Exploring Character Recognition through Modeling:
Cognitive Reverse Engineering

Jeffrey Cochran, Jason VanBuiten, Michael Smith

Abstract

What makes letters harder or easier to read? Moreover, what factors are most important to the brain for successful character recognition? These were the concepts investigated in this study. A "critical point theory" was developed that suggested that the brain uses intersection points and points of inflection in letters for pattern recognition. An experiment was developed in which letters were shown to subjects that had certain factors controlled such as occlusion of the letter, skew, and frequency of the letter in English. Data was collected as the time it took for the subjects to name each letter. A multiple linear regression model was developed, analyzed, and refined based on this data. The model was intended to be used to recognize what factors were significantly important for character recognition, and to predict response time of hypothetical letters. Statistical evidence of the validity of the critical point theory was found. However, the study suffered from poor experimental design, and the resulting regression model was unsatisfactory for use in prediction. A follow-up study with a revised experiment is suggested, and our results are discussed.

Most of the higher-level functionalities of the brain are still mysterious unknowns. We know that the brain can do a "mind-boggling" (so to speak) variety of things, such as storing memories, learning to swim, and making snap-decisions. The real question to be asked is *how* does it do these things? These functionalities for the most part are "black boxes." We know generally what they do, but not how they do it. The goal of this study is to focus on the letter recognition black box, and rather than "opening it" by figuring out how it works inside the brain, we aim to make initial steps toward reverse engineering a math and computer programming based emulation. The mechanisms of letter recognition should be governed by the theories of pattern recognition and linguistics, assuming that these theories are true. One such theory of pattern recognition is Irving Biederman's Recognition-By-Components theory (Biederman, 1987), which describes a method of pattern recognition in which visual stimuli are broken down into corresponding three-dimensional building blocks called "geons." The observed geons and their interrelations are matched against existing models. The models which most accurately fit the visual stimuli are what the actual objects being seen are recognized as. A three-dimensional method of parsing two-dimensional characters seems to be somewhat illogical. Instead, we propose a new structural analysis-based theory specific to the recognition of characters. Our theory suggests that characters are recognized by the brain via the points on the character where the strokes intersect, and the vertices and sometimes points of inflection of curves. These “critical points” and their layout are used as the data that comparisons to stored critical point information are based on. This study aimed to prove or disprove this new critical point theory, and to develop a model that can be used to predict the difficulty of character recognition based on several factors.

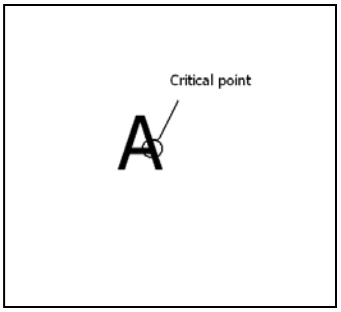

Fig. 1: An example of a critical point

In a general black box model, there are three parts. The input is the part of the model which is a list of objects that are passed to the black box for it to act upon. The second part is the black box itself, which performs some mysterious operation whose implementation and algorithm(s) are unknown. After performing this task, there is usually some end-result that comes out the other end of the black box. This is the third part: the output. The first problem we are facing in this study is that our black box model of character recognition has one part missing: the input. If we are to make an emulation of this function of the brain, we need to first determine what information the mystery function acts upon.

In order to investigate which factors are used as input, we designed an experiment that consisted of showing people letters which had several factors controlled. For example, some letters were occluded by use of an image manipulation program. Some had critical points occluded, and some were occluded without covering critical points. By judging how long it took subjects to identify letters, we could investigate which factors were used by the brain in character recognition. The logic behind this is that if a certain factor’s “quality” is decreased, or is not present, and the factor is indeed used by the brain to recognize a letter, the difficulty of reading the letter will increase. Such an increase in difficulty should be reflected in an increase in response time.

Method

Participants

The participants consisted of family, friends, and students. Age, gender and ethnicity were not controlled; however the participants were from a variety of different locations.

Fig. 2: Sample of images used. Left to right: Control, Critical Occlusion, Noncritical Occlusion, Critical Bend, Noncritical Bend

