THE EFFECT OF FONT DESIGN CHARACTERISTICS ON FONT LEGIBILITY

Yan Zhang

A Thesis

In

The Department

of

Computer Science and Software Engineering

Presented in Partial Fulfillment of the Requirements

For the Degree of Master of Computer Science at

Concordia University

Montreal, Quebec, Canada

August 2006

Your file Votre référence
ISBN: 978-0-494-20785-7
Our file Notre référence
ISBN: 978-0-494-20785-7

ABSTRACT

The effect of font design characteristics on font legibility

Yan Zhang

Human eyesight has deteriorated over the past decades. One of the main reasons is that we have to read a lot. The long-term goal of our research is to identify legible print-fonts or invent some new print-fonts which will sooth our eyes and prevent damages from too much reading. In this thesis, from a pattern recognition point of view, we focus on the study of the effect of font design characteristics on font legibility.

First, the effects of typeface on legibility are investigated in an experiment. Studied are nine sans serif typefaces and nine serif typefaces. OCR systems, one of the most successful applications of machine reading, were used in legibility measurements. The results show how the legibility is affected by different fonts.

Second, the detailed font design characteristics were studied. They include serif and sans serif design, x-height, ascender and descender of the typeface, the inter-letter spacing setting, the weight of the stroke and the distinctive character features. In order to measure them mathematically and objectively, each characteristic is represented by one or two legibility factors and the extraction methods of some factors are proposed.

In addition, the statistics on the common recognition errors of the eighteen fonts are studied. How the legibility factors affect and contribute to legibility is analyzed. Moreover, design suggestions are made on typeface characteristics to maximize typeface legibility.

Finally, one of the important font characteristics, inter-letter spacing, is studied. We addressed the issue of how the condensed or expanded inter-letter spacing setting contributes to legibility. By measuring the legibility of different inter-letter spacing settings for a given typeface, and analyzing their corresponding touching errors, better inter-letter spacings for reading are suggested.

Acknowledgements

First and foremost, I would like to extend my sincere gratitude to my supervisor, Dr. Ching Y. Suen, for his guidance, enthusiastic support, valuable suggestions and encouragement throughout my Master's study.

I would further like to thank all of my colleagues and friends, especially those at CENPARMI. Thanks to Dr. Qizhi Xu for her guidance and helpful suggestions; thanks to Nicola Nobile for all his thoughtful suggestions and for providing his Dynamic String Matching function DLL, used in my program; thanks to Chunlei He for all her encouragement and useful suggestions. Thanks also go to: Dr. Wumo Pan, Dr. Jianxiong Dong, Dr. Ping Zhang, Ning Wang, Yun Li, Katherine Zannibi, Wu Ding, Sapargali Kambar, Guiling Guo, Sara Izadi, Javad Sadri, etc., who helped me in one way or another. I would also like to thank Ms. Shira Katz for her assistance in correcting the grammar in my thesis and in enhancing my English writing skills.

Finally, I would like to thank my husband, Yu Wang, and my parents, for being with me, loving me, and believing in me. Without their continuous support and encouragement, I would not have been able to complete this thesis.

CONTENTS

List of Figures

List of Tables

Chapter 1

Introduction

In this Chapter, the motivation and objectives of this thesis are introduced. We also review the state of the art in terms of typography terminologies, legibility versus readability and the legibility factors on font characteristics. Finally, the outline of this thesis is presented.

1.1 Motivation and Objectives

In this thesis, we focus on the effect of font design characteristics on font legibility. The legibility of different fonts is measured and compared by experiments. Some typical font characteristics that have been commonly researched in previous legibility studies are described. We propose measurement and extraction methods for some of the main design characteristics. By using statistical analysis, the correlations between the designed font characteristics and font legibility are drawn. After a general study between the font characteristics and legibility, one of the most important characteristics, inter-letter spacing setting, is examined in detail.

People read every day. Engineers read manuals; students read textbooks; people read newspapers and magazines. Reading is unavoidable in most people's lives. Seeing, perceiving, and recognizing lines and dots as a form of language is a process that is

extremely complicated, yet necessary. Unfortunately such activities have brought unprecedented demand on our eyes. We can easily see that many children and adults in this modern society have damaged their eyes. Scientists have studied many aspects of the visual reading process and one of their concerns is in the field of legibility. A minimum requirement for text type to be read is that it must be legible [2]. Legibility, according to the dictionary, mindful of the Latin root of the word, means the quality of being easy to read [1]. In typography, the definition is drawn a litter closer to scientific research, where legibility is defined as the quality of being decipherable and recognizable. The long-term goal of our research is to identify legible print-fonts or invent some new print-fonts which will sooth our eyes and prevent them from damages of too much reading.

Optimal legibility of print is affected by several different aspects, such as the spatial arrangement of a printed page, the color of print and background, printing surfaces, special printing situations, reading situations, or typographical factors. Tinker [11] mentioned that "Legibility . . . is concerned with perceiving letters and words, and with the reading of continuous textual material. The shapes of letters must be discriminated, the characteristic forms perceived, and continuous text read accurately, rapidly, easily, and with understanding. In the final analysis, one wants to know what typographical factors foster ease and speed of reading." It has been well accepted that typographical factors affect legibility [2] [3] [4] [7] [11].

Several typeface characteristics such as color, weight, size, serif design, character spacing, distinctive features and others have been studied and they have been considered as factors that contribute to legibility [2] [8] [10] [12]. Although there are many research studies on font characteristics with legibility, this topic is still an exciting issue for two main reasons.

First, since too many research results on this topic seem counter-intuitive and even conflict with each other, more studies are still needed. For instance, in the typographic literature, some researchers give different views on whether serif fonts versus sans serif fonts are more legible (see section 2.1). Second, so far, most of the studies of how the font characteristics affect legibility have been limited to the fields of psychology, physiology and typography design. For psychological or physiological specialists, by using visual acuity, reading speeds, eye movement, or even frequency of eye blinks as the legibility criteria and using human beings as subjects, they constructed experiments to study the relations between legibility and some font characteristics. For typography designers, as Walter Tracy mentioned in his book, "Letters of Credit", designing legible fonts is a proper balancing of function and aesthetics. Type designers have given their advice or suggestions on legible font design mainly based on their design experience and taste.

In our study, we examine how font characteristics contribute to legibility from a new point of view. Firstly, an experiment on comparing font legibility is constructed and eighteen typical typefaces in two sizes are studied. Different from other legibility experiments, in this thesis, Optical Character Recognition (OCR) systems, one of the most successful applications of machine reading, have been used instead of humans, to read printed documents in different fonts.

After comparing the legibility of different fonts, our next step is to analyze how the font characteristics affect legibility. In order to measure font characteristics mathematically and objectively, each characteristic is represented as one or two legibility factors and the extraction methods of those factors are proposed. Then each factor is evaluated for every studied font.

The relationship between the extracted typical font characteristics and font legibility are then analyzed statistically. In this thesis, one of the most important characteristics, inter-letter spacing, is studied further. The issue of how the condensed or expanded inter-letter spacing setting contributes to legibility is also addressed.

1.2 State of the Art

In this section, we will review the state of the art, which includes the terminologies on typography, legibility and readability, and legibility factors on font characteristics.

1.2.1 Terminologies on Typography

Letterforms are sets of letters, numbers and other symbols. A typeface is a design for a set of characters. Times New Roman, Helvetica and Courier are three examples of typefaces. There are two general categories of typefaces: serif and sans serif. Sans serif typefaces are composed of simple lines; while serif typefaces have formal strokes (these ornaments are called serifs) on the ends of some letters. Helvetica is a sans serif typeface and Times Roman is a serif typeface. A font is a particular example of a typeface, which often represents a typeface in a particular size (such as point 8 or point 10), weight (such as light, normal or bold) and angle (such as normal or italic). The individual symbols in a font are called glyphs.

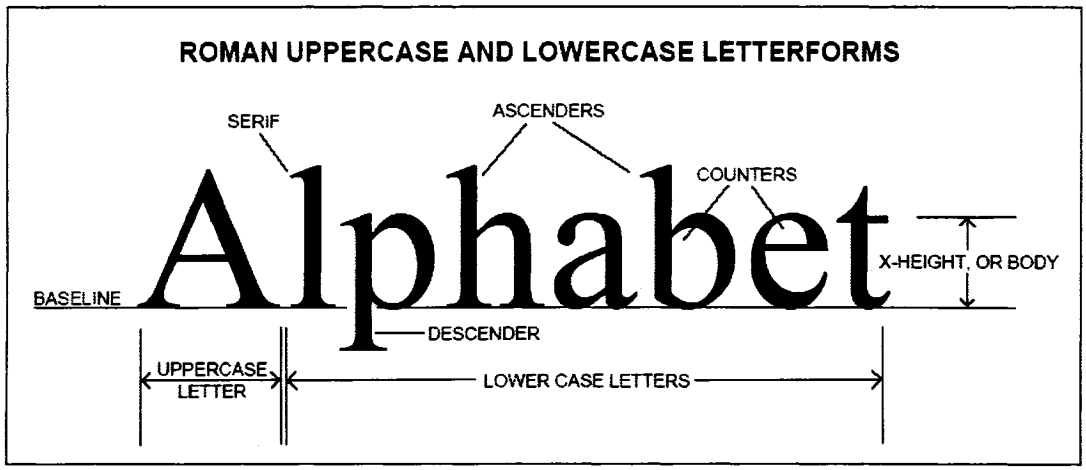


Figure 1.1 Letterform (adopted from Cralg, 1980) [29]

x-height: is the basic height of the lowercase letter x. The x-height can vary greatly from typeface to typeface at the same point size.

Baseline: is the level of the feet of H and x.

Ascender: is the part of the letter that ascends above the x-height.

Descender: is the part of the letter that descends below the baseline.

Counters: is the white space within characters.

Contrast: With a particular font, contrast often refers to the variety of stroke thicknesses that make up the characters.

Color: is the blackness of a text. Color is a function of the relative thickness of the strokes that make up the characters.

Weight: is the relative darkness of the characters.

Leading: refers to the vertical space between lines of text.

Font Measurement Unit:

Different measurement units are used by different typographers, but most measurements in typography are specified in picas and points. Pica is an absolute measurement. The measurement equivalents are as follows:

1 pica = 12 point 1 point = 1/72 inch

Resolution is commonly measured in DPI (dots per inch), which means that the number of pixels (dots) fit into one inch. Therefore, if we have text in a certain point size, we can convert this size to a number of pixels by using the formula below:

Pixel size = point size * resolution/72

Apart from the absolute measurements, some measurements are relative, which means that they vary from font size to font size. Typical relative measurements are: em space, en space and thin spaces, defined as [2]:

em space = a space equal in width to the size of the typeface

en space = a space half of that width

Thin space = a quarter of an em.

em is traditionally defined as the width of the uppercase M in the current face and point size. It is more simply defined as the current point size. For example, in 10-point type, em space is a distance of 10 points, en space is 5 points and the thin space is 2.5 points.

1.2.2 Legibility and Readability

Most people, even those professionally involved in typography, think legibility is equal to readability. Actually, although legibility and readability are connected aspects of type, they are separate concepts. In [2], legibility type is defined as large enough and distinct enough

for a reader to discriminate among individual words and/or among letters. However, readability takes legibility one step further. Text may be legible but not readable if the reader is unable to read easily and smoothly and becomes quickly tired and bored. Readability is the quality that makes the page easy to read, inviting, and pleasurable to the eye. However, we can say that legibility is mainly concerned about the ease with which we can recognize a character or distinguish it from another similar character. For example, the improper character spacing setting can always make "rn" look very similar to "m". Or we may say that "I" and "l" tend to be illegible in some typefaces, and that bold font is not always legible in small sizes.

Although legibility and readability have always been studied together, in our study, we mainly focus on the font characteristics that make the printed text legible or easy to distinguish and to recognize. Those characteristics that make the eye tired and uncomfortable are outside the scope of our study.

1.2.3 Legibility Factors of Characteristics

Many studies have suggested visual design principles regarding typography's impact on legibility. In the book "Legibility of Print", Tinker [11] mentioned that "Optimal legibility of print, therefore, is achieved by a typographical arrangement in which shapes of letters and other symbols, characteristic word forms, and all other typographical factors such as type size, line width, leading, etc., are coordinated to produce comfortable vision and easy and rapid reading with comprehension. In other words, legibility deals with the coordination of those typographical factors inherent in letters and other symbols, words, and connected textual material which affect ease and speed of reading." He found that all uppercase print greatly retards the speed of reading in comparison with lowercase print. He claimed that the

legibility of certain letters can be increased through the use of serifs, heaviness of stroke, delineation of distinguishing characteristics, simplification of outline, and white space within a letter. He also mentioned that the leading has a significant influence on the legibility of type; especially on type less than 12 points in size.

Binns [2] demonstrated that the design of the typeface, line spacing, word spacing and letter spacing affect legibility. She mentioned that the factors of typeface design, which affect legibility, include the thickness of the basic stroke weight, the evenness of the stroke and the distribution of the internal and external spaces within the letter. Victor Gaultney [12] stated that in order to create legible text, the type size, the contrast and weight of the stroke, serif design, distinctive character features and counter shape familiar forms should be carefully considered by the type designer.

Based on a literature review of the basic typography research, Degani [29] provided recommendations for designing legible text. He suggested the following: that sans-serif fonts are usually more legible than serif fonts; that the font design without distinctive features will reduce the print legibility; that when specifying the font height, the distinction between the "x" height and the overall size should be made; that for a font which is viewed in front of the observer, the recommendation of height-to-width ratio is 5:3; that the vertical spacing between lines should be greater than 25-33% of the overall font size; and he recommended that the horizontal spacing between characters should be 25% of the overall font size and not smaller than one stroke width.

In this thesis, typical typeface characteristics are studied, which include serif and sans serif design, x-height, ascender and descender of the typeface, the inter-letter spacing setting, the weight of the stroke and the distinctive character features.

1.3 Outline of Thesis

This first chapter introduced the motivation and objectives of our study. The state of the art included a review of the terminologies on typography, an introduction to legibility and legibility factors on font characteristics.

In Chapter 2, the legibility of eighteen typical typefaces in two sizes is compared. After a review of common legibility measurement methods, we propose a newly designed legibility measurement experiment, which uses machine reading and OCR technology. The legibility comparison of studied fonts is then examined.

Chapter 3 will introduce five of the typical font design characteristics. Then the font legibility factors, which are designed based on font characteristics, are proposed, extracted, and measured on the 18 studied fonts.

In Chapter 4, the correlations between font legibility and font design characteristics are examined. Based on the legibility measurement results in Chapter 2 and the extractions of legibility factors in Chapter 3, the correlations of how the different characteristics contribute to legibility are analyzed statistically.

Chapter 5 will focus mainly on one of the most important font legibility factors, inter-letter spacing setting. Newly designed experiments show the relationship between inter-letter spacing setting and legibility. The analysis of how the condensed or expanded inter-letter spacing setting contributes to legibility will be addressed.

Finally, based on the results of our investigation, Chapter 6 will draw the conclusions of this thesis and suggest some topics for future research.

Chapter 2

Font Legibility Measurement and Comparison

In this chapter, our investigation compares the legibility of the eighteen commonly used typefaces. This Chapter begins with a review of font legibility measurement methods followed by our proposed method. Instead of using human reading, machine reading is used in our proposed legibility measurement. The character recognition rate is used as a legibility comparison criterion. After measuring and comparing, we found that the legibility of some typefaces was comparatively high regardless of print size; that some typefaces always had relatively lower legibility in both font sizes; and that the legibility of some typefaces was dramatically increased when the font size became bigger.

2.1 Review of Font Legibility Measurement Methods

In the psychology and physiology fields, legibility has been studied and measured by many criteria, such as the distance threshold, speed of reading, comprehension, eye movements and so on. Based on [3] [11] [21] [30], the main legibility measurement methods can be grouped into the following:

1) Distance of perception-measures

The distance method is used to measure the distance from the eyes at which printed symbols

can be accurately perceived. This method is probably the most common method to assess text legibility, and it is widely used to assess the relative legibility of individual symbols or of letters which are intended to be read at a distance (such as road signs, street lettering and car number plates).

2) Reading speed of perception method

The reading speed method is used to measure the time taken by the subject to read the stimuli. In [30], two reading speed methods are mentioned. One reading speed measurement uses Rapid Serial Visual Presentation (RSVP) to read letters on computer monitor and another uses a subject's continuous reading of scrambled text passages to measure the reading speed of conventional text, printed on paper.

3) Eye movement measurement

In the eye-movement method, a number of techniques are employed to record eye movements, ranging from direct observation to more precise and reliable methods using mechanical or electrical control.

4) Peripheral vision measurement

The peripheral vision method is used to measure the horizontal distance from a fixation point at which a sign can be accurately perceived. This method is mainly used to determine the relative legibility of single characters and to compare the legibility of black letters on white backgrounds with white on black.

5) Frequency of eye blinks

The frequency of eye blinks method counts the number of blinks made by the subject while reading. The assumption of this method is that the reader's blink rate will increase as legibility is reduced.

6) Comprehension method by Suen and Komoda

With the aid of computerized optical scanner, characters in different font styles can be read and converted into digital images. Suen et al. [27] used these binary images to conduct experiments and investigated effects of font-styles on legibility on three studies fonts, Letter Gothic, Courier and DECwriter. In their experiment, reading comprehension is used as the legibility criteria. Their experimental results show that there is a main effect of font on legibility and the DECwriter font give the overall poorer performance.

2.2 Proposed Legibility Measurement Method

In the common legibility measurement methods, humans are always used as subjects to test font legibility. However, human beings have different cultures, education, interests, or other human characteristics that may affect their reading processes in some experiments. For example, Rayner and Pollatsek note in "The Psychology of Reading" (Englewood Cliffs: Prentice-Hall 1989) that users will read at a faster level than normal if interested in the subject matter. Therefore, these human differences may affect the accuracy of legibility measurements.

In our study, we designed our font legibility measurement experiment by using machine reading because it may avoid the problem of human characteristics affecting study results and it is a faster and cheaper way to measure font legibility. One successful application of machine reading, OCR technology, is used. OCR has been researched for more than 50 years. There has been very substantial research efforts in OCR devoted to automatic processing and recognition of machine printed characters. Although OCR technology studies are still facing some challenges, in machine printed character recognition, some commercial OCR packages are already very successful and are commonly used. Since our study focuses on the effects of

different font designs on legibility and we do not need to be concerned about the human being feedback, OCR systems can objectively achieve our target. The character recognition rate is used as a criterion in our legibility comparison. The assumption of the legibility measurement is that if the character recognition rate of the font is higher, the legibility should be better.

2.2.1 Procedure

In the experiments, four commercial OCR systems were used to read the machine printed documents. Eight different document sets were tested. In total, thirty-six fonts, consisting of eighteen different typefaces in two different sizes, were evaluated.

The procedure of the experiment is described below:

1) Eight texts with 2244 words and 7859 characters in total were used as the test sets. The texts were set in eighteen typefaces and two font sizes.
2) The text samples were converted to binary images at 200 x 200 dpi resolution.
3) The legibility of the created images was tested by using four common commercial OCR systems: ABBYY FineReader 6.0 professional, TypeReader 6.0 Professional, TextBridge Pro millennium and Omnipage Pro 12.0.
4) By using a dynamic string matching method, the recognition rates were calculated and recorded.

2.2.2 Studied Fonts

Nowadays, there are thousands of typefaces and many font classifications. Fig. 2.1 lists the eighteen typefaces of our study. We chose these typefaces because of their common use and their special properties. There are nine sans serif typefaces (AlbertusMedium, Arial,

FrutigerLinotype, Haettenschweiler, Helvetica, Impact, Letter Gothic, OCRB, MicrosoftSansSerif) and nine serif typefaces (Batang, BerkerlyBook, Century, Century SchoolBook, CourierNew, Garamond, Georgia, Palatino and TimesNewRoman).Also, these eighteen typefaces consist of various design characteristics, such as x-height, ascender and descender, stroke weight and inter-letter spacing setting, etc. *Appendix A* presents these eighteen typefaces more clearly.

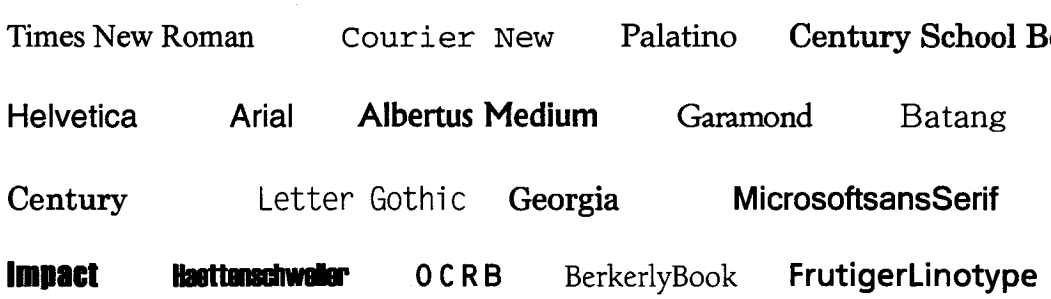


Figure 2.1 18 studied typefaces

2.2.3 Test Database

In order to get the general and reasonable recognition performance, 8 different kinds of documents were tested. In our test data sets, not only the commonly used words, such as some high frequency words or words from elementary books but also some low frequency words or nonsense words were used. The contexts of each document are shown completely in *Appendix B*. Here are the lists of the eight test sets in our database.

- 110 High Frequency Words :

This test set included the first 110 most frequently used words, which are listed in Computation analysis of present-day American English [35].

- 126 Low Frequency Words :

This test set included 126 low frequency words, listed in Computation analysis of present-day American English as well.

- Bigram Words

This test set included all the 676 (26 x 26) bigram words, such as aa, ab, bb, zy, zz etc.

- High Frequency N-Gram Words

This test set included the top 50 2-Gram, 3-Gram, 4-Gram and 5-Gram words, as shown in [37].

- NonSense Words by using Chinese pinyin syllables:

This test set includes 404 Non-Sense words, which was generated by Chinese pinyin syllables.

- Documents from Elementary Book:

This test set included one paragraph and two short letters from the elementary book.

