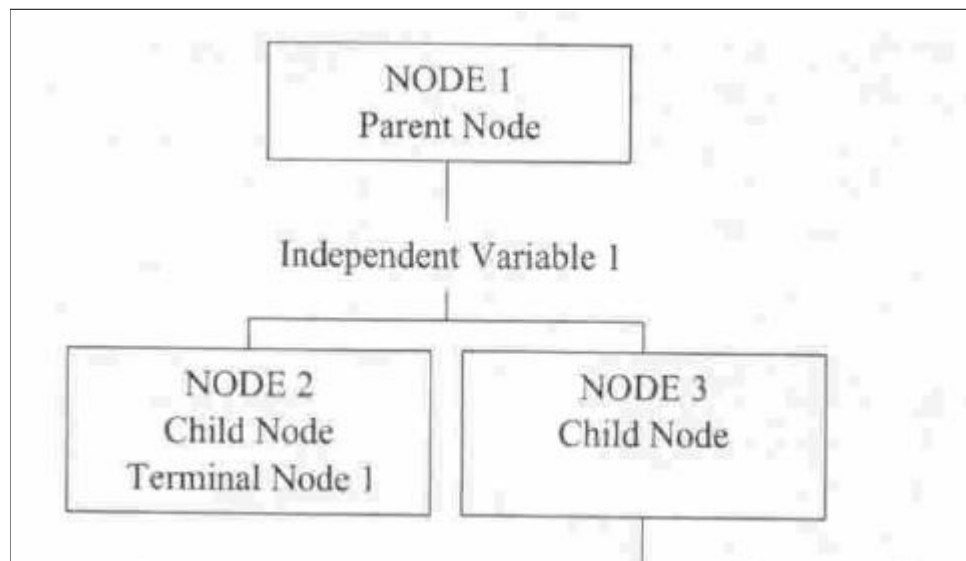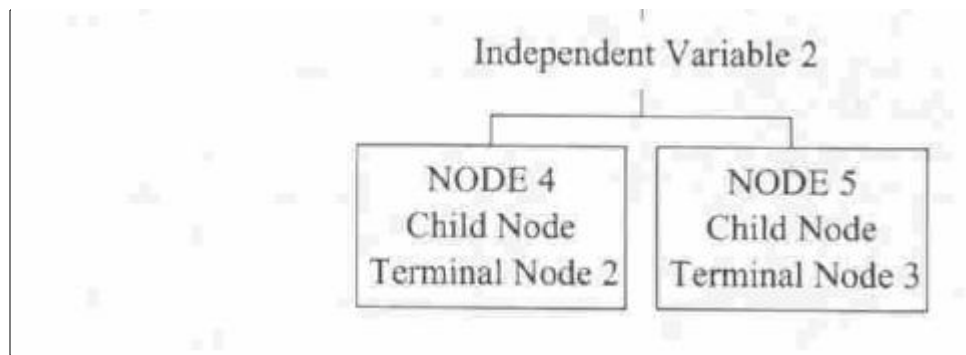
Figure 1. Example of classification tree (Lemon et al., 2003).

The Division Symbol

Classification tree results for the division symbol (see Figure 2) suggest that the height between the stem and dot was the most influential feature. Typefaces that had a height between the stem and dot greater than or equal to 2.5 pixels were correctly identified 497 trials out of 540 presentations which yields a 92% correct identification rate. Conversely, if the height between the stem and dot was less than 2.5 pixels it was correctly identified 30 trials out of 60 presentations which yields a 50% correct identification rate.

When the height between the stem and dot was less than 2.5 pixels, weight was the next most influential feature. When weight was greater than or equal to .91, the division symbol was correctly identified 17 out of 30 trials yielding a 57% correct identification rate. When weight was less than .91 there were 13 correct classifications out of 30 which yields a 43% correct identification rate.

