# DISCOVERING LEGIBLE AND READABLE CHINESE TYPEFACES FOR READING DIGITAL DOCUMENTS

BING 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 2011

# CONCORDIA UNIVERSITY

## School of Graduate Studies

This is to certify that the thesis prepared

By: Bing ZHANG

Entitled: Discovering Legible and Readable Chinese Typefaces for Reading Digital Documents

and submitted in partial fulfillment of the requirements for the degree of

**Master of Computer Science**

complies with the regulations of the University and meets the accepted standards with respect to originality and quality.

Signed by the final examining committee:

______________________________Chair
Dr. R. Jayakumar

______________________________Examiner
Dr. T. Kasvand

______________________________Examiner
Dr. C. Lam

______________________________ Supervisor
Dr. Ching Y. Suen

Approved by ________________________________________
Chair of Department or Graduate Program Director

__________ 20_____ ________________________________________
Dr. Robin A. L. Drew, Dean
Faculty of Engineering and Computer Science

# ABSTRACT

DISCOVERING LEGIBLE AND READABLE CHINESE TYPEFACES FOR READING DIGITAL DOCUMENTS

BING ZHANG
Concordia University

In recent years, more and more fonts have been implemented in the digital publishing industry and in reading devices. In this thesis, we focus on the methods of evaluating digital Chinese fonts and their typeface characteristics. Our goal is to seek a good way to enhance the legibility and readability of Chinese characters displayed on digital devices such as cell phones, tablets and e-book devices. To accomplish this goal, we have combined methods in data mining, and pattern recognition with psychological and statistical analyses. Our research involved an extensive survey of the distinctive features of eighteen popular Chinese digital typefaces. Survey results were tabulated and analyzed statistically. Then, two objective experiments were conducted, using the best six fonts derived from the survey results. These experimental results have revealed an effective way of choosing legible and readable Chinese digital fonts that are most suitable for the comfortable reading of books, magazines, newspapers, and for the display of texts on cell-phones, e-books, and digital libraries. Results also helped us find out the features for improving character legibility and readability of different Chinese typefaces. The relationships among legibility, readability, eye-strain, and myopia, will be discussed. Moreover, digital market requirements and analyses will be provided.

# ACKNOWLEDGEMENTS

It is a pleasure to thank many people who helped me accomplish this thesis a lot. Their helps to me not only is to finish my current study, but helping my life's study as well. What I learnt from my supervisor are not only the expertise and research ideas, but also the research among human, including communication, negotiation, cooperation and so on.

Thus, firstly I would like to appreciate the helps sincerely from my dear supervisor, Dr. Ching Y. Suen, for his suggestive instruction, persistent guidance, warm encouragement and continuous support. Without his help and support, my study would not have been accomplished. I feel extremely lucky to have been supervised by Dr. Suen, and it is my honor to work with him.

Moreover, I have to say thank you to all my colleagues and staff in the Center for Pattern Recognition and Machine Intelligence (CENPARMI), for providing a positive, optimistic and studious environment in which to learn and grow. Especially, I am really grateful to Guiling Guo, Ying Li, Xiaoxiao Niu, Xiaohua Xie, and Chao Yao. They are all my "family members" in CENPARMI, and they taught me a lot in both study and living. A special thanks to the research manager of CENPARMI, Mr. Nicola Nobile, for his technical support and assistance, and to Ms. Shira Katz, for her editing and proofreading of this thesis. Also I want to say thank you to our office assistants, Ms. Marleah Blom, for her administrative assistance.

Then, I wish to say thank you to my parents, for their constant moral support. I am very thankful to them for all they have given me in my life, for sharing ups and downs, and for reminding me, when necessary.

Last but not least, I am very thankful to everyone who has directly and indirectly helped me, in any way, to accomplish this research.

## Table of Contents

## List of Figures

## List of Tables

# Chapter 1. Introduction

In this chapter, the background, motivation, objectives and structure of this thesis are introduced. Section 1.1 includes the background information on the current digital market trend. In Sections 1.2 and 1.3, the motivation and objectives are presented. Then, in Section 1.4, we will discuss the current Chinese typefaces, and the outline is described in Section 1.5.

## 1.1 Background

Hundreds of Chinese fonts exist around us nowadays, such as SongTi, HeiTi, KaiTi, calligraphy etc. Different kinds of fonts have different designs, e.g. some are designed to display formal documents, and some are created by the authors' subjective mood or from aesthetical points of view. As the readers of Chinese texts, thus, people sometimes may have trouble of understanding or may misunderstand an article's real meaning if the document font is unsuitable or incorrect, which may invoke feelings such as impatience and nervousness. When lots of electronic products become available to people, the problem of how to provide a good Chinese typeface to be displayed on screens seems to be more critical.

With more and more people accepting and purchasing electronic book devices and tablets, international companies, such as Apple, Amazon, Google, Sony, etc., have all invested huge amounts of money on developing more of these new products not only making them easier to carry, but also having extra functionalities, such as a friendly interface display, clear typeface display, etc. in order to capture a bigger percentage of the electronic market. Because of the fierce competition in marketing, there has been a battle of price and quality over the various products among different brands.

Currently, pure e-book products and tablets have totally beaten the original and ordinary laptop and the once prevailing web laptop, and being pursued by the public, especially by the office ladies and stock investors according to the popular and fashionable new concepts in digital usage, new ways of reading (e.g. tablets could be used as a book), better looking designs, easy reading of digital documents and easy to carry designs, friendly usage, and acceptable prices.

Therefore, the original book market has encountered an unprecedented serious challenge, because many previous users prefer to download digital documents and books online, and use their laptops or tablets to read them. Compared to digital documents, the outstanding paper printing quality of original books has become a burden to carry and extra expenses. To accommodate the change of the market, more and more publishers have produced digital versions of their publications.

Moreover, the appearance of a variety of tablets from different companies brings this digital market more competitions and more opportunities. To compete with Apple Company's products (IPAD I and IPAD II), Apple's rivals have all advertised their new products and have put them on the market. For example, colorful devices include: Google Android, RIM BlackBerry PlayBook, Palm WebOS / HP, Samsung Galaxy Tab, etc., and there is also electronic "paper" used as normal ebooks made from Amazon, Sony, etc.

DisplaySearch is the Worldwide Leader in Display Market Research and Consulting. In 2005, DisplaySearch became an independently operated subsidiary of the NPD Group, the leading global provider of consumer and retail market research. Backed by its parent company, DisplaySearch provides a true end-to-end view of the display supply chain, delivering the most accurate reporting and forecasting that the industry has to offer [1]. DisplaySearch has published many reports about the tablet market competitions. According to the DisplaySearch Q1'11 Touch Panel Market Analysis update, touch screen shipments for tablet (or slate) PCs are forecast to reach 60M units in 2011. Apple will likely continue to account for the majority of tablet PC touch screens in 2011 and 2012, yet other brands could catch up in 2012 and beyond. DisplaySearch forecasts total touch panels for tablet PCs to reach 260M units in 2016, up 333% from 2011[2]. Based on this report, it is obvious that the fierce competition in the digital market has been started.

On the other hand, myopia has become a serious problem all over the world. The rise in myopia in Asian populations is causing much concern in China, where 50% of teenagers today are nearsighted compared to only 15% in the 1970's [3]. According to a preliminary survey from children's myopia prevention and cooperation research project, the incidence of myopia in China was 33% of the population, which included almost 400 million domestic people with myopia, 1.5 times higher compared to the world average of 22%. The highest incidence of myopia group has been traced to the juveniles of Chinese, whose myopia incidence rate is now 50% to 60%. This number has made Chinese youth place first in the world of myopia [4].

From a medical point of view, myopia mainly has two causes: environmental and genetic factors. Other data [42, 43, 44] show that the genetic factor though affects the disease in young people who are suffering from myopia, it is not the major factor. However the environment reason plays an important role. For a long time people indulge in watching TVs, staring on the computer screens to play games or work, people would get more eye fatigue and aggravate the myopia disease. Another causative environmental factor is from schooling. Since the academic burden of primary and secondary schools is too heavy, students have to concentrate on their homework and assignments for a long time and have no breaks to relax their eyes to relieve fatigue. So this is another leading environmental cause of myopia [4].

Along with the acceptance of tablets and e-book devices and partial substitution of original books, people spend more time on electronic screens for reading, including news, commercials, magazines, novels and so on. Actually, reading on screens causes more eye fatigue than reading on paper, as declared by Microsoft psychologist Kevin Larson teams with typographers and computer engineers [5]. Larson's team conducted a bunch of studies to find out what kind of muscles were causing fatigue, and they found that it was related to the orbicularis oculi, a large muscle around the eye that is responsible for blinking and squinting. Whenever people are reading a text that is too light in contrast to the background or too small, e.g. anything below 12 points, the orbicularis oculi becomes more active and the blink rate decreases. If we can give a higher contrast text and larger text sizes, then that is going to eliminate at least one of the causes of eye fatigue, or at least reduce it.

Thus, e-book devices and tablets manufacturers, while developing new products with new technologies, they have also invested lots of money to add additional functionality for eyestrain relief. These decisions help manufacturers match the market requirement and get more market shares. As well parents today often focus more on that whether the product is good for their children's health and growth than on product price. Many adults prefer to live healthy with the health consumption concept when they see new electronic products. Also if their salaries are limited, then within the limited salary, people often prefer to make a comparison based on quality, software design, products looking and price, among these wanted products before purchasing one, and then choose a better one. Therefore, using eye strain relief as a basis, we need to immediately figure out how to decrease the relative reading time, how to provide a relative clear and easy information displayer for reading, how to improve the comfortable feelings when reading. All of these factors, including the economical benefits mentioned above, point to a potential solution: typeface improvement.

## 1.2 Motivation

First of all, the usage of typefaces should be defined and explained to the public, in order to relieve people's confusion, anxiety, and worries when they are reading, typing, or designing a document.

Then, to increase the competitive power of tablets and e-book readers market, improving the software GUI (graphic user interface) design and typeface display design are good ways except the hardware update and price decrement methods, and by these ways designers of these products could make them more attractive for users in order to get more of a share of the digital market. Moreover, compared with hardware improvement, software improvement is more controllable and easier to implement. Because as we know, the hardware is provided by some assigned brands manufacturers, and there is one manufacturer supports several tablet designs. And the price of hardware must be fit for the fluctuation of hardware market, which is out of the tablet producers' control.

Thus, compared to the hardware, the software GUI design could be the key competition item among different tablet providers. And as the most information describer and carrier of GUI and documents, typeface display plays an incredibly important role in this new market war. Moreover, if the typeface has a high legibility and readability, and is really easy to read for a relatively long time with less eyestrain compared with other existing fonts that have been published already, any brand device with this outstanding typeface would directly catch the public's eyes. This means that a good typeface display may contribute to more of the digital market share.

Pattern recognition, which is "the act of taking in raw data and taking an action based on the category of the data" [6], is an innate ability of animals. It has been studied in many fields, including Psychology [7] and Ethnology [8].

While Artificial Intelligence (AI) has achieved its greatest successes in the 1990's and early 21st century, pattern recognition /machine learning in Computer Science [9, 10, 11, 12] arose as a field of interest to researchers. These researchers endeavored to design and develop the algorithms that allow computers to simulate human beings by recognizing (classifying) patterns based either on a priori knowledge or on statistical information extracted from the patterns. [13]

And due to the successes of these researchers, many applications of pattern recognition systems and techniques are available, and they cover a broad scope of fields, such as Engineering, Agriculture, Biology, Economics, Medicine, and so forth. These techniques have even been and applied to studies in Psychology/Cognitive Science and Ethnology [13].

Based on all the successful research, it is possible to combine pattern recognition and data mining with psychology, according to the bridges which have been built. Followed by their creative ideas and opinions on combining fields, a huge amount of researchers have concentrated on this junction area, and have published their research results, which have gotten the public attention for awhile. The relative typeface analysis is a very common topic in these areas of research [14, 15].

Currently, the research on typefaces has two different directions, one is to analyze the persona of fonts, and another one is to discuss the font display performance and influence to human eyes, and most of the research is based on English typeface analyses. On the persona of fonts part, for example, Eva R. Brumberger [16] made a discussion on the rhetoric of typography and analyzed the persona and characteristics of the current familiar English fonts with statistical methods, Dawn Shaikh [17] made an analysis of 40 onscreen typefaces, covering serif, san-serif, display, and handwriting classes to know their semantic differential presentations, and Ying Li [18] also linked the visual images of typefaces with their design characteristics. Her paper discussed the relationship between typefaces and their personality traits by using statistical analyses on the data.

On the topic of the character display performance, researchers seem to have had more interest on the typeface's serif features rather than on the font categories and on local features at the very beginning of touching display performance area. Normally there are three kinds of opinions when comparing serif and sans serif fonts: better, equivalent and worse. Actually, serif fonts were used and accepted by human beings much earlier than sans serif fonts. Thereby, people are more familiar with serif fonts than with sans serif fonts. However, with the prevalence of sans serif fonts, researchers have noted that the category of sans serif fonts is much more legible when displayed on a screen because of the screen display features [21].

However, according to Tinker [21], Zachrisson [22] and Bernard et. al, after years of analyses and comparison between serif and sans serif fonts, there is almost no difference between them in legibility when shown on a screen or a web site. It is of course possible that serifs or the lack of them have an effect on legibility, but it is very likely that they are so peripheral to the reading process that this effect is not even worth measuring [23]. And indeed, a greater difference in legibility can easily be found within members of the same type family than between a serif and a sans serif typeface. There are also other factors such as x-height, counter size, letter spacing and stroke width that are more significant for legibility than the presence or absence of serifs [24, 25].

Moreover, the researchers mentioned above provided the evidence from many different angles to support their conclusion, including the statements which opposed some advantages of serif features which had been accepted by the public. For example, they against that Serifs are used to guide the horizontal "flow" of the eyes; The lack of serifs contributes to a vertical stress in sans serifs, which is supposed to compete with the horizontal flow of reading [26], and Serifs are used for body text because sans serif causes fatigue. On the other hand, they also said that there has not been borne out by recent evidence that has shown no difference in legibility between serif and sans serif font on the web, because of the constraint of screen display.

SURL (Software Usability Research Laboratory) has made an experiment (By Michael Bernard & Melissa Mills) to test which type of font and size should be used on websites [46]. They compared two different fonts (Arial and Times New Roman) in two different sizes (12 and 14 points), with and without anti-aliased technology [45] and draw a conclusion that letters displayed in bigger size (14 pts) would be better than small size (12 pts). However they could not cover all the conditions of comparison and the only two fonts' comparisons were insufficient. However, the font size influenced the typefaces shown on screen was verified, which also meant that people preferred 12 points size other than 10 points size. In another experiment of "Determining the Best Online Font for Older Adults" [47], researchers also concluded that font size affected the legibility of fonts.

According to these researchers' results, suspicions and arguments the obscure direction was found from the darkness, to some extent. We are confident that our idea of studying typefaces used on tablet and e-book devices is feasible and reasonable at least.

## 1.3 Objective

To help readers better understand legibility and readability experiments, we present the definitions of legibility and readability at the beginning of this section. Quoted from Walter Tracy, an English typographer and writer and designer of books, magazines, and newspapers [49]. "Legibility and readability are separate, though connected, aspects of type. Properly understood, and used in the meanings appropriate to the subject, the two terms can help to describe the character and function of type more precisely than legibility alone. Legibility, says the dictionary, mindful of the Latin root of the word, means the quality of being easy to read. Readability is a different thing. The dictionary may say that it, too, means easy to read. Legibility, then, refers to perception, and the measure of it is the speed at which a character can be recognized; if the reader hesitates at it the character may be poorly designed. Readability refers to comprehension, and the measure of that is the length of time that a reader can give to a stretch of text without strain" [19].