- Documents from Better Type:

This test set included paragraphs from [2].

- Documents from Devoghelaere Report

This test set included paragraphs from a report written by Cedric Devoghelaere [36]

In Table 2.1, the number of testing words and characters in each testing set is listed.

Table 2.1 legibility test sets

Test set	No. of words	No. of characters (no space)
110 High Frequency Words	110	389
126 Low Frequency Words	126	826
Bigram Words	676	1352
High Frequency N-Gram Words	146	431
Non-Sense Words by using Chinese pinyin syllables	404	1308
Documents from Elementary Book	186	772
Documents from Better Type	354	1639
Paragraphs from Devoghelaere Report	242	1142
Total	2244	7859

Figure 2.2 and Figure 2.3 show the legibility measurement results of 18 different fonts on the eight test sets in size 8 and size 10, respectively. From these two figures, we can see that for all the fonts, the recognition results on Non-Sense words and Bigram words test sets are lower.

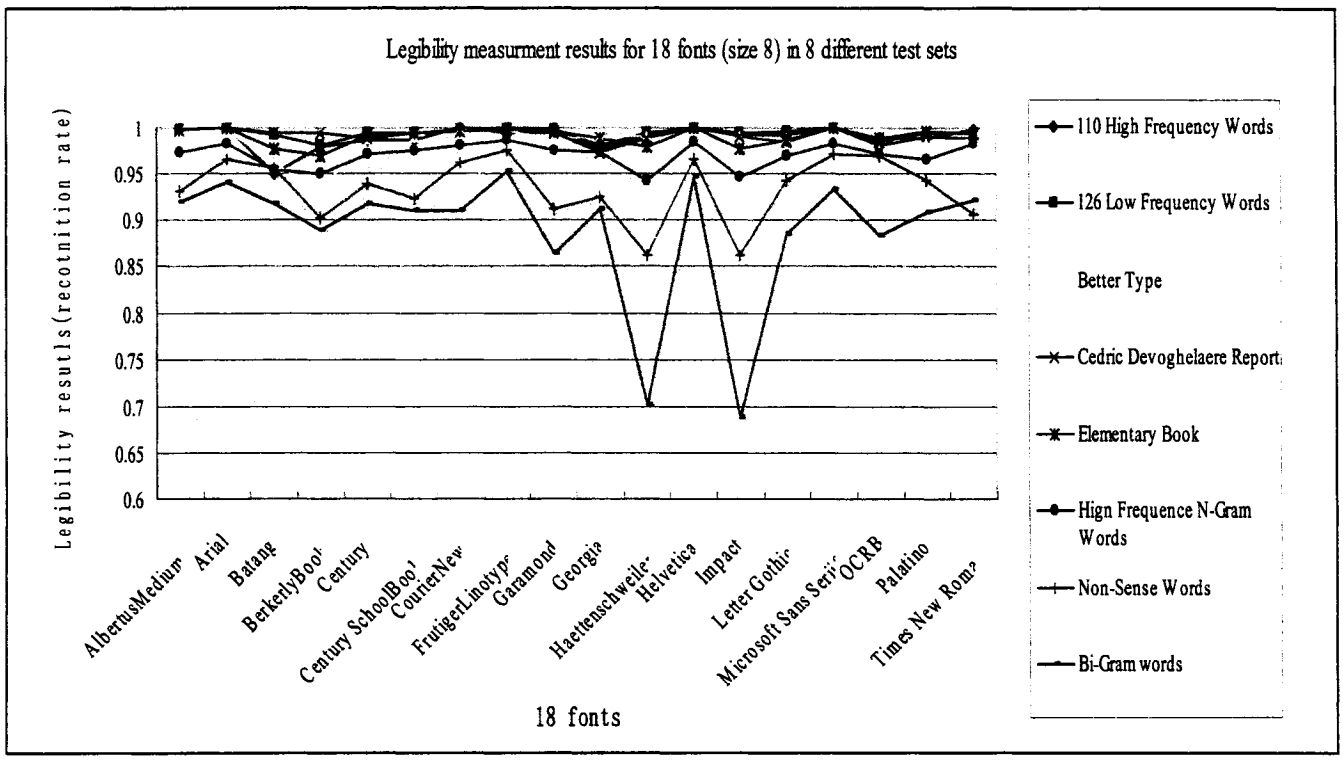


Figure 2.2 18 font (size 8) legibility measurement results in 8 different test sets

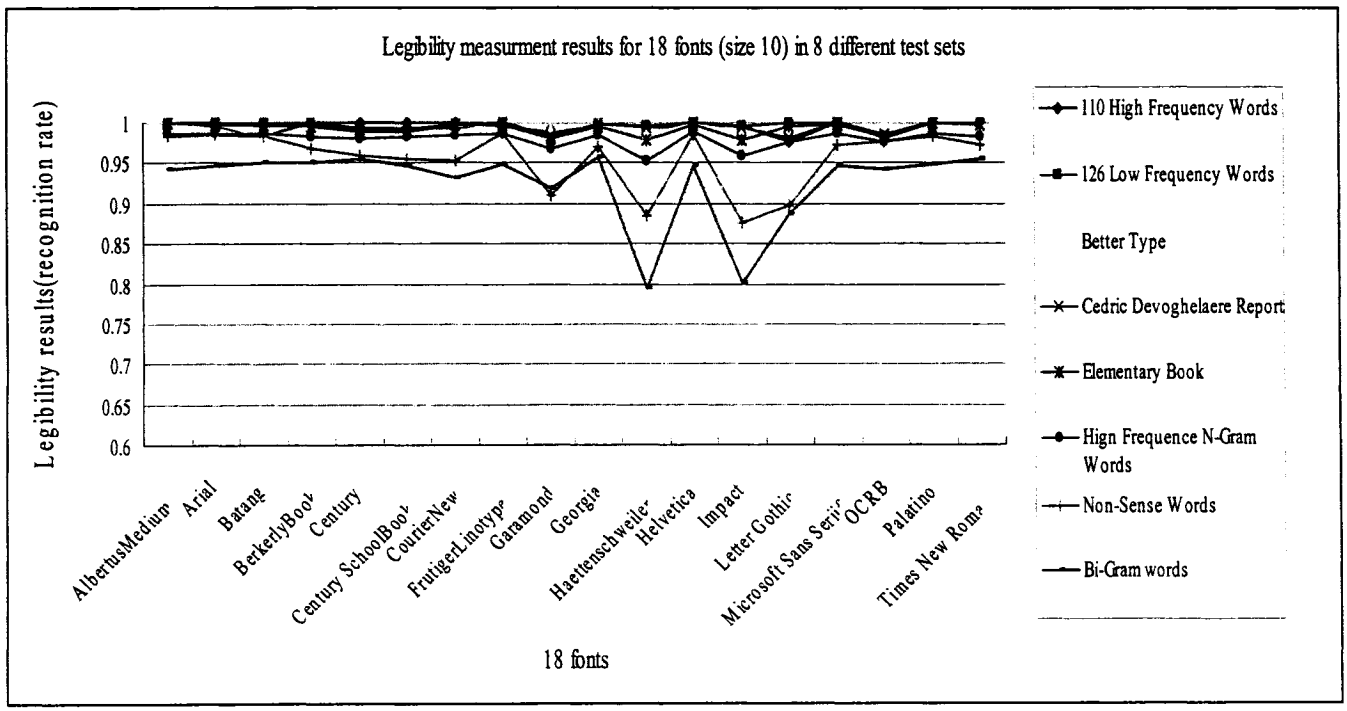


Figure 2.3 18 font (size 10) legibility measurement results in 8 different test sets

2.2.4 Dynamic String Matching Algorithm

In order to calculate the recognition rate, one of the dynamic string matching algorithms, the Levenshtein distance algorithm, was used in our study. The Levenshtein distance algorithm is designed to measure the similarity of two strings. The distance is based on the number of insertions, deletions, or substitutions required for transforming one string to the other. This algorithm was devised by Russian scientist Vladimir Levenshtein in 1965 [28], and is used in applications that need to verify the similarity of two strings, such as spell checking, and DNA analysis. In our analysis, we used this algorithm to compare an OCR output with the corresponding label.

2.3 Legibility Comparison Results of 18 Typefaces

After measuring the legibility of the selected fonts, the mean character recognition rate of each font was evaluated by averaging the character recognition results of each OCR system and each test set. The mean recognition rates of eighteen typefaces in two sizes are shown in Figure 2.4 and Table 2.2.

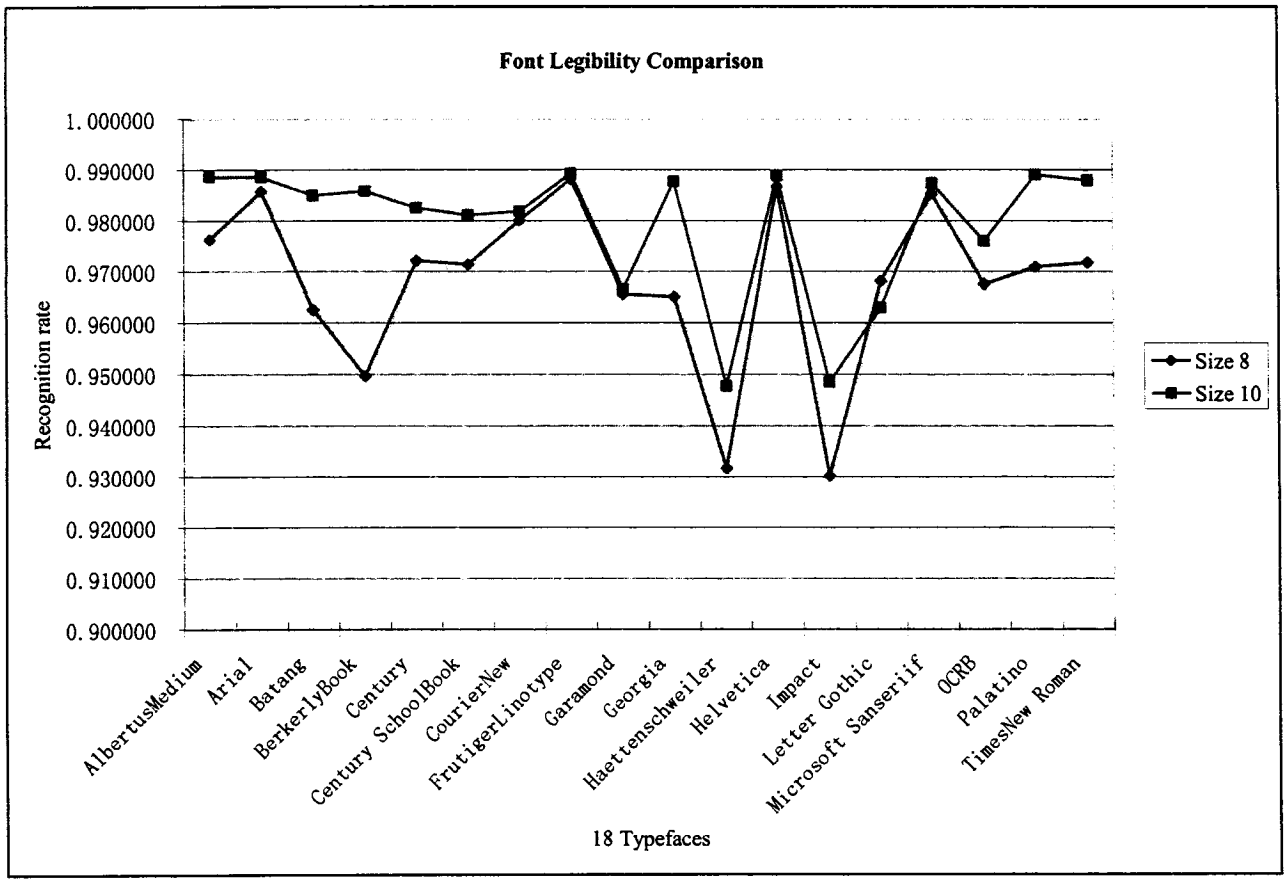


Figure 2.4 Font legibility trend in 18 typefaces with size 8 and size 10

This Figure shows that the legibility of Letter Gothic drops slightly from size 8 to size 10, but all other typefaces have better recognition rates in size 10.

From the recognition rates showed on Table 2.2, we found that:

1) The legibility of typefaces Arial, Frutiger Linotype, Helvetica and Microsoft Sans serif are comparatively better in both size 8 and size 10.

Table 2.2 Font mean recognition rate in 18 typefaces with size 8 and size 10

Fonts	Size 8	Size 10
Albertus Medium	0.976243	0.988606
Arial	0.985792	0.988584
Batang	0.962604	0.985012
Berkerly Book	0.949724	0.985849
Century	0.972113	0.982563
Century School Book	0.971381	0.981142
Courier New	0.980052	0.981875
Frutiger Linotype	0.988181	0.989329
Garamond	0.965612	0.966589
Georgia	0.965115	0.987725
Haettenschweiler	0.931667	0.947700
Helvetica	0.986730	0.988943
Impact	0.930213	0.948522
Letter Gothic	0.968187	0.963028
Microsoft Sans serif	0.985356	0.987441
OCRB	0.967557	0.975992
Palatino	0.970953	0.989118
Times New Roman	0.971784	0.987956

2) The legibility of typefaces Garamond, Haettenschweiler, Impact and Letter Gothic are worse in both size 8 and size 10.

3) When the size increases from size 8 to size 10, it makes a dramatic increase in legibility for typefaces Batang, BerkerlyBook, Georgia, Palatino and Times New Roman.

We will measure the detailed font characteristics of each individual font in the next chapter and analyze the effect of font characteristics on legibility in Chapter 4.

Chapter 3

Font Legibility Factors

In this chapter, we describe the main font characteristics which affect legibility. In order to measure them mathematically and objectively, each characteristic is represented by one or two font legibility factors. The extraction methods are proposed for each font legibility factor.

3.1 Font Characteristics

With the increased use and design of font types, studies about the effects of various font characteristics on legibility have been markedly increased over the years. Psychologists, engineers, ophthalmologists, typographers, educators and even newspaper publishers have studied from various aspects of this issue related to reading. Based on the review of legibility factors on font characteristics in Chapter 1, our study focuses on five particular typeface characteristics, which are the most common ones discussed in the literature. They include serif and sans serif design, x-height, ascender and descender of the typeface, the inter letter character spacing setting, the weight of the stroke and the distinctive character features. In this section, we will introduce these characteristics individually, in detail.

3.1.1 Serif and Sans Serif Designs

Serif and sans serif are two of the main categories of typeface design. Serif fonts have curls and small appendices at the end of each letter. From Merriam-Webster's online dictionary, a serif is defined as "any of the short lines stemming from and at an angle to the upper and lower ends of the strokes of a letter". Typefaces that do not have serifs are called sans serif typefaces (taken from the French word "sans" meaning "without"). Figure 3.1 shows one of the typical serif faces "Times New Roman" and one of the typical sans serif faces "Helvetica".

Serif typeface: Times New Roman Sans Serif typeface: Helvetica

Figure 3.1 Serif and Sans serif fonts

Although the designs of serif are different in each typeface, serifs are categorized into several basic groups based on the distinction of stroke weight, shape and attachment to the main stroke. In Figure 3.2, several basic groups of serif design, which include wedge serif, bracketed serif, hairline serif and slab serif, are shown. Wedge serifs have a straight angular fill and bracketed serifs have a curved fill between the main strokes of a letter and the serif. Hair serifs have a very fine "hairline" stroke weight. Slab serifs, or Egyptian serifs, have visually square serifs such that the thickness of the serif is about equal to its extension from the main stroke.

Wedge serif Bracket serif Hair serif Slab serif

Figure 3.2 Different serif designs

From the typographic literature, serifs are generally believed to be a factor that impacts legibility. However, the debate on whether serif versus sans serif fonts is more legible always exists.

Common wisdom, often repeated, is that serif typefaces are inherently more legible than sans-serif ones. In [30], there are two main reasons cited to explain why serifs should enhance legibility. First, serifs are believed to increase letter discrimination because the serif designs make the spatial code of letter forms more complex. Mclean [9] wrote that "Sans-serif type is intrinsically less legible than serif type. . .because some of the letters are more like each other than letters that have serifs, and so the certainty of decipherment is diminished". Second, serifs increase the visibility of the end of strokes, which increases the salience of the main strokes of letters, so that serifs may enhance the legibility of letters by providing an additional cue to the location of stroke ends. Rubinstein [15] mentioned that "Serifs have an important role in the readability of type, providing. . .accentuation to the ends of strokes that may help the reader read faster and avoid fatigue."

However, some researchers, such as Paterson and Tinker [16], Poulton [17], Smither and Braun [18], or Moriarty and Scheiner [19], show that there is no difference between serif and sans serif fonts in terms of legibility. In [30], possible reasons why serifs may have little effect on legibility are also explained. The author states that since serifs are generally designed as ornamental rather than essential parts of the letter form, the serifs are small

relative to the letter size. Therefore, one might suspect that serifs should have little impact on letter identification.

Apart from whether the serif or sans-serif fonts are more legible, the discussions also consider how the shape of serifs can influence legibility. Tinker's analysis suggests that the legibility of individual letters is diminished by the use of hair lines or of long or heavy serifs [11]. However, by contrast, in certain production environments, particularly phototypesetting, bracketed serifs retain their shape best, increasing letter distinction and legibility [22].

3.1.2 The x-height, Ascender and Descender

The x-height is the height of a lower case "x". The ascender is the part of the letter that reaches above the x-height and the descender is the part of the letter that falls below the baseline, as shown in Figure1.1 of Chapter 1. According to Barnhurst [13], and Poulton [14], the x-height is a somewhat more realistic way to denote the size of a typeface. Research has shown that typefaces with larger but moderate x-heights are generally more legible and it seems that the increased x-height gives increased legibility like that of a larger print size [12]. In the same type size (e.g. 10 points or 12 points), one typeface may look bigger than another if it has a longer x-height. As shown in Figure 3.3, the typefaces Helvetica and Garamond are both set in size point 10. Notice how the Garamond with its small x-height appears much smaller than the Helvetica with its high x-height.

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'.

10 point Helvetica

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'.

10 point Garamond

Figure 3.3 Variations in x-height of Helvetica and Garamond

When Lucie Lacava, a well-known Canadian newspaper designer and consultant, talked about newspaper redesign, she said that "Bigger is better; Disregard the point size, it is the x-height (visual appearance) that counts; never give them a visually smaller font than they currently are accustomed to".

Although a bigger x-height can bring more information to distinguish one letter from another, the ascender and descender are also important factors for font legibility. Studies have shown that reading relies heavily on the tops of letters. As shown in Figure 3.4, we can see that the string is more easily recognized by its top half than by its bottom half. Also, other studies have shown that insufficiently distinct ascenders and descenders always cause confusion of one letter with another, such as h and n, i and j, and n and p. Therefore, the balance between a moderately large x-height and clear ascenders and descenders is very important for a legible font.

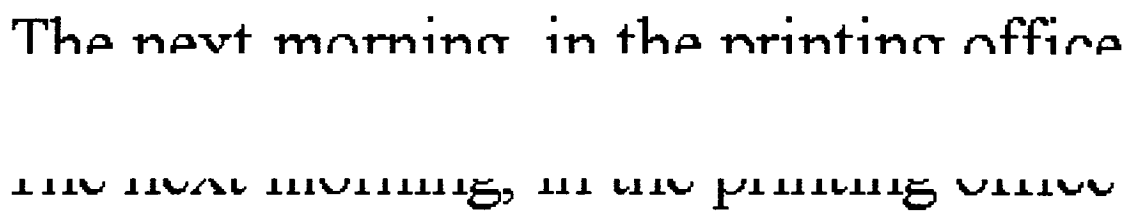


Figure 3.4 The top half and bottom half parts of a string

3.1.3 Inter-Letter Spacing

To design a type successfully, getting a good balance of white space inside and outside the letters is really important. The 'fitting' of letters - the allocation of the correct amount of space to each side of them, so that when they are associated together into words they have a balanced relationship, without unsightly gaps or congestion - is a process fundamental to the success of a type design [1]. In digital type, the spacing on both sides of a letter is represented by numerical values for the space to the left of a character - the Left Side Bearing (LSB) - and the space to the right of the character – the Right Side Bearing (RSB), as shown in Figure 3.5.

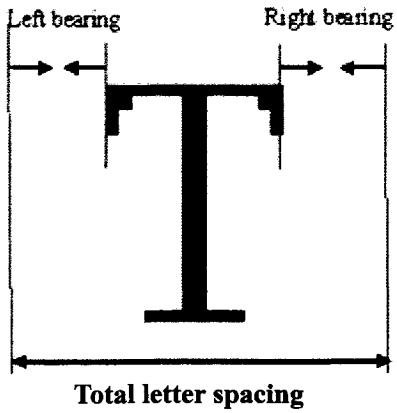


Figure 3.5 Character T with its space setting

In other words, the spacing between two consecutive letters is equal to the right bearing of

the first letter plus the left bearing of the second letter. For example, Figure 2.7 (b) shows the result when letter "T" and letter "a" are put together in a word without any spacing modification. However, in order to get the best design result, another concept, "Kerning", is also very important in space setting. Kerning defines the modification of the space between specific letter pairs. Figure 3.6 (a) shows one of the examples of kerning design in typeface Arial. From this example, we can see that kerning, when applied to specific letter pairs, reduces the space on one side of a letter to enable it to stand closer to the adjacent letter.

(a) Design with kerning (b) Design without kerning

Figure 3.6 Font Arial string "Ta", with and without kerning

We should keep in mind that the objective of good kerning and character spacing is not about how much closer we can put the letters together, but is about how rhythmically the letters are entwined together. In the combination of letters, some of them need to be kerned slightly, such as "Ta" or "Fa", and some of them need to be kerned further, like "rn" (since they may be confused with letter "m"). The amount of kerning letter pairs is not "the more, the better." Actually, the average amount of kerning for letter pairs of a given typeface should be around 200-500 [26].