Figure 3 shows an example of a division symbol that was legible and one that was not legible. The Bodoni division symbol (left) has a height less than 2.5 pixels between the bottom of the dot and top of the stem, while the Rockwell division symbol has a height greater than 2.5 pixels between the dot and stem (right). Division symbols that had a height between the stem and dot greater than 2.5 pixels were more legible than those that did not.
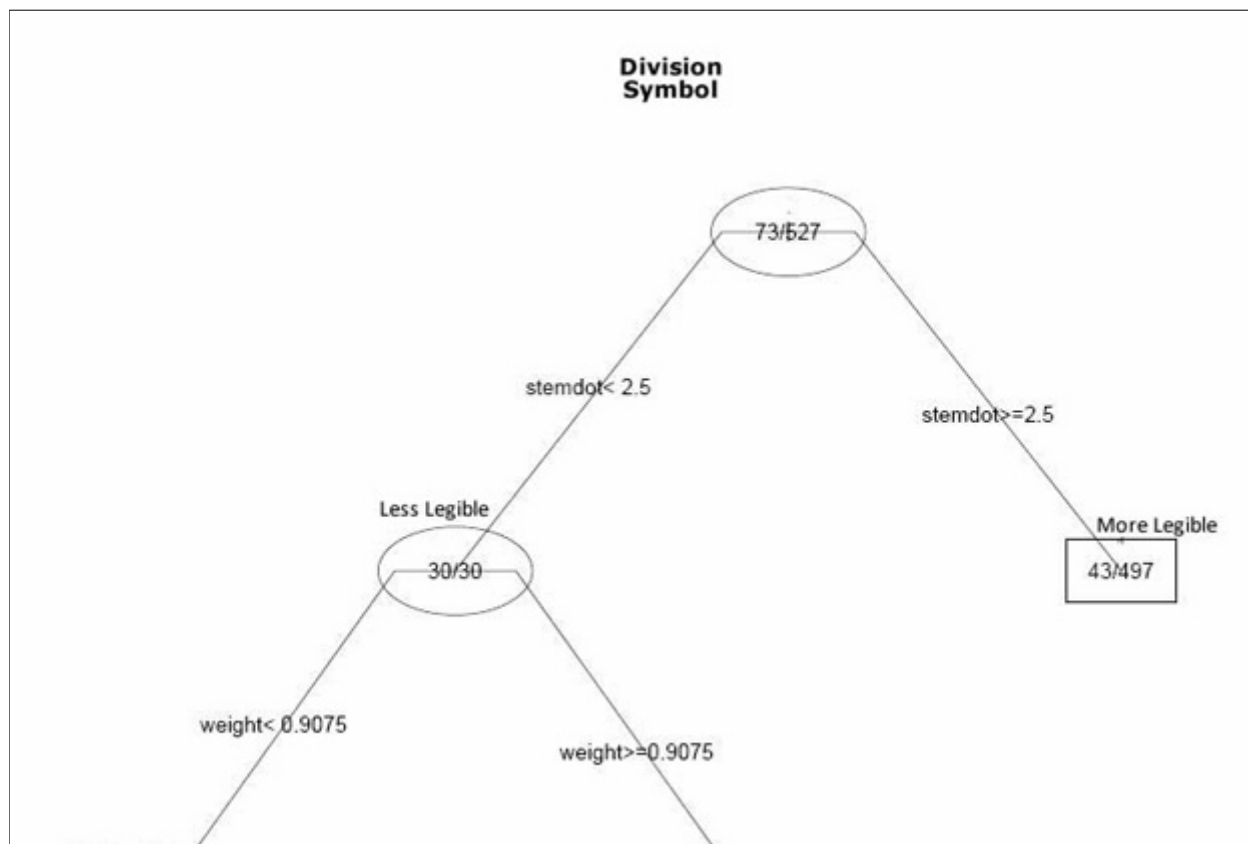
Figure 2. Classification tree for the division symbol.