From this explanation, we would understand that the legibility is concerning the single character recognition, and could be measured by recognition speed of single character. The readability focuses on the comprehension of contexts.

Many famous researchers have published a huge number of papers related to the typeface all over the world. However most of the research on typeface is based on the English language and systematical research on Chinese is limited. Thus there exists a need for us to find out more about it.

Our objective is to make a particular analysis and comparison for current existing typefaces covering all the categories and famous Chinese typefaces from different angles. We attempt to discover and understand the legibility and readability influences made by different typefaces and search for the features that could improve legibility and readability. By solving these real-life problems pertaining the popular tablets and e-book devices, we may bring a good and healthy usage concept to the public.

Firstly, Chinese typeface has been targeted and focused on, because the population of Chinese readers is huge, and there are very few publications that discuss the Chinese typeface displays performance. Also compared with English, the Chinese typeface is much more complicated either in structure or shape. Moreover, unlike English fonts, Chinese typefaces have several main font categories, such as Song Ti, Hei Ti, Yuan Ti, etc., and Chinese people are familiar with these main categories of fonts. Chinese font manufacturers all have their own typeface designs for the main font categories besides their special fonts. This means, for example, that we have several Song Ti typefaces with tiny differences between every pair of fonts made by different companies. Thus, starting with the current typefaces and analyses, the relative tiny differences is a good way to find out the features which would produce a high legibility and readability.

Secondly, we have set our typeface display performance research on tablets and e-book devices specifically. This decision narrows down our research direction in order to make it more target-oriented rather than being too general to fit the market requirement of the electronic products. We attempt to solve the real display problem that is intractable recently in the fierce electronic market competition.

Thirdly, to satisfy all Chinese customers with health purposes, a combination of knowledge about computer science and psychology has to be built, to better understand Chinese readers' favor. With the help of psychology, we would analyze the confused data collected from participants by statistical methods for searching the unobvious evidence to reflect the human's choice.

Moreover, the comparison between human OCR and machine OCR must be carried out. Although this thesis focuses on the typeface usage of tablets and e-book devices to improve legibility and readability, we are still concerned with the future work on digital documents. If our research result on humans matches the machine OCR result, then the OCR software customers will not need to update their tools and directly apply the typeface we recommend.

However, it is nonsense to analyze all the Chinese typefaces in current typeface market. Thus a survey must be done to find people's preference of typefaces at the beginning of our research and to filter the unacceptable typefaces out. Then we can concentrate on the remaining typefaces to get more accurate results.

## 1.4 Current Chinese typefaces

There are several frequently used Chinese font categories named as mainstream fonts: SongTi, HeiTi, YuanTi, LiTi, KaiTi, Weibei Ti, etc., and the visual difference among the categories is obvious. Meanwhile, many reformative typefaces based on these main categories have been accepted by the public as well, such as Fangsong Ti, Xihei Ti and so on. Additionally, a huge number of calligraphies designed by font manufacturers and amateur designers appeared in front of customers eyes. Adding to frequently used mainstream fonts, the number of fonts now far exceeds 500 kinds, making this typeface market wide and complicated.

However, we cannot tabulate all the fonts that have appeared in the market in this thesis, because not all of them are popular or prevailing fonts. In this study, we have conducted an extensive survey of various typefaces that are commonly used in textbooks, newspapers, e-books and other digital devices. Out of more than 100 typefaces that have been produced by different font companies, we have chosen eighteen different Chinese fonts for in-depth studies, using five characteristics which are pertinent to reading. The sample of the 18 typefaces is shown below.

微软简楷体 方正卡通 华文中宋 微软雅黑 方正宋三

迷你简毡笔黑 方正准圆 方正隶书 方正仿宋

汉仪综艺体简 微软简中圆 迷你简丫丫

方正魏碑 方正黑体 迷你简雪君 微软简标宋

汉仪凌波体简 经典平黑简

**Figure 1 The 18 different typefaces under study (14 pt).**

In Figure 1, we can easily see the difference between every two fonts, even they are belonging to the same category. To better understand the differences in the design of different fonts, we have measured the fundamental attributes of them.

### 1.4.1 Character shape

Unlike Roman letters which have a characteristic shape (e.g. lower-case letters mostly occupy the x-height, and there are ascenders or descenders on some letters), Chinese characters occupy a more or less square area, in which the components of every character are written inside in order to maintain a uniform size and shape. This is especially true for small printed characters in Ming and sans-serif styles. Because of this area special characteristic, beginner writers often practice writing on a squared graph paper, and people sometimes use the term "Square-Block Characters" (方块字) in reference to Chinese characters [20]. However, after the carefully examination of Chinese characters, we have found that Chinese characters are not exactly square. In order to examine the typographic characteristics of Chinese fonts and to decide whether these characteristics will influence the font legibility and the viewer’s response to Chinese fonts directly, we have analyzed the eighteen typefaces by looking at three parts: Character Shape, Character Blackness and Character Spacing. We have chosen the Chinese characters with a square outer border as the font experimental samples shown below (Figure 2).

**Figure 2 Samples of Chinese characters with the square outer border in different fonts (from top to bottom: FZKT(方正楷体), MNJZBH(迷你简毡笔黑) and WRYH (微软雅黑), see abbreviation explanation in Table 1).**

We calculated the font height, font width and the ratio between height and width of our eighteen Chinese characters (Table 1, Figures 3, 4 and 5). This helped us to decide the character shape, in which the eighteen typefaces were set to 22 points in measurement. To keep the original design of these fonts, we did not do any normalization on them.

The font height, width and the ratio between the height and the width of Chinese characters were measured by using the horizontal and vertical projection profiles. And the height and the width were measured in millimeters.

| Abbreviation | Font | Height | Width | Ratio between height and width |
|---|---|---|---|---|
| WRJKT | 微软简楷体 | 26.1 | 21.8 | 1.2168 |
| FZKT | 方正卡通 | 24.42 | 24.25 | 1.017 |
| HWZS | 华文中宋 | 27.33 | 25.08 | 1.0979 |
| WRYH | 微软雅黑 | 28 | 27.17 | 1.0387 |
| FZSS | 方正宋三 | 27.25 | 24.58 | 1.1337 |
| MNJZBH | 迷你简毡笔黑 | 26.67 | 25 | 1.0728 |
| FZZY | 方正准圆 | 26.33 | 27 | 0.9812 |
| FZLS | 方正隶书 | 17.8 | 24 | 0.7186 |
| FZFS | 方正仿宋 | 26.3 | 21.2 | 1.2581 |
| HYZYTJ | 汉仪综艺体简 | 27.08 | 26.83 | 1.0138 |
| WRJZY | 微软简中圆 | 26.17 | 24.67 | 1.0654 |
| MNJYY | 迷你简丫丫 | 23.33 | 21.5 | 1.0913 |
| FZWB | 方正魏碑 | 22.58 | 20 | 1.1427 |
| FZHT | 方正黑体 | 26.67 | 24.92 | 1.078 |
| MNJXJ | 迷你简雪君 | 26.17 | 19.92 | 1.337 |
| WRJBS | 微软简标宋 | 27.75 | 22.83 | 1.2934 |
| HYLBTJ | 汉仪凌波体简 | 26.17 | 24.42 | 1.0803 |
| JDPJH | 经典平黑简 | 27.25 | 25.42 | 1.078 |

**Table 1 Font height, font width and the ratio between the height and the width of eighteen fonts. Height and width unit is millimeter.**

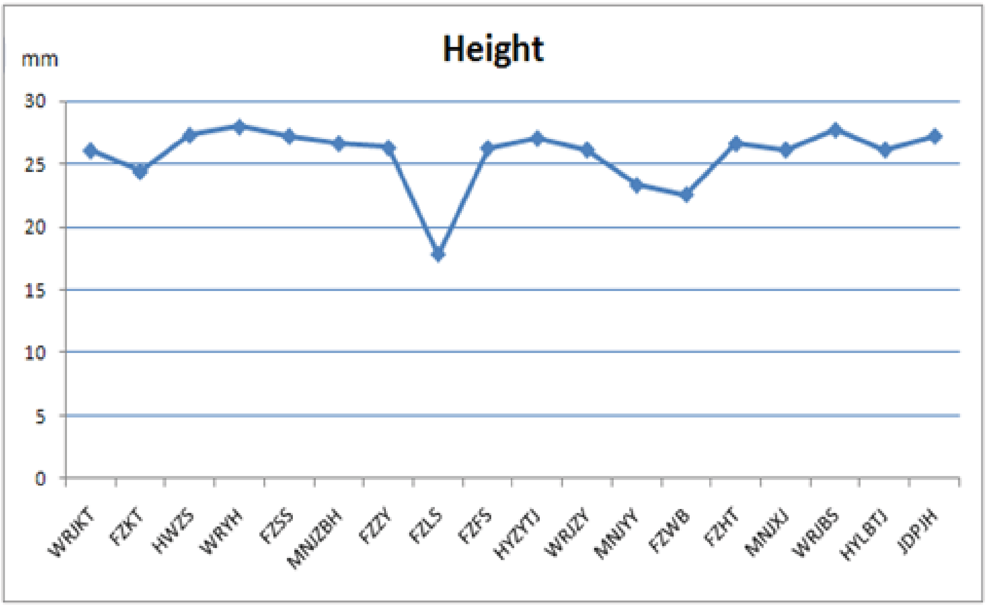



**Figure 3　　Font heights for the eighteen fonts.**

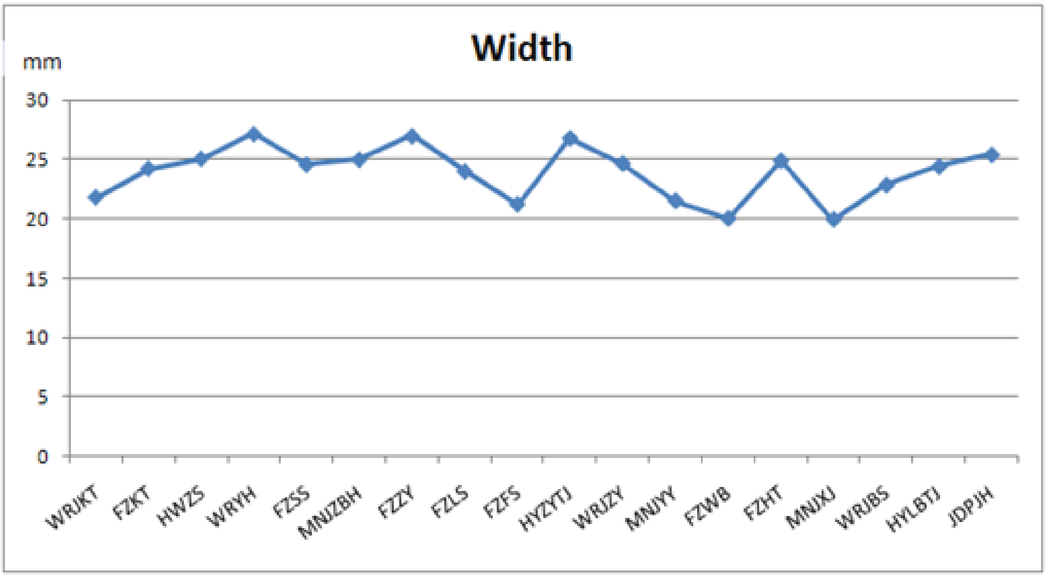



**Figure 4　　Font widths for the eighteen fonts.**

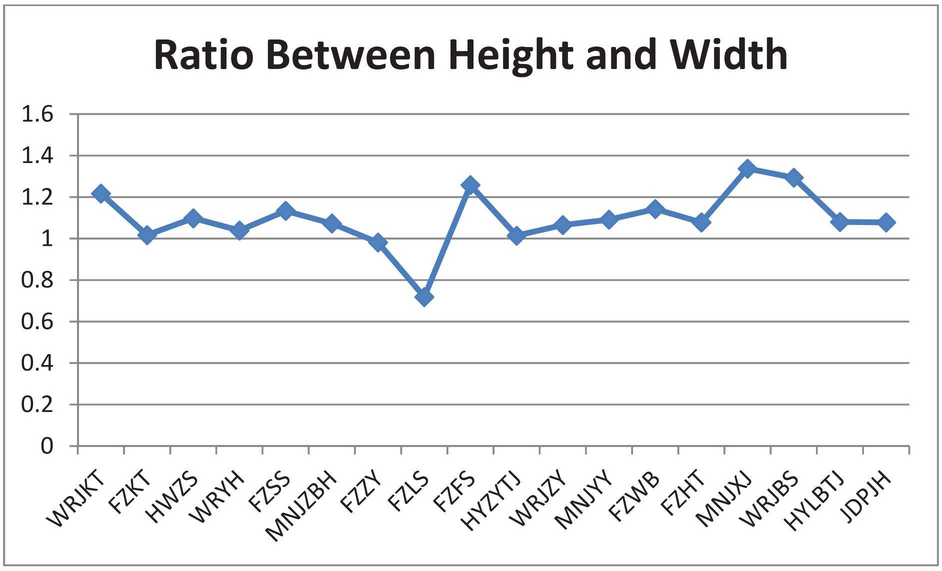



**Figure 5 Ratio between the height and the width for the eighteen fonts.**

From Table 1, Figure 3, Figure 4 and Figure 5 we find that:

1. All the Chinese characters in these eighteen fonts were not a real square, and all the ratios between height and width were not equal to 1. These figures show that all the eighteen fonts were always in rectangles. Although most fonts were in the vertical rectangular shape, only FZZY and FZLS were in the flat rectangular form. The following fonts had ratios of height and width closest to a true square shape : HYZYTJ, FZKT, FZZY and WRYH.
2. In the eighteen fonts, the “shortest” one was FZLS while the “tallest” one was WRJBS. The “thinnest” one was MNJXJ while the “fattest” one was FZLS.
3. Five typefaces that were the most associated with the characteristic “Legible” were WRJKT，WRYH, HWZS，FZSS 和 FZZY, whose ratios were between 1.01 and 1.22.

4. And compared with the five illegible typefaces were MNJZBH，HYZYTJ，MNJXJ，MNJYY and HYLBTJ, whose ratios were in a wide range from 1.01 to 1.34.

5. Based on figure 3 and 4, we found that ratio between height and width was not an important factor with respect to font legibility.