Therefore, we can say that the inter-letter spacing of two letters can computed from side

bearings and pairwise kerning by the following equation [20]:

$$d = RB_1 + LB_2 + kerning_{12} \qquad (3.1)$$

where

d is the inter-letter space,

RB_1 is the right bearing of the first letter,

LB_2 is the left bearing of the second letter,

$Kerning_{12}$ is the pairwise kerning of the two letters; and $kerning_{12} = 0$ if there is no kerning between these two letters.

Most printed material today uses variable (or proportional) spacing. With this type of space setting, the horizontal space that a character occupies depends on the width of the individual character, such that letter i takes less space than letter w. With a fixed space setting, the same amount of horizontal space is assigned to each letter. In order to avoid overlapping of the characters, this amount of space has to be at least as great as that of the widest character, usually the uppercase W [20]. Fixed space fonts are mainly used on typewriters. In Figure 3.7, test strings are shown by using a proportional (variable) width font, Times New Roman, and a fixed width font, Courier New. We can see that in Times New Roman, letters i and j take different spaces from letters m and w; while in Courier New, each letter takes the same amount of space.

(a) Proportional width font (Times New Roman) (b) Fixed width font (Courier New)

Figure 3.7 Test strings in proportional (variable) width font and fixed width font

Some studies have shown that the proportional spaced fonts are easier to read [33], [34]. By comparing reading performance (speed) with fixed and variable (proportional) character widths, Arditi et al. [31] found variable pitch to yield better performance at medium and large character sizes and fixed pitch to be superior for small character sizes approaching the acuity limit. Skottun and Freeman [33] studied the inter-letter spacing and apparent letter size in relation to acuity. Acuity levels for letters with five different inter-letter spaces were measured. They found that when letters were closely spaced, acuity was poorest and this effect diminished as the letter spacing increased, up to a separation of about three minarc (measurement used in psychology), after which acuity was relatively unaffected. When the perceived letter size was tested, the critical value of three minarc also occurred. Their studies showed that widely spaced letters, of the same height, tended to appear larger than letters that were closely spaced. Liu and Arditi [32] studied how inter-letter spacing would affect the kind of errors made by human subjects. By using the 1.0 and 0.1 letter height for inter-letter spacing, they found that the narrow and wide spacing letter strings produced different letter confusion matrices. Apart from the letter confusions that were shared by both wide and narrow spacing strings, narrow spacing strings produced more

random confusions and a set of unique letter confusions. Degani [29] recommended that the horizontal spacing between the characters should be 25% of the overall font size and not smaller than one stroke width.

3.1.4 Color and Weight

Color represents the blackness of a text. The thickness of the basic stroke weight is the primary factor that determines the lightness or darkness of the type. Therefore, we can consider color as a function of the relative thickness of the strokes that make up the characters. The term weight is used in a comparative sense. We can use the different weights to express the colors of the characters, and the weight is indicated by relative terms such as thin, light, bold, extra-bold and black.

Various studies did not show a clear difference in legibility between regular and bold weights, but it has still been suggested that extreme weights in strokes should be avoided. Tinker performed one experiment about bold face design. Although in his experiment, the bold face was read at the same speed as normal ones, the majority of the subjects (70%) commented about the unpleasantness of the text as compared to the plain font. Some studies [24] have suggested that "The optimal stroke width for individual letters should be about 18% of the total width or height of the letter".

3.1.5 Distinctive Character Features

Distinctive character features are also important factors that affect legibility. Strong distinctive features in font design mean that the letters of a font have strong distinctive character shapes, which can improve the rapid and accurate discrimination of letters. For example, many studies show that in most of the font designs, letters "c" and "e" or "i" and "l" have low legibility because of their similar design features; while letters "d", "m", "p", "q" and "w" have high legibility because of their distinctive features. It is easy to understand that with strong distinctive character design features, legibility is higher in typefaces.

Various studies cite commonly confused characters according to their similar character shapes: "C" and "G"; "H" and "N"; "E" and "F"; "c" and "e"; "i" and "l" (lower case I and lower case L); "i" and "j"; "b" and "h"; and "e", "a", and "s". Therefore, in order to have distinctive character features in letters and to improve the legibility of the font design, the dissimilarity of the characters' shapes in a font design should be emphasized.

3.2 Proposed Font Legibility Factors

In order to evaluate each font design, we need to measure each characteristic of each given font. In order to measure them mathematically and objectively, each characteristic is represented as one or two legibility factors and the extraction methods of some factors are proposed.

3.2.1 Proportion of x-height, Ascender and Descender

As discussed in section 3.1.2, the moderate balance of x-height, ascender and descender are important for font legibility. In order to examine the type design of these three parts, the proportion of these three parts for each font design is measured. As shown in Equations 3.2 and 3.3, XA and XD, which measure the ratios among x-height with ascender, and x-height with descender, respectively, are used as two legibility factors.

$$XA = \text{x - height} / Ascender \quad (3.2)$$

$$XD = \text{x - height} / Descender \quad (3.3)$$

In order to get the values of x-height, ascender and descender, the typographical structure of text lines (top line, upper line, base line and bottom line) are determined from the vertical projection profile, based on [23]. The vertical projection profile VP[i] represents the sum of black pixels of the *ith* scanline. According to the main peaks of VP, the upperline (ul) and baseline (bl) are detected by using equations (3.4) and (3.5).

$$ul = i \quad \text{if} \quad i \in \left[to, to + \frac{1}{2}|bo - to| \right] \& \max(VP[i+1] - VP[i]) \quad (3.4)$$

$$bl = i \quad \text{if} \quad i \in \left[to + \frac{1}{2}|bo - to|, bo \right] \& \max(VP[i-1] - VP[i]) \quad (3.5)$$

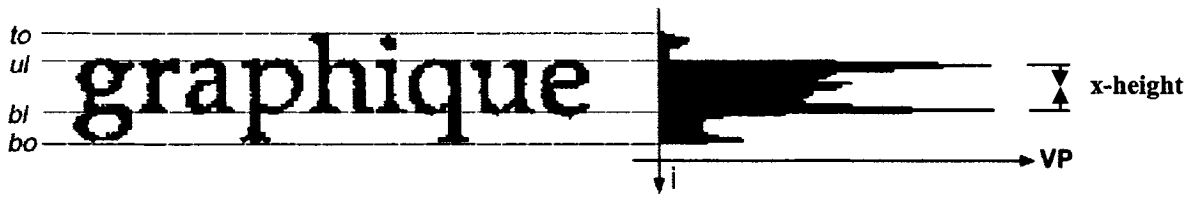


Figure 3.8 Four typographical lines from vertical projection profile [23]

In order to measure the proportion of x-height, ascender and descender of each studied font, test string images, which were set with each font with size 10, were used to extract the legibility factors XA and XD in Table 3.1.

Table 3.1 Legibility factors XA and XD for each font design in size 10

Fonts (size 10)	XA	XD	Fonts (size 10)	XA	XD
AlbertusMedium	2.50	3.00	Georgia	2.00	2.33
Arial	3.00	2.50	Haettenschweiler	4.50	6.00
Batang	1.86	3.25	Helvetica	3.00	2.50
BerkerlyBook	1.50	1.71	Impact	4.50	6.00
Century	1.86	2.17	Letter Gothic	2.50	3.00
Century SchoolBook	1.86	2.17	Microsoft Sanseriif	3.00	2.50
CourierNew	2.40	2.40	OCRB	2.00	2.80
FrutigerLinotype	2.00	2.33	Palatino	1.86	1.63
Garamond	1.57	1.57	TimesNew Roman	1.86	2.17

Table 3.1 shows that font Haettenschweiler and Impact have very large XA and XD values, which almost double the average XA and XD values of the 18 fonts; while font BerkerlyBook and font Garamond all have smaller XA and XD values compared with other fonts. Also, the XD value for font Palatino is comparably smaller.

3.2.2 Vertical and Horizontal Stroke Thickness

In order to compare the weights of a given font, vertical stem widths and horizontal stem heights within characters should be considered. Also, the design size of a font should also be considered. In our study, the ratios of vertical or horizontal stroke thickness with x-height of a font are used to measure the font weight, as follows:

$$Weight_v = {}^{T_v}\!/_{\text{x - height}} \quad (3.6)$$

$$Weight_h = {}^{T_h}\!/_{\text{x - height}} \quad (3.7)$$

where T_v and T_h are the average vertical and horizontal stem thicknesses of a given font, respectively. x-height is the average x-height of this font.

Tv and T_h are computed by considering the most frequent horizontal and vertical black-runs in each given letter. Figure 3.9 shows the extracted horizontal and vertical stroke thickness of character "o" in font Arial with size 10 point. In our study, for each font, 26 lower case letters were considered in order to get the average vertical and average horizontal stem thicknesses.

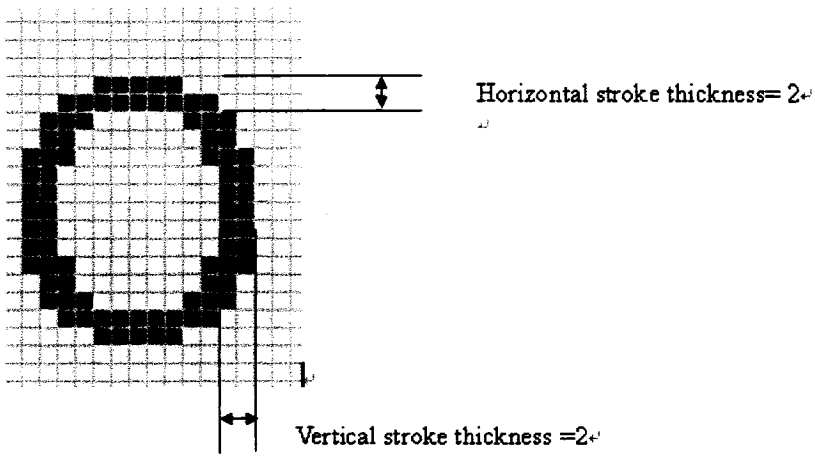


Figure 3.9 Horizontal and vertical stroke thicknesses for character "o" in font Arial with size 10 points

Table 3.2 Legibility factors Weightv and Weighth for each font design in size 10

Fonts (size 10)	$Weight_v$	$Weight_h$	Fonts (size 10)	$Weight_v$	$Weight_h$
AlbertusMedium	0.20	0.13	Georgia	0.21	0.07
Arial	0.13	0.13	Haettenschweiler	0.28	0.17
Batang	0.15	0.08	Helvetica	0.13	0.13
BerkerlyBook	0.08	0.08	Impact	0.28	0.17
Century	0.15	0.08	Letter Gothic	0.07	0.07
Century SchoolBook	0.15	0.08	Microsoft Sanseriif	0.20	0.13
CourierNew	0.08	0.08	OCRB	0.21	0.14
FrutigerLinotype	0.21	0.14	Palatino	0.15	0.08
Garamond	0.18	0.09	TimesNew Roman	0.15	0.08

Table 3.2 shows that the font BerkerlyBook, CourierNew and Letter Gothic all have very

small $Weight_v$ and $Weight_h$ values, which means that they all have much thinner vertical and horizontal strokes (or weights).

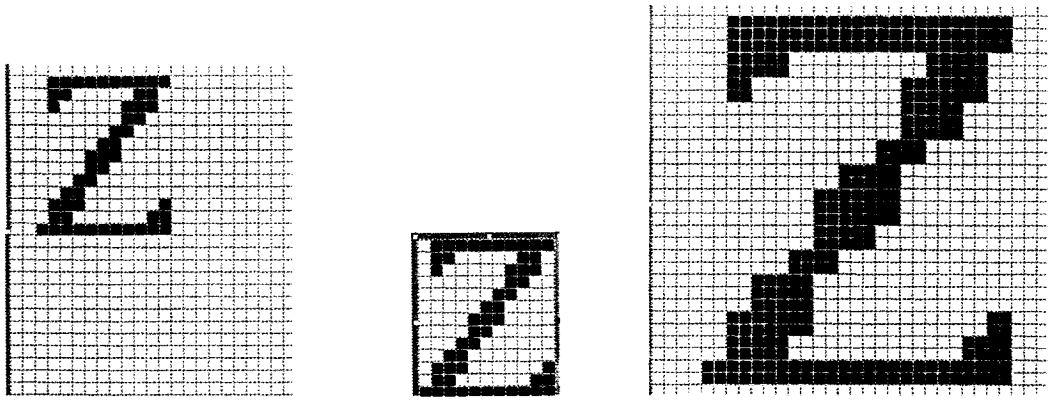
3.2.3 Character Shape Similarity

In order to measure the distinctive features (Section 3.1.5) of a given font design, (which is used to detect the character shape design similarity) gradient features are extracted from the each character image in a given font. The gradient measures the magnitude and direction of the greatest change in intensity in a small neighborhood of each pixel. Since the object contour and structure are encoded in the gradient direction and magnitude at each pixel of the image, features extracted from the gradient representation are very powerful for pattern recognition.

Given each character image, the following operations are used for comparing the similarity of the character shapes in a given font design:

1) **Normalization of character image**

In order to keep the same dimensionality of each feature vector and make the feature invariant to size variation and shape distortion of characters, the image is normalized to a fixed size. First, we binarize and cut the original image (Figure 3.10 (a)) of a font character into a rectangle with the same height and width of the original pattern (Figure 3.10 (b)). After that, by using a linear interpolation algorithm, we enlarge the image to size 32 * 32 (Figure 3.10 (c)) for the extraction of gradient features.



(a) (b) (c)

Figure 3.10 Sample images in size normalization:

(a) the original image, (b) a cut image, and (c) an enlarged image for feature extraction

In the linear interpolation algorithm method, the cut image of size I*J has been normalized to the image of size M*N. For any given original image pixel (*i, j*), its corresponding pixel in the normalized image is (*m, n*), as shown in Figure 3.11. The relationship is given by:

$$i - \frac{I}{2} = \lambda(m - \frac{M}{2}) \qquad (3.8)$$

$$j - \frac{J}{2} = \lambda(n - \frac{N}{2}) \qquad (3.9)$$

where λ = max (I/M, J/N),

I = cut image width,

J = cut image height,

M = normalized image width,

N = normalized image height.

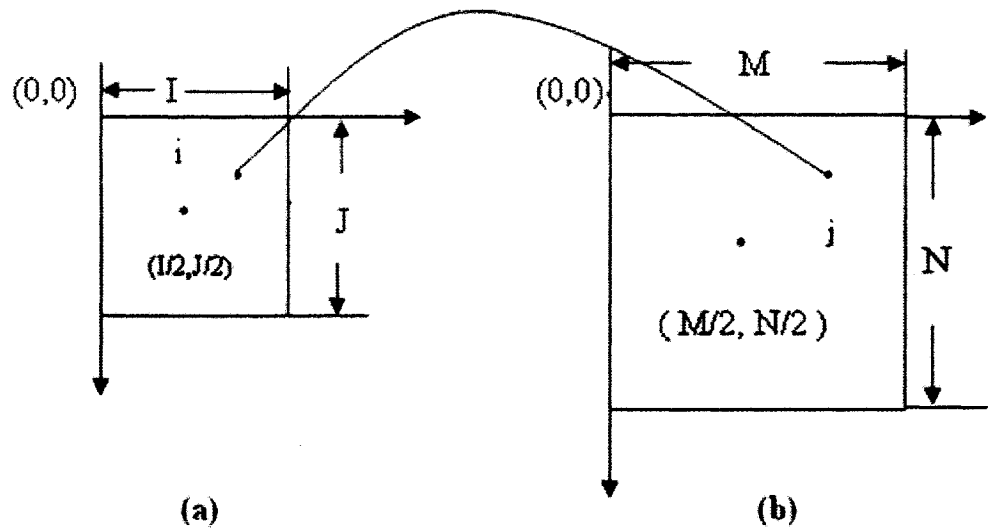


Figure 3.11 The relationship between (a) a cut image and (b) a normalized image

2) Gradient feature extraction

In our study, a gradient feature extraction method [38] has been implemented and applied to represent the distinctive features of each character in a given font. Using the gradient feature extraction method, a feature vector of size 400 was produced for each character. The feature vector consisted of 5 horizontal, 5 vertical and 16 directional resolutions.

In order to extract the gradient features, a smoothed gray-scaled image is obtained by applying a size 3*3 mask $\frac{1}{16}\begin{bmatrix}1 & 1 & 1\\ 1 & 8 & 1\\ 1 & 1 & 1\end{bmatrix}$ to the character image three times. The gray-scaled normalized image is standardized such that its mean and maximum values are 0 and 1.0, respectively. A Robert filter [5] is then applied to each pixel g(i, j) of the normalized image to calculate the gradient strengths and directions, as shown in equation (3.10):

$$\begin{aligned}
\Delta u &= g(i+1, j+1) - g(i, j) \\
\Delta v &= g(i+1, j) - g(i, j+1) \\
\theta &= \arctan\left(\Delta v / \Delta u\right) \\
|\nabla g| &= \sqrt{(\Delta u)^2 + (\Delta v)^2}
\end{aligned} \qquad (3.10)$$

where θ and $|\nabla g|$ denote the gradient direction and strength, respectively.

The direction of the detected gradients is quantized to 32 levels with a $\frac{\pi}{16}$ interval. The normalized character image is divided into 81 (9 horizontal, 9 vertical) blocks. The strength of the gradient is accumulated separately in each of the 32 directions, in each block, to produce 81 local spectra of direction.

The spatial resolution is reduced from 9*9 _ to 5*5 _ by down sampling every two horizontal and every two vertical blocks with a Gaussian filter $\frac{1}{2\pi}\exp\left(-\frac{x^2+y^2}{2}\right)$. Similarly, the directional resolution is reduced from 32 to 16 by down sampling with a weight vector $[1 \quad 4 \quad 6 \quad 4 \quad 1]^T$, to produce a feature vector of size 400 (5 horizontal, 5 vertical, and 16 directional resolutions). Finally, the variable transformation $(y = x^{0.4})$ is applied to make the distribution of the features Gaussian-like.

3) Measuring similar design pairs

After getting the feature vector of size 400 for each character in a given font, the most similar character pairs are found by calculating the distance between two characters in the Euclidean space, as shown in equation (3.11):

$$d = \|x - y\| = \sqrt{\sum_{i=1}^{400} |x_i - y_i|^2} \qquad (3.11)$$

where x_i and y_i are the ith feature vectors of two compared images, respectively.

i varies from 1 to 400.

For each font design, our goal is to find the degree of shape similarity between the 26 lower case character designs. Therefore, for each lower case letter in a given font, gradient feature vectors are extracted from the normalized image and the values of d between each of the 26 lower case letters and each of the 52 lower case and upper case letters are computed. That is, for each font, low case letter "a" is compared with lower case letters "a" to "z" and upper case letters "A" to "Z". Similarly, lower case letters "b" to "z" of each font are also compared with 52 letters (all the lower case and upper case letters). The legibility factor of character shape similarity in a given font is then measured using the average d of 26 * 52 pairs.

Table 3.3 Character shape similarity of each font

Fonts	Average d	Fonts	Average d
AlbertusMedium	111.7182	Georgia	92.4798
Arial	130.4506	Haettenschweiler	91.46082
Batang	93.02068	Helvetica	130.6329
BerkerlyBook	79.88246	Impact	93.47969
Century	92.65556	Letter Gothic	76.54783
Century SchoolBook	92.70812	Microsoft Sanseriif	130.6888
CourierNew	78.55499	OCRB	119.2179
FrutigerLinotype	126.009	Palatino	94.5982
Garamond	90.62936	TimesNew Roman	94.15444

From Table 3.3, we can see that Arial, FrutigerLinotype, Helvetica and Microsoft Sans Serif

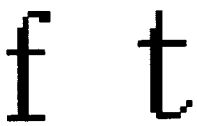
less similar.

The proposed extraction method for this legibility factor can also detect the character pairs which have similar design shapes. In Appendix C, the character pairs, which have similar design shapes and a distance d value of less than 20, are listed for each font. From these statistics, we find that the most similar character pairs are "l→I", "j→i", "l→i", "f→t", "g→p", "c→e", "b→h" etc (as shown in Figure 3.12).

(a) Font Arial "l"→"I"

(b) Font Albertus Medium"I"→ "j"

(c) Font Haetenschweiler "l"→ "I"



(d) Font Batang "f"→ "t"

(e) Font Courier New "g"→ "q"

(f) Font Georgia "c"→ "e"

(g) Font Impact "b"→"h"

Figure 3.12 Character pairs with the most similar design shapes

3.2.4 Serif versus Sans serif Legibility Factor

In our study, serif or sans serif design is used as one legibility factor. Since this factor can be easily detected from images by human being, we give the values of this legibility factor of each font directly without using any extraction method.

Table 3.4 Serif versus sans serif legibility factor for each font design

Fonts	Serif (1) /Sans serif (0)	Fonts	Serif (1) /Sans serif (0)
AlbertusMedium	0	Georgia	1
Arial	0	Haettenschweiler	0
Batang	1	Helvetica	0
BerkerlyBook	1	Impact	0
Century	1	Letter Gothic	0
Century SchoolBook	1	Microsoft Sanseriif	0
CourierNew	1	OCRB	0
FrutigerLinotype	0	Palatino	1
Garamond	1	TimesNew Roman	1

From Table 3.2, we can see that in the eighteen studied fonts, there are nine serif fonts and nine sans serif fonts.

3.2.5 Fixed versus variable spacing Legibility Factor

In this section, for different character spacing designs, the legibility factor of spacing is only considered if a given font has fixed or variable spacing. A more detailed study of how different spacing settings affect legibility will be discussed in Chapter 5. The value of this legibility factor is detected by checking the font metric information with FontLab 4.5.

Table 3.5 Fixed versus variable spacing legibility factors for each font design

Fonts	Fix(0)/Proportional(1) width	Fonts	Fix(0)/Proportional(1) width
AlbertusMedium	1	Georgia	1
Arial	1	Haettenschweiler	1
Batang	1	Helvetica	1
BerkerlyBook	1	Impact	1
Century	1	Letter Gothic	0
Century SchoolBook	1	Microsoft Sanseriif	1
CourierNew	0	OCRB	0
FrutigerLinotype	1	Palatino	1
Garamond	1	TimesNew Roman	1

In the eighteen fonts, only Courier New, Letter Gothic and OCRB have fixed space settings and all other fonts have variable space settings.