Materials and Design

The experiment was designed as a slide-show presentation on a laptop. The laptop's resolution was set at 1920 by 1200 pixels, and the backlight was set to full brightness. The laptop was using 32-bit colors. The slide show consisted of five sets of ten letters. The letters were chosen at random by a pseudorandom number generator. The case of each letter was chosen by a coin flip. If the letter chosen by the number generator already was present in the set, the new letter was discarded and a new letter generated so that there was no repetition of letters within each set. The design allowed for repetition across sets, however. Each letter was then typed in the image editing program GIMP-2.0 (The GIMP Development Team, (No centralized location)). The letters were typeset in Sans, with a font size of 18 pixels. Each image was 640 by 640 pixels in size, and the letters were centered on each image, with one letter per image. For all images, the background was white (0xFFFFFF)[1] and the font was black (0x000000). The five different sets contained these images of letters with certain factors controlled. The first set was control, and the letters were not modified. The second set was the critical occlusion set (CO). Each letter in the critical occlusion set had one or more critical points covered. A blue paintbrush tool was used in the image editor to occlude these points. The number of points occluded, and amount of surface area covered was not very random, as the paintbrush tool was applied manually. The blue (0x0000EE) "paint" was applied such as to be amorphous in shape, and to not contain any negative space within the boundaries of the "blob." The third set of letters was the noncritical occlusion set (NO). The letters were occluded in the same method as in the critical occlusion set, with the difference being that none of the letters' critical points were occluded. The fourth set of letters was the critical bend set (CB). In this set, the relative orientations of some or all of the critical points were altered. This was achieved by placing marks on the ends of the brush strokes for each brush stroke in the already-drawn font. The marks were connected using either a linear vector tool (for straight or angled lines), or a Bézier curve vector tool (for curves). The lines and curves were then modified such that the critical points would have a relative orientation that was neither the original orientation in the font,

[1] All color mixing values are RGB, and represent additive color mixing.

nor an orientation that is a scaled equivalent of said orientation. The vectors were then stroked in black (0x000000) with a stroke weight of 300 pixels, and the original font was deleted. The fifth set was the noncritical bend set (NB). The noncritical bend set was created and “bent” in the same manner by which the critical bend set was, with the difference being that the critical point relative orientation was left intact, and other parts of the letter were “bent.” The images were exported into the portable network graphic (png) format. A sample letter from each set can be seen in Figure 2 (note that the images were scaled down from their original sizes in order to fit in this paper). Occlusion levels in each image from the occlusion sets were quantified by dividing the number of black pixels in the image used by the number of black pixels in the image before the “paint” was applied. The resultant values for each image were recorded as occlusion percentages. Each of the control letters were recorded as having 0% occlusion. The frequency of each letter in English was obtained from Tom Linton’s old cryptology class home page from when he was employed at Moravian College (Linton, 2001)[2], and recorded. A length-fifty number sequence was entered into a website which randomizes sequences using atmospheric noise in its randomizing function (www.random.org). The images were then entered into a Microsoft Office Power Point 2007 Version 12.0.6307.5 (Microsoft, Washington) slide show presentation in the random order generated from the website. There was one image per slide. The images were centered on their slides. The slides had white (0xFFFFFF) backgrounds. Between each slide, a grey (0xBFBFBF) slide was inserted.

Procedure

The subjects and the experimenters sat at a table in front of a laptop. The first slide in the experiment read “You will be shown a letter, and then a gray screen. The letter may be bent, or have paint covering it, or have nothing wrong with it. Tell the experimenter what letters you see. You do NOT need to say if the letter is capital or lowercase.” The experimenter held a cell phone with a timer function and a notebook. Both were kept out of sight of the subject. The experiment was performed immediately, and there was no “practice” segment of the experiment. When the subject said he or she was ready, the experimenter changed the slide to the first character, and pressed the start key on the timer at the same time. As soon as the subject responded, the experimenter changed slides to the grey screen and pressed the stop key on the timer. The timer stopped as soon as the subject started speaking; the experimenter did not verify that the answer was correct before hitting the stop key. If the subject’s response was correct, the experimenter wrote down the time taken to respond and then moved on to

[2] Note: This page is from Moravin College's Math 390 cryptography course site. The class was offered in the spring of 2001. The table the data was drawn from is titled "Relative Frequencies of Letters in General English Plain text"

the next slide. If the letter said was incorrect, the experimenter wrote "FAIL." The experimenter made sure that the subject was not able to see their time on the timer, or any of the times written down on the notebook. After the experiment, the experimenter subtracted his own mean reaction time from all of the times recorded. All these times, along with the data described in the "Materials and Procedures" section of this paper, were entered into a spread sheet for analysis.

Results

The data was divided into two parts for analysis: bend and occlusion. We will first discuss the occlusion data, and then the bend data.

Occlusion

The mean times for the critically occluded data (CO) and the noncritically occluded data (NO) (mean = 0.6559524, SD = 0.1065119; mean = 0.6171905, SD = 0.09884119) were found to differ significantly from each other through an ANOVA F-test (F = 497.89, p < 2.2e-16, DF = 2, 627). The interaction between the type of occlusion (CO, NO, or control) and occlusion percentage was found to be insignificant (p = 0.09467). This discounts the possibility that a difference in occlusion could have falsely suggested that CO has higher overall reaction time than NO. Via stepwise regression, a multiple linear regression model was derived. The dependent variable was response time in seconds. The explanatory variables were occlusion type (b_{CO} = 0.2775859, SD = 0.0188549, p < 2e-16; b_{NO} = 0.2369551, SD = 0.0188404, p < 2e-16); position in the experiment (b = 0.0005275, SD = 0.0002326, p = 0.0237); and occlusion (b = -0.0007018, SD = 0.0004284, p = 0.1019). As can be seen best in Figure 3, the occlusion slope is practically zero. Additionally, the p-value does not reject the null hypothesis that the slope is zero. Upon removal of this term, R^2 only drops .16% (with: R^2 = 61.86%; without: R^2 = 61.70%). Therefore, occlusion was removed. This left us with only occlusion type and position.