### 1.4.2 Character blackness

The second factor for testing was the character blackness. Here, the ratio between the sum of black pixels inside the square block of each Chinese character and the entire square block was calculated (Table 2 and Figure 6).

| Abbreviation | Font | Character blackness ratio |
|---|---|---|
| WRJKT | 微软简楷体 | 34.45% |
| FZKT | 方正卡通 | 44.81% |
| HWZS | 华文中宋 | 41.73% |
| WRYH | 微软雅黑 | 48.24% |
| FZSS | 方正宋三 | 36.46% |
| MNJZBH | 迷你简毡笔黑 | 55.89% |
| FZZY | 方正准圆 | 50.59% |
| FZLS | 方正隶书 | 57.07% |
| FZFS | 方正仿宋 | 33.47% |
| HYZYTJ | 汉仪综艺体简 | 67.46% |
| WRJZY | 微软简中圆 | 50.45% |
| MNJYY | 迷你简丫丫 | 32.86% |
| FZWB | 方正魏碑 | 48.78% |
| FZHT | 方正黑体 | 51.65% |
| MNJXJ | 迷你简雪君 | 48.32% |
| WRJBS | 微软简标宋 | 40.45% |
| HYLBTJ | 汉仪凌波体简 | 48.71% |
| JDPJH | 经典平黑简 | 51.65% |

**Table 2　　　Character blackness ratios for eighteen fonts.**

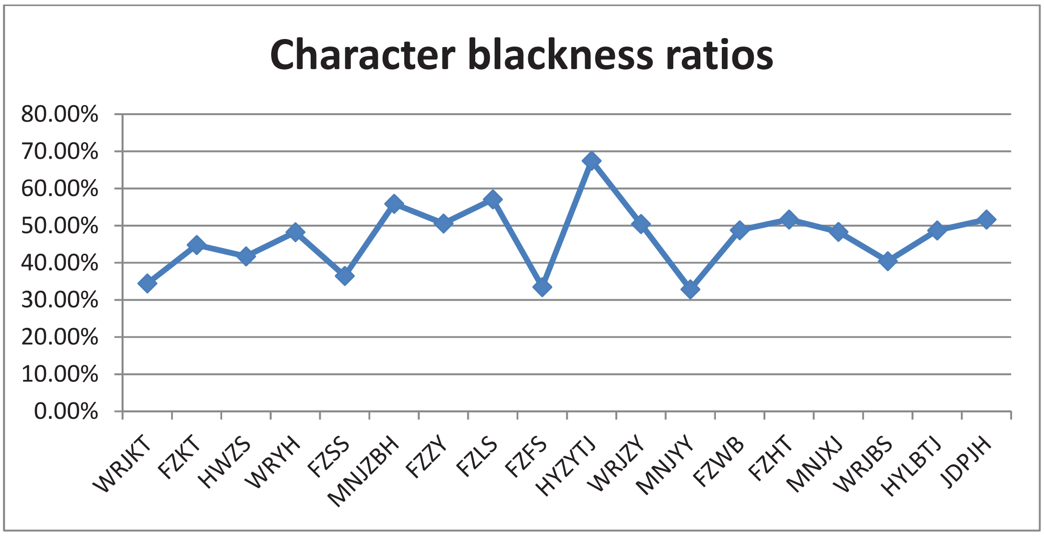



**Figure 6 Character blackness ratios for eighteen fonts.**

From Table 2 and Figure 6, we found that:

1. The Character Blackness of all the eighteen fonts fell within the range from 32% to 68%.

2. Of the eighteen fonts, the "lightest" one was MNJYY while the "strongest" one was HYZYTJ. These two fonts were both belonging to calligraphy fonts. The blackness of calligraphy fonts varied greatly from 32% to 68%. However, the blackness of mainstream fonts fell within the range of 30% to 50%.

3. The blackness of SongTi fell in the range of 35% to 40%. The blackness of HeiTi and YuanTi fell around 50%. While, the blackness of Kai Ti fell around 35%.

4. Combined character blackness result with the results of character, we found that the shapes of HeiTi and YuanTi were close to a square and their blackness fell in the range of 50%.

## 1.5 Thesis Outline

In this thesis, we focus on the discovery of legibility and readability of the current Chinese fonts when they are displayed on screen. We concentrate on the good applications for a Chinese typeface, combining psychology with pattern recognition knowledge to make a reasonable, authentic, satisfied and acceptable analysis. The goal is to help font developers and readers understand more about Chinese typefaces. This thesis has been organized into six chapters, with the next five of them described below.

- In Chapter 2, we describe the details of the survey on typeface characteristics. We use a psychological method to analyze the survey data, filter the results and retain the six favorite typefaces from the 18 different fonts, then focus on them.

- In Chapter 3, we present the legibility experiment. We also provide a detailed analysis on the single characters to find out which font performs best in single character display.
- In Chapter 4, we describe the readability experiment in detail. The experimental results are analyzed and compared with the legibility experiment result for searching for one or more typefaces performed well in both legibility and readability aspects.
- In Chapter 5, we test the connection between human recognition and machine OCR. We look at the single character recognition and integral recognition to verify whether our potential recommended typeface is fit for the current digital market.
- Finally, we summarize this thesis in Chapter 6 with some reasonable conclusions, and then we make a plan for the future work.

# Chapter 2. Survey of Human Preference of Fonts

In this chapter, we applied a psychological method to complete a survey and analysis to search for people's favorite Chinese fonts. Our goal was to detect people's implicit preference of fonts when reading. Additionally, we verified the influence of Chinese typefaces on Chinese readers and their performance on screens. Moreover, by applying this survey, we filtered the current 18 typefaces and kept the top six typefaces for the following research, which would make our research result more accurate and more suitable according to current people's preference.

## 2.1 Survey Description

In this survey, there were a total of 18 different fonts displaying the Chinese pangram. There were two different kinds of questions in this survey for collecting the related information. One was a relative subjective question concerning the candidates' feelings such as, "Do you think it is legible?" or "Do you think it is formal?" The candidates had to choose their answers from multiple choice options, representing different degrees of different characteristics, such as, 0~20%, 20%~40%, etc. The second kind of questions was relatively objective, by comparison. It was more like an experiment, so we just named it as an experiment here. Firstly, the candidates had to read a paragraph of instruction about a neck exercise. After understanding the whole meaning of this instruction, candidates had to estimate how long it would take to finish this exercise as required. These two parts of the survey had different purposes. The subjective question's purpose was to search for the fonts that people prefer, and the objective part was used to discover whether different typefaces would influence participants' reading speed and comprehension. After finishing this survey, we collected all the information and data for statistical analysis. At last, we filtered these 18 different fonts and the six best remaining fonts were the ones for further investigation.

## 2.2 Structural Design

Firstly, we divided all the participants into two groups, and we prepared two versions of survey for these two groups. One version was experiment 1 (easy font) and pangrams of 18 fonts, and the other one was experiment 2 (curved font) and pangrams with the same fonts with version 1. We attempted to compare the results of time consumption estimation from these two experiments.

Secondly, in order to make sure that all candidates would give a fair choice for all typefaces, we regulated the appearance of different fonts. We resized the pangrams' area height of different fonts to the same height after setting them on the same line space, making them look like they were of the same size, meanwhile keeping their original font designs and structures untouched. Actually Chinese characters are not similar to English characters. The Chinese character is a square character, unlike the English combination of letters. Moreover Chinese characters of different structures are seldom of same size, even in the same structure. Thus, a Chinese character doesn't have X height, so we cannot resize these Chinese characters to the standard X height as in the English letter resize setting. The only way to resize characters after testing is to resize the pangram area height to a standard height, which makes the pangrams all perform clearly and easier to read. Firstly, we set a bounding box on the pangram's outer pixels, to delete the empty space.

Then, we measure the height of the box. Lastly, we resize the height to the standard height, keeping the original ratio of height and width.

Thirdly, we put the objective experiment at the top of the question list, and then we randomly chose the display sequence of these pangrams for the 18 fonts. Then we collected the participants' answers, using a statistical method to find the font preference of people.

## 2.3 Detailed Design

In this section, we present the designs of this survey in detail, including the objective and subjective parts separately. In Section 2.3.1, we describe the subjective questions including the fonts' selection, pangram design and questions design. In Section 2.3.2, we provide the detail information about the objective experiment design.

### 2.3.1 Subjective Questions Design

- 18 Fonts

  These 18 different Chinese fonts have been displayed in Chapter 1 (Figure 1). We refer to that figure.

- Pangram design:

  2010 年，房地产调控政策频频出炉，但一线城市的房价依然“高烧不退”。
  的一是在了不和有大这主中人上为们地个用工绸矣忧啡芥蚂萃髅鼾馗
  米日品森思华霜花基想意褒裹村联伟搞刚郭街坳滩傲圆国医庆尾匀句遍建闻函
  床前明月光，疑是地上霜。举头望明月，低头思故乡。

**Figure 7 Sample of the Chinese pangram.**

In this pangram, there are 110 characters, 4 digits, and 9 punctuations in total. The material on every line has been derived from different usages. The first line was a piece of news quoted from a Chinese newspaper. The second line was chosen from a Chinese character usage frequency table, such that: the first 20 characters were made of high frequent usage characters; 7 moderate frequent usage characters were placed in the middle; and lastly, we used 3 low frequency characters at the end of the second line [27]. The third line was made of 19 different structural compositions of character which form the Chinese characters, and each composition consists of different kinds of radicals in different positions and sizes. For example, two parts are placed side by side in character “村”; the second part is placed below the first part in character “霜”. Finally the last line was a famous poem called “Jing Ye Si” (静夜思, this poem described the author′s homesick emotion induced by the moonlight when he was looking at the moon) which was written by the famous Chinese poet, Li Bai, in Tang Dynasty.

- Survey Questions:

  After reading the pangram displayed in one of these 18 fonts, participants had to answer 5 questions about the pangram with respect to 5 different characteristics:

legible, attractive, comfortable, artistic and formal. These five characteristics are the fundamental aspects of typeface design, and here we used them to directly reflect the candidates' subjective feeling. Meanwhile, we could also learn whether these typefaces are fit for screen display from participants' responses. Each question was a multiple choice question. The candidates had to choose their own answers from the five degrees in each characteristic question. The options are shown below.


**Figure 8 Sample of the voting survey.**

### 2.3.2 Objective Experiment Design:

People misread the ease of processing instructions as indicative of the ease with which the described behavior can be executed [28] (refer the behavior in the reference). In reading materials, as predicted, participants estimated that the exercise would take less time and feel quicker and more fluent when the font was easy to read than when the font was difficult to read. Accordingly, they reported a higher willingness to make the exercise part of their daily routine when it was described in an easy-to-read font than when it was described in a difficult-to-read font.

Research by Hyunjin Song and Norbert Schwarz [29] shows that the way we perceive textual information can be affected dramatically by how simple or complex the font is. In particular, they found that a simple font was more likely to get the readers to make a commitment. Moreover, Song and Schwarz performed a similar experiment involving a sushi recipe. Subjects who saw the instructions in the simple font, Arial, estimated that preparation would take 5.6 minutes, while those who read the directions in Mistral, a more complicated font, expecedt it to take 9.3 minutes [29]. These results confirmed their conclusion. The method described is a great way to verify the differences caused by two different typefaces, to some extent. However, it is obvious to see the differences between these two different fonts. However, we are not sure whether this method is fit for Chinese typefaces as well.

Thus, we modified this experiment in our survey to test whether it would be effective for Chinese fonts, and to test how people would be influenced by them. Firstly, we translated the English material into Chinese, and then we used the same method for questioning. We used two different styles of typefaces to conduct this experiment. The first one was WRYH, a mainstream typeface, which is the default display font in the operating system for Windows Vista and Windows 7. The second one was MNJYY, belonging to a new calligraphy group, which is a very prevalent font used on internet chatting tools, such as, QQ and MSN, because of its fancy and cute design. Samples are shown below.

低头，将下巴尽可能靠近胸口，抬头，将下巴尽可能向上扬，重复 6 次。
摆正头部，然后向左歪头，使左耳尽量靠近左肩，回正头部，然后向右歪头，
使右耳尽量靠近右肩，重复 6 次。

低头，将下巴尽可能靠近胸口，抬头，将下巴尽可能向上扬，重复 6 次。
摆正头部，然后向左歪头，使左耳尽量靠近左肩，回正头部，然后向右歪头，
使右耳尽量靠近右肩，重复 6 次。

**Figure 9 Samples of the exercise material, the first one was printed in WRYH, the second one was printed in HYLBTJ.**

- Survey Question:

  Firstly, participants had to read this paragraph of instruction concerning a neck exercise. After understanding the whole meaning of this instruction, participants had to estimate how long it would take for him or her to finish this exercise as the instruction described.

## 2.4 Survey Results and Analysis

After the data collection, we calculated the mean value of the rating score of each of the five typeface characteristics. We examined the minimum values, maximum values and standard deviations of the rating scores of each typeface based on each characteristic. Table 3 shows the mean values of rating scores for eighteen typefaces related to the five characteristics. We summarized the results from the top five typefaces that were the most and least associated with each of the five characteristics in Tables 4 and 5.

| Abbreviation | Font | Legible | Comfortable | Attractive | Artistic | Formal |
|---|---|---|---|---|---|---|
| WRJKT | 微软简楷体 | 3.89 | 3.62 | 3.52 | 3.25 | 3.46 |
| FZKT | 方正卡通 | 3.23 | 3.08 | 3.30 | 3.15 | 2.00 |
| HWZS | 华文中宋 | 3.95 | 3.38 | 3.36 | 2.80 | 4.23 |
| WRYH | 微软雅黑 | 3.97 | 3.23 | 3.36 | 2.92 | 3.54 |
| FZSS | 方正宋三 | 4.00 | 3.39 | 3.57 | 3.10 | 4.21 |
| MNJZBH | 迷你简毡笔 | 3.08 | 2.89 | 2.84 | 3.15 | 2.10 |
| FZZY | 方正准圆 | 3.93 | 3.44 | 3.39 | 3.00 | 3.33 |
| FZLS | 方正隶书 | 3.56 | 3.89 | 3.52 | 3.87 | 3.18 |
| FZFS | 方正仿宋 | 3.64 | 3.18 | 3.38 | 2.95 | 3.80 |
| HWZYTJ | 汉仪综艺体 | 3.03 | 2.62 | 2.59 | 2.74 | 2.75 |
| WRJZY | 微软简中圆 | 3.62 | 3.30 | 3.41 | 3.10 | 3.28 |
| MNJYY | 迷你简丫丫 | 2.07 | 1.89 | 2.02 | 2.33 | 1.33 |
| FZWB | 方正魏碑 | 3.98 | 3.93 | 3.77 | 3.74 | 3.59 |
| FZHT | 方正黑体 | 3.93 | 3.44 | 3.36 | 2.98 | 3.77 |
| MNJXJ | 迷你简雪君 | 2.93 | 3.31 | 2.79 | 3.84 | 2.10 |
| WRJBS | 微软简标宋 | 3.59 | 2.92 | 2.72 | 2.61 | 3.34 |
| HYLBTJ | 汉仪凌波体 | 1.82 | 1.97 | 1.90 | 2.57 | 1.23 |
| JDPHJ | 经典平黑简 | 3.82 | 3.18 | 3.30 | 2.84 | 3.56 |

**Table 3 Mean values of rating scores for 18 typefaces related to five characteristics in the survey.**

| Characteristics | Font | | | | |
|---|---|---|---|---|---|
| Legible | FZSS | FZWB | WRYH | HWZS | FZZY/FZHT |
| | 4.00 | 3.98 | 3.97 | 3.95 | 3.93 |
| Attractive | FZWB | FZLS | WRJKT | FZZY | FZHT |
| | 3.93 | 3.89 | 3.62 | 3.44 | 3.44 |
| Comfortable | FZWB | FZSS | WRJKT | FZLS | WRJZY |
| | 3.77 | 3.57 | 3.52 | 3.52 | 3.41 |
| Artistic | FZLS | MNJXJ | FZWB | WRJKT | FZHT/MNJZ |
| | 3.87 | 3.84 | 3.74 | 3.25 | 3.15 |
| Formal | HWZS | FZSS | FZFS | FZHT | FZWB |
| | 4.23 | 4.21 | 3.80 | 3.77 | 3.59 |