Chapter 4

Font Legibility Analysis

We compared the legibility of different font designs in Chapter Two. We measured their main font characteristics in Chapter Three. In this Chapter, we will analyze the common recognition errors of the eighteen typefaces; we will consider their legibility ranks and compare the similarities and differences of their design characteristics. These analyses will be used to summarize font design characteristics on font legibility.

4.1 Legibility Analysis of Different Typefaces

For each typeface, its main design characteristics, its legibility and its common recognition errors are discussed and analyzed in this section. The main design characteristics are examined by using the proposed extraction methods described in Chapter Three. The legibility is measured by using our legibility measurement method described in Chapter Two. The common error statistics are taken from the recognition results of four commercial OCRs and eight different test sets. The recognition errors discussed in this section consist of the common misclassifications of single lower case letters. The upper case letter and punctuation classification errors are outside the scope of our study.

4.1.1 Albertus Medium

Albertus Medium is a sans serif and a proportional width space typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.5

Ratio of x-height/Descender(XD): 3.0

Ratio of $T_v / x-height$ ($Weight_v$): 0.2

Ratio of $T_h / x-height$ ($Weight_h$): 0.13

Compared with other typefaces, Albertus Medium has relatively short ascenders and descenders. The legibility of Albertus Medium is better in size 10 than in size 8. It has an average recognition rate of 97.6% in size 8 and 98.8% in size 10. The lower case and upper case letters of typeface Albertus Medium are shown in Figure 4.1.

Figure 4.1 Lower case and upper case letters of typeface Albertus Medium

Tables 4.1 and 4.2 show the common recognition errors in Albertus Medium with size 8 and size 10. "Actual Text" means the original image label. "Recognition Result" means the recognition result by using the OCR systems. "Error Count" is the total number of errors of a given character. "Error Rate" is equal to the error count divided by the total number of occurrences of a given character. From these recognition results, several main design problems were detected as follows:

1) One of the main errors in this typeface design was the confusion between the lower case

"l (l)" and upper case " I (I)". By using our proposed similarity measurement method in Chapter 3, the shape dissimilarity distance was 2.0. This means that these two letters have very similar designs. We also found that apart from the slight difference in the very upper parts of these two letters, they are exactly the same.

2) Another significant recognition error occurs between letter "j " and letter "i ". The relatively short descender and the inconspicuous curve tail in letter "j" increase its misclassification errors with letter "i". As shown in Table 4.1, these two letters of Albertus Medium in size 8 are especially difficult to distinguish, which yields a 16.4% error rate.

3) The misclassification between letters "l " and "i " is common in typeface Albertus Medium. This is also a common error for almost all the other typefaces.

Table 4.1 Common recognition errors of Albertus Medium with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l](l)	[i]	84	7.0%
[j](j)	[i](i)	46	16.4%
[l](l)	[I](I)	46	3.9%
[c]	[C]	18	1.9%
[c]	[e]	13	1.4%
[l]	[t]	10	0.8%

Table 4.2 Common recognition errors of Albertus Medium with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l](l)	[i]	99	8.3%
[l](l)	[I](I)	38	3.18%
[j](j)	[i](i)	20	7.1%%
[i]	[l]	18	0.8%
[c]	[C]	12	1.2%
[c]	[e]	12	1.2%

4.1.2 Arial

Arial is a sans serif typeface with a proportional space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 3.0

Ratio of x-height/Descender(XD): 2.5

Ratio of $T_v / x-height\,(Weight_v)$: 0.13

Ratio of $T_h / x-height\,(Weight_h)$: 0.13

Arial has relatively short ascenders and descenders. $Weight_v$ and $Weight_h$ show that the vertical and horizontal stroke thicknesses are even. The legibility of typeface Arial is relatively better than other typefaces in both size 8 and size 10. It has an average recognition rate of 98.5% in size 8 and 98.8% in size 10. Figure 4.2 shows the images of the lower case and upper case letters in typeface Arial.

a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 4.2 Lower case and upper case letters of typeface Arial

Tables 4.3 and 4.4 show the statistics of common recognition errors in typeface Arial. From these common errors, we found the following design problems for this typeface:

1) The misclassification of "l" (l) and "I" (I) is caused by their identical designs. By using our shape similarity measurement, we found that the dissimilarity distance of these two characters is 0. This misclassification caused an 8.8% recognition error rate,

the highest in size 8 for typeface Arial.

2) The misclassification between letter "l" and letter "i" is a common recognition error for many typefaces. Although the legibility of Arial is comparably better, this type of error is still unavoidable. In our test set with the typeface Arial and size 10, 7.7% of the time letter "l" was been misclassified as letter "i".

3) Because of the short descender curve tail of letter "j", it is easily confused with the letter "i".

4) In typeface Arial and size 10, 3.9% of the time letter "g" is misclassified as letter "q". There are two types of letter "g" in typeface designs. One is with a two-storey form [1], such as the "g" in Times New Roman, and another is with a single-storey form with a tail, such as the "g" in Arial. The single-storey form with a tail makes letter "g" easily confused with letter "q".

Table 4.3 Common recognition errors of Arial with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l](I)	[l](I)	105	8.8%
[i](i)	[l](I)	20	0.9%
[e]	[a]	19	0.6%
[j](j)	[i] (i)	15	5.4%
[a]	[e]	14	0.6%
[c]	[e]	10	1.0%

Table 4.4 Common recognition errors of Arial with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	92	7.7%
[l](I)	[l](I)	32	2.7%
[q]	[g]	11	3.9%
[c]	[C]	11	1.2%
[c]	[e]	10	1.1%
[i]	[l]	10	0.4%

4.1.3 Batang

Batang is a serif typeface with a proportional space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.86

Ratio of x-height/Descender(XD): 3.25

Ratio of $T_v / x-height$ ($Weight_v$): 0.15

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

From the values of XA and XD, we can detect that the typeface Batang has a very short descender design compared with its ascender and x-height parts. $Weight_v$ and $Weight_h$ show that the horizontal strokes are very thin and that the contrast between vertical and horizontal strokes is great. In the legibility measurement, it has an average recognition rate of 96.26% in size 8 and 98.50% in size 10. Figure 4.3 shows the lower case and upper case letters of typeface Batang.

a b c d e f g h i j k l m n o p q r s t u v w x y z
A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 4.3 Lower case and upper case letters of typeface Batang

Tables 4.5 and 4.6 show the statistics of common recognition errors in typeface Batang. Similar to other typefaces, there are many misclassification errors between "l" and "i", and "j" and "i". The recognition of typeface Batang also resulted in some unique errors, (which can be used to explore its design defects) as follows:

1) Misclassification errors between "g "and "s " are common in Batang, with size 8. The letter "g" in Batang is a two-storey design and the descender in this typeface is very short. Therefore, in small size, letter "g" is easily confused with the center character "s".

2) In size 8, the short descender design makes letter "p " easily confused with letter "D".

3) In size 8, there are many types of common recognition errors with letter "h ", such as "h"→ "il", "h"→ "li" and "h" → "ll". The main reason is that in small print size, compared with the vertical strokes, the thin horizontal stroke in letter "h" is difficult to detect.

4) In small print size, like size 8, the great contrast between vertical and horizontal strokes in Batang causes the common recognition errors between the lower case letter "z " and the lower case letter "s ".

Table 4.5 Common recognition errors of Batang with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[z]	[s]	54	14.7%
[h]	[il]	33	2.0%
[h]	[li]	29	1.7%
[i]	[l]	22	0.9%
[h]	[ll]	19	1.1%
[l]	[i]	17	1.4%
[g]	[s]	14	5%
[i]	[r]	13	0.6%
[e]	[c]	13	0.4%
[c]	[e]	13	1.4%
[c]	[o]	11	1.2%
[l]	[1]	11	0.9%
[p]	[D]	10	1.3%

Table 4.6 Common recognition errors of Batang with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	70	5.9%
[j]	[i]	27	9.6%
[c]	[e]	15	1.5%
[h]	[ll]	14	0.8%
[u]	[n]	10	0.8%

4.1.4 BerkerlyBook

BerkerlyBook is a serif typeface with proportional width space design. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.50

Ratio of x-height/Descender(XD): 1.71

Ratio of $T_v / x-height$ ($Weight_v$): 0.08

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

BerkerlyBook has relatively longer ascenders and descenders compared with other typefaces. Both the vertical and horizontal strokes of BerkerkyBook are very thin. The legibility drops dramatically, when the print size decreases from 10 point to 8 point. It has an average recognition rate of 94.97% in size 8 and 98.58% in size 10. Figure 4.4 shows the lower case and upper case letters of typeface BerkerlyBook

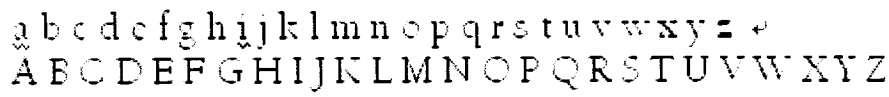


Figure 4.4 Lower case and upper case letters of typeface BerkerlyBook

In Tables 4.7 and 4.8, the statistics of common recognition errors of typeface BerkerlyBook in size 8 and size 10 show that:

1) In size 8, lower case "h" is frequently misrecognized as upper case "H". This common error dramatically drops the legibility of font BerkerlyBook in size 8, especially when comparing with other fonts. The main reason for this error is that BerkerlyBook has very thin strokes. The color of this typeface design is very light, especially in small print size.

2) The misclassification errors between the letter "l" and the letter "I" are common, in size 10 and size 8.

Table 4.7 Common recognition errors of BerkerlyBook with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[h]	[H]	252	14.6%
[l]	[i]	71	6.0%
[c]	[C]	19	2.0%

Table 4.8 Common recognition errors of BerkerlyBook with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	47	3.9%
[c]	[e]	18	1.9%

4.1.5 Century

Century is a serif typeface and a proportional space typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.86

Ratio of x-height/Descender(XD): 2.17

Ratio of $T_v/x-height$ ($Weight_v$): 0.15

Ratio of $T_h/x-height$ ($Weight_h$): 0.08

Compared with other serif typefaces, we found that the serif strokes in Century are relatively longer. It has an average recognition rate of 97.2% in size 8 and 98.3% in size 10. The images of lower case and upper case letters of typeface Century are shown in Figure 4.5.

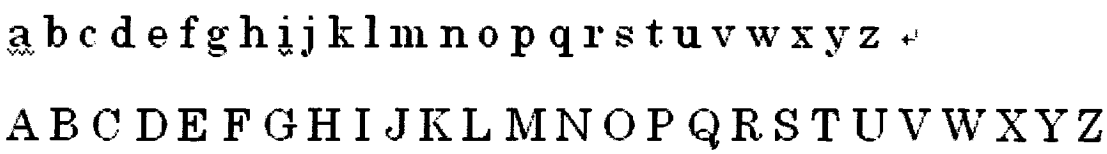


Figure 4.5 Lower case and upper case letters of typeface Century

From common recognition errors of typeface Century in size 8 and size 10, as shown in Tables 4.9 and 4.10, we can see that:

1) Similar to other typefaces, the classification errors between the letter "l" and the letter "i" are common both in size 8 and size 10. The recognition errors between the letter "j" and the letter "i" are especially common with smaller print sizes (in our case, we used 8 points).

2) Misclassifications between the letter "b" and the letter "h", and the letter "u" and the letter "n" are common. We detected a great contrast between horizontal and vertical strokes, and the relatively longer horizontal serifs may be the main cause of these errors.

Table 4.9 Common recognition errors of Century with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	59	4.9%
[j]	[i]	33	11.7%
[c]	[e]	27	2.9%
[b]	[h]	16	3.0%
[u]	[n]	15	1.2%
[c]	[C]	10	1%

Table 4.10 Common recognition errors of Century with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	59	4.9%
[i]	[l]	18	0.80%
[c]	[e]	16	1.7%
[u]	[n]	11	0.85%

4.1.6 Century School Book

Century School Book is a serif typeface and a proportional space typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.86

Ratio of x-height/Descender(XD): 2.17

Ratio of $T_v / x-height$ ($Weight_v$): 0.15

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

The main design characteristics and the common recognition errors of typeface Century School Book and typeface Century are very similar. The legibility of typeface Century School Book is very similar to typeface Century. It has an average recognition rate of 97.1% in size 8 and 98.1% in size 10. The images of lower case and upper case letters are shown in

Figure 4.6. The common errors of typeface Century School Book are shown in Tables 4.11 and 4.12.

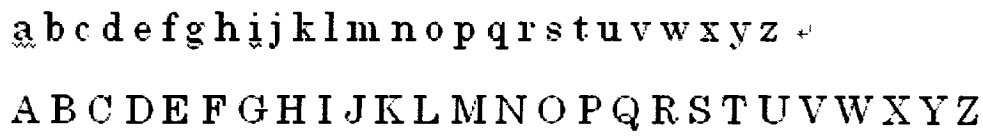


Figure 4.6 Lower case and upper case letters of typeface Century School Book

Table 4.11 Common recognition errors of Century with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	61	5.1%
[c]	[e]	43	4.5%
[u]	[n]	32	2.5%
[b]	[h]	14	2.6%
[j]	[i]	11	3.9%
[c]	[C]	10	1.1%

Table 4.12 Common recognition errors of Century with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	54	4.5%
[c]	[e]	42	4.4%
[i]	[l]	18	0.79%
[e]	[c]	17	0.56%
[u]	[n]	11	0.85%

4.1.7 Courier New

Courier New is a serif typeface with a fixed width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.4

Ratio of x-height/ Descender(XD): 2.4

Ratio of $T_v / x-height$ ($Weight_v$): 0.08

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

XA and XD show that the ascender and descender are relatively shorter in typeface Courier New. Compared with other typefaces, $Weight_v$ and $Weight_h$ show that the color of typeface Courier New is relatively light. The recognition rates of typeface Courier New in size 8 and size 10 are 98.0% and 98.1%, respectively. The images of lower case and upper case letters are shown in Figure 4.7.

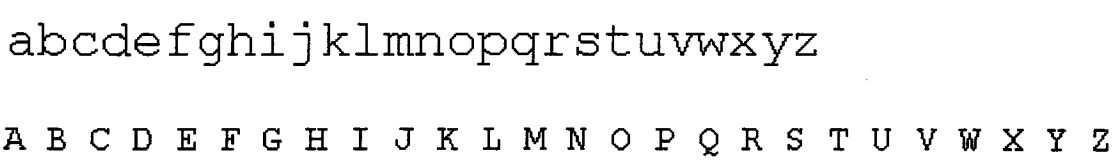


Figure 4.7 Lower case and upper case letters of typeface Courier New

From common errors of typeface Courier New in size 8 and size 10, as shown in Tables 4.13 and 4.14, we can see that:

1) Similar to other typefaces, "l" and "i", and "j" and "i", are easily misclassified in typeface Courier New.
2) The similar design, and the longer foot serif under the vertical stroke in letter "q" make the letter "g" and the letter "q" easily confused.
3) In the letter "c", the two ends of the curved stroke are too close, which makes it confused with the letter "o" in some cases.
4) The longer foot serif and the thinner strokes are the main causes of the misclassification between the letter "o" and the letter "n".

Table 4.13 Common recognition errors of Courier New with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	45	3.8%
[g]	[q]	43	4.2%
[z]	[Z]	22	5.9%
[j]	[i]	22	7.8%
[s]	[S]	18	1.1%
[w]	[W]	13	1.9%
[o]	[r]	13	0.66%
[a]	[s]	11	0.46%
[c]	[C]	11	1.2%
[i]	[l]	11	0.48%
[e]	[c]	10	0.33%
[c]	[e]	10	1.1%

Table 4.14 Common recognition errors of Courier New with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[o]	[n]	78	3.9%
[l]	[i]	56	4.7%
[c]	[e]	13	1.4%
[c]	[C]	12	1.3%
[e]	[c]	10	0.33%
[c]	[o]	10	1.1%
[z]	[Z]	10	2.7%

4.1.8 Frutiger Linotype

Frutiger Linotype is a serif typeface with a proportional width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.0

Ratio of x-height/ Descender(XD): 2.33

Ratio of $T_v / x-height$ ($Weight_v$): 0.21

Ratio of $T_h / x-height$ ($Weight_h$): 0.14

Among the eighteen typefaces, Frutiger Linotype has the best recognition rates in both size 8

and size 10, i.e. 98.81% and 98.93% respectively. Figure 4.8 shows the lower case and upper case letters in typeface Frutiger Linotype.

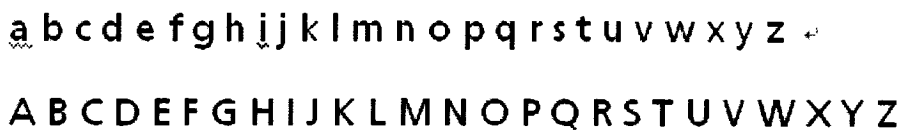


Figure 4.8 Lower case and upper case letters of typeface Frutiger Linotype

From the common errors of typeface Frutiger Linotype in size 8 and in size 10 (as shown in Tables 4.15 and 4.16), we can see that:

1) In typeface Frutiger Linotype, the design of lower case letter "l (l)" and upper case letter "I (I)" are exactly the same, which is the cause of most errors in both size 8 and size 10.

2) Single-storey and a tail design in letter "g "make it easily confused with the letter "q " in smaller print sizes (8 points in our case).

Table 4.15 Common recognition errors of Frutiger Linotype with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[I]	79	6.6%
[q]	[g]	21	7.5%
[c]	[C]	21	2.2%
[i]	[l]	17	0.7%
[c]	[e]	10	1.1%

Table 4.16 Common recognition errors of Frutiger Linotype with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[I]	97	8.1%
[c]	[e]	12	1.3%

4.1.9 Garamond

Garamond is a serif and a proportional width space setting typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.57

Ratio of x-height/Descender(XD): 1.57

Ratio of $T_v / x-height\,(Weight_v)$: 0.18

Ratio of $T_h / x-height\,(Weight_h)$: 0.09

Of the eighteen typefaces, Garamond has one of the worst legibilities. The average recognition rates of Garamond are 96.56 % and 96.65%, in size 8 and size 10, respectively. Among the 18 typefaces, Garamond has the shortest x-height. In size 10, among the eighteen typefaces, the average x-height is 14 pix, but the x-height of Garamond is only 11 pix. Figure 4.9 shows the lower case and upper case letters in typeface Garamond.

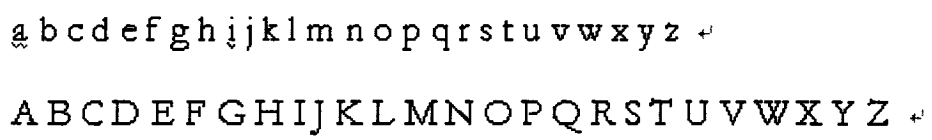


Figure 4.9 Lower case and upper case letters in typeface Garamond

Tables 4.17 and 4.18 show the common errors of typeface Garamond in size 8 and size 10. From these errors, we can see that:

1) The short x-height makes Garamond look smaller than other fonts when they are set in the same print size. This characteristic causes some common recognition errors among center letters, such as "e" and "c", "o" and "r", "c" and "r", "a" and "s".

Therefore, the short x-height is the main reason for lower legibility of the typeface Garamond.

2) The contrast between the thickest/thinnest strokes in the letter "f" causes the misclassification of letter "i" in small print size 8.

3) Misclassifications between the letters "i" and "l" are common.

Table 4.17 Common recognition errors of Garamond with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	79	6.6%
[o]	[r]	65	3.3%
[c]	[r]	34	3.6%
[f]	[i]	29	4.7%
[i]	[l]	23	1.1%
[e]	[c]	22	0.7%
[c]	[e]	12	1.3%

Table 4.18 Common recognition errors of Garamond with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	61	5.1%
[a]	[s]	31	1.3%
[c]	[C]	24	2.5%
[i]	[r]	17	0.8%
[s]	[S]	14	0.9%
[e]	[c]	13	0.4%
[c]	[e]	13	1.4%
[i]	[l]	11	0.5%

4.1.10 Georgia

Georgia is a serif typeface with a proportional space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.00

Ratio of x-height/Descender(XD): 2.33

Ratio of $T_v / x-height$ ($Weight_v$): 0.21

Ratio of $T_h / x-height$ ($Weight_h$): 0.07

The serif is relatively long in typeface Georgia. $Weight_v$ shows that the color of Georgia is relatively dark. The contrast of vertical and horizontal strokes is the greatest among the eighteen typefaces. Figure 4.10 shows the lower case and upper case letters in typeface Georgia.

abcdefghijklmnopqrstuvwxyz

A B C D E F G H I J K L M N O P Q R S T U V W X Y Z

Figure 4.10 Lower case and upper case letters in typeface Georgia

Tables 4.19 and 4.20 show the common recognition errors of Georgia with size 8 and with size 10. There are several typical errors caused by its design features. They are:

1) The uneven contrast between vertical and horizontal strokes in letter "b", and the two longer foot serifs in letter "h", make them easily confused.

2) Common errors between "u" and "n" are caused by the unevenness of thick/thin strokes and by the longer serif design.

3) Uneven thick/thin strokes in letter "o" make the width of the horizontal curve parts very narrow. This design feature causes misclassifications between the letter "o" and letters "u" or "n".

Table 4.19 Common recognition errors of Georgia with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	46	3.86%
[c]	[e]	34	3.62%
[u]	[n]	26	2.03%
[b]	[h]	17	3.22%
[i]	[l]	14	0.62%
[o]	[n]	11	0.56%
[u]	[o]	11	0.86%
[z]	[Z]	11	2.99%%
[e]	[c]	10	0.33%

Table 4.20 Common recognition errors of Georgia with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	70	5.87%
[c]	[e]	24	2.55%
[u]	[o]	20	1.56%
[e]	[c]	15	0.49%
[u]	[n]	15	1.17%
[b]	[h]	13	2.46%
[o]	[n]	10	0.51%

4.1.11 Haettenschweiler

Haettenschweiler is a sans serif typeface with a proportional space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 4.5

Ratio of x-height/Descender(XD): 6.0

Ratio of $T_v / x-height$ ($Weight_v$): 0.28

Ratio of $T_h / x-height$ ($Weight_h$): 0.17

XA and XD show that Haettenschweiler has a really short descender part compared with

other typefaces. $Weight_v$ is much bigger in Haettenschweiler than those in the other typefaces. This characteristic makes Haettenschweiler look very dark and narrow. As a result, the white spaces inside the letters are relatively much smaller. The thickness between vertical strokes and horizontal strokes is not even. The legibility of typeface Haettenschweiler is low in both size 8 and size 10. The average recognition rates of Haettenschweiler in size 8 and size 10 are 93.17% and 94.77, respectively. Figure 4.11 shows the lower case and upper case letters in typeface Haettenschweiler.