height between
stem and dot
< 2.5 pixels

height between
stem and dot
> 2.5 pixels
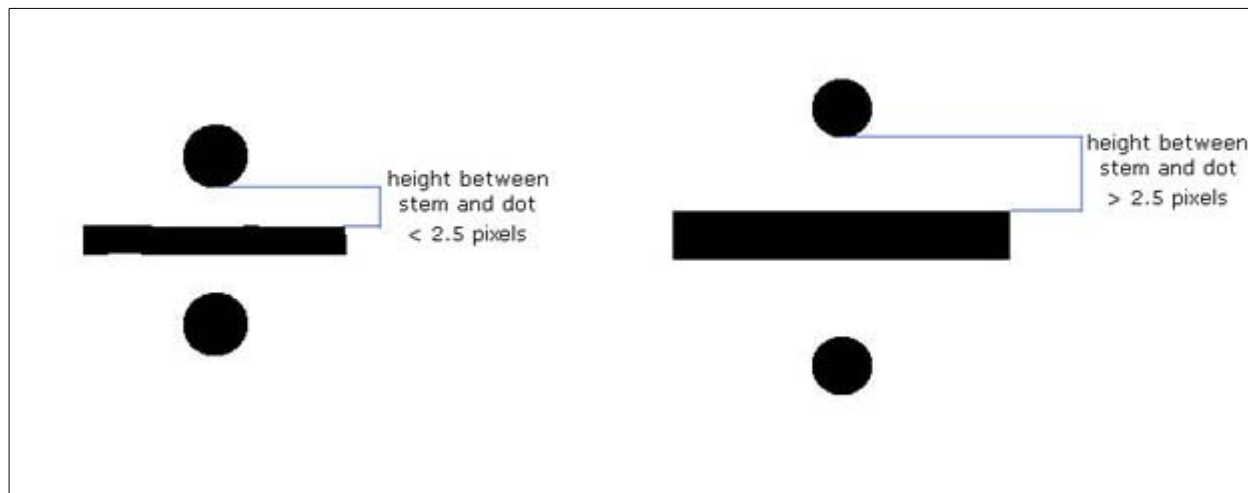
Figure 3. The Bodoni division symbol (left) has a height less than 2.5 pixels between the bottom of the dot and top of the stem, while the Rockwell division symbol has a height greater than 2.5 pixels between the dot and stem (right). Classification tree results suggest that had height less than 2.5 pixels were confused more.
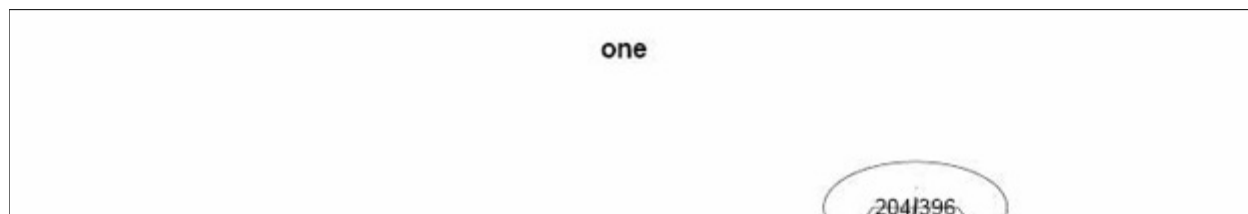
The Number One

Classification tree results for the number one (see Figure 4) suggest that perimeter is the most influential feature. A perimeter measure of 136 or greater had 163 correct identifications out of 180 which yields a 91% correct identification rate. On the other hand, a perimeter less than 136 had 233 correct identifications out of 420 which yields a 55% correct identification rate. Perimeter is the feature that measures the outside of the character. It is similar to the perimeter of a geometric shape, such as a square.

If perimeter was less than 136 then the apex of the one was influential in misclassification. If the apex was less than 1 then there were 165 correct classifications out 270 which yields a 61% correct identification rate. Typefaces that presented a one with an apex value greater than or equal to 1 had 68 correct identifications out of 150 presentations which yields a 45% correct identification rate.

Height was the next influential feature for number ones with an apex less than one. A height greater than or equal to 14.5 pixels had 17 correct identifications out of 30 which yields a 57% correct identification rate. A height less than 14.5 pixels had 51 correct identifications out of 120 which yields a 43% correct identification rate.