**Table 4 Five typefaces that were the most associated with each of the five characteristics and their means.**

| Characteristics | Font | | | | |
|---|---|---|---|---|---|
| Legible | HYLBTJ | MNJYY | MNJXJ | HYZYTJ | MNJZBH |
| | 1.82 | 2.07 | 2.93 | 3.03 | 3.08 |
| Attractive | MNJYY | HYLBTJ | HYZYTJ | MNJZBH | WRJBS |
| | 1.89 | 1.97 | 2.62 | 2.89 | 2.92 |
| Comfortable | HYLBTJ | MNJYY | HYZYTJ | WRJBS | MNJXJ |
| | 1.90 | 2.02 | 2.59 | 2.72 | 2.79 |
| Artistic | MNJYY | HYLBTJ | WRJBS | HYZYTJ | HWZS |
| | 2.33 | 2.57 | 2.61 | 2.74 | 2.80 |
| Formal | HYLBTJ | MNJYY | FZKT | MNJZBH | MNJXJ |
| | 1.23 | 1.33 | 2.00 | 2.10 | 2.10 |

**Table 5 Five typefaces that were the least associated with each of the five characteristics and their means.**

The purpose of this thesis is to find out how to improve the Chinese fonts that can be much better for people's reading, and to find the typeface that is the most fit to be displayed on screens for e-book devices and tablets. Moreover, we need good fonts for the adult and the student target market for their daily usage, so we focused on three of the five characteristics: legible, comfortable and formal. These three main aspects are shown in Table 6. We filtered the typefaces whose sum of scores was not in the top five high ranking (refer Table 6, last column). Meanwhile, we kept the typefaces which had at least two characteristics with scores belonging to the top 5 candidates' choices. We thereby avoided the effect of a single high score based on one characteristic, which would make the typefaces with several characteristics outstanding overlooked. Finally, the remaining five fonts after filtering were: Fzss(5), Hwzs(3), Fzwb(13), Fzht(14), wrjkt(1).

| Abbreviation | Font | Legible | Comfortable | Formal | Total sum value |
|---|---|---|---|---|---|
| WRJKT | 微软简楷体 | 3.89 | 3.52 | 3.46 | 10.87 |
| FZKT | 方正卡通 | 3.23 | 3.30 | 2.00 | 8.53 |
| HWZS | 华文中宋 | 3.95 | 3.36 | 4.23 | 11.54 |
| WRYH | 微软雅黑 | 3.97 | 3.36 | 3.54 | 10.87 |
| FZSS | 方正宋三 | 4.00 | 3.57 | 4.21 | 11.78 |
| MNJZBH | 迷你简毡笔 | 3.08 | 2.84 | 2.10 | 8.02 |
| FZZY | 方正准圆 | 3.93 | 3.39 | 3.33 | 10.65 |
| FZLS | 方正隶书 | 3.56 | 3.52 | 3.18 | 10.26 |
| FZFS | 方正仿宋 | 3.64 | 3.38 | 3.80 | 10.82 |
| HWZYTJ | 汉仪综艺体 | 3.03 | 2.59 | 2.75 | 8.37 |
| WRJZY | 微软简中圆 | 3.62 | 3.41 | 3.28 | 10.31 |
| MNJYY | 迷你简丫丫 | 2.07 | 2.02 | 1.33 | 5.42 |
| FZWB | 方正魏碑 | 3.98 | 3.77 | 3.59 | 11.34 |
| FZHT | 方正黑体 | 3.93 | 3.36 | 3.77 | 11.06 |
| MNJXJ | 迷你简雪君 | 2.93 | 2.79 | 2.10 | 7.82 |
| WRJBS | 微软简标宋 | 3.59 | 2.72 | 3.34 | 9.65 |
| HYLBTJ | 汉仪凌波体 | 1.82 | 1.90 | 1.23 | 4.95 |
| JDPHJ | 经典平黑简 | 3.82 | 3.30 | 3.56 | 10.68 |

**Table 6 Typefaces ranked according to the sum values of legible, comfortable and formal.**

We found that the top five typefaces which people thought were the most suitable for reading were distributed into 4 font categories: KaiTi, SongTi, HeiTi, WeiBeiTi. Through this survey, we found that people did not like the WRYH font which is the default display typeface of the operating system for Windows Vista and Windows 7. This default typeface was set by Microsoft Company, and has been used by billions Chinese customers. We made a decision to add the WRYH to the best fonts list for further research and analysis, and we would make an elaborate comparison between WRYH and FZHT, both produced by Founder Company in Beijing, China. Thus, we have six typefaces for further research: FZSS, HWZS, FZWB, FZHT, WRJKT, and WRYH.

方正黑体字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

方正宋三字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

微软简楷体字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

华文中宋字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

方正魏碑字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

微软雅黑字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

**Figure 10 Display of six remaining fonts. After the first comma, characters of different typefaces are consistent. The font display sequence from top to bottom is FZHT, FZSS, WRJKT, HWZS, FZWB, WRYH.**

On the other hand, after collecting the experimental data in this survey, we found that in Chinese, the typeface effect exists and is almost similar to the result of the experiment applied in English. The mean estimated time of reading the legible font is 4 minutes and 23 seconds, and the mean estimated time of reading the difficult font is around 6 minutes and 37 seconds. Although the estimation could not deduce a conclusion about which font is the best and which one is the worst, it verified that the mainstream typeface was much easier to read and understand for human brains. According to the participants' preferences, the mainstream typeface is more suitable for people's daily lives than the calligraphy fonts.

# Chapter 3. Legibility Analysis

After researching on the human's preference of fonts, we kept the top six typefaces for the further research. In this chapter, we attempt to discuss the legibility of fonts. To find out which features of a single character will improve legibility or decrease legibility, we use single characters to conduct this experiment. To better understand this legibility experiment described in this chapter, we summarize its content here: In Section 3.1, we provided the detail about experiment data collection. In Sections 3.2 and 3.3, we introduced the experiment participants' information and experiment preparations. In Section 3.4, we presented the core of this experiment, including the experimental results processing and comparison of results. Lastly, we made an extensive discussion to analyze the results in Section 3.5.

## 3.1 Data Collection

In this experiment, we collected the character data from the five fundamental structures in Chinese characters.

- Five fundamental structures of Chinese characters

  There are several ways to define the structures of Chinese characters. For example, for the 19 character structures that were defined in Chapter 2. Each structure could always be subdivided and then sub-defined, recursively. For instance, character "滩", can be defined as a left-right structure at the first level, and then the right part, "难", could be divided into two parts again as a left-right structure at the second level, since it is a left-right structure character as well. A detailed illustration is shown below

(Figure 11). Considering this condition, we used the five very fundamental structures to group Chinese characters, which can totally cover all Chinese character components without any exceptions: left-right, up-down, half - cover, full - cover and single structure. (See Fig. 12) These very fundamental structures were defined as the very first level (one level) in character division, based on the major division of features, and ignoring the remaining features. Character structural division sample is shown below.

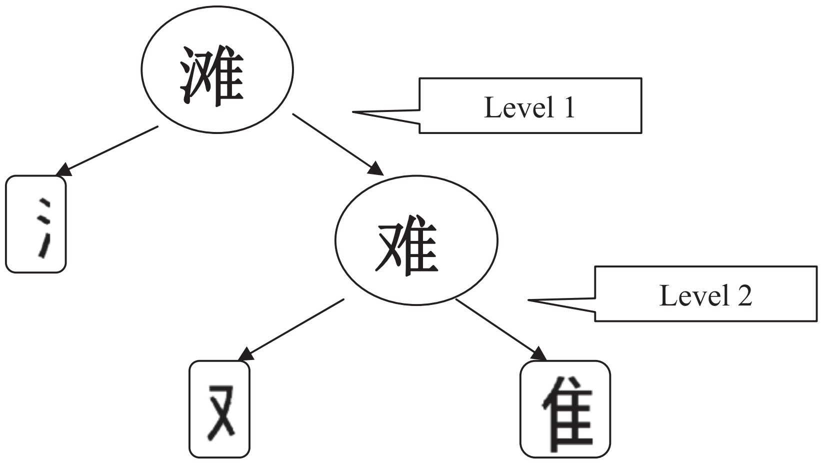



**Figure 11 Two levels dividing a Chinese character.**

| | | | | | |
|---|---|---|---|---|---|
| Right-left | 加 | Up-down | 导 | Single | 之 |
| Half-cover | 适 旬 | | Full-cover | 回 囵 | |

**Figure 12 Five main categories of Chinese character structures.**

To guarantee the accuracy of this legibility experiment, we must use enough characters to balance and to increase the precision of the experimental result. Meanwhile, every character must represent and display one font structure. According to these five fundamental structures, we began to filter the Chinese characters by following the four rules below:

1. There should be five groups of characters and each group represents one typeface structure.
2. Each group must include the characters with different strokes, which means that in each group, we have various kinds of characters that may have different sub-structures, representing one main structure.
3. Each font covers all of these five different structures of characters.
4. There should be an equal quantity of characters with the same structures and the same strokes distributed among the six selected fonts.

According to the rules above, we could easily control the experiment's progress, and guarantee the quality of the experiment. Thus, we prepared 600 characters for these six different typefaces which had been selected in Chapter 2, which meant that each font had 100 characters. Then, for every 100 characters in the same font, we used 5 different font structures. Each font structure had the same number of different characters. Thus, in this experiment's database, we had 5*20*6 = 600 characters, which meant in each structure, we had 20 characters. As the amount of full-cover characters was low, we allowed character duplication in full-cover structure. The percentage of full-cover characters was lower than 5%.

## 3.2 Participants

Fifty-one participants (16 males and 35 females) were recruited for this experiment. They were all students from Beijing Normal University, with ages ranging from 18-28 years old (the mean age was 22 years old). Their educational background covered undergraduates, masters and doctorate students. Every participant was qualified for the requirement of good eye sight (twenty-twenty vision, and no problem with color perception). All participants were paid after test.

## 3.3 Materials and Equipment

In this part, we introduce the experimental data preprocessing in detail in Section 3.3.1, and describe the environment of applying this legibility experiment in Section 3.3.2.

### 3.3.1 Data Preprocessing

Different fonts have different styles, for example, same characters in different fonts may have different character heights and widths. To maintain the design of different fonts, and to reduce the unnecessary visual effects when reading, we adopted the linear normalization method. This method brings all fonts to the same height, because the Chinese character's point weight is measured by its height.

Moreover, we calculated the black to white contrast and the height and width contrast for further analyses. In Chinese, full-cover structure characters are most with representative and easier to see the differences; hence we used full-cover characters to obtain the values of these two types of contrasts. In Table 7 below, the two different contrasts are shown for the six different typefaces.

| | FZHT | FZWB | HWZS | WRJKT | WRYH | FZSS |
|---|---|---|---|---|---|---|
| H/W[a] | 1.0780 | 1.1427 | 1.0979 | 1.2168 | 1.0387 | 1.1337 |
| B/W[b] | 51.65% | 48.78% | 41.73% | 34.45% | 48.24% | 36.46% |

a. H/W means height/ width contrast. b. B/W means black/ white contrast

**Table 7 Two Different Contrast Values for six Different Typefaces.**

### 3.3.2 Equipment

This experiment was computerized, and all the computers had a CPU with P4 2.8GHz and a 17 inches CRT monitor. The screen resolution was 1024*768 pixels and the vertical refresh frequency was 85Hz. During the experiment, the data was displayed by a software tool called E-prime 1.0 (made by Microsoft) [30].

Based on the explanation on legibility mentioned in Chapter 1, section 1.3, we decided to use the Single Character Flashing Response Test (SCFRT) method in this experiment. The SCFRT method will be described in the next section.

## 3.4 SCFRT procedure

The SCFRT method is described below: Firstly, a focus point is displayed at the center of the screen for 500 ms. Then this point disappears, and the stimulator (recognition target of a single character, which is displayed in a random order, chosen from our database of 600 characters with different typefaces) appears at that position for 100 ms. Then the response stimuli (a pair of characters with similar structure and partially similar look, such as “澡” and “燥”) are displayed in the same typeface which is different from all these six fonts listed in Table 7. This process is to guarantee that there is no hint caused by font’s global features to remind and affect the participants’ recognition results. The response stimulus is displayed on screen, replacing the stimulator for 1500 ms after a 100 ms time interval. This is used to avoid the masking effect stimulator on the response stimulus. All the participants were asked to respond to the single stimulator as soon as possible while keeping a high accuracy, i.e. if the participants think that the previous single stimulator character is displayed in the left position in the response stimulus (a pair of partially similar characters), he or she just presses the Q button, and if that stimulator is shown in the right position, press the P button. Moreover, to help participants keep a high accuracy and to reduce visual fatigue, there was a free time of 1000 ms between every two questions. The whole experiment is under a time control that is hidden from all participants, to make them feel relaxed.

After finishing this experiment as described above, we have to accomplish the following processing to ensure the experimental results accurate and effective. In Subsection 3.4.1, we introduce the data processing methods. In Subsection 3.4.2, we present the tests of within-subjects effects analyses to make sure whether the influence caused by fonts and structures existed or not. In Subsection 3.4.3, we provide a pairwise comparison to test whether the difference could be ignored or not.

### 3.4.1 Data Processing

To ensure the effectiveness of the experimental result, we set a threshold. If the participant's accuracy was lower than 95%, we just ignored this data. After filtering, we got 47 effective results for analysis (See Table 8). Meanwhile, we measured all participants' response times pertaining to recognizing every single stimulator character.

Then, we calculated the mean recognition time of different typefaces and different structures. A descriptive Statistical analysis is presented in Table 9:

| Num | Minimum | Maximum | Mean | Std. Deviation |
|---|---|---|---|---|
| 47 | 0.95 | 1 | 0.971702 | 0.010697 |

**Table 8 Mean value and standard deviation value of the total 47 data, which have a maximum value of 1 and a minimum value of 0.95.**

| Name | Mean Time (ms) | Std. Deviation | N |
|---|---|---|---|
| FZHT-fullcover | 522.1157 | 54.59896 | 47 |
| FZHT-halfcover | 472.3074 | 46.75997 | 47 |
| FZHT-rrghtleft | 474.9648 | 50.02214 | 47 |
| FZHT-single | 447.326 | 35.82618 | 47 |
| FZHT-updown | 472.2412 | 45.80617 | 47 |
| FZWB-fullcover | 532.9306 | 50.98345 | 47 |
| FZWB-halfcover | 469.2197 | 42.02677 | 47 |
| FZWB-rightleft | 468.7825 | 48.09553 | 47 |
| FZWB-single | 444.2429 | 38.34053 | 47 |
| FZWB-updown | 471.4212 | 43.78109 | 47 |
| HWZS-fullcover | 533.9610 | 71.49120 | 47 |
| HWZS-halfcover | 508.7778 | 56.63352 | 47 |
| HWZS-rightleft | 516.376 | 63.43634 | 47 |
| HWZS-single | 499.2978 | 56.82770 | 47 |
| HWZS-updown | 497.882 | 54.73623 | 47 |
| WRJKT-fullcover | 512.7419 | 58.12920 | 47 |
| WRJKT-halfcover | 512.5829 | 58.12659 | 47 |
| WRJKT-rightleft | 536.7780 | 60.17536 | 47 |
| WRJKT-single | 503.7253 | 60.76482 | 47 |
| WRJKT-updown | 516.4334 | 59.48079 | 47 |
| WRYH-fullcover | 553.8212 | 71.59743 | 47 |
| WRYH-halfcover | 513.0593 | 54.46279 | 47 |
| WRYH-rightleft | 514.7989 | 54.54844 | 47 |
| WRYH-single | 499.9706 | 57.23684 | 47 |
| WRYH-updown | 506.3140 | 50.94732 | 47 |
| FZSS-fullcover | 552.4442 | 63.89894 | 47 |
| FZSS-halfcover | 518.6544 | 55.90874 | 47 |
| FZSS-rightleft | 570.5710 | 66.1541 | 47 |
| FZSS-single | 538.8540 | 60.47079 | 47 |
| FZSS-updown | 549.1565 | 67.02490 | 47 |

**Table 9 Detailed fonts and structures for information display.**

### 3.4.2 Tests of Within-Subjects Effects

To make sure whether different fonts, different structures, and different fonts with different structures would influence the speed of recognizing characters by humans, we conducted the Tests of Within-Subjects Effects. We used four different analytical methods to analyze and compare results, and fortunately, we got consistent results. The results showed that those fonts and structures' main effects and interactions between fonts and structures were all above the significance level. This meant that they all had the capacities to affect the speed of recognition. A detailed analysis will be shown later in this section. To better analyze the experimental data and obtain better results, we introduced the ANOVA method and multiple factors analysis method below.