Figure 4.11 Lower case and upper case letters in typeface Haettenschweiler

Tables 4.21 and 4.22 show the common errors of typeface Haettenschweiler in size 8 and size 10. From these errors, we can see that:

1) The short descender design makes some letters, like "j ", "g ", "y ", difficult to be recognized correctly, especially in small print sizes. By analyzing the common errors of Haettenschweiler, we found some common misclassification pairs caused by the short descender design. They are "i " and "j ", "v " and "y ", "g " and "u ", and "q " and "u ".

2) The heavy vertical stroke weight makes the white space inside of letters "e" "o" and "a" smaller. This characteristic causes common misclassifications for "e " and "o ", or "a " and "e "

3) In Haettenschweiler, common recognition errors occur between letters "b" and "h" and letters "u" and "v". One of the main causes of errors is the thinner horizontal stroke in the lower part of letters "b" and "u". Although, by examining the ratio of average thickness of horizontal stroke and x-height, we found the average horizontal stroke is not very thin. However, we found that the thicknesses of some horizontal strokes varied, such as the lower horizontal part in the letter "b". Also, the narrow space between the two vertical strokes makes the thinner horizontal strokes between them easily ignored.

4) In Haettenschweiler, the letter "l (l)" has an identical design to the letter "I (I)". This comparison has been identified by using our character shape similarity comparison method, which shows zero distance between these two letters. Because of these designs, misclassifications between the letter "l" and the letter "I" are very common.

Table 4.21 Common recognition errors of Haettenschweiler with size 8

Actual Text	Recognition Result	Error Count	Error Rate
(e)	(o)	100	3.2%
(v)	(u)	75	18.0%
(b)	(h)	52	9.8%
(l)	(I)	49	4.1%
(f)	(t)	46	7.6%
(k)	(R)	30	6.1%
(y)	(v)	21	2.9%
(i)	(l)	17	0.8%
(e)	(a)	15	0.5%
(u)	(o)	13	1.0%
(j)	(i)	13	4.6%

Table 4.22 Common recognition errors of Haettenschweiler with size 10

Actual Text	Recognition Result	Error Count	Error Rate
(b)	(h)	100	18.9%
(e)	(o)	75	2.4%
(l)	(I)	61	5.11%
(y)	(v)	32	4.4%
(i)	(l)	32	1.5%
(f)	(t)	31	5.1%
(j)	(i)	30	10.7%
(g)	(u)	20	2.0%
(e)	(a)	19	0.6%
(t)	(L)	17	0.8%
(q)	(u)	14	5.0%
(o)	(g)	12	0.6%
(c)	(e)	11	1.2%
(a)	(e)	11	0.5%
(v)	(y)	10	2.4%

4.1.12 Helvetica

Helvetica is a popular sans serif typeface with a fixed width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 3.0

Ratio of x-height/Descender(XD): 2.5

Ratio of $T_v / x-height$ ($Weight_v$): 0.13

Ratio of $T_h / x-height$ ($Weight_h$): 0.13

Helvetica has a very similar design to typeface Arial. The lower case and upper case letters are shown in Figure 4.12.

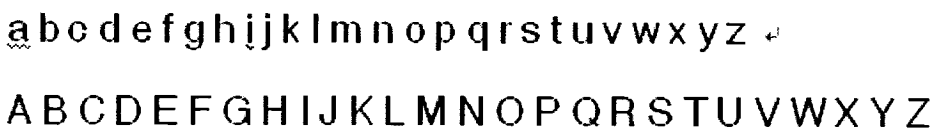


Figure 4.12 Lower case and upper case letters in typeface Helvetica

Tables 4.23 and 4.24 show the common errors of typeface Helvetica in size 8 and size 10. Similarly with Arial, the common errors of Helvetica are as follows:

1) The identical design causes the common misclassification between "l" (l) and "l" (L).

2) The misclassification between letter "l" and letter "i" is also common.

3) Letters "j" and "i" are easily confused because of the short descender curve tail of the letter "j" and the sans serif design.

4) Similar to typeface Arial and Frutiger Linotype, in typeface Helvetica, letter "g", which has single-storey form and a tail, and letter "q" are commonly confused in size 10.

Table 4.23 Common recognition errors of Helvetica with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l](I)	[l](I)	93	7.8%
[i]	[l]	26	1.2%
[e]	[a]	19	0.6%
[j](j)	[i] (i)	14	5.0%
[a]	[e]	14	0.58%
[c]	[e]	10	1.1%

Table 4.24 Common recognition errors of Helvetica with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	90	7.6%
[l] (l)	[l] (I)	29	2.4%
[q]	[g]	11	3.9%
[i]	[l]	10	0.44%
[c]	[e]	10	1.1%

4.1.13 Impact

Impact is a sans serif typeface with a proportional space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 4.5

Ratio of x-height/Descender(XD): 6.0

Ratio of $T_v / x-height$ ($Weight_v$): 0.28

Ratio of $T_h / x-height$ ($Weight_h$): 0.17

We found that the typeface Haettenschweiler and typeface Impact have very similar design characteristics. The ratio of x-height and ascender is one of the largest among the eighteen typefaces. The ratio of x-height and descender is almost double the average ratio of the 18 typefaces. Impact has a dark stroke weight compared with other typefaces. Similar to typeface Haettenschweiler, the legibility of typeface Impact is low compared to the 18 typefaces both in size 8 and size 10. The average recognition rates of Impact in size 8 and size 10 are 93.02% and 94.85%, respectively. Figure 4.13 shows the lower case and upper case letters in typeface Impact.

Figure 4.13 Lower case and upper case letters in typeface Impact

From Tables 4.25 and 4.26, we found that typeface Impact has common errors very similar to typeface Haettenschweiler, as follows:

1) A short descender design causes common errors between letters "i " and "j ", "v " and "y ", "g " and "u ", and "q " and "u ".

2) A dark stroke design results in the smaller white space in typeface Impact, which decreases the legibility of letters "e ", "o", "c "and "a ".

Table 4.25 Common recognition errors of Impact with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[e](e)	[o](o)	98	3.2%
[v](v)	[u](u)	75	18.0%
[b](b)	[h](h)	51	9.7%
[l](l)	[I](I)	47	3.9%
[f](f)	[t](t)	43	7.0%
[k](k)	[R](R)	33	6.7%
[l](i)	[l](l)	18	0.8%
[e](e)	[a](a)	15	0.5%
[y](y)	[v](v)	14	1.9%
[u](u)	[o](o)	13	1.0%
[j](j)	[i](i)	12	4.2%
[v](v)	[y](y)	11	2.6%
[c](c)	[e](e)	10	1.1%
[l](i)	[t](t)	10	0.4%

3) The dark vertical strokes and the unevenness of strokes cause errors among the letters "b " and "h " and the letters "u " and "v ".

4) The identical design between letters "  (l)" and "  (I)" is also a defect in typeface Impact.

Table 4.26 Common recognition errors of Impact with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[b] (b)	[h] (h)	103	19.5%
[e] (e)	[o] (o)	89	2.9%
[l] (l)	[I] (i)	63	5.2%
[I] (i)	[b] (b)	45	2.0%
[y] (y)	[v] (v)	41	5.6%
[j] (j)	[i] (i)	30	10.7%
[q] (q)	[u] (u)	27	9.6%
[g] (g)	[u] (u)	25	2.5%
[i] (i)	[l] (l)	23	1.0%
[f] (f)	[t] (t)	21	3.5%
[e] (e)	[a] (a)	19	0.6%
[t] (t)	[L] (L)	15	0.7%
[o] (o)	[g] (g)	12	0.6%
[a] (a)	[e] (e)	11	0.5%
[c] (c)	[e] (e)	11	1.2%
[i] (i)	[u] (u)	10	0.4%
[v] (v)	[y] (y)	10	2.4%

4.1.14 Letter Gothic

Letter Gothic is a common sans serif and a fixed width inter-letter space setting typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.5

Ratio of x-height/Descender(XD): 3.0

Ratio of $T_v / x-height$ ($Weight_v$): 0.07

Ratio of $T_h / x-height$ ($Weight_h$): 0.07

$Weight_v$ and $Weight_h$ show that Letter Gothic has the lightest stroke weight among the 18 typefaces. Letter Gothic is one of the typefaces whose legibility is comparatively worse both in size 8 and size 10. The average recognition rates of letter Gothic are 96.8% and 96.3% in size 8 and size 10, respectively. Figure 4.14 shows us the lower case and upper case letters in typeface Letter Gothic. The common errors of typeface Letter Gothic in size 8 and size 10 are listed in Table 4.27 and Table 4.28.

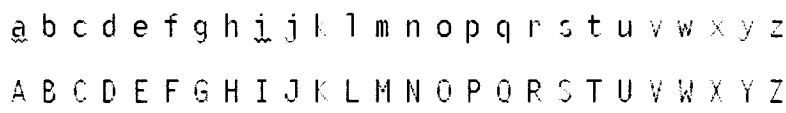


Figure 4.14 Lower case and upper case letters in typeface Letter Gothic

Compared with other typefaces, we found that:

1) In Letter Gothic, the misclassification between the letter "f" and the letter "t" is very serious. The error rates between letter "f" and letter "t" is 12.3% in size 8 and 6.2% in size 10. By evaluating the characteristics of typeface Letter Gothic, we found that the thinner stroke is the main cause of this type of error.
2) In Letter Gothic, the width of letter "m" is almost the same as letter "n". This may be the main cause of the errors between them in small size 8.
3) The curved stroke in letter "r" has an extended vertical stroke. This may be the main reason for misclassification with letter "n" in small print size.
4) The recognition errors between letters "i" and "l", and between letters "j" and "i" are also common in typeface Letter Gothic.

Table 4.27 Common recognition errors of Letter Gothic with size 8

Actual Text	Recognition Result	Error Count	Error rate
[i]	[l]	82	3.6%
[f]	[t]	75	12.3%
[i]	[I]	67	3.0%
[l]	[i]	53	4.4%
[c]	[C]	20	2.1%
[r]	[n]	19	1.3%
[m]	[n]	13	1.6%
[j]	[i]	13	4.6%
[c]	[e]	12	1.3%
[z]	[Z]	10	2.7%

Table 4.28 Common recognition errors of Letter Gothic with size 10

Actual Text	Recognition Result	Error Count	Error rate
[i]	[l]	121	5.4%
[f]	[t]	42	6.2%
[l]	[i]	38	3.2%

4.1.15 Microsoft Sans Serif

Microsoft Sans Serif is a sans serif and a proportional width space setting typeface. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 3.0

Ratio of x-height/Descender(XD): 2.5

Ratio of $T_v / x-height$ ($Weight_v$): 0.20

Ratio of $T_h / x-height$ ($Weight_h$): 0.13

Compared with other typefaces, the heights of ascender and descender are relatively short in typeface Microsoft Sans Serif. Among the 18 typefaces, the legibility of typeface Microsoft Sans Serif is comparatively good. Its average recognition rates are 98.53% in size 8 and 98.74% in size 10. Figure 4.15 show the images of lower case and upper case letters in

Microsoft Sans Serif.

Figure 4.15 Lower case and upper case letters of typeface Microsoft Sans Serif

Tables 4.29 and 4.30 show the common recognition errors of typeface Microsoft Sans Serif in size 8 and size 10. From the design characteristics and the common recognition errors of Microsoft Sans Serif, we found that:

1) The identical design of lower case letter "l" (l) and upper case "I" (I) cause common recognition errors.

2) The short curve parts in letters "f" and "t" make them easily confused.

3) The short curve tail in letter "j" makes it very similar to letter "i" in small print sizes.

Table 4.29 Common recognition errors of Microsoft Sans Serif with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	90	7.6%
[l](I)	[l](I)	49	4.1%%
[c]	[C]	39	4.1%
[f]	[t]	30	4.9%
[j](j)	[i]	28	10.0%
[e]	[a]	19	0.6%
[i]	[l]	17	0.7%
[c]	[e]	14	1.5%
[a]	[e]	13	0.5%
[z]	[Z]	12	3.3%

Table 4.30 Common recognition errors of Microsoft Sans Serif with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l] (l)	[l] (I)	81	6.8%
[i]	[l]	61	2.7%
[f]	[t]	28	4.6%
[c]	[e]	12	1.2%

4.1.16 OCRB

OCRB is a sans serif typeface, with a fixed width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 2.0

Ratio of x-height/Descender(XD): 2.8

Ratio of $T_v / x-height$ ($Weight_v$): 0.21

Ratio of $T_h / x-height$ ($Weight_h$): 0.14

XA and XD show that the ascender and descender parts of this typeface are relatively short. It has an average recognition rate of 96.7% in size 8 and 97.5% in size 10. The images of lower case and upper case letters of typeface OCRB are shown in Figure 4.16.

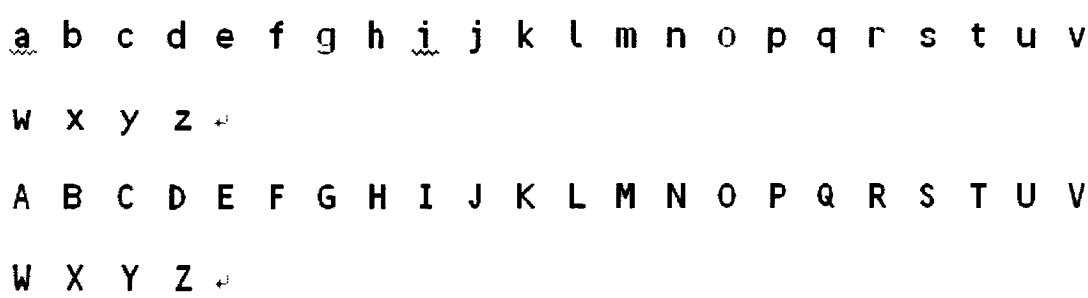


Figure 4.16 Lower case and upper case letters of typeface OCRB

Tables 4.31 and 4.32 show the common recognition errors of typeface OCRB in size 8 and size 10. From these common errors, the following design defects were found:

1) In both sizes 8 and 10, the misclassifications between lower case "l" and upper case "L" are very serious. In comparison with other typefaces, we found that OCRB is the only typeface that has a horizontal tail design in lower case letter "l". It is a very uncommon design.
2) Normally, the height in the middle vertex of the lower case 'w' is the same as the two diagonal bars. However, in OCRB, the height in the middle vertex of lower case "w" is only half of the height of the two diagonal bars. This unfamiliar design may be the cause of common recognition errors between the letter "w" and the letter "u".
3) Among the eighteen typefaces, a comparatively short descender height for OCRB may be the main cause of misclassification between the letter "j" and the letter "i".

Table 4.31 Common recognition errors of OCRB with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[L]	499	41.8%
[w]	[u]	46	6.7%
[j]	[i]	37	13.2%
[c]	[C]	30	3.2%
[l]	[i]	20	1.6%
[v]	[V]	13	3.1%
[c]	[e]	11	1.2%

Table 4.32 Common recognition errors of OCRB with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[L]	205	17.2%
[j]	[i]	33	11.8%
[w]	[u]	27	3.9%
[l]	[I]	24	2.0%
[c]	[e]	11	1.2%

4.1.17 Palatino

Palatino Roman is a serif typeface with a proportional width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.86

Ratio of x-height/Descender(XD): 1.63

Ratio of $T_v / x-height$ ($Weight_v$): 0.15

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

Among the 18 typefaces, Palatino has a comparatively longer ascenders and descenders. The descender part in Palatino is long enough to distinguish between letter "y" and letter "v", and letter "i" and letter "j". Palatino has an uneven contrast between vertical and horizontal strokes. The recognition rate of typeface Palatino in size 8 is 97.10%, and in size 10 it is 98.91%. The images of lower case and upper case letters of typeface Palatino are shown in Figure 4.17.

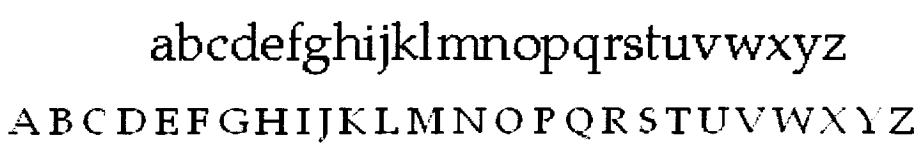


Figure 4.17 Lower case and upper case letters of typeface Palatino

In the common errors (as shown in Tables 4.33 and 4.34), we found that:

1) Similar to other typefaces, the misclassification between the letter "l" and the letter "i" is common.

2) In typeface Palatino with small print size 8, the misclassification between "j" and "J" is common because of their similar designs.

Table 4.33 Common recognition errors of Palatino with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	55	4.6%
[i]	[l]	23	1.0%
[j]	[J]	17	6.0%
[s]	[S]	13	0.8%
[c]	[e]	12	1.3%
[e]	[c]	10	0.3%

Table 4.34 Common recognition errors of Palatino with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	78	6.5%
[e]	[c]	14	0.5%
[c]	[e]	11	1.1%

4.1.18 Times New Roman

Times New Roman is a serif typeface with proportional width space setting. In legibility factors, the important ratios include:

Ratio of x-height/Ascender (XA): 1.86

Ratio of x-height/Descender(XD): 2.17

Ratio of $T_v / x-height$ ($Weight_v$): 0.15

Ratio of $T_h / x-height$ ($Weight_h$): 0.08

Compared to other typefaces, Times New Roman has a relatively moderate proportion among x-height, ascender and descender. The contrast of vertical and horizontal strokes is relatively great when compared with other typefaces. The recognition rate of typeface Times

New Roman in size 8 is 97.18%, and in size 10 it is 98.79%. The images of lower case and upper case letters of typeface Times New Roman are shown in Figure 4.17.

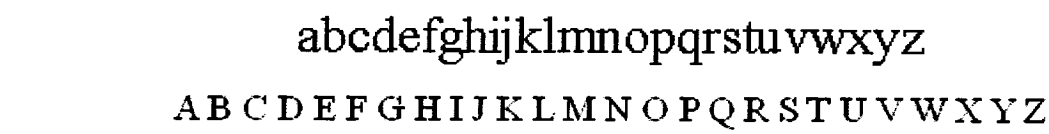


Figure 4.18 Lower case and upper case letters of typeface Times New Roman

In the common errors (as shown in Tables 4.35 and 4.36), we found that:

1) The misclassifications between the letter "l" and the letter "i" are common in typeface Times New Roman.

2) The uneven stroke contrast in the letter "b" and the longer foot serif under vertical strokes in the letter "h" make them easily confused. Similarly, the letter "u" and letter "n" are confused in some cases for the same reasons.

Table 4.35 Common recognition errors of Times New Roman with size 8

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	74	6.2%
[i]	[l]	20	0.88%
[c]	[C]	17	1.8%
[u]	[n]	15	1.17%
[c]	[e]	12	1.27%

Table 4.36 Common recognition errors of Times New Roman with size 10

Actual Text	Recognition Result	Error Count	Error Rate
[l]	[i]	63	5.28%
[c]	[e]	18	1.91%
[l]	[I]	12	1.0%
[b]	[h]	12	2.27%

4.2 Effect of the Legibility Factors

By analyzing the legibility of each typeface, we conclude with several summaries about typeface design. These summaries are mainly based on the typeface characteristics and the design of some individual letters. These characteristics include serif design, x-height, ascender and descender, typeface stroke weight, character shape similarity and unfamiliar character design, as described below:

4.2.1 Serif design

According to the analysis of typeface legibility and their common errors, we found that the long serif design affect the legibility of individual letters. The long serif design mainly cause the misclassifications on "b" and "h", "u" and "n", "o" and "n", or "o" and "u". This result is consistent with Tinker's analysis, which suggests [11] that the legibility of individual letters is diminished by the use of hair lines or of long or heavy serifs.

4.2.2 x-height, ascender and descender

Our experimental results proved that the balance between a moderately large x-height and noticeable ascenders and descenders is very important for a legible font. When the x-height is very short, many misclassification errors of center letters occur, especially for smaller print sizes. Of the eighteen typefaces, typeface Garamond has the shortest x-height. Therefore, the errors between "e" and "c", "o" and "r", "c" and "r", and "a" and "s" are typical errors in typeface Garamond with smaller print size 8. The short ascender and descender can also

lower the typeface legibility. In both typeface Haettenschweiler and typeface Impact, the x-heights are 4.5 times larger than their ascenders and the x-heights are 6 times larger than their descenders. These letter proportions make the ascender and descender parts unclear and cause many recognition errors for ascender and descender letters. These letter proportions can easily cause the recognition errors on "l"(i) and "j"(j), "v"(v) and "y"(y), "g"(g) and "u"(u), "q"(q) and "u"(u), "o"(o) and "g"(g), and "f"(f) and "t"(t).

4.2.3 Typeface stroke weight

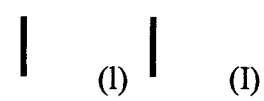
Among the eighteen typefaces, the typefaces Haettenschweiler and Impact have the darkest stroke weights. In these two typefaces, the ratio between average vertical stroke thickness and x-height is 0.28 and the ratio between average horizontal stroke thickness and x-height is 0.17. These two ratios are the largest among the eighteen typefaces. In particular, the ratios between vertical stroke thickness and x-height of typefaces Haettenschweiler and Impact are 1.75 times bigger than the average ratio of the eighteen typefaces. Therefore, the two typefaces look very dark and narrow and the white spaces inside the letters are very small. As a result, common errors for "e" (e) and "c" (c), "e" (e) and "o" (o), and "a" (a) and "e" (e) occur and the legibility of these two typefaces dropped dramatically. On the other hand, the very light stroke weight also affects typeface legibility. Among the eighteen typefaces, BerkerlyBook, Courier New and Letter Gothic all have very light stroke weights in both vertical and horizontal strokes. These characteristics decrease the typeface legibility, especially at a smaller print size.