In Figure 5, the one on the left had a perimeter less than 136, and the one on the right had perimeter greater than 136. The Candara number one had a perimeter less than 136, and was perceived as less legible. The Clearview Text number one on the right had a perimeter greater than 136 and was perceived as more legible.
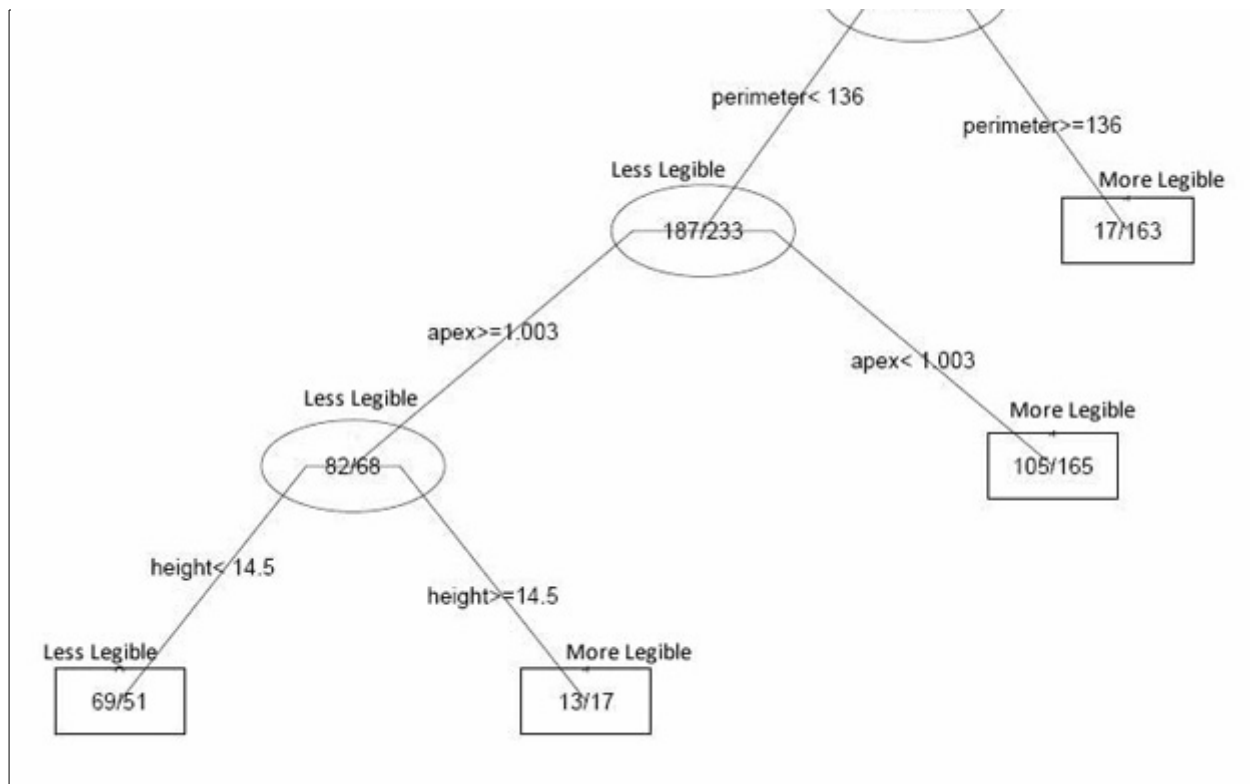
one

204/396

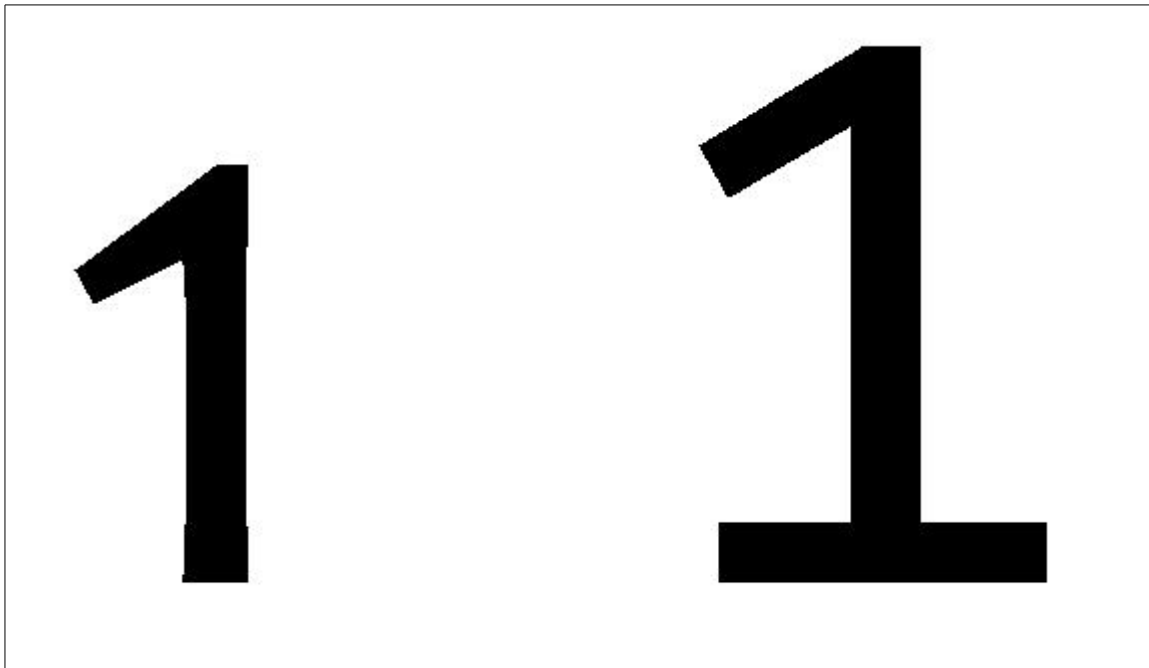**Figure 4. Classification tree for the number one.**

Figure 5. The Candara number one (left) has a perimeter value less than 136 compared to the Clearview Text number one (right) that had a perimeter value greater than 136. Classification tree results suggest that a perimeter less than 136 results in a less legible"1".