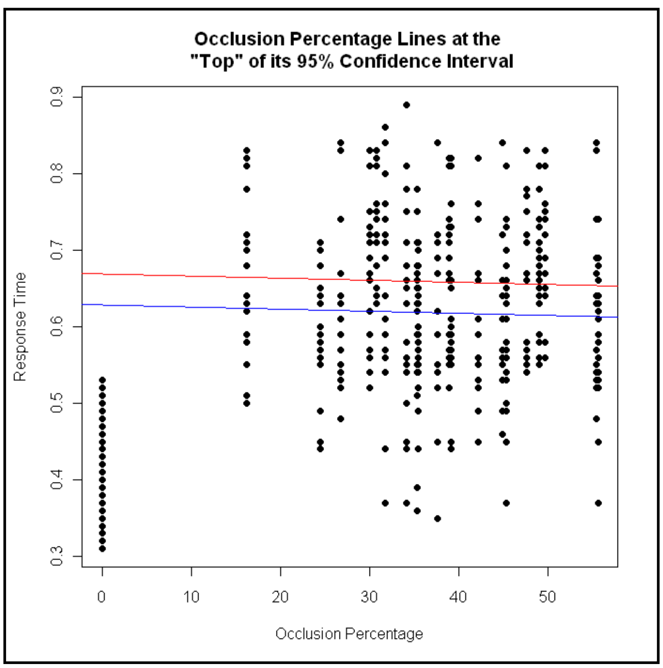

Fig 3: Occlusion Percentage Lines

Bend

Unfortunately, we were unable to devise a method of quantifying how bent a letter is, so we couldn't analyze bend with the occlusion data for a more complete model. An ANOVA F-test was performed on the bend data. It was shown that critical bend was a significant factor in predicting response time ($F = 52.719$, $p = 2.046e-12$, $DF = 1, 397$). The mean response time for critical bend data, as predicted, averages higher (see Figure 4).

Discussion

Our study suffered from a lack of randomness in the experiment across participants. The model that was created only had two factors: occlusion type and position. It did not have enough predictive power to be useful. As can be seen in Figure 3, there is a "hole" in the occlusion percentage data between zero and fifteen percent occlusion, and the occlusion percentage goes no higher than about 57% or so. We hypothesize that this is the cause of the poor occlusion model. We did show evidence supporting our claim of this critical point theory, and it may be valuable to perform a revised experiment that creates a better critical point model, and compare the predictive power against other theories, such as Recognition-by-Components theory. A better experimental design would have letters, occlusion, and sequence generated on-line at the time of each experiment. This way we would have a more representative data set. This study was designed to be exploratory, and in that it can be deemed successful. With a properly revised study, it may be possible to start scratching away at the black box of pattern recognition, and ultimately determine how pattern recognition occurs at an algorithmic, if not implementational level.

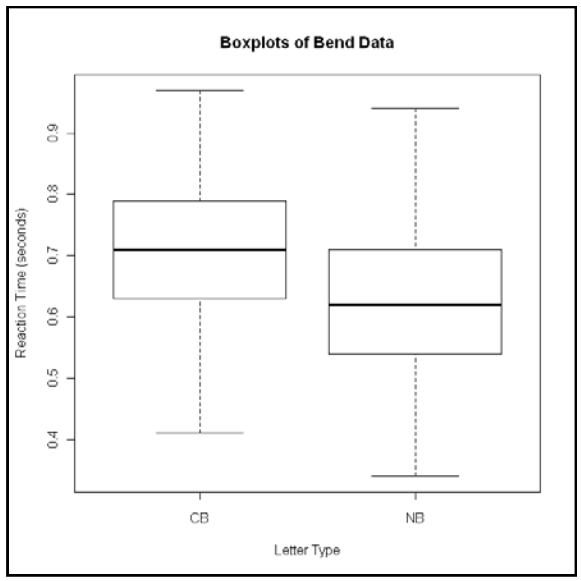

Fig. 4: Bend Data

References

Biederman, I. (1987). Recognition-by-components: A theory of human image understanding. *Psychological Review , Volume 92* (2), 115-147.

Linton, T. (2001, Unkown Unknown). *Letter Frequency.* Retrieved December 18, 2009, from Cryptography: http://pages.central.edu/emp/LintonT/classes/spring01/cryptography/letterfreq.html