The contrast between vertical and horizontal stroke weights also shows the importance of

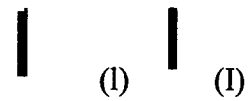
typeface legibility. Among the eighteen typefaces, some have very great contrasts between horizontal and vertical stroke weights. For example, in typeface Georgia, the weight of the vertical stroke is 3 times darker than the weight of the horizontal stroke. This design feature plus the longer serif design results in the misclassifications between "b" and "h", "u" and "n", "o" and "u", or "o" and "n" in typeface Georgia.

4.2.4 Character Shape Similarity

The similar character shape design is an important factor that affects legibility. Of the eighteen typefaces, the high legibility typefaces all have larger average dissimilarity distance d values. Therefore, the dissimilarity of character shape design is a requirement for high typeface legibility. In our study, we also found some common character pairs that have similar character shape designs, which decrease the individual character's legibility. One of the serious and typical examples is between the lower case letter "l" and the upper case letter "I" in sans serif typeface design. In our study, there are nine sans serif typefaces among eighteen typefaces. In the sans serif typefaces, there are six typefaces that use identical designs for letters "l" and "I": Arial, FrutigerLinotype, Haettenschweiler, Helvetica, Impact, and Microsoft Sans Serif (as shown in Figure 4.19 (a)). Although the "l" and "I" designs are not exactly the same for typeface Albertus Medium, they are very similar (as shown in Figure 4.19 (b)). Of the nine sans serif typefaces, only Letter Gothic and OCRB use a different design for letters "l" and "I". They use the serif typeface design for the letter "I" and add small horizontal bar for the letter "l" (as shown in Figure 4.19 (c), (d)).



(a) Identical design between "l" and "I"



(b) "l" and "I" in typeface Albertus Medium

(c) "l" and "I" in typeface Letter Gothic

(d) "l" and "I" in typeface OCRB

Figure 4.19 Character designs of "l" and "I" in sans serif typefaces

4.2.5 Unfamiliar character design

Although dissimilarity of letters is a requirement for legibility, familiarity is also important. Of the eighteen typefaces, OCRB has a very particular design of the lower case letter "l", which has a horizontal stroke (as shown in Figure 4.20(c)). This unfamiliar design makes lower case letter "l" easily confused with its upper case letter "L". We also show the designs of the lower case letter "l" and the upper case letter "L" of typeface Times New Roman and typeface Arial in Figure 4.20 (a), (c). These two typefaces use familiar designs for the lower case letter "l". Therefore, in typeface design, trying to use familiar forms for character design is important to maximize legibility. These results are consistent with Watts and Nisbet's statement in [6], "familiar forms are more legible than unfamiliar ones".

(a) Letters "l" and "L" in Times New Roman

(b) Letters "l" and "L" in Arial

(c) Letters "l" and "L" in OCRB

Figure 4.20 Designs in lower case letter "l" and upper case letter "L"

Chapter 5

The Effect of Inter-letter Spacing on Legibility

In previous chapters, several typical characteristics were measured and analyzed based on font legibility. This chapter will mainly focus on one of the most important font legibility factors, inter-letter spacing. Firstly, we will show the design principles regarding how a type designer sets the inter-letter spacing of a typeface. Secondly, we will address the issue of how the different inter-letter spacing settings affect legibility. For each studied typeface, we create texts, which differ only in terms of their inter-letter spacing. We found that some typefaces had relatively larger spacing settings, which could be condensed without affecting their legibility. Moreover, some typefaces had relatively tight spacing settings and the expanded inter-letter spacing settings could increase their legibility.

5.1 Basic Design Principles of Character Spacing Setting

In the past, mathematicians and software engineers developed algorithms in attempts to automate the process of determining letter spacing, but the results they achieved were always very poor [25]. Ultimately, what really counted in type design is what was perceived by the human eye. The type designers believed that the spacing of those letters relied upon the rules

inherent in their optical appearance. In [26], two aspects of good letter fit were mentioned. First, the letters must be placed so that the spaces between them are consistent. A common analogy here is to look upon the spaces between letters as a sort of volume that is filled with sand: you ought to be able to put about the same amount of sand between any pair of letter. Second, good letter fit is a function of the absolute amount of space between the letters. That is, for different sizes of a typeface, we should adjust the letter spacing.

In the type industry, the most common methods of spacing design start by considering the letters with simple straight stems and simple curved stems. As a result, the letters "n" and "o", and "H" and "O" are the best candidates. The different spacing setting methods in the type industry have very similar principles. Among them, the most common and typical theory comes from Tracy [1].

In his method, the letters are grouped into four categories:

1) Letters with a straight upright stroke:

   Upper cases: B D E F H I J K L M N P R U

   Lower cases: b d h i j k l m n p q r u

2) Letters with a round stroke:

   Upper cases: C D G O P Q

   Lower cases: b c d e o p q

3) Triangular letters

   Upper cases: A V W X Y

   Lower cases: v w z y

4) The odd letters

Upper cases: S T Z

Lower cases: a f g s t z

In Tracy's method, upper case inter-letter spacing settings are mainly designed by using the following steps:

1) Start with basic letter "H". Four "H"s, which should be set together in a row, are required in order to adjust the side bearings , which include left side bearing and right side bearing (see definition on page 26) of H, as follows:

During the adjustment, the weight of the stroke, the width of its interior space and the shape and length of its serifs are all considered for a harmonious design. The next step is to measure the width between the uprights of the letter and set half of that amount on each side of the letter on all four "H"s. Since the crossbar of the "H"s makes the vertical stroke look smaller, the distance between the "H"s should be set smaller than the interior space of H. However, for the bold font, the interior space should equal to the side space. The serifs of four adjacent letter "H"s make a linking effect, so that the side spaces of sans serif H should be smaller than those of serif H. After getting a harmonious look, the half width of the space between two H's is considered as a side bearing for H and for all the other uppercase letters that have a straight vertical stroke.

2) Once the side bearing of H has been allocated, the next important letter, "O", should be considered. In order to set the balance of the side space of "O", two test strings are used. After balancing the spaces between the letters in each test string, the space between "H"

and "O" is measured. After subtracting the side bearing amount belonging to "H", the side bearing of "O" is determined.

HOHOH HHOOHH

3) After balancing the basic letters "H" and "O", the side space of other uppercase letters can be set by considering the four groups of letters listed earlier. The principle is described below:

4A4	1B3	5C3	1D5	1E3
1F3	5G2	1I1	4J1	1K4
1L4	2M1	2N2	1P5	5Q5
1R4	6S6	4T4	1U2	4V4
4W4	4X4	4Y4	3Z3	

1 = same as side bearing of H

2 = slightly less than the side bearing of H

3 = about half of side bearing of H

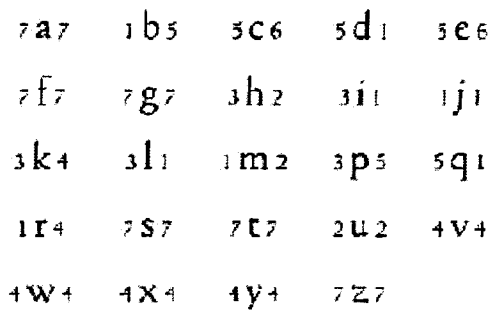
4 = Minimum space

5 = same as side bearing of O

6 = must be spaced visually, between standards

Similar to the upper case letters, for setting the side bearing of the lower case letters, two standard letters are chosen. They are "n" and "o". By measuring the width of the uprights of

"n", half of the space is set to the left side bearing of n. The right side bearing of n is slightly less than the left side bearing of n since the arched corner seems to add the space. Next, four copies of "n" should be put in a row in order to adjust the side spacing of n. The serif or sans serif and bold or normal fonts should be considered as upper case letters. After deciding on the side spacing of "n", the letter "o" should be considered. Some test strings, which include several "n"s and "o"s should be used to balance the spacing of "o", such as "nnonn", "nnonon" or "nnoonn". After getting the balanced spacing settings of "n" and "o", the side spacing of other lowercase letters are examined. The scheme is shown below:



1 = same as left side bearing of n

2 = same as right side bearing of n

3 = slightly more than left side bearing of n

4 = Minimum space

5 = the same side bearing of o

6 = slightly less than the side bearing of o

7 = must be spaced visually, between standards

Designers always check the spacing of their typefaces by looking at proofs of each character as they are set between every other character. Every possible letter combination is scrutinized.

The basic inter-letter design principles are set towards a readable typeface and they are followed by subtle fine-tuning to get as consistent a rhythm as possible [25]. After the designers have spaced their typefaces satisfactorily, it is time to consider kerning. The kerning routine allows a letter to intrude into the 'air space' of another letter when the circumstances make it desirable. The familiar examples are the closing up of T, V, W, and Y with non-ascender lower case letters.

5.2 Studies of Inter-letter Spacing on Legibility

The objective of our study is to discover if the typeface design has reasonable inter-letter spacing and if the adjustments of inter-letter spacing have positive, negative, or no impact on legibility for different typeface deigns. Therefore, we can determine how the different inter-spacing settings affect font legibility. In this thesis, we chose six typical study typefaces. These six typefaces include different characteristics. For each typeface, we created different fonts, which differ only in terms of their inter-letter spacing settings. By measuring the legibility of different inter-letter spacing settings and analyzing their corresponding touching errors, suggestions for inter-letter spacing design are given.

5.2.1 Method

In our study, in order to examine the effect of inter-letter spacing setting on legibility, we chose six typefaces as the targets. They were Helvetica, Century, Microsoft Sans Serif and Garamond, OCRB and Frutiger Linotype. For each typeface, 9 different fonts were constructed, distinctly based on their inter-letter spacing settings. A base font was constructed by using a 10 point print size and the inter-letter spacing setting of the original design. The other eight fonts were constructed with expanded or condensed by 0.2 point, 0.4 point, 0.6

point, 0.8 point inter-letter spacing settings. The adjustments of inter-letter spacing settings were made in Microsoft Word. These spaces were added only at the edges of each letter glyph so that each letter shape was unaltered by the manipulation of spacing. Appendix D shows the image examples of the nine different inter-letter spacing settings for each typeface.

In the legibility tests, different character strings, high frequency words, low frequency words, non-sense words and bi-gram words were designed as test strings (as shown in Appendix B). The test strings included 1316 words and 3875 characters. All the nine fonts of each of the six typefaces with different spacing settings were used for all texts and were employed in the experiments. The test strings were all converted to images with a resolution of 200*200 dpi. Two OCR systems, FineReader and TextBridge, were used as test tools and the recognition results are recorded. By using a dynamic string matching method, the recognition rates were calculated.

5.2.2 Legibility Measurement of Six Typefaces with Different Spacing Settings

For each typeface, the recognition rates of different inter-letter spacing setting were calculated. By examining their different spacing settings and corresponding legibilities, we grouped the six typefaces into three categories, as follows:

Category one: Original inter-letter spacing is large; condensed inter-letter spacing does not affect legibility

Of the six typefaces, when the inter-letter spacing setting is condensed, two typefaces, OCRB and Frutiger Linotype have similar legibility trends. As shown in Figure 5.1, for the typeface OCRB, the condensed inter-letter spacing settings, 0.2 point, 0.4 point, 0.6 point and even 0.8 point, have nearly no effect on legibility. Similarly, the legibility of typeface FrutigerLinotype have nearly negligibly impact with the condensed inter-letter spacing settings 0.2 point, 0.4 point, and 0.6 point, as shown in Figure 5.2. Therefore, for economic purpose, the inter-letter spacing of these two typefaces can be set more tightly (such as condensed 0.6 point) , yet retaining their legibility.

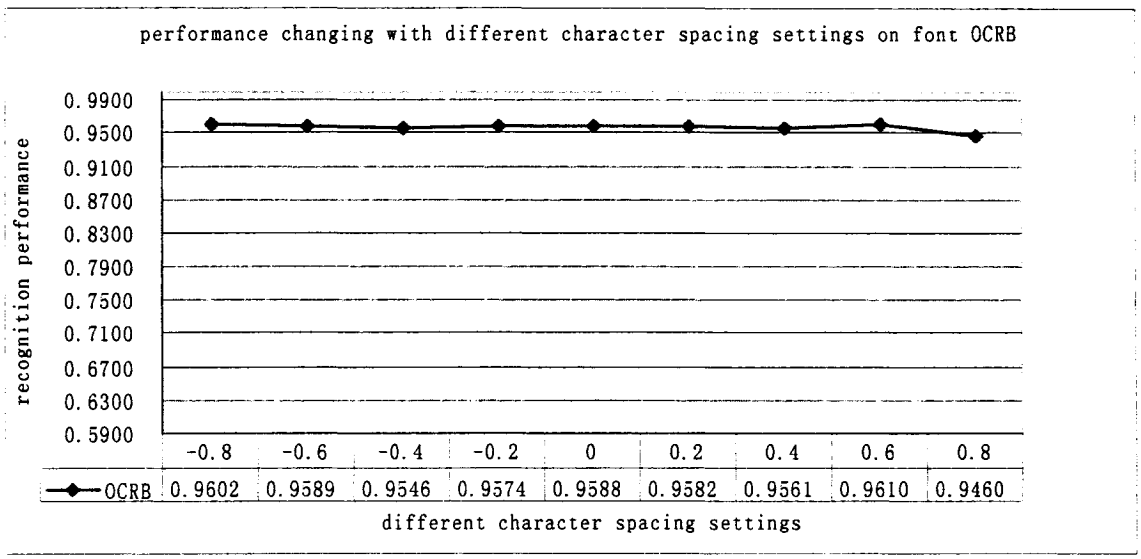


Figure 5.1 Recognition rate of typeface OCRB with different inter-letter spacing settings

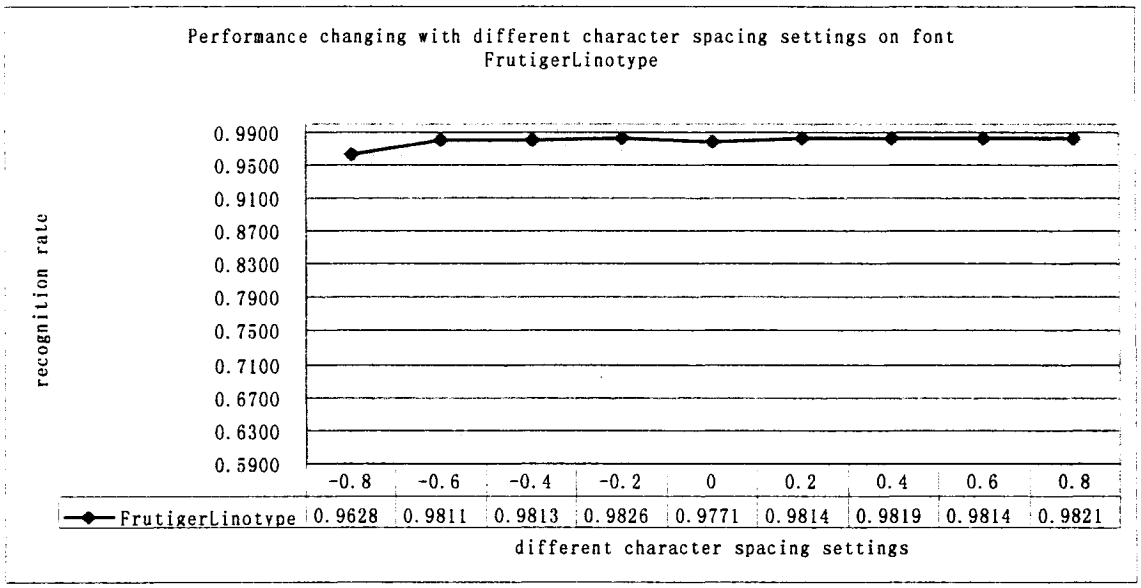


Figure 5.2 Recognition rate of typeface Frutiger Linotype with different inter-letter spacing settings

Category Two: Original inter-letter spacing is small; expanded inter-letter spacing increases its legibility

The experimental results showed that expanding the inter-letter spacing setting could increase the legibility of the typeface Garamond. As shown in Figure 5.3, the expanded inter-letter spacing settings 0.2 point, 0.4 point and 0.6 point, could increase the legibility of typeface Garamond. By contrast, the condensed spacing setting could drop the legibility of the typeface Garamond dramatically. Therefore, the inter-letter spacing can be expanded in order to maximize the legibility of typeface Garamond.

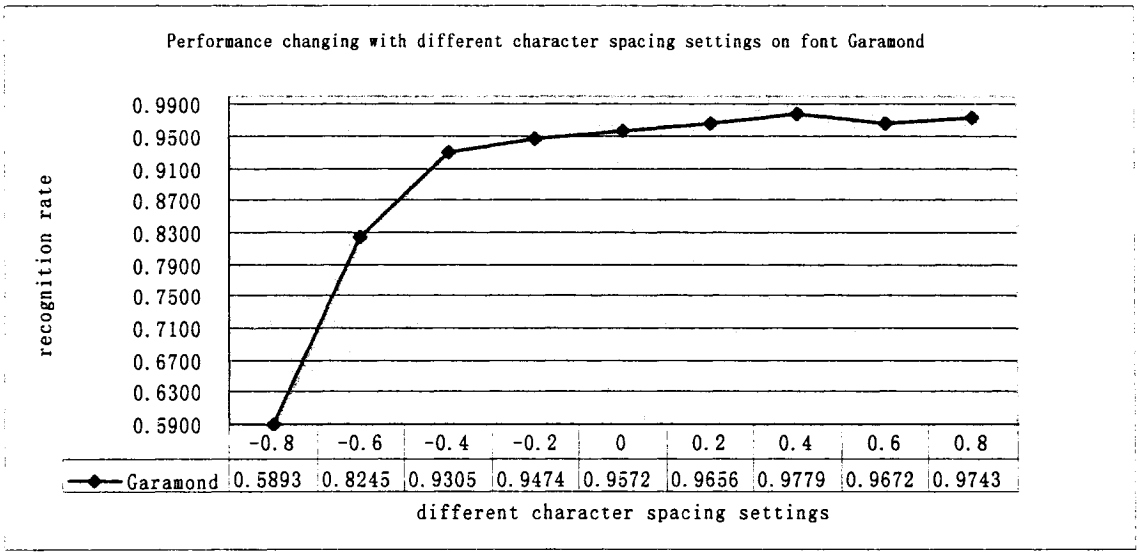


Figure 5.3 Recognition rate of typeface Garamond with different inter-letter spacing settings

Category Three: Compared with condensed and expanded spacing, original spacing has higher legibility

In the typefaces Helvetica and Microsoft Sans Serif, as shown in Figures 5.5 and 5.6, the expanded inter-letter spacing settings do not increase their legibilities and the condensed inter-letter spacing settings decreased their legibilities. In the typeface Century, the expanded inter-letter spacing has nearly negligibly impact on legibility, and condensed inter-letter spacing drops its legibility, as shown in Figure 5.4. Therefore, for these three typefaces, the original spacing settings have higher legibilities compared with condensed and expanded spacing settings.