- ANOVA method

In statistics, Analysis of Variance (ANOVA) is a collection of statistical models, and their associated procedures, in which the observed variance in a particular variable is partitioned into components, attributable to different sources of variation [32]. In its simplest form, ANOVA provides a statistical test of whether or not the means of several groups are all equal, and therefore it generalizes the *t*-test. The *t*-test is the most commonly used method to evaluate the differences in means between two groups, or to among more than two groups [31]. By multiple two-sample *t*-tests, it could result in an increased chance of committing a type I error. For this reason, ANOVAs are useful when comparing two, three or more means. In general, the purpose of analysis of variance is to test the differences in means (for groups or variables) for a statistical significance (SS). This can be accomplished by analyzing the variance, that is, by partitioning the total variance into the component that is due to a true random error (i.e., within-group SS) and by partitioning the components that are due to the differences between means. These latter variance components are then tested for the statistical significance, and, if significant, we reject the null hypothesis of no differences between means and accept the alternative hypothesis that the means (in the population) are different from each other [32].

- Multiple factors

The world is complex and multivariate in nature, and instances when a single variable completely explains a phenomenon are rare. For example, when trying to explore how to grow a bigger tomato, we would need to consider factors that have to do with the plant's genetic makeup, soil conditions, lighting, temperature, etc. Thus, in a typical experiment, many factors are taken into account. One important reason for using the ANOVA methods rather than multiple two-group studies, analyzed via *t* tests, is that the former method is more efficient, and with fewer observations we can gain more information. Thus, we used the normal ANOVA method to analyze the fonts and influence of structures, and used the multiple ANOVA to find the effect caused by fonts and structures.

Moreover, we used SPSS, which is a statistical tool that was developed by IBM Company. SPSS is used for solving data mining, text analytics, statistical analysis, and collaboration & deployment problems [35]. It was used to finish all calculations about the main effects. A detailed calculation and analysis is shown below in Table 10:

| | | MAIN EFFECT | | | | | |
|---|---|---|---|---|---|---|---|
| source | | Type III Sum of Squares | df | Mean | F | Sig. | Partial Eta Squared |
| Font (F) | Sphericity | 814663.1953 | 5 | 162932.6391 | 63.5468 | 1.93605E- | 0.5801 |
| | Greenhouse- | 814663.1953 | 2.6697 | 305142.1378 | 63.5468 | 2.70493E- | 0.5801 |
| | Huynh-Feldt | 814663.1953 | 2.8497 | 285872.9708 | 63.5468 | 1.06669E- | 0.5801 |
| | Lower-bound | 814663.1953 | 1 | 814663.1953 | 63.5468 | 3.25655E- | 0.5801 |
| Error(F) | Sphericity | 589714.6787 | 230 | 2563.9768 | | | |
| | Greenhouse- | 589714.6787 | 122.81 | 4801.8456 | | | |
| | Huynh-Feldt | 589714.6787 | 131.08 | 4498.6178 | | | |
| | Lower-bound | 589714.6787 | 46 | 12819.88432 | | | |
| Structures(S) | Sphericity | 344197.1603 | 4 | 86049.2900 | 115.752 | 3.8411E-49 | 0.7156 |
| | Greenhouse- | 344197.1603 | 3.0527 | 112750.7551 | 115.752 | 4.05853E- | 0.7156 |
| | Huynh-Feldt | 344197.1603 | 3.2942 | 104484.9169 | 115.752 | 6.26085E- | 0.7156 |
| | Lower-bound | 344197.1603 | 1 | 344197.1603 | 115.752 | 3.77741E- | 0.7156 |
| Error(S) | Sphericity | 136783.8848 | 184 | 743.3906 | | | |
| | Greenhouse- | 136783.8848 | 140.42 | 974.0680 | | | |
| | Huynh-Feldt | 136783.8848 | 151.53 | 902.6583 | | | |
| | Lower-bound | 136783.8848 | 46 | 2973.5627 | | | |
| F * S | Sphericity | 221109.3409 | 20 | 11055.4670 | 21.8020 | 3.3696E-64 | 0.3215 |
| | Greenhouse- | 221109.3409 | 11.585 | 19084.6244 | 21.8020 | 3.015E-38 | 0.3215 |
| | Huynh-Feldt | 221109.3409 | 15.764 | 14025.6292 | 21.8020 | 3.82798E- | 0.3215 |
| | Lower-bound | 221109.3409 | 1 | 221109.3409 | 21.8020 | 2.64247E- | 0.3215 |
| Error(F*S) | Sphericity | 466516.7752 | 920 | 507.0834 | | | |
| | Greenhouse- | 466516.7752 | 532.94 | 875.3585 | | | |
| | Huynh-Feldt | 466516.7752 | 725.17 | 643.3165 | | | |
| | Lower-bound | 466516.7752 | 46 | 10141.6690 | | | |

**Table 10 Main effect analysis information displayed in detail.**

As illustrated, firstly, we used the Sphericity Assumption method to analyze the relative data. By this method, we analyzed three different aspects - different fonts, different structures and different fonts with different structures. We wanted to see whether they affected the speed of recognizing characters by humans. We used these formulas:

Main effect F (different font, different error font) = f (value);

p = sig ; ( where p is the significant possibility level in a statistical treatment)

effect size = Partial Eta Squared value;

Normally in statistics, if value $p < 0.05$, it means that this event has reached the minimum possibility event threshold and that the significant effect happened. Otherwise, there is no statistical difference.

Based on the formulas, we could deduce the relation between the different fonts and the Chinese character recognition speed. For example, $F (4, 230) = 63.547$, and $p = 0.000 < .001$, where the effect size ( power ) = 0.58, this meant that different fonts had effect on the recognition speed, and the significance level was below 0.001. We used the same way to calculate and analyze the influence of different structures and the effect of different fonts and structures. Then we found that different structures affected the speed of character recognition significantly, and different fonts with different structures also impacted that speed.

Secondly, to guarantee the accuracy of our finds, we used three other analysis methods (Greenhouse-Geisser, Huynh-Feldt and Lower-bound) to compare, analyze and verify the existence of influences based on consistent data. Fortunately, the conclusions were accordant.

According to the information that was displayed in Table 10 and the analysis of the test of within-subjects effects, we can find the existing differences in recognition speed pertaining to different fonts and different structures. To display the important information clearly and understandably in a direct way, charts are shown below in Figures 13 and 14. Although the recognition speed differences could be seen clearly in these two figures, whether the differences among the recognition speed of different fonts were significant or not and whether these differences should be kept or ignored between every two typefaces were unclear. Normally from the practical angle, if the difference is not significant, or the difference is too tiny to influence the final result, it should be excluded. Thus, another analysis named as the pairwise comparison has to be conducted to verify the truth of the influences mentioned above and the differences between the existing values.

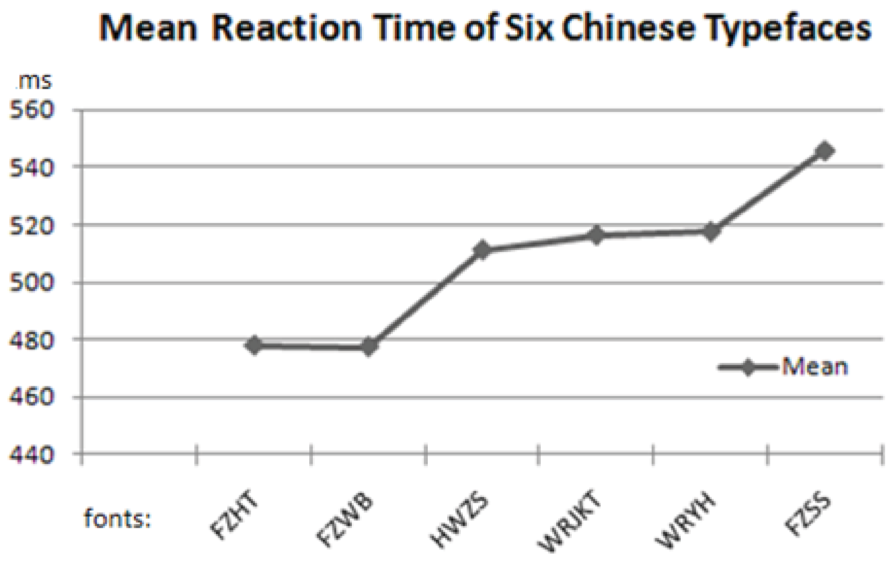



**Figure 13 Mean recognition Times of different fonts. FZHT – 方正黑体, FZWB – 方正魏碑, HWZS – 华文中宋, WRJKT – 微软简楷体, WRYH – 微软雅黑, FZSS – 方正宋三.**

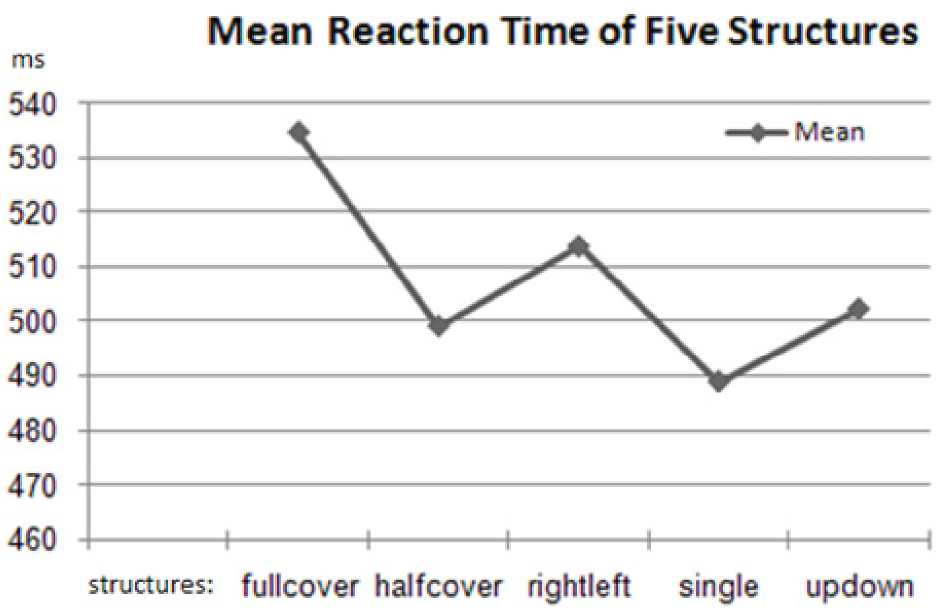



**Figure 14 Mean recognition Times of different structures.**

### 3.4.3 Pairwise Comparisons

A pairwise comparison generally refers to any process of comparing entities in pairs, to judge which of each entity is preferred, or has a greater amount of some quantitative property. The method of pairwise comparison is used in the scientific study of preferences, attitudes, voting systems, social choice, public choice, and multi-agent AI systems. In the psychology literature, it is often referred to as a “paired” comparison [33].

After conducting the within-subjects effects analysis, it is clear that fonts, structures, and fonts with combining structures affect the human recognition speed. However, we are not sure about whether these recognition differences are significant and could be considered as different. In statistics, the data distribution can affect the effectiveness of the comparison result. Also we had to find which font would lead to a higher legibility and which one would have a reduced legibility. Thus, a reasonable and reliable comparison was of huge importance. Therefore, we conducted the following tests to verify the reliability of the difference.

In this part, we conducted two different comparisons. One was based on comparing every pair of fonts, and another one was based on comparing every two different structures. We adopted the MANOVA method to process this comparison.

- Multivariate ANOVA

There are two situations in which a MANOVA is used. The first is when there are several correlated dependent variables, and the researcher desires a single, overall statistical test on this set of variables instead of performing multiple individual tests. The second, and in some cases, the more important purpose is to explore how independent variables influence some patterning of response on the dependent variables. Here, one literally uses an analogue of contrast codes on the dependent variables to test hypotheses about how the independent variables differentially predict the dependent variables. MANOVA also has the same problems of multiple post hoc comparisons as ANOVA. An ANOVA gives one overall test of the equality of means for several groups for a single variable. The ANOVA will not tell you which groups differ from which other group (Of course, with the judicious use of a priori contrast coding, one can overcome this problem). The MANOVA gives one overall test of the equality of mean vectors for several groups. But it cannot tell you which groups differ from which other groups on their mean vectors (As with ANOVA, it is also possible to overcome this problem through the use of a priori contrast coding.) In addition, the MANOVA will not tell you which variables are responsible for the differences in mean vectors. Again, it is possible to overcome this with proper contrast coding for the dependent variables [34].

Simply speaking, Multivariate analysis of variance (MANOVA) is an extension of analysis of variance (ANOVA) methods. MANOVA covers cases where there is more than one dependent variable and where the dependent variables cannot simply be combined. As well as identifying whether changes in the independent variables have a significant effect on the dependent variables, the technique also seeks to identify the interactions among the independent variables and the association between dependent variables.