DISCUSSION

Results from the classification trees suggest that the height between the stem and dot was the most influential feature for the legibility of the division symbol. A division symbol that had a height greater than or equal to 2.5 pixels was more legible than a division symbol with a height less than 2.5 pixels (see Table 3). Thus, clear separation between the dot and stem is needed for the division symbol to be

legible. Otherwise, it will more than likely be mistaken for an addition symbol or possibly the letter "t".

Table 3. Percent correct for identification for all typeface renderings of the division symbol. Division symbols that had a height between the dot and stem less than 2.5 pixels were less legible. The bold bar separates the division symbols that had a height less than 2.5 pixels and those had a height greater than 2.5 pixels.

Typeface	Character	Percentage correct	Height b/w stem & dot
Bodoni	÷	43.3	2
Gulliver	÷	56.67	2
Cambria	÷	73.3	4
Candara	÷	83.3	8
Constantia	÷	83.3	3
Corbel	÷	83.33	4
Baskerville	÷	86.7	6
Clearview Text	÷	86.7	4
Calibri	÷	90.0	4
DIN 1451 Mittleschrift	÷	93.33	4
Franklin Gothic ITC	÷	93.33	5
Frutiger	÷	96.67	4
Futura	÷	96.67	4
Helvetica	÷	96.67	5
Consolas	÷	96.7	4
Centaur	÷	100.0	6
Garamond	÷	100.00	4
Georgia	÷	100.00	6
Rockwell	÷	100.00	5
Verdana	÷	100.00	4

In addition to the height between the stem and dot, weight was an influential feature for the division symbol. A weight greater than .91 was more legible than a weight less than .91. This suggest that the thickness and darkness of the bar and dot should be bolder to be perceived more legible. However, this is only a suggestion for the minimum of what the weight should be; the maximum weight that will yield a legible division symbol was not considered.

For the number one, perimeter was the most influential feature (see Table 4). A one with a perimeter greater than or equal to 136 was more legible than a one with a perimeter less than 136. Even though Centaur was the least legible typeface representation for the number one, the legibility of this character could not be contributed to the perimeter (even though it was less than 136) but rather the unusual shape of the character. However, the other representations of ones seem to follow the pattern that legibility is influenced by the perimeter. Results suggested that two features that contribute to the perimeter of the one, apex and height, influence legibility. Thus, it seems a number one that has larger perimeter will make it more distinguishable from characters like the letter "l".

Table 4. Percent correct for identification for all typeface renderings of the number one. A number "1"with a perimeter less than 136 was less legible. The bold bar separates the number "1"'s that had a perimeter less than 136 and those had a had a perimeter value greater than 136.

Typeface	Character	Percentage correct	Perimeter
Centaur	I	36.7	123

Candara	1	43.3	87
DIN 1451 Mittleschrift	1	43.33	114
Rockwell	1	43.33	128
Constantia	1	46.7	96
Georgia	1	50.00	121
Garamond	1	53.33	118
Frutiger	1	56.67	96
Futura	1	56.67	120
Corbel	1	60.00	111
Gulliver	1	66.67	133
Baskerville	1	66.7	123
Bodoni	1	73.3	132
Helvetica	1	80.00	116
Calibri	1	83.3	139
Cambria	1	83.3	139
Consolas	1	93.3	151
Franklin Gothic ITC	1	93.33	157
Verdana	1	93.33	154
Clearview Text	1	96.7	144

The results from the classification tree analysis are helpful by defining the features that influence legibility the most. This is also helpful to practitioners who may be deciding which typeface to use for applications where users need to discern critical information that is displayed in alphanumeric codes.

Note: A paper based on this work will be presented at the Human Factors and Ergonomics Society's 52nd (2008) Annual Meeting in New York, NY.

REFERENCES

Fox, D., Chaparro, B., & Merkle, E. (2007). Examining the Legibility of the Letter "e" and Number "0" Using Classification Tree Analysis. Usability News, 9.2. Retrieved May 27, 2008 from http://www.surl.org/usabilitynews/92/legibility.asp.

Lemon, S. C., Roy, J., Clark, M. A., Friedmann, P. D., & Rakowski, W. (2003). Classification and regression tree analysis in public health: Methodological review. Annals of Behavioral Medicine, 26(3), 172-181.

Unclear displays lead to air traffic controller errors. (2002, May 23). Retrieved January 15, 2008, from NewScientistTech Website: http://technology.newscientist.com/article/dn2318

SUBSCRIBE to *Usability News*!