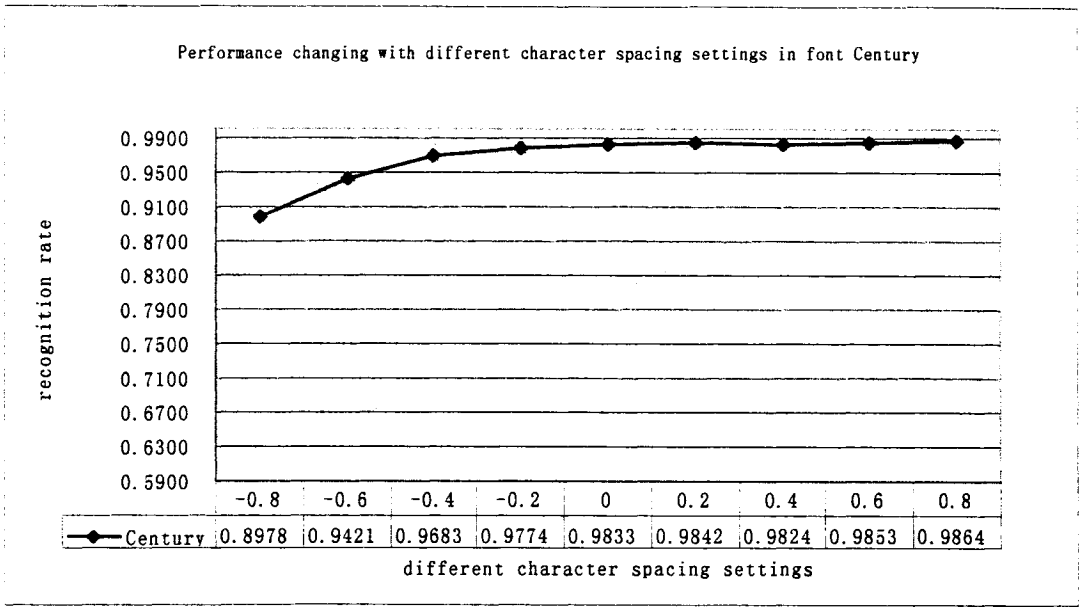


Figure 5.4 Recognition rate of typeface Century with different inter-letter spacing settings

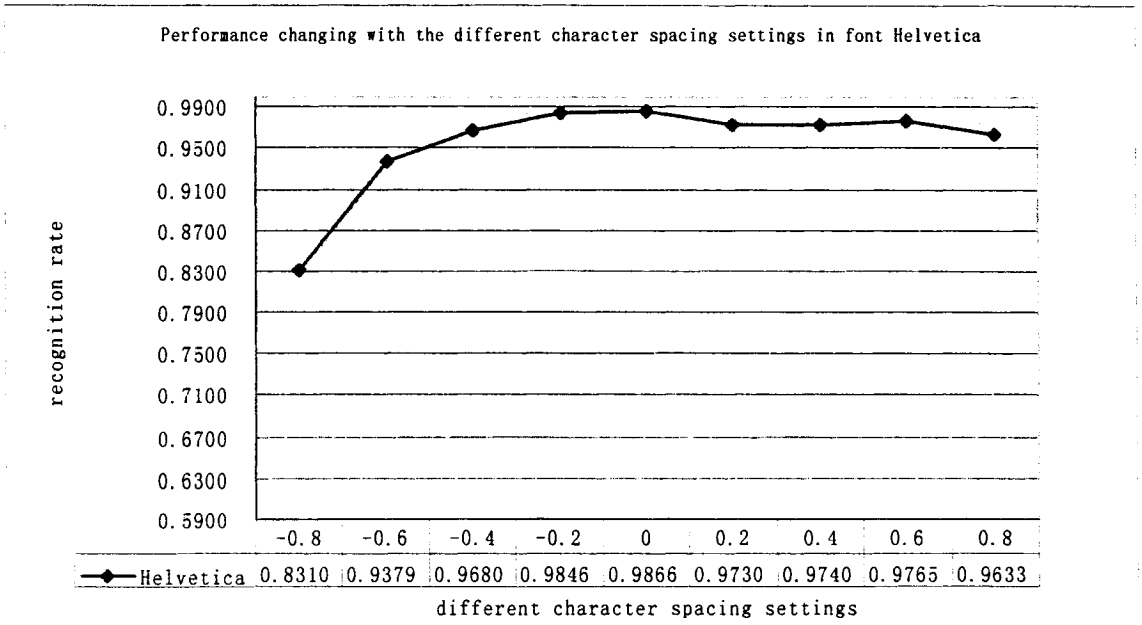


Figure 5.5 Recognition rate of typeface Helvetica with different inter-letter spacing settings

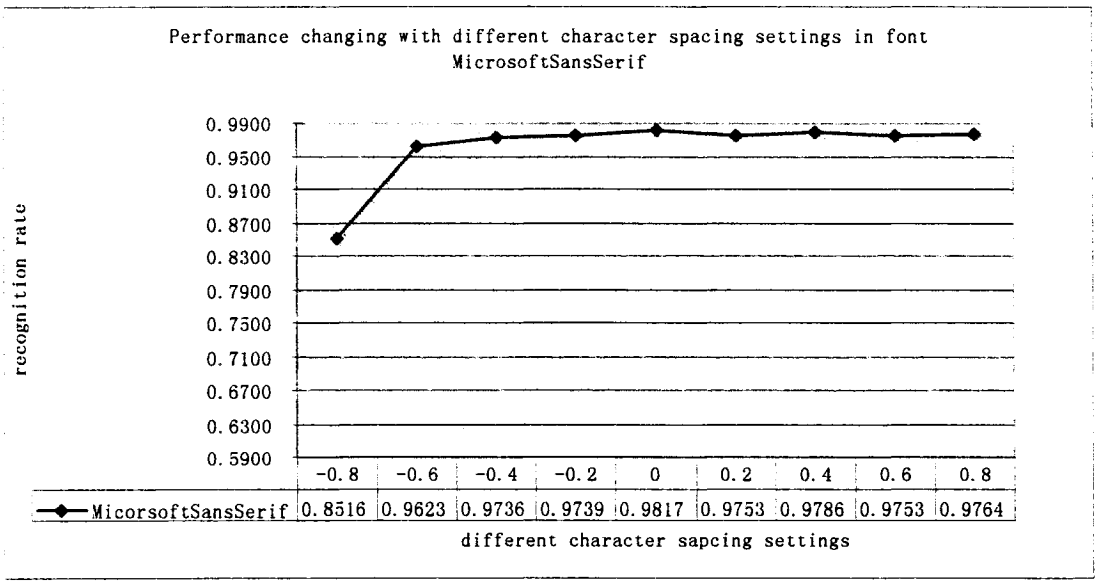


Figure 5.6 Recognition rate of typeface Microsoft Sans Serif with different inter-letter spacing settings

5.2.3 Inter-letter Spacing Analysis

In this section, the effect of inter-letter spacing on legibility is analyzed from three perspectives, which are: a) the basic spacing setting of standard letters, b) the side bearings of 26 lower case letters, and c) the touching errors of different spacing settings.

1) Basic spacing setting of standard letters in six typefaces

As in Section 5.1, some standard letters were chosen for upper case and lower case letter designs. They were upper case letters "H" and "O" and lower case letters "n" and "o". Table 5.1 shows the left side bearings and the right side bearings of these standard letters in the six studied typefaces. We found that the inter-letter spacing settings of the standard letters differ considerably from one typeface design to another. In typeface OCRB, the side bearings of the four standard letters are much bigger that those of the other typefaces. On the contrary, in typeface Garamond, the side bearings of the standard letters are comparatively smaller than other typefaces. These basic spacing settings affected the side bearing settings of the

non-standard letters. The side bearing information was detected by using Fontlab 4.5.

Table 5.1 Left side bearings (LSB) and right side bearings (RSB) of letters "H", "O", "n" and "o" in six typefaces

Typefaces	H	O	n	o
	(LSB,RSB)	(LSB,RSB)	(LSB,RSB)	(LSB,RSB)
OCRB	(295,294)	(244,243)	(279,278)	(240,239)
Frutiger Linotype	(193,194)	(109,108)	(164,164)	(94,94)
Garamond	(39,53)	(94,98)	(35,21)	(73,74)
Helvetica	(158,156)	(80,80)	(133,133)	(72,72)
Century	(96,96)	(111,110)	(64,42)	(69,69)
Microsoft Sans Serif	(164,164)	(96,94)	(139,135)	(72,72)

In Figures 5.7 and 5.8, the typical inter-letter spacings of test strings in typeface OCRB and typeface Garamond are shown. In Tray's inter-letter spacing design method, he mentioned that "the serifs of four adjacent letter 'H's make a linking effect, so that the side spaces of sans serif H should be smaller than those of serif H". However, the side spaces of sans serif H in typeface OCRB are much larger than those of serif H in typeface Garamond. That means the adjustments of inter-letter spacing settings in these two typefaces are necessary.

HHHH HOHOH

Figure 5.7 Typical testing strings of inter-letter spacing settings in typeface OCRB

HHHH HOHOH

Figure 5.8 Typical testing strings of inter-letter spacing settings in typeface Garamond

2) Lower case letters inter-letter spacing settings

In Table 5.3, the inter-letter spacing setting information of 26 lower case letters, which

includes character width, left side bearing and right side bearing, is shown for six typefaces. From these font metrics, we calculated the average left and right side bearings of the lower case letters for each typeface (as shown in Table 5.2 below). These average side bearings provide the basic inter-letter designs of each typeface. From these values, it is evident that OCRB has the largest average side bearings, which makes the legibility unchanged when condensed to a 0.8 point character spacing setting. The average side bearings of typeface Frutiger Linotype are smaller than OCRB, but still comparatively larger than the other typefaces. Typeface Garamond has the smallest average side bearings, so the expanded spacing settings can increase its legibility and the condensed spacing setting can dramatically drcrease its legibility. Typefaces Helvetica, Century and Microsoft Sans Serif have comparatively moderate average spacing settings, which make nearly no change in legibility with the increase of character spacing. For each typeface, the weight of the stroke, the width of its interior space and the shape and length of its serifs all need to be considered for a harmonious inter-letter spacing design. Therefore, there were no fixed optimal side bearing values for all of the typeface designs. However, the analysis of the average left and right side bearings and the legibility for the six typefaces were used to evaluate their original inter-letter spacing designs.

Table 5.2 Average side bearings of lower case letters for six typefaces

Typeface	Average LSB	Average RSB bearing
OCRB	299	296
Frutiger Linotype	99	104
Garamond	42	21
Helvetica	81	78
Century	61	49
Microsoft Sans Serif	79	77

Table 5.3 Character width and side bearings of lower case letters in six typeface designs

(LSB means left side bearing and RSB means right side bearing)

	Character width	LSB	RSB	Character width	LSB	RSB	Character width	LSB	RSB
	Helvetica			Century			MicrosoftSansSerif		
a	1139	74	53	1139	101	46	1139	78	140
b	1139	119	80	1139	16	101	1139	139	72
c	1024	61	47	909	61	61	1024	72	27
d	1139	72	117	1176	109	33	1139	72	140
e	1139	82	82	1024	87	70	1139	72	72
f	569	29	33	682	47	-137	569	16	-41
g	1139	82	117	1100	99	15	1139	72	140
h	1139	133	133	1251	43	46	1139	139	135
i	455	137	134	645	49	44	467	139	139
j	455	-32	137	606	-64	174	467	-72	139
k	1024	137	-1	1241	51	39	1024	139	8
l	455	137	138	645	50	48	467	139	139
m	1706	133	131	1821	101	70	1706	139	133
n	1139	133	133	1251	67	42	1139	139	135
o	1139	72	72	1024	69	69	1139	72	72
p	1139	119	80	1176	61	96	1139	139	72
q	1139	72	127	1139	107	30	1139	72	140
r	682	158	2	909	52	35	682	139	-8
s	1024	66	74	948	113	86	1024	63	80
t	569	29	43	797	51	46	569	35	18
u	1139	139	137	1251	56	53	1139	135	140
v	1024	16	16	1100	26	22	1024	27	27
w	1479	29	27	1593	53	50	1479	10	11
x	1024	23	20	1100	48	31	1024	14	14
y	1024	23	23	1100	43	22	1024	27	27
z	1024	63	63	985	98	77	1024	41	43

	Character width	LSB	RSB	Character width	LSB	RSB	Character width	LSB	RSB
	Garamond			OCRB			FrutigerLinotype		
a	832	66	13	1474	256	342	1139	94	154
b	1045	34	80	1474	274	233	1251	164	94
c	853	78	53	1474	336	301	909	78	57
d	1024	67	25	1474	336	278	1251	94	164
e	853	78	49	1474	276	278	1139	94	95
f	661	96	-164	1474	330	352	797	43	29
g	917	13	-27	1474	258	301	1251	94	164
h	1045	30	27	1474	319	321	1251	164	164
i	469	0	15	1474	383	550	1251	176	176
j	469	42	154	1474	387	497	569	-33	176
k	960	52	-17	1474	367	182	1139	164	39
l	469	10	4	1474	498	395	569	188	188
m	1579	36	35	1474	213	215	1821	162	160
n	1045	35	21	1474	279	278	1251	164	164
o	1045	73	74	1474	240	239	1251	94	94
p	1045	24	74	1474	274	235	1251	164	94
q	1003	70	-18	1474	236	276	1251	94	164
r	683	37	3	1474	393	288	797	164	49
s	747	113	88	1474	301	301	797	70	70
t	597	56	-8	1474	328	358	797	14	62
u	1003	33	12	1474	279	278	1251	164	164
v	960	-12	-18	1474	213	215	1024	23	23
w	1365	-22	-18	1474	236	235	1706	25	25
x	939	28	28	1474	233	235	1024	20	20
y	853	7	-28	1474	201	184	1024	14	33
z	875	54	78	1474	319	321	1024	88	88

3) Touching error analysis

Touching error analysis is another way to analyze the effect of different character spacing settings on typeface legibility. Because of the space limitation, we only show the different touching errors on Bi-gram test strings recognized by FineReader OCR system. In this part, two typefaces, Frutiger Linotype and Garamond, with 10 point print size, were chosen as examples. We selected these two typefaces because Frutiger Linotype was one of the typical typefaces that the condensed inter-letter spacing settings do not affect legibility and Garamond was one of the typical typefaces that the expanded inter-letter spacing settings increase its legibility.

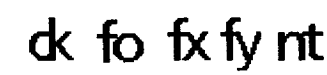
Frutiger Linotype

Table 5.4 shows the touching errors with different spacing settings in typeface Linotype. We can see that compared with the normal spacing setting, condensed 0.2, 0.4 and 0.6 point spacing settings do not produce any new touching errors. However, when the condensed 0.8 point spacing is applied, more touching errors occur, which decreases the overall legibility of this typeface. These touching errors are: "ck→d", "fo→ib", "fxfy→kiy" and "nt→rrt". On the other hand, the touching error "vv->w" can be avoided when the expanded 0.4 point, 0.6 point, and 0.8 point spacings applied.

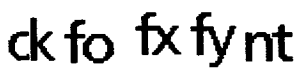
Table 5.4 Different touching errors with different spacing settings of typeface Frutiger Linotype

	Condensed 0.8 Point	Condensed 0.6 Point	Condensed 0.4 Point	Condensed 0.2 Point	Normal	Expanded 0.2 Point	Expanded 0.4 Point	Expanded 06 Point	Expanded 0.8Point
ck->d<	✓								
fo->ib	✓								
fxfy->kiy	✓								
nt->rrt	✓								
vv->w	✓	✓	✓	✓	✓	✓			

Figures 5.9 (a) and (b) show the images of touching strings "ck", "fo", "fx", "fy" and "nt" in typeface Frutiger Linotype with condensed 0.8 point and condensed 0.6 point character spacing setting. Figures 5.9 (c) and (d) show the images of string "vv" in typeface Frutiger Linotype with normal and expanded 0.4 point spacing settings.



(a) Condensed 0.8 point character spacing for "ck", "fo", "fx", "fy" and "nt"



(b) Condensed 0.6 point character spacing setting for "ck", "fo", "fx", "fy" and "nt"

(c) Normal character spacing for "vv"

(d) Expanded 0.4 point character spacing for "vv"

Figure 5.9 Touching string images of typeface Frutiger Linotype

From the error analysis of typeface Frutiger Linotype, we found that the condensed 0.2, 0.4, and 0.6 point spacing settings do not cause extra touching errors. Therefore, the inter-letter spacing of typeface Frutiger Linotype can be set tight for economic consideration of space.

Garamond

Table 5.5 shows the different touching errors with different spacing settings in typeface Garamond. Since with the condensed 0.6 and 0.8 point spacing settings, the legibility of typeface Garamond is very low and there are many touching errors, we do not list those errors. We can see that with the normal character spacing setting, there are three touching errors: "vv->w", "rn->m" and "ri->n". Touching error "rn->m" can be avoided with the expanded 0.4 point spacing setting; "vv->w" can be avoided with the expanded 0.6 point spacing setting; and "ri->n" can be avoided with the expanded 0.8 point spacing setting.

Moreover, in typeface Garamond, the condensed 0.2 point spacing setting causes four extra touching errors: "lx->k", "ll->u","li->H", and "im->nn" and the condensed 0.4 spacing setting produces six more touching errors: "ru->m", "lr->k", "la->k", "iu->m", "ci->a", and "cj->q". The images of these touching error strings are shown in Figure 5.10.

Table 5.5 Different touching errors with different spacing settings of typeface Garamond

	Condensed 0.4 Point	Condensed 0.2 Point	Normal	Expanded 0.2 Point	Expanded 0.4 Point	Expanded 0.6 Point	Expanded 0.8Point
ru->m	✓						
lr->k	✓						
la->k	✓						
iu->m	✓						
ci->a	✓						
cj->q	✓						
lx->k	✓	✓					
ll->u	✓	✓					
li->H	✓	✓					
im->nn	✓	✓					
rn->m	✓	✓	✓	✓			
vv->w	✓	✓	✓	✓	✓		
ri->n	✓	✓	✓	✓	✓	✓	

(a) String "rn" with normal spacing

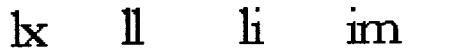
(b) with expanded 0.4 point spacing

(c) String "vv" with normal spacing

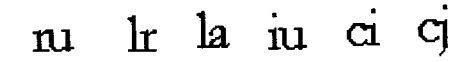
(d) with expanded 0.6 point spacing

(e) String "ri" with normal spacing

(f) with expanded 0.8 point spacing



(g) Strings "lx->k", "ll->u", "li->H", and "im->nn" with condensed 0.2 point spacing



(h) Strings "ru->m", "lr->k", "la->k", "iu->m", "ci->a", and "cj->q" with condensed 0.4 point spacing

Figure 5.10 Touching string images of typeface Garamond

The error analysis of typeface Garamond shows that the expanded inter-letter spacing settings can solve many touching errors caused in normal spacing setting, and the condensed spacing settings cause more extra touching errors. Therefore, in typeface Garamond, expanding inter-letter spacings can be a way to increase the character legibility.

5.2.4 Discussion and Suggestions

In this Chapter, by using our constructed legibility measurement experiments, the legibility of a given typeface with different character space settings was examined. The basic letter spacing settings, the side bearings of lower case letters and the touching errors were analyzed in order to detect the inter-letter spacings that cause design defects of different typefaces. In the six studied typefaces, we found that:

1) The inter-letter spacing settings in typeface OCRB and Frutiger Linotype are too wide. With the condensed 0.8 point spacing setting, the legibility of OCRB does not drcrease and no extra touching errors are produced. Similarly, the legibility of typeface Frugiter Linotype does not decrease when the condensed 0.6 point spacing setting is applied. Moreover, the inter-letter spacing settings of basic letters and the average side bearings of 26 lower case letters of typefaces OCRB and Frugiter Linotype are comparatively larger

than other typefaces. Therefore, for economic purposes, the inter-letter spacing of these two typefaces can be set more tightly, yet retaining their legibility.

2) The inter-letter spacing settings in typeface Garamond are relatively tighter. The expanded inter-letter space setting can avoid some touching errors and can increase the overall typeface legibility. On the contrary, the condensed spacing settings will cause many extra touching errors and the overall typeface legibility will decrease dramatically. Therefore, increasing the inter-letter spacing settings in typeface Garamond is a way to maximize its legibility.

3) For typefaces Helvetica, Century and Microsoft Sans Serif, the expanded spacing settings would not increase their legibilities and the condensed spacing settings would decrease their legibility. Therefore, the adjustments of inter-letter spacing in these three typefaces should not be a way to maximize their legibility.

Chapter 6

Conclusion

In this chapter, we will summarize the contributions of this thesis and present some future work in this research direction. By using a newly designed legibility measurement method, the legibilities of eighteen typefaces were examined. The main design characteristics of the eighteen typefaces were extracted mathematically. We examined the effect of different characteristics on legibility. Moreover, the relationship between the typeface legibilities and the inter-letter spacing settings, were investigated and analyzed in detail.

6.1 Summary

The basic concept of this thesis was to investigate the design of font characteristics on font legibility. Since most of the studies on this topic have been limited to the fields of psychology, physiology and typography design, the goal of this study was to address this study from a new point of view.

Firstly, an investigation of legibility among eighteen commonly used typefaces was addressed. Different from other legibility measurement methods, in our study, OCR systems, one of the most successful applications of machine reading, were used in the legibility measurements. The character recognition rate was used as a legibility comparison criterion. After measuring and comparing, the legibility trends of these eighteen typefaces were detected.

Secondly, the main typeface characteristics, serif and sans serif design, x-height, ascender and descender of the typeface, the inter-letter character spacing setting, the weight of the stroke and the distinctive character features, were studied. In order to measure these font characteristics mathematically and objectively, each characteristic was represented by one or two font legibility factors. The extraction methods were proposed for each font legibility factor. Using the proposed extraction methods, the main characteristics of each studied typeface were evaluated mathematically.

Based on the legibility comparison of eighteen typical typefaces, the evaluation of their contrast or unity characteristics and the common recognition errors caused by each typeface, and the effect of font design characteristics on legibility were discussed. For each of the eighteen typefaces, their design defects were analyzed. Moreover, design suggestions were made on typeface characteristics, which maximize typeface legibility.

In our study, one of the most important characteristics, inter-letter spacing, was examined. We first included the main design principles of inter-letter spacing from the type designers' point of view. Then, by using constructed experiments, with the aid of machine reading, we elevated the inter-letter spacing settings of different typeface designs. For six typefaces, we created different fonts, which differ only in terms of their inter-letter spacing, and we analyzed how the condensed or expanded inter-letter spacing settings contributed to its legibility. By investigating the inter-letter spacing design of each typeface and its common touching errors caused by different spacing settings, we found that some typefaces had relatively wide spacing settings, which could be condensed without affecting their legibility and that some typefaces had relatively tight spacing settings that could be expanded to increase their legibility.

6.2 Future Research

Based on the traditional research ideas in the fields of psychology, physiology and typography design, our study offers a new point of view of font legibility analysis. However, the work is far from finished, and more research related to the following aspects may be conducted in the future:

1. In our study, the effect of separate font design characteristics on legibility is discussed based on the legibility comparison of eighteen typical typefaces. It includes the evaluation of their contrast or unity characteristics and the common recognition errors caused by each typeface. In the future, the degree of association or correlation between these font characteristics and legibility could be studied in detail by using statistics, such as regression analysis.
2. In our thesis, we chose eighteen typical typefaces because of their common usage and their special properties. In the future, the font legibility analysis could be applied to particular applications. For example, we may analyze the legibility of the commonly used typefaces in newspapers, so that particular suggestions could be made for the specific design fields.

References

1. Walter Tracy, *Letters of credit,* David R. Godine, Publisher, 1986, pp. 70-80.

2. B. Binns, *Better type*, New York: Watson-Guptill Publications, 1989.

3. B. Zachrisson, *Legibility of printed text,* Stockholm: Almqvist and Wiksell,1965.

4. J. Ryder, *The case for legibility*, Moretus Press, Incorporated, 1979.

5. L. G. Roberts, *Machine perception of three-dimensional solids*, Optical and Electro-Optical Information Processing, J. T. Tippet, Ed. MIT Press, Cambridge, MA, 1965.

6. L. Watts and J. Nisbet, *Legibility in children's books: a review of research,* NFER Publishing Company, Ltd., 1974, pp. 10.

7. J. J. Foster, *Legibility research 1972–1978: a summary,* London: Graphic Information Research Unit, Royal College of Art, 1980, pp. 14.

8. R. Reason, "Design with reason: newspaper body text," *Presstime Magazine*, 1999.

9. R. McLean, "The Thames and Hudson manual of typography," Vol. 1, London, UK: Thames and Hudson Ltd, 1980.

10. K. Tinkel, "What make type easy to read and why," *Adobe Magazine*, March/April 1996.

11. M. Tinker, *The legibility of print,* Lowa State University Press, 1963.

12. V. Gaultney, "Balancing typeface legibility and economy: practical techniques for the type designer," 2001 [Online]. Available: http://www.sil.org/~gaultney/research.html.