In comparing different fonts, we calculated the mean speed of single character recognitions and standard error values as the preprocessing firstly. Then we used these data to do the pairwise comparison. The data is shown in Table 11:

| No. | Mean (ms) | Std.Error | 95% Confidence Interval | | Typeface |
|---|---|---|---|---|---|
| | | | Lower Bound | Upper Bound | |
| 1 | 477.7911 | 6.366754 | 464.9755 | 490.6068 | FZHT |
| 2 | 477.3194 | 6.071598 | 465.098 | 489.5409 | FZWB |
| 3 | 511.2592 | 8.362228 | 494.4269 | 528.0915 | HWZS |
| 4 | 516.4523 | 8.127638 | 500.0923 | 532.8124 | WRJKT |
| 5 | 517.5929 | 7.709742 | 502.0739 | 533.1118 | WRYH |
| 6 | 545.9361 | 8.439718 | 528.9478 | 562.9244 | FZSS |

**Table 11 Mean recognition time per character and standard error values of the six remaining fonts.**

Then, we used font pairs to compare, calculate and analyze whether there exists a difference between every instance of two fonts. Then we used the value of Sig. to find out whether the difference between these two fonts was significant, according to the P threshold mentioned earlier. Normally, $P < 0.05$ meant the difference is significant. The detailed information is shown in Table 12:

| (I) | (J) | Mean Difference (I-J) | Std.Error | Sig. | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 0.471702 | 2.3130 | 0.8397 | -4.1961 | 5.1394 |
| | 3 | -33.4681 | 5.5730 | 2.91E-07 | -44.6980 | -22.2382 |
| | 4 | -38.6612 | 5.4516 | 6.64E-09 | -49.6347 | -27.6877 |
| | 5 | -39.8017 | 5.2083 | 1E-09 | -50.2855 | -29.3179 |
| | 6 | -68.1449 | 4.2738 | 2.31E-20 | -76.7476 | -59.5422 |
| 2 | 1 | -0.4717 | 2.3190 | 0.839708 | -5.13945 | 4.1960 |
| | 3 | -33.9398 | 5.3225 | 7.87E-08 | -44.6534 | -23.2261 |
| | 4 | -39.1329 | 5.2921 | 2.34E-09 | -49.7854 | -28.4804 |
| | 5 | -40.2734 | 5.0169 | 2.69E-10 | -50.3719 | -30.1749 |
| | 6 | -68.6166 | 3.9053 | 4.94E-22 | -76.4776 | -60.7556 |
| 3 | 1 | 33.46809 | 5.5790 | 2.91E-07 | 22.23821 | 44.6979 |
| | 2 | 33.93979 | 5.3225 | 7.87E-08 | 23.22613 | 44.6534 |
| | 4 | -5.19311 | 2.5710 | 0.0492 | -10.3683 | -0.0178 |
| | 5 | -6.33362 | 4.2306 | 0.1412 | -14.8494 | 2.1821 |
| | 6 | -34.6769 | 5.2921 | 4.28E-08 | -45.3292 | -24.0244 |
| 4 | 1 | 38.66119 | 5.4516 | 6.64E-09 | 27.68773 | 49.6346 |
| | 2 | 39.13289 | 5.2921 | 2.34E-09 | 28.48041 | 49.7853 |
| | 3 | 5.193106 | 2.5710 | 0.0492 | 0.017882 | 10.3683 |
| | 5 | -1.14051 | 3.3564 | 0.7356 | -7.89652 | 5.6154 |
| | 6 | -29.4837 | 4.9871 | 3.92E-07 | -39.5223 | -19.4451 |
| 5 | 1 | 39.8017 | 5.2083 | 1E-09 | 29.31792 | 50.2854 |
| | 2 | 40.2734 | 5.0170 | 2.69E-10 | 30.17493 | 50.3718 |
| | 3 | 6.333617 | 4.2306 | 0.141198 | -2.18213 | 14.8493 |
| | 4 | 1.140511 | 3.3564 | 0.7356 | -5.61549 | 7.8965 |
| | 6 | -28.3432 | 5.5041 | 5.31E-06 | -39.4225 | -17.2639 |
| 6 | 1 | 68.14494 | 4.2738 | 2.31E-20 | 59.54228 | 76.7475 |
| | 2 | 68.61664 | 3.9053 | 4.94E-22 | 60.75566 | 76.4776 |
| | 3 | 34.67685 | 5.2921 | 4.28E-08 | 24.02449 | 45.3292 |
| | 4 | 29.48374 | 4.9871 | 3.92E-07 | 19.44515 | 39.5223 |
| | 5 | 28.34323 | 5.5041 | 5.31E-06 | 17.26397 | 39.4224 |

**Table 12 Results of the pairwise comparison among the six different typefaces. The mean difference is significant at 0.05 level, and adjustment for multiple comparisons: least significant difference (equivalent to no adjustment).**

Based on the significant main effects and the Pairwise comparison result, we found that for FZHT and FZWB, participants read them faster than any other typefaces. From Figure 13 shown above, there were three recognition speed levels for typefaces. The span of the recognition time between every two levels was obvious. Moreover, any fonts belonging to the same speed level could not be compared by faster and slower recognition speeds, because the significant difference sig>0.05, was equivalent to no difference. In other words, if we do not consider the influence of structures on single character recognition speed, then either FZHT or FZWB will lead to the best legibility for a single character displayed on screens, in the similar size as displayed in this experiment.

In the other study where structures were compared, we also calculated the mean speed and standard error values firstly. Then based on the data, we did the pairwise comparison as well. The processing details were almost like a pairwise comparison of different typefaces. The detailed information is shown in Table 13:

| No. | Mean (ms) | Std.Error | 95% Confidence Interval | | Typeface |
|---|---|---|---|---|---|
| | | | Lower bound | Upper | |
| 1 | 534.6691 | 7.9258 | 518.7152 | 550.6231 | fullcover |
| 2 | 499.1003 | 6.7802 | 485.4524 | 512.7482 | halfcover |
| 3 | 513.7120 | 7.2150 | 499.1889 | 528.2351 | rightleft |
| 4 | 488.9029 | 6.5954 | 475.6270 | 502.1787 | single |
| 5 | 502.2416 | 6.8883 | 488.3760 | 516.1070 | updown |

**Table 13 The mean recognition time per character with one structure and standard error values of the five main structures of characters.**

| (I) | (J) | Mean Difference (I-J) | Std.Error | Sig. | 95% Confidence Interval for Difference | |
|---|---|---|---|---|---|---|
| | | | | | Lower Bound | Upper Bound |
| 1 | 2 | 35.5688 | 2.6354 | 1.3148E-17 | 30.2640 | 40.8736 |
| | 3 | 20.9571 | 2.3874 | 2.16443E-11 | 16.1516 | 25.7626 |
| | 4 | 45.7663 | 3.0691 | 3.08991E-19 | 39.5884 | 51.9441 |
| | 5 | 32.4276 | 2.5873 | 1.94301E-16 | 27.2195 | 37.6357 |
| 2 | 1 | -35.5689 | 2.6354 | 1.3148E-17 | -40.8736 | -30.2641 |
| | 3 | -14.6117 | 1.6220 | 1.01008E-11 | -17.8765 | -11.3469 |
| | 4 | 10.1974 | 1.8925 | 2.36274E-06 | 6.3880 | 14.0068 |
| | 5 | -3.1412 | 1.8112 | 0.0896 | -6.7884 | 0.5060 |
| 3 | 1 | -20.9571 | 2.3874 | 2.16443E-11 | -25.7626 | -16.1516 |
| | 2 | 14.6117 | 1.6212 | 1.01008E-11 | 11.3469 | 17.8765 |
| | 4 | 24.8091 | 2.2569 | 1.84299E-14 | 20.2662 | 29.3520 |
| | 5 | 11.4705 | 2.0923 | 1.71521E-06 | 7.2589 | 15.6820 |
| 4 | 1 | -45.7663 | 3.0691 | 3.08991E-19 | -51.9441 | -39.5885 |
| | 2 | -10.1974 | 1.8925 | 2.36274E-06 | -14.0068 | -6.3881 |
| | 3 | -24.8091 | 2.2569 | 1.84299E-14 | -29.3520 | -20.2662 |
| | 5 | -13.3387 | 2.2355 | 3.24859E-07 | -17.8386 | -8.8388 |
| 5 | 1 | -32.4276 | 2.5873 | 1.94301E-16 | -37.6356 | -27.2195 |
| | 2 | 3.1412 | 1.8119 | 0.0896 | -0.5060 | 6.7885 |
| | 3 | -11.4705 | 2.0922 | 1.71521E-06 | -15.6820 | -7.2590 |
| | 4 | 13.3387 | 2.2355 | 3.24859E-07 | 8.8388 | 17.8386 |

**Table 14 Results of the pairwise comparison among the five main different structures of characters. The mean difference is significant at 0.05 level, and adjustment for multiple comparisons: least significant difference (equivalent to no adjustment).**

According to the data collection and analyses, we found that the single structure led to the fastest speed in recognition. In this structural experiment, there were three recognition speed levels for structures, and the time span between every two levels was obvious. The recognition time of any structure belonging to the same speed level had the least significant differences with each other, and we could not tell which structure may have led to a slightly faster reading speed.

After these two different pairwise comparisons, to better understand the advantages of a specific Chinese font compared with others, we made an interactive analysis. We analyzed the results of the font comparison and the structure comparison interactively, and a crossing comparison result was drawn to show the interaction. The relationship is displayed in Fig. 15, below.

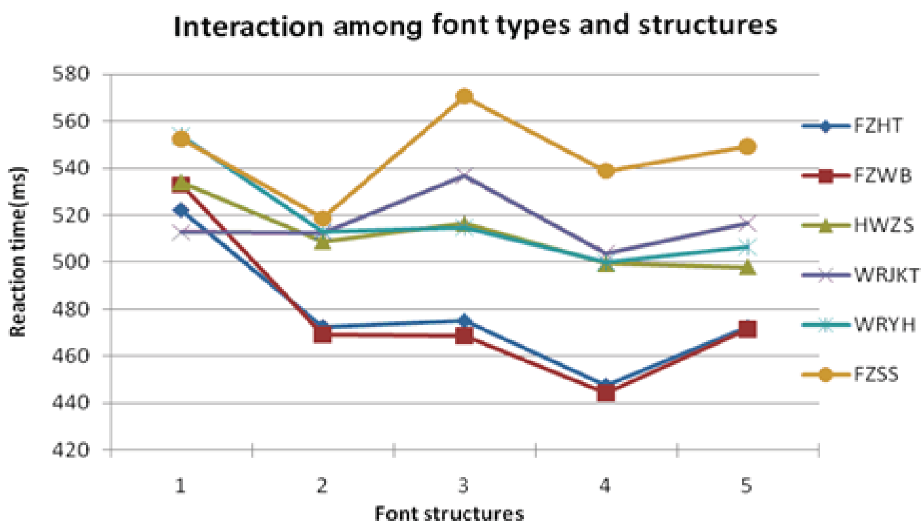



**Figure 15 Crossing analyses among different fonts and different structures. The x-axis represents the different structures, y-axis represents the recognition speed, and the nodes in this figure represent the different typefaces. Font structures representation: 1-fullcover, 2-halfcover, 3-rightleft, 4-single, 5-updown.**

From this figure, we can easily see that FZHT and FZWB performed best among these six different typefaces, though WRJKT performed a little bit better in the full-cover structure. However, this little flaw can be ignored under the outstanding general performance of the FZHT and FZWB.

## 3.5 Discussion

When combining the two results for the two different pairwise comparisons, we can easily draw the conclusion that FZHT and FZWB performed best in terms of the overall situation, even though when they were in full-cover structure, their reading times were a little bit slower than WRJKT.

Because FZHT is a kind of standard sans serif HeiTi typeface, FZHT makes the characters look much clearer when the characters are not big, being displayed on screens. Sans serif fonts focus on single character display rather than on the connections between characters. The WRYH, belongs to the category of HeiTi as well, is a modification of the standard structure of Chinese character. The inner part of WRYH character was much bigger, keeping it clear enough to read when it was displayed on a small size [52]. However, obviously this modification did not catch people’s attention (because it was not ranked in the top five), and it was not superior in recognition speed as well, thus we can deduce that people prefer the ordinary structures of characters. Detailed analysis would be presented in Chapter 4.

SongTi, has the serif characteristics are obvious and a character structure that is similar to the handwritten KaiTi font. KaiTi was the first character model for almost all Chinese children to start learning how to write Chinese characters, and it is considered to be the perfect font for elementary books. Thus, as the branches of SongTi font, the FZSS and HWZS still maintain the SongTi's design style. However, because SongTi emphasizes on the Heng(横) and Shu(竖) contrast in structure, the horizontal line is weakened when it is seen from a distance, which leads to a lower rate of character recognition. Though the hinting, anti-aliasing, and subpixel rendering techniques [53] have partially mitigated the legibility problem of serif fonts displayed on screens, the basic constraint of screen resolution which is typically 100 pixels per inch or less, and the small font size display problem both continue to limit the legibility of fonts on a screen, especially pertaining to fast reading.

Because of the low black to white contrast (34.45%), WRJKT cannot perform as well as the other five fonts. When reading fast, the strokes in WRJKT are overlooked. However, for the two most outstanding typefaces, FZHT and FZWB, their black to white contrasts were both above 45%.

FZWB, a special typeface, is neither a serif nor a sans serif font. It has a high black to white ratio value, which is enough to attract people's eyesight. Moreover, it has enough spaces among its different strokes. Although in this font, characters look short and a little bit fat, both subjective and objective proofs show that this typeface indeed performs well in a single character display. Thus, we can say that FZWB is quite suitable for titles.

Thus, FZHT and FZWB got the highest legibility scores in single character displays. They cause less fatigue and eyestrain relatively, when compared among the four remaining typefaces.

# Chapter 4. Readability Experiment

After analyzing the legibility experimental results, we got the result that the FZHT and FZWB performed best in the single characters display on screens. However, single character recognition cannot satisfy people's reading purpose, because the main goal of reading documents, newspapers, magazines etc. is to comprehend the text content. Thus, in this chapter, we present the readability experiment to verify whether FZHT and FZWB perform well in comprehension or not, and whether any other font would perform better than these two typefaces.

In Section 4.1, we list the fonts that would be used for readability experiment. In Section 4.2, we discuss the detail about the selection of the experimental materials and the experiment implementation. Then, we present the experimental results and analyses in detail in Section 4.3. Lastly, we provid a discussion about this experiment.

## 4.1 Studied Chinese Fonts

The six different Chinese fonts which were used in previous legibility experiment, described in previous chapter, were used in this readability experiment. These six typefaces were the most legible, formal and comfortable fonts that had been filtered out in our previous survey. To clearly see the on screen display performance of sentences and to better understand the difference between legibility and readability, we prepared another Chinese poem that was different from the one in the pangram presented in Chapter 2. This poem was displayed in all six fonts. An example is shown below:

方正黑体字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

方正宋三字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

微软简楷体字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

华文中宋字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

方正魏碑字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

微软雅黑字体显示，春眠不觉晓，处处闻啼鸟，夜来风雨声，花落知多少

**Figure 16 Six Chinese fonts used in this experiment are similar to the samples displayed in Chapter 2, Fig. 10.**

We used short paragraphs to conduct this readability experiment. Readability refers to the comprehension of sentences, as explained at the very beginning of this thesis (Chapter 1 includes the definitions of legibility and readability). Since the focus of this experiment was different from the legibility experiment, and to maintain a high accuracy measurement, we measured the time to read each Single Sentence and conduct the whole paragraph comprehension test (TTRSPC test) in this experiment. (Detailed description of TTRSPC test would be provided in Section 4.2) The purpose of this experiment is to collect information pertaining to the readability of different fonts. Moreover, good experimental materials, as described below in Section 4.2, were needed to increase the effectiveness and accuracy of this test.

## 4.2 Materials

The paragraphs used in the experiment were chosen from the Reading Comprehension exams of the HSK (the Chinese version of TOEFL) test paper, Intermediate level. The reading comprehension section contained reading passages and questions about the passages. The questions were related to the information that was stated or implied in the passage, and there were questions about some of the specific words displayed in the passages. The passages were drawn from different subject matters, including the humanities, social sciences, biological sciences and physical sciences, containing just the general knowledge on those subjects. In order to make the experimental results more accurate, we chose 18 different reading comprehension paragraphs that were in almost the same difficulty level in our experiment. We made every three paragraphs in one Chinese font, and provided two general questions for every paragraph. Based on some preliminary experimental results, the passages were all offered in moderate, neither too difficult nor too easy.

Each paragraph included approximately 165-175 Chinese characters that were distributed into 5 sentences. Two questions came with each paragraph. Participants were instructed to read and analyze each passage a little more carefully before answering the accompanying questions. At last, we measured the font readability through participants' reading times and analyzed the accuracy in their understanding of the short passages. One passage and questions samples are illustrated in Fig. 17.

商标的设计，要注意各民族和地区的不同风俗习惯。中国销往欧洲国家的"大象牌"手电筒销售情况一直不理想，由于欧洲人认为大象呆头呆脑，所以不愿意购买。但如果在印度，情况就不同了，大象在印度人看来是美好的象征，所以情况恰恰相反法国人以孔雀为祸鸟，意大利人忌讳菊花，英国人忌讳用人像作为商品的商标。日本人喜欢樱花，忌讳用荷花作商标图案，美国人喜欢富有生机的图案，如梅花、兰花等。

a, 从这段文字看，意大利人比美国人更讨厌：

A、菊花 B、荷花 C、梅花 D、樱花

b, 根据本文，设计商标时最好是：

A、用花卉作图案 B、用人像作图案

C、注意消费者心理 D、根据设计者喜好

**Figure 17 Sample of one passage and two questions. In this paragraph, the author describes the design of brand, and gives examples that different races and regions have different customs. There are five sentences in this paragraph. Question A asks which flower Italians dislike. Question B asks what the best idea to design brands is. All the question answers can be found in this short paragraph.**

Firstly, we divided these 18 different passages into six sets, where each set contained three passages. Then we assigned one display font to one set of passages. To balance the influence of contents in different paragraphs, we swapped the typefaces with different sets of paragraphs (keeping the content of the paragraph in the set untouched), this meant that our data was expanded into 108 combinations of different fonts with different contents, guaranteeing that all the paragraphs were written in the six different fonts. There was enough variability to evaluate the effect of readability pertaining to different fonts. At last, these 108 combinations were distributed into six groups, and each group included six sets of paragraphs. Every set was displayed in one particular font. Below, we describe the TTRSPC method in detail.

- TTRSPC Method

We programmed a tool to conduct this experiment by using JAVA. By this method, we just focused on the speed of reading paragraphs and the accuracy of understanding them. Therefore, a timer was used to measure the reading time, and we also measured the accuracy of answers to the questions to test the understanding of these 18 paragraphs in 6 different fonts. Firstly, all participants were divided into six groups, and each group had its own test materials. All participants were told that the reading time was being measured, and they had to read as fast as they could while keeping an understanding of the content of the materials. Each paragraph was displayed sentence by sentence, and every sentence was shown separately after the previous sentence disappeared. This strategy forced participants to read each sentence once, with no reading backward, and no whole paragraph review, making the time measurement more accurate and effective. When the reading finished, two questions showed up (refer Fig. 17). Based on the answers to these questions, we could easily deduce how much they understood.

## 4.3 Experimental Results

When the participants finished their experiment, we asked them which paragraph they wanted to cancel. Although the difficulty of the passages was almost similar, because the paragraphs were selected from the intermediate level exam of HSK, there still existed differences, and some topics might not have been familiar with many of the participants.

After statistical analysis, almost all participants said that they would like to cancel the first paragraph because they thought they hardly learned anything at the very beginning. To keep this experiment accurate, we cancelled the first paragraph and counted the others in the last step. After collecting all the data, we analyzed them in terms of speed and comprehension.

We analyze the experimental results for two aspects separately. In Subsection 4.3.1, we provide the detailed analysis of the reading speed aspect. In Subsection 4.3.2, we present the comprehension accuracy analysis in detail.

### 4.3.1 Analysis of reading speed

We found that different people had their own reading patterns, which meant that they had their own reading speeds. If we were to use their reading speeds directly, then huge errors would be caused by different individual reading patterns. Therefore, data normalization was necessary for processing and analysis. Below, we discuss the data normalization, data filtering, data calculation and pairwise comparison separately in detail.

- Data normalization:

Although every participant read paragraphs according to his/her own speed, the percentage of the reading time of one content was at a specific percentage level of the total reading time of all the contents. Thus, we set all the reading times to a reading time percentage for all 18 paragraphs to guarantee that there would be no individual reading factor influence. For instance, if participant A read the first paragraph in 30 s, and he used 500 s to finish reading all the 18 paragraphs, then his percentage of reading time of the first paragraph should be 30/500 = 0.06 = 6%, therefore we would use this 6% to replace his real reading time. We used this method for the subsequent steps of processing. In reality each subject took 411.796 to 821.954 s to finish reading the 18 paragraphs.

- Data filtering:

After analyzing these normalized data, there still was a problem: the data may have contained useless information. Since participants sometimes could not focus 100% on every paragraph without any distractions, and these distractions affected the reading time accuracy directly. Thus we had to filter out the useless data.

We filtered the data using this formula: Data collected from the same paragraph should be distributed as a Gaussian distribution in probability theory, so we should use sig. to help us filter these useless data. The Gaussian distribution is a continuous probability distribution that is often used as a first approximation to describe real-valued random variables that tend to cluster around a single mean value. The graph of the associated probability density function is "bell"-shaped, and is known as the Gaussian function or the bell curve [36]. The formula is shown below:

$$f(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}$$

In a normal distribution, if $x < 3\sigma$, then x should be filtered, because the probability of the values of all $x < 3\sigma$ counted together is smaller than 0.03. This means that these data are definitely far away from the normal data distribution, and should be ignored. In this experiment, we set the threshold at $x<2.5\sigma$ to better filter our data, and to guarantee the effectiveness of the data. In each group of data (10 participants in one group) we calculated and screened the data for each paragraph, and all of these filtered data would be set 0 instead of the previous value. Then, repeated filtering and setting were carried out until no value changed.

- Calculation:

After filtering the data pertaining to each paragraph in each group, we calculated the mean percentage value of normalized reading time value pertaining to each group of paragraphs, and then we listed them, as in table 15 below.

| | WRYH | FZWB | FZSS | FZHT | WRJKT | HWZS |
|---|---|---|---|---|---|---|
| P1 | 0.0733 | 0.0834 | 0.0778 | 0.0894 | 0.0738 | 0.0906 |
| P2 | 0.0604 | 0.0703 | 0.0635 | 0.0697 | 0.0678 | 0.0635 |
| P 3 | 0.0684 | 0.0623 | 0.0619 | 0.0556 | 0.0533 | 0.061 |
| P 4 | 0.0526 | 0.0555 | 0.0591 | 0.0522 | 0.0528 | 0.0501 |
| P 5 | 0.0415 | 0.0492 | 0.0541 | 0.0482 | 0.0458 | 0.0486 |
| P 6 | 0.0569 | 0.0564 | 0.0505 | 0.0552 | 0.0518 | 0.061 |
| P 7 | 0.0718 | 0.0734 | 0.0666 | 0.0656 | 0.0695 | 0.0636 |
| P 8 | 0.0557 | 0.0587 | 0.0562 | 0.0513 | 0.0524 | 0.0526 |
| P 9 | 0.0504 | 0.0534 | 0.0539 | 0.0525 | 0.0541 | 0.0565 |
| P 10 | 0.0594 | 0.0694 | 0.0664 | 0.0622 | 0.0689 | 0.0599 |
| P 11 | 0.0425 | 0.0509 | 0.0472 | 0.043 | 0.0412 | 0.0449 |
| P 12 | 0.0538 | 0.0612 | 0.0622 | 0.0514 | 0.0624 | 0.0585 |
| P 13 | 0.0491 | 0.0455 | 0.0474 | 0.0487 | 0.0426 | 0.0474 |
| P 14 | 0.0519 | 0.052 | 0.0523 | 0.0512 | 0.0521 | 0.0499 |
| P 15 | 0.0506 | 0.0506 | 0.0477 | 0.0476 | 0.0482 | 0.0438 |
| P 16 | 0.047 | 0.0493 | 0.0498 | 0.0446 | 0.0454 | 0.0499 |
| P 17 | 0.0592 | 0.0639 | 0.0511 | 0.058 | 0.0579 | 0.0577 |
| P 18 | 0.0453 | 0.0442 | 0.0463 | 0.0437 | 0.0444 | 0.0462 |
| Average | 0.0539 | 0.0567 | 0.0551 | 0.0530 | 0.0536 | 0.0538 |

**Table 15 Mean values of normalized percentage of reading time. Paragraph 1 data has been ignored. WRYH-微软雅黑, FZWB-方正魏碑, FZSS-方正宋三, FZHT-方正黑体, WRJKT-微软简楷体, HWZS-华文中宋.**

Then, we chose all 18 paragraphs in the same font and calculated the mean reading speed percentage for that font. This percentage covered all 18 paragraphs, and avoided the effect of little difficulty level differences, making the result of the comparison more accurate. The first paragraph's reading speed data was ignored, as mentioned at the beginning of this paragraph. We used the remaining 17 paragraphs for analysis. A comparative result is shown below, where the smaller the value is, the faster it is to read. To display the differences among these six different fonts, we used a chart to show the reading time percentage, the chart is shown in Fig. 18 below:

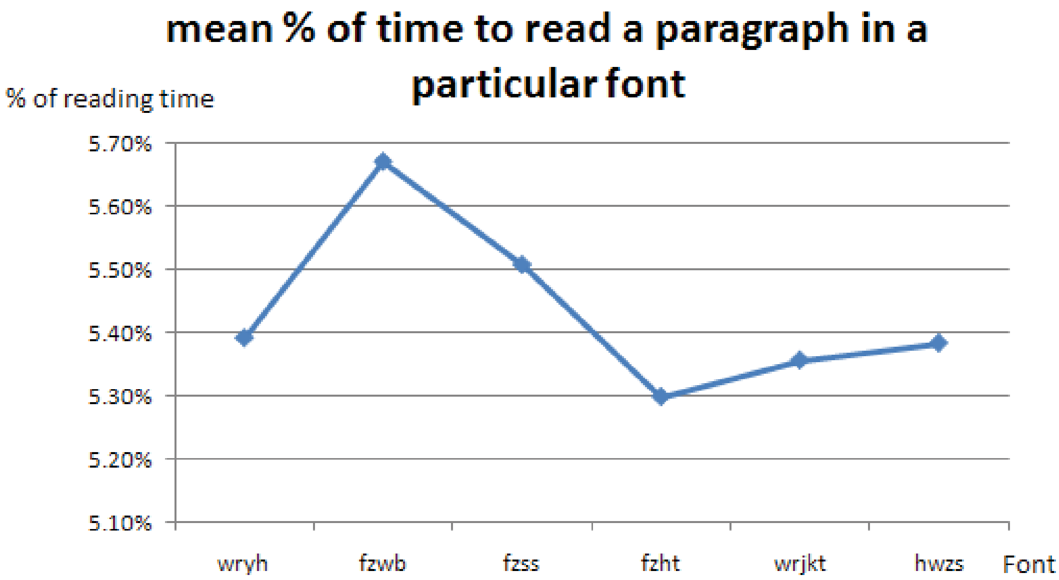



**Figure 18 Distribution of the mean percentages of time to read a paragraph in different fonts, where the vertical ordinate represents the percentage of reading time.**

After obtaining this result, we had to use the multiple comparison method to verify its effectiveness. In statistics, some tiny differences will not affect the conclusion significantly and should be overlooked. Here we used the same method as that in the previous experiment, described in Chapter 3.

- Pairwise Comparison

Here, we still had to make a pairwise comparison to find whether the difference between any two fonts was significant. The reason is the same as the one described in the previous chapter. We used another method to finish this Multivariate ANOVA step, which is called Least Significant Difference method.

Fisher's Least Significant Difference (LSD) procedure is a two-step testing procedure for pairwise comparisons of several treatment groups [37]. In the first step, a global test is performed for the null hypothesis that the expected means of all treatment groups under study are equal. If this global null hypothesis can be rejected at the pre-specified level of significance, then in the second step of the procedure, one is permitted in principle to perform all pairwise comparisons at the same level of significance (although in practice, not all of them may be of primary interest). Fisher's LSD procedure is known to preserve the experimentwise type I error rate at the nominal level of significance, if (and only if) the number of treatment groups is three. The procedure may therefore be applied to phase-III clinical trials when comparing two doses of an active treatment against the placebo in the confirmatory sense (while in this case, no confirmatory comparison has to be performed between the two active treatment groups). The powerful properties of this approach are examined. It has been shown that the power of the first step is a global test, and therefore the power of the overall procedure may be relevantly lower than the power of the pairwise comparison between the more favorably active dose group and the placebo. Achieving a certain overall power for this comparison by applying Fisher's LSD procedure which is irrespective of the effect size of the less favorable dose group, may require slightly larger treatment groups than sizing the study with respect to the simple Bonferroni alpha adjustment. Therefore, if Fisher's LSD procedure is used to avoid an alpha adjustment for phase III clinical trials, then the potential loss of power due to the first step of the global test should be considered at the planning stage [38, 39]. The detailed information about LSD is shown in Table 16 below.

| Multiple Comparisons | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (I) font | (J) font | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
| | | | | | | Lower Bound | Upper Bound |
| LSD | 1 | 2 | -.29* | .112 | .011 | -.52 | -.07 |
| | | 3 | -.12 | .112 | .304 | -.34 | .11 |
| | | 4 | .09 | .112 | .409 | -.13 | .32 |
| | | 5 | .03 | .112 | .757 | -.19 | .26 |
| | | 6 | .01 | .112 | .942 | -.21 | .23 |
| | 2 | 1 | .29* | .112 | .011 | .07 | .52 |
| | | 3 | .18 | .112 | .119 | -.05 | .40 |
| | | 4 | .39* | .112 | .001 | .16 | .61 |
| | | 5 | .33* | .112 | .005 | .10 | .55 |
| | | 6 | .30* | .112 | .009 | .08 | .52 |
| | 3 | 1 | .12 | .112 | .304 | -.11 | .34 |
| | | 2 | -.18 | .112 | .119 | -.40 | .05 |
| | | 4 | .21 | .112 | .066 | -.01 | .43 |
| | | 5 | .15 | .112 | .182 | -.07 | .37 |
| | | 6 | .12 | .112 | .271 | -.10 | .35 |
| | 4 | 1 | -.09 | .112 | .409 | -.32 | .13 |
| | | 2 | -.39* | .112 | .001 | -.61 | -.16 |
| | | 3 | -.21 | .112 | .066 | -.43 | .01 |
| | | 5 | -.06 | .112 | .604 | -.28 | .16 |
| | | 6 | -.08 | .112 | .451 | -.31 | .14 |
| | 5 | 1 | -.03 | .112 | .757 | -.26 | .19 |
| | | 2 | -.33* | .112 | .005 | -.55 | -.10 |
| | | 3 | -.15 | .112 | .182 | -.37 | .07 |
| | | 4 | .06 | .112 | .604 | -.16 | .28 |
| | | 6 | -.03 | .112 | .814 | -.25 | .20 |
| | 6 | 1 | -.01 | .112 | .942 | -.23 | .21 |
| | | 2 | -.30* | .112 | .009 | -.52 | -.08 |
| | | 3 | -.12 | .112 | .271 | -.35 | .10 |
| | | 4 | .08 | .112 | .451 | -.14 | .31 |
| | | 5 | .03 | .112 | .814 | -.20 | .25 |

Based on observed means.

*. The mean difference is significant at the .05 level.

**Table 16 Detailed pairwise comparison result information based on font pairs. Font 1-FZHT, 2-FZWB, 3-HWZS, 4-WRJKT, 5- WRYH, 6- FZSS.**

In Table 16, it is obvious to see that FZHT is a much better font than the others, with a less percentage of reading time of at least 0.058% (refer Table 18 below). Although the differences among these six fonts were so tiny, the longer the text was, the greater the differences would be seen. Therefore time was saved in reading the font FZHT. After the pairwise comparison, we could easily find that FZWB had a significant difference with other fonts (where the 0.1 level should be assigned to a weak significant difference). Thus, FZWB performed the worst in this experiment. We had to do another analysis to figure out whether the comprehension of the paragraph would affect the final result as well.