13. K. G. Barnhurst, *Seeing the newspaper,* New York: St. Martin's Press, 1994.

14. E. C. Poulton, "Size, style, and vertical spacing in the legibility of small typefaces," *Journal of Applied Psychology,* Vol. 50, No. 2, 1972, pp. 156-161.

15. R. Rubinstein, *Digital typography: An introduction to type and composition for computer system design,* Boston, MA: Addison Wesley, 1988.

16. D. G. Paterson and M. A. Tinker, "Studies of typographical factors influencing speed of reading," *Journal of Applied Psychology*, Vol. 16, 1932, pp. 605-613.

17. E. D. Poulton, "Letter differentiation and rate of comprehension in reading," *Journal of Applied Psychology,* Vol. 49, No. 5, 1965, pp. 358-362.

18. J. A. Smither and C. C. Braun, "Readability of prescription drug labels by older and younger adults," *Journal of Clinical Psychology in Medical Settings*, Vol. 1, No. 2, 1994, pp. 149-159.

19. S. E. Moriarty and E. Scheiner, "A study of close-set text type," *Journal of Applied Psychology,* Vol. 69, No. 4, 1984, pp. 700-702.

20. E. Gray, "Least -squares font metric estimation from images," *IEEE Transaction on image processing,* Vol. 2, No. 4, 1993.

21. H. Spencer, *The visible word,* London: Lund Humphries, 1969.

22. K. Weidemann, "Biblica — designing a new typeface for the Bible," *Baseline*, Vol. 6 1985, pp. 7–11.

23. A. Zramdini and R. Ingold, "Optical font recognition using typographical features," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* Vol. 20, No. 8, Aug 1998, pp. 877-882.

24. R. F. Rehe, *Typography: how to make it most legible*, Carmel, Indiana: Design Research International, 1974, pp. 31; pp. 24.

25. M. Jamra, "Basic character spacing in type design," 2004 [Online]. Available:www.typeculture.com.

26. S. Moye, *Type by design,* MIS: Press Publisher, 1995, pp. 75-107.

27. C. Y. Suen and M. K. Komoda , "Legibility of digital type-fonts and comprehension in reading," *Text Processing and Document Manipulation Proceedings of the International Conference,* University of Nottingham, UK, 1986.

28. V. I. Levenshtein. "Binary codes capable of correcting deletions, insertions and reversals," *Sov. Phys. Dokl.*, Vol. 6, 1966, pp. 707-710.

29. A. Degani, "On the typography of flightdeck documentation," NASA Contractor Report 177605, Moffett Field, CA: NASA Ames Research Center, 1992.

30. A. Arditi, and J. Cho, "Serifs and font legibility," *Vision Research,* Vol. 45, No. 23, 2005, pp. 2926-2933.

31. A. Arditi and K. Knoblauch , I. Grunwald, "Reading with fixed and variable pitch," *J Opt Soc Am A,* Vol. 7, 1990, pp. 2011–2015.

32. L. Liu and A. Arditi, "How crowding affects letter confusion," *Optometry & Vision Science,* Vol. 78, No. (1), 2001, pp. 50-55.

33. B.C. Skottun and R.D. Freeman, "Perceived size of letters depends on inter-letter spacing: a new visual illusion," *Vision Research*, Vol. 23, 1983, pp. 111–112.

34. J. A. Smither and C. C Braun "Readability of prescription drug labels by older and younger adults," *Journal of Clinical Psychology in Medical Settings*, Vol. 1, No. 2, 1994, pp. 149-159.

35. H. Kucera and W. N. Francis, *Computation analysis of present-day American English*, Brown University Press, 1970.

36. C. Devoghelaere, "Analysis of font recognition," CENPARMI, Concordia, University, Montreal, QC, Tech. Rep., May 2003.

37. C. Y. Suen, "n-Gram statistics for natural language understanding and text processing," *IEEE Trans. on Pat. Anal. and Mach. Intel.,* Vol. PAMI-1, No. 2, Apr.1979, pp. 164-172.

38. M. Shi, Y. Fujisawa, T. Wakabayashi, and F. Kimura, "Handwritten numeral recognition using gradient and curvature of gray scale image," *Pattern Recognition,* vol. 35, No. 10, 2002, pp. 2051 – 2059.

Appendix A 18 different Typefaces

Times New Roman Courier New Palatino Century School Book

Helvetica Arial Albertus Medium Garamond Batang

Century Letter Gothic Georgia MicrosoftsansSerif

Impact Haettenschweiler OCRB BerkerlyBook FrutigerLinotype

Times New Roman

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Courier New

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Palatino

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Century School Book

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Helvetica

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Arial

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Albertus Medium

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Garamond

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Batang

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Century

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Letter Gothic

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Georgia

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Microsoft Sans Serif

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Impact

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Nasttanschweiler

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

OCRB

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

BerkerlyBook

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

FrutigerLinotype

The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

Appendix B Tested database

From "Computational analysis of present-day American English"
Frequency words

110 high frequency words (from frequency level 1)

the of and to a in that is was he for it with as his on be at by I this had not are but from or have an they which one you were her all she there would their we him been has when who will more no if out so said what up its about into than them can only other new some could time these two may then do first any my now such like our over man me even most made after also did many before must through back years where much your way well down should because each just those people how too little state

126 low frequency words (from frequency level 8)

abolish aborigines absurdity accelerometers accidents alley ankle arcs
backing bacteria baked bites boom bosom bugs
charging claimant clamped concealed courtyard cyclist cushion
deed diminishing directional dispelled dock
elect evolved excluded exemption expose exterior
faulty feeble flatness flu foe frail
gait glue gazing greasy gracefully
heed horns hypothetical hostess humidity hurricane
illiterate immediacy inaugural iodide irrational
jammed junk
kern kills kilometer kindly kinetic knelt
lakes leagues liberated liberties logs lone
merry midway mock mouths
negotiating niece nominated nostalgia notify
obey odors offense outsiders oval
pad parameters peaks peer pig
qualification
radiant recipe relates repel rot rumor ruins rusty
sailors sane scandal scrap slum
tactical tangle testify thief tricks trophy
ultrasonic umbrella uncommon undergo upside usable
varieties veil vines
wander weaker whereof wreath wreck witches wins
yourselves
zeal

From the better Type Book
The next morning, in the printing office, Edwin came upon Big James giving a lesson in composing to the younger apprentice, who in theory had 'learned his cases'. Big James held the composing stick in his great left hand, like a match-box, and with his great right thumb and index picked letter after letter from the case, very slowly in order to display the movement, and dropped them into the stick. In his mild, resonant tones he explained that each letter must be picked up unfalteringly in a particular way, so that it would drop face upward into the stick without any intermediate manipulation. And he explained also that the left hand must be held so that the right hand would have to travel to and fro as little as possible.

He was revealing the basic mysteries of his craft, and was happy, making the while the broad series of stock pleasantries which have probably been current in composing rooms since printing was invented. Then he was silent, working more and more quickly, till his right hand could scarcely be followed in its twinklings, and the face of the apprentice duly spread in marvel. When the line was finished he drew out the rule, clapped it down on the top of the last row of letters, and gave the composing stick to the apprentice to essay.

The apprentice began to compose with his feet, his shoulders, his mouth, his eyebrows — with all his body except his hands, which nevertheless traveled spaciously far and wide. 'It's not in seven year, nor in seventy, as you'll learn, young sun of a gun!' said Big James. And, having unsettled the youth too his foundations with a bland thwack across the head, he resumed the composing stick and began again the exposition of the unique smooth movement which is the root of rapid typesetting. 'Here!' said Bid James, when the apprentice has behaved worse than ever. 'Us'll ask Mr. Edwin to have a go. 'Us'll see what he'll do.' And Edwin, sheepish, had to comply. He was in pride bound to surpass the apprentice, and did so.

From Cedric Devoghelaere's report

I am the voice of today, the herald of tomorrow. I am type! Of my earliest ancestry neither history nor relics remain. The wedge-shaped symbols impressed in plastic clay in the dim past by Babylonian builders foreshadowed me: from them, on through the hieroglyphs of the ancient Egyptians, down to the beautiful manuscript letters of the medieval scribes, I was in the making.

The path of the righteous man is beset on all sides by the inequities of the selfish and the tyranny of evil man. Blessed is he who, in the name of charity and good will, shepherds the weak through the valley of darkness, for he is truly his brother's keeper and the finder of lost children. And I will strike down upon thee with great vengeance and furious anger those who attempt to poison and destroy my brothers. And you will know my name is the Lord when I lay my vengeance upon you.

The x-height is the height of the lowercase letter exclusive of ascenders and descenders. Although this is not a unit of measurement, it is significant because it is the x-height of the letter which conveys the visual impact of the type size. Therefore, typefaces which are the same point size may appear smaller or larger because of variations in the x-height. Study these five samples closely: Garamond, with its small x-height, appears much smaller than Century Expanded or Helvetica with their larger x-heights.

From the elementary book

Pets are like friends. They cannot talk, but sometimes they seem to talk. A dog barks. A cat meows. A bird sings. Most children like to have a pet.

Dear Aunt Margaret and Uncle Tim,
Thank you for Amy's coat and my coat. Amy loves the coat. My mother put the coat on Amy. She did not want to take it off. She kept it on for a few minutes. Then my mother took the coat off.
I got a horse, paints, and a pencil case with an eraser, tape, pencils and clips.
Happy New Year.
Love, Lynn.

Dear Pupil,
I hope you enjoyed writing in the first two books. Now you should be ready to learn to write in

cursive writing as grown-up people write. That's what you will be taught to do in Book 3.

Manuscript writing is very easy to read. You will want to use it often even after you learn another style of writing. Keep up your practice in manuscript writing.

Someone once said that YOUR LETTER YOU ON PAPER. Make a good impression.

Your far-away teacher,
Margarte B. Parke

Bigram

aa ab ac ad ae af ag ah ai aj ak al am an ao ap aq ar as at au av aw ax ay az
ba bb bc bd be bf bg bh bi bj bk bl bm bn bo bp bq br bs bt bu bv bw bx by bz
ca cb cc cd ce cf cg ch ci cj ck cl cm cn co cp cq cr cs ct cu cv cw cx cy cz
da db dc dd de df dg dh di dj dk dl dm dn do dp dq dr ds dt du dv dw dx dy dz
ea eb ec ed ee ef eg eh ei ej ek el em en eo ep eq er es et eu ev ew ex ey ez
fa fb fc fd fe ff fg fh fi fj fk fl fm fn fo fp fq fr fs ft fu fv fw fx fy fz
ga gb gc gd ge gf gg gh gi gj gk gl gm gn go gp gq gr gs gt gu gv gw gx gy gz
ha hb hc hd he hf hg hh hi hj hk hl hm hn ho hp hq hr hs ht hu hv hw hx hy hz
ia ib ic id ie if ig ih ii ij ik il im in io ip iq ir is it iu iv iw ix iy iz
ja jb jc jd je jf jg jh ji jj jk jl jm jn jo jp jq jr js jt ju jv jw jx jy jz
ka kb kc kd ke kf kg kh ki kj kk kl km kn ko kp kq kr ks kt ku kv kw kx ky kz
la lb lc ld le lf lg lh li lj lk ll lm ln lo lp lq lr ls lt lu lv lw lx ly lz
ma mb mc md me mf mg mh mi mj mk ml mm mn mo mp mq mr ms mt mu mv mw mx my mz
na nb nc nd ne nf ng nh ni nj nk nl nm nn no np nq nr ns nt nu nv nw nx ny nz
oa ob oc od oe of og oh oi oj ok ol om on oo op oq or os ot ou ov ow ox oy oz
pa pb pc pd pe pf pg ph pi pj pk pl pm pn po pp pq pr ps pt pu pv pw px py pz
qa qb qc qd qe qf qg qh qi qj qk ql qm qn qo qp qq qr qs qt qu qv qw qx qy qz
ra rb rc rd re rf rg rh ri rj rk rl rm rn ro rp rq rr rs rt ru rv rw rx ry rz
sa sb sc sd se sf sg sh si sj sk sl sm sn so sp sq sr ss st su sv sw sx sy sz
ta tb tc td te tf tg th ti tj tk tl tm tn to tp tq tr ts tt tu tv tw tx ty tz
ua ub uc ud ue uf ug uh ui uj uk ul um un uo up uq ur us ut uu uv uw ux uy uz
va vb vc vd ve vf vg vh vi vj vk vl vm vn vo vp vq vr vs vt vu vv vw vx vy vz
wa wb wc wd we wf wg wh wi wj wk wl wm wn wo wp wq wr ws wt wu wv ww wx wy wz
xa xb xc xd xe xf xg xh xi xj xk xl xm xn xo xp xq xr xs xt xu xv xw xx xy xz
ya yb yc yd ye yf yg yh yi yj yk yl ym yn yo yp yq yr ys yt yu yv yw yx yy yz
za zb zc zd ze zf zg zh zi zj zk zl zm zn zo zp zq zr zs zt zu zv zw zx zy zz

Non-Sense Words by using Chinese pinyin syllable:

a ai an ang ao
ba bai ban bang bao bei ben beng bi bian biao bie bin bing bo bu
ca cai can cang cao ce cen ceng cha chai chan chang chao che chen cheng chi chong chou chu
chuai chuan chuang chui chun chuo ci cong cou cu cuan cui cun cuo
da dai dan dang dao de dei den deng di dian diao die ding diu dong dou du duan dui dun duo
e ei en eng er
fa fan fang fei fen feng fiao fo fou fu
ga gai gan gang gao ge gei gen geng gong gou gu gua guai guan guang gui gun guo
ha hai han hang hao he hei hen heng hong hou hu hua huai huan huang hui hun huo
ji jia jian jiang jiao jie jin jing jiong jiu ju juan jue jun
ka kai kan kang kao ke ken keng kong kou ku kua kuai kuan kuang kui kun kuo
la lai lan lang lao le lei leng li lia lian liang liao lie lin ling liu lo long lou lu luan lun luo
ma mai man mang mao me mei men meng mi mian miao mie min ming miu mo mou mu

na nai nan nang nao ne nei nen neng ng ni nian niang niao nie nin ning niu nong nou nu nuan nuo
o ou
pa pai pan pang pao pei pen peng pi pian piao pie pin ping po pou pu
qi qia qian qiang qiao qie qin qing qiong qiu qu quan que qun
ran rang rao re ren reng ri rong rou ru rua ruan rui run ruo
sa sai san sang sao se sen seng sha shai shan shang shao she shei shen sheng shi shou shu shua shuai
shuan shuang shui shun shuo si song sou su suan sui sun suo
ta tai tan tang tao te teng ti tian tiao tie ting tong tou tu tuan tui tun tuo
wa wai wan wang wei wen weng wo wu
xi xia xian xiang xiao xie xin xing xiong xiu xu xuan xue xun
ya yan yang yao ye yi yin ying yo yong you yu yuan yue yun
za zai zan zang zao ze zei zen zeng zha zhai zhan zhang zhao zhe zhen zheng zhi zhong zhou zhu zhua
zhuai zhuan zhuang zhui zhun zhuo zi zong zou zu zuan zui zun zuo

High Frequency N-Gram Words (top 50 2-Gram, 3-Gram, 4-Gram and 5-Gram words)

e th he s d t n in an er re r y on nd at en f o or of to it ed is es ha te st ti
ar ou as al hi nt ng l se ve a me le h g ea co ne de ro the and ing ion ent
her for tio ly hat tha his ere ll was ce ter ut ati all ith ts thi ch tion that ther
with atio ted ould nce here are ment uld this had not one ave out from but rom
ght have ore een ight hen ver ers they ich ons hey ill ough hich ation which
were other ions there would tions ction ting their heir ence been when ally more
hing will ning thing what ding said ical ever state ents ring about bout ound into
than them only

Appendix C Similar Designed Character Pairs

In the following tables, the character pairs, which have similar designed shapes and distance d value is less than 20, are listed for each font.

Albertus Medium

Character	Most similar character	d	Character	Most similar character	d
l	I	2.069841	l	i	16.43236
j	i	7.023564	j	J	18.78137

Arial

Character	Most similar character	d	Character	Most similar character	d
f	t	13.990101	i	l	14.434844
g	q	18.276383	l	I	0.000000
i	I	14.434844	v	V	13.044067
i	j	17.542134			

Batang

Character	Most similar character	d	Character	Most similar character	d
c	e	19.360485	i	l	15.605187
f	i	18.514600	l	I	14.076594
f	t	9.381393	o	O	17.758753
i	j	15.231309	v	V	19.964713
w	W	14.701788			

BerkerlyBook

Character	Most similar character	d	Character	Most similar character	d
f	i	18.241252	j	J	14.850810
f	t	16.231891	l	I	16.204735
i	I	11.168136	p	P	14.951054
i	l	17.078059	v	V	19.494447
i	t	17.145962	w	W	18.512914

Century

Character	Most similar character	d	Character	Most similar character	d
f	i	18.960105	o	O	16.792934
f	t	17.338489	p	P	17.264227
i	I	17.419753	v	V	17.021447
i	l	14.400158	w	W	19.847218
l	I	7.105575			

Century School Book

Character	Most similar character	d	Character	Most similar character	d
f	i	18.960105	o	O	16.792934
f	t	17.338489	p	P	17.264227
i	I	17.419753	v	V	16.204507
i	l	14.400158	w	W	11.892529
l	I	7.105575			

Courier New

Character	Most similar character	d	Character	Most similar character	d
c	C	19.383939	l	I	16.504401
c	o	17.932176	v	V	10.428263
g	q	9.432009	y	Y	18.626295
i	l	13.334106			

FrutigerLinotype

Character	Most similar character	d	Character	Most similar character	d
i	I	14.359434	l	I	0.000000
i	j	14.998881	s	S	16.722094
i	l	14.359434	v	V	14.088150
j	J	16.530881	w	W	17.184616

Garamond

Character	Most similar character	d	Character	Most similar character	d
i	j	18.530747	p	P	17.824247
i	l	13.851279	v	V	16.141903
j	J	17.470726	w	W	15.212257

Georgia

Character	Most similar character	d	Character	Most similar character	d
c	e	15.382852	p	P	16.213090
l	I	11.843985	w	W	19.219336
o	O	18.100357			

Haettenschweiler

Character	Most similar character	d	Character	Most similar character	d
b	h	12.127675	j	I	12.432717
c	C	14.489476	j	J	10.198792
f	I	19.872422	j	l	12.432717
f	i	13.675935	k	K	13.522261
f	j	18.842764	l	I	0.000000
f	l	19.872422	l	J	10.257624
f	r	19.581408	l	j	12.432717
g	q	18.234321	o	O	16.060223
i	I	7.109762	s	S	18.831831
i	J	15.291264	v	V	13.371235
i	j	7.490110	w	W	18.777046
i	l	7.109762	z	Z	16.883228

Helvetica

Character	Most similar character	d	Character	Most similar character	d
f	t	13.990101	i	l	14.434844
g	q	18.276383	l	I	0.000000
i	I	14.434844	v	V	13.044067
i	j	17.542134			

Impact

Character	Most similar character	d	Character	Most similar character	d
b	h	12.127675	i	l	7.109762
c	C	14.489476	j	I	12.432717
f	I	19.872422	j	J	10.198792
f	i	13.675935	j	l	12.432717
f	j	18.842764	k	K	13.522261
f	l	19.872422	l	I	0.000000
f	r	19.581408	l	J	10.257624
g	q	18.234321	s	S	18.831831
i	I	7.109762	v	V	13.371235
i	J	15.291264	w	W	18.777046
i	j	7.490110	z	Z	16.883228

Letter Gothic

Character	Most similar character	d	Character	Most similar character	d
f	t	17.015434	k	K	15.816191
i	l	14.498312	v	V	19.007682

MicrosoftsansSerif

Character	Most similar character	d	Character	Most similar character	d
f	i	17.812019	j	I	17.319327
f	t	13.255176	j	l	17.319327
g	q	17.125903	l	I	0.000000
i	I	12.734793	w	W	15.850396
i	j	12.760502	x	X	18.997897
i	l	12.734793			

OCRB

Character	Most similar character	d	Character	Most similar character	d
g	q	14.042019	i	j	8.157390

Palatino

Character	Most similar character	d	Character	Most similar character	d
f	i	16.755812	l	I	15.540674
f	t	13.530569	v	V	14.762760
i	l	15.012773	w	W	18.264117

Times New Roman

Character	Most similar character	d	Character	Most similar character	d
c	e	17.244552	l	I	15.540674
f	t	19.420763	p	P	14.567583
i	j	18.884469	s	S	17.318747
i	l	14.702661			

Appendix D six typefaces with different inter-letter spacing setting

1) Helvetica

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded 0.2pt the of and to a in that

Expanded 0.4pt the of and to a in that

Expanded 0.6pt the of and to a in that

Expanded 0.8pt the of and to a in that

1) Microsoft Sans Serif

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded 0.2pt the of and to a in that

Expanded 0.4pt the of and to a in that

Expanded 0.6pt the of and to a in that

Expanded 0.8pt the of and to a in that

2) Century

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded 0.2pt the of and to a in that

Expanded 0.4pt the of and to a in that

Expanded 0.6pt the of and to a in that

Expanded 0.8pt the of and to a in that

3) Garamond

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded 0.2pt the of and to a in that

Expanded 0.4pt the of and to a in that

Expanded 0.6pt the of and to a in that

Expanded 0.8pt the of and to a in that

4) OCRB

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded0.2pt the of and to a in that

Expanded0.4pt the of and to a in that

Expanded0.6pt the of and to a in that

Expanded0.8pt the of and to a in that

5) Frutiger Linotype

Condensed 0.8pt the of and to a in that

Condensed 0.6pt the of and to a in that

Condensed 0.4pt the of and to a in that

Condensed 0.2pt the of and to a in that

Normal the of and to a in that

Expanded 0.2pt the of and to a in that

Expanded 0.4pt the of and to a in that

Expanded 0.6pt the of and to a in that

Expanded 0.8pt the of and to a in that