### 4.3.2 Comprehension Analysis

To compare the differences in comprehension, we used the second part of the experimental result, which was collected from the answers to the two questions of every paragraph. We ignored the reading speed temporarily. Firstly, we calculated the error rate (number of mistakes / total number of questions) of each paragraph in each group (refer the group definition in Subsection 4.2), counting two questions together, because these two questions all responded to the same content. Then, we put the six groups together and chose the 18 paragraphs in the same font to calculate the mean error rate. Then, we ignored the first two questions belonging to paragraph 1 as Section 4.3.1 did, because participants' reading quality could affect the comprehension directly. After comparing average comprehension error rate of the six different fonts, found the differences that are shown in Fig. 19 below. Then, we did the same pairwise comparison as the one described in Section 4.3.1, and the processed result is shown in Table 17 below.

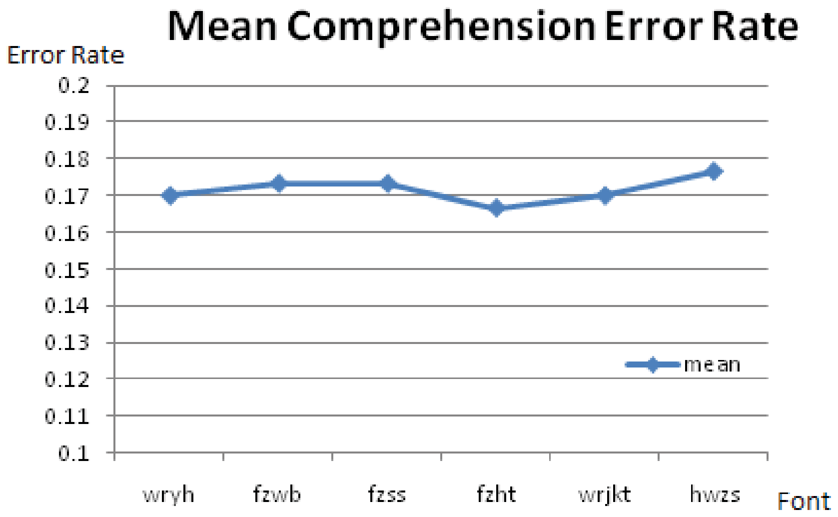



**Figure 19** **Mean comprehension error rate for the six typefaces.**

| Multiple Comparisons | | | | | | | |
|---|---|---|---|---|---|---|---|
| | (I) font | (J) font | Mean Difference (I-J) | Std. Error | Sig. | 95% Confidence Interval | |
| | | | | | | Lower Bound | Upper |
| LSD | 1 | 2 | .00 | .029 | .910 | -.06 | .06 |
| | | 3 | .00 | .029 | .910 | -.06 | .06 |
| | | 4 | .00 | .029 | .910 | -.06 | .06 |
| | | 5 | .00 | .029 | 1.000 | -.06 | .06 |
| | | 6 | -.01 | .029 | .821 | -.07 | .05 |
| | 2 | 1 | .00 | .029 | .910 | -.06 | .06 |
| | | 3 | .00 | .029 | 1.000 | -.06 | .06 |
| | | 4 | .01 | .029 | .821 | -.05 | .07 |
| | | 5 | .00 | .029 | .910 | -.06 | .06 |
| | | 6 | .00 | .029 | .910 | -.06 | .06 |
| | 3 | 1 | .00 | .029 | .910 | -.06 | .06 |
| | | 2 | .00 | .029 | 1.000 | -.06 | .06 |
| | | 4 | .01 | .029 | .821 | -.05 | .07 |
| | | 5 | .00 | .029 | .910 | -.06 | .06 |
| | | 6 | .00 | .029 | .910 | -.06 | .06 |
| | 4 | 1 | .00 | .029 | .910 | -.06 | .06 |
| | | 2 | -.01 | .029 | .821 | -.07 | .05 |
| | | 3 | -.01 | .029 | .821 | -.07 | .05 |
| | | 5 | .00 | .029 | .910 | -.06 | .06 |
| | | 6 | -.01 | .029 | .735 | -.07 | .05 |
| | 5 | 1 | .00 | .029 | 1.000 | -.06 | .06 |
| | | 2 | .00 | .029 | .910 | -.06 | .06 |
| | | 3 | .00 | .029 | .910 | -.06 | .06 |
| | | 4 | .00 | .029 | .910 | -.06 | .06 |
| | | 6 | -.01 | .029 | .821 | -.07 | .05 |
| | 6 | 1 | .01 | .029 | .821 | -.05 | .07 |
| | | 2 | .00 | .029 | .910 | -.06 | .06 |
| | | 3 | .00 | .029 | .910 | -.06 | .06 |
| | | 4 | .01 | .029 | .735 | -.05 | .07 |
| | | 5 | .01 | .029 | .821 | -.05 | .07 |

Based on observed means.
The error term is Mean Square(Error) = .006.

**Table 17 Detailed pairwise comparison result for six typefaces with statistics on the answers to the 18 paragraphs. Font 1-FZHT, 2-FZWB, 3-HWZS, 4-WRJKT, 5-WRYH, 6- FZSS.**

According to the results presented in Table 18, we find that FZHT performs best and leads to the best comprehension, and HWZS has the highest error rate in comprehension. The distance between the best and worst fonts is 1%. However, after a pairwise comparison, we found that the difference between any two fonts was not significant and should be ignored, which means that people were all had a good understanding of the contents while doing this experiment.

| Font | Mean % of Reading Time | Mean Error Rate |
|---|---|---|
| WRYH | 5.39 | 0.17 |
| FZWB | 5.67 | 0.1733 |
| FZSS | 5.51 | 0.1733 |
| FZHT | 5.30 | 0.1666 |
| WRJKT | 5.36 | 0.17 |
| HWZS | 5.38 | 0.1767 |

**Table 18 Combination results including mean percentage of reading time and mean error rate for both reading time and comprehension.**

When combining the two aspects together as Table 18, we found that FZHT got the fastest reading speed with the highest accurate comprehension rate, which should be ignored in the statistical analyses. However, FZWB did not perform as well in both reading speed and comprehension in this readability experiment, as it did in the single character recognition experiment. A detailed analysis about FZWB will be provided in the next section.

## 4.4 Discussion

In this section, we present the comparisons between FZHT and FZWB and between FZHT and WRYH in detail. The purpose of making these comparisons is to extract the features of typefaces which would lead to the higher readability. Then, detailed analyses would be provided as well.

### 4.4.1 FZHT vs. FZWB

In the single character recognition experiment, we had two best fonts that produced the fastest recognition response: FZHT and FZWB. However, after the reading speed and comprehension experiment, these two different fonts produced totally different results. For instance, FZHT still brought the fastest reading speed and the highest comprehension, while FZWB had the slowest reading speed and the second to last position in comprehension. Thus, we can deduce that the factor making the worse performance of FZWB in readability experiment is related to the spaces between every two characters rather than the design of single characters, because there was no difference in the performance of these two typefaces (FZHT and FZWB) in the single character recognition experiment. To verify the accuracy of this deduction, we provided a method to measure the spaces between characters.

- Space measurement between characters

We measured the default space between every pair of characters of all these six fonts, and measured their character widths and heights as well by the projection method. Chinese printing characters can be segmented easily without any complicated segmentation method, according to the neat character spaces between them.

The projection method used the 1-D signal to represent the 2-D signals. This method counts the number of black pixels in each row and each column.

To measure the spaces, widths and heights of Chinese characters in different typefaces, we used the full cover structures, for the characters with this structure could cover all the main features of Chinese characters, while maintaining the font's original design. The character boundary was just what we needed to measure. Because the heights and widths of Chinese characters were not similar, and the spaces between every two characters were different as well, the space value itself would not reflect any useful information. Therefore, we chose the character to space ratio to represent the differences in space. The formula for this is described below:

$$\text{space distance} = {}^{\text{character width}}\!/_{\text{character space}}$$

By applying this equation, we could see how big the space was by comparing the space distance. The bigger the distance was, the smaller the space would be. We also found the differences in space of different typefaces listed below in Table 19. As the table shows, FZWB has the biggest default space between two characters, and WRJKT gets the second place. WRYH characters have the nearest neighbors. This means that the space distance value of WRYH is the smallest. By now, we are sure that there is an influence on character space according to the performance of FZHT and FZWB in both legibility and readability experiments. That is because in the single character displays, they both performed best.

| Font | Height (mm) | Width (mm) | character space (mm) | character area ($mm^2$) | space area ($mm^2$) | character /space |
|---|---|---|---|---|---|---|
| WRJKT | 26.10 | 21.80 | 10.27 | 568.98 | 268.05 | 2.12 |
| HWZS | 27.33 | 25.08 | 6.55 | 685.44 | 179.01 | 3.83 |
| WRYH | 28.00 | 27.17 | 4.64 | 760.76 | 129.92 | 5.86 |
| FZSS | 27.25 | 24.58 | 5.73 | 669.81 | 156.14 | 4.29 |
| FZWB | 22.58 | 20.00 | 11.82 | 451.60 | 266.90 | 1.69 |
| FZHT | 26.67 | 24.92 | 6.82 | 664.62 | 181.89 | 3.65 |

**Table 19 Different statistics related to typeface space between character pairs.**

### 4.4.2 FZHT vs. WRYH

Both FZHT and WRYH belong to the same category of Chinese fonts, the HeiTi category, and both of them are produced by the same company, Founder Co. As the branches of HeiTi, both FZHT and WRYH have to keep the HeiTi features, which means that characters displayed in these two typefaces all have relatively similar stroke designs. However, they do not have the tail decorations which are the most obvious feature of the serif design. Normally, different typefaces from the same category look more or less the same, with tiny difference at the peripheral spots, just like HWZS and FZSS, which both belong to the SongTi category, a Serif font. We can see in Fig. 20 below the detailed difference between FZHT and WRYH. These differences will be shown by the method described later in this chapter.

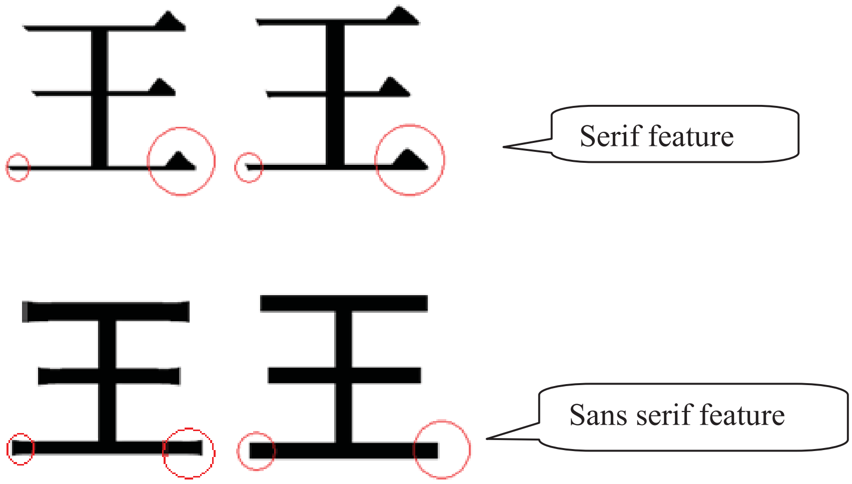



**Figure 20 Differences between Serif font and Sans serif fonts, with peripheral serif differences between FZSS (top left) and HWZS (top right), and sans serif differences between FZHT (bottom left) and WRYH (bottom right). Red marked circles refer to these characters.**

However, there is a design difference between WRYH and FZHT. The purpose of the WRYH design is to make the character clear when being displayed in a small size. Thus, the designer has modified the structure of Chinese characters while keeping the HeiTi features. We compared the differences between WRYH and FZHT with analyses for hundreds times, we found that the structures of half-cover and full-cover which were included in the characters in WRYH had been changed, enlarging the space of the inner parts, while keeping the other parts almost untouched. Below, we present a method to verify the finds in detail.

Firstly, we used two groups consisting of the same characters which included more than 100 different characters. The characters in both groups covered the 5 principal structures that were used in Chapter 3, not only singly including the different 5 structures within characters, but with different structure combinations in one single character as well.

Then, we normalized the same characters to the same height, without touching the design of these fonts, and marked these pairs of characters with different colors, such as WRYH with black and FZHT with red. Then, we aligned these pairs from the very left vertical column, overlapping them, and moved the slimmer character to the right side until its right side was aligned. This way, we could easily find the difference between the same characters in different fonts along with the moving character. A detailed comparison is shown in Table 20 below.

We found that the inner area of WRYH was much more spacious than that of FZHT (partial samples are shown in Table 20). The differences are especially obvious for the half-cover and full cover structures inside the inner space. These differences make the character look a little bit wider, and display better in a small size. That is because they make the strokes look more separate, avoiding the connected or overlapping strokes caused by displaying them in a small size. However, on a normal tablets or e-book reader screens, fonts are displayed with a size of 10 pt. to 12 pt., WRYH's advantage of design does not seemed to make it readable, and this special design may not be totally accepted by the public as well.

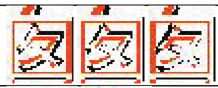
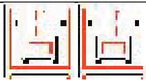
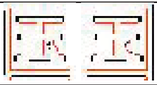
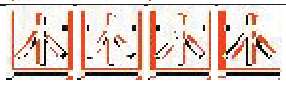
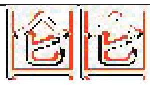
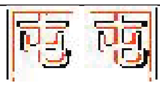
| Original WRYH | Original FZHT | Difference |
|---|---|---|
| 图 | 图 |  |
| 固 | 固 |  |
| 国 | 国 |  |
| 囫 | 囫 |  |
| 圜 | 圜 |  |
| 圈 | 圈 |  |
| 困 | 困 |  |
| 囵 | 囵 |  |
| 囤 | 囤 |  |
| 圆 | 圆 |  |

**Table 20 Detailed character comparison between FZHT and WRYH is applied by character overlapping method. All samples are with full-cover structure.**

# Chapter 5. Simulation and Verification of Machine OCR

To find the connection between machine OCR and human recognition, we have to simulate the process of human recognition. This includes test data preparation, preprocessing, feature extraction, similarity comparison, and sentence comprehension. Simulation processing is necessary to realize the goal of this chapter. By this processing, we can find out which fonts people think are more legible or readable, also lead to a more precise result in machine recognition. Moreover, we can also find out which typefaces are good for displaying on tablets and e-book readers, would be recognized well by current OCR tools.

We conduct this simulation processing in two different ways. One is to simulate the single character recognition progress, which is described in Section 5.1. The other one is to refer the practical content recognition in human's daily lives, which is presented in Section 5.2. Then, a detailed discuss will be provided in Section 5.3.

## 5.1 Simulation of the Single Character Recognition

In this section, we focus on the simulation of human recognition of single characters and search for the relation between human recognition and machine OCR. In Subsection 5.1.1, we present the test data selections. In Subsection 5.1.2, we present the necessary data preprocessing methods to process our test data. Then in Subsection 5.1.3, feature extraction methods are provided. Lastly, the Euclidean distance would be used to measure the similarity between Chinese characters in Subsection 5.1.4.

### 5.1.1 Test data

We used the entire 3755 Chinese characters of the first level national standard characters in GB2312, which is a database which represented the Chinese national standard of simple characters. This database was created and published by the Chinese state administration of standards in May 1st, 1981. The 3755 first level national standard characters were filtered by usage frequency and covered more than 99.8% Chinese character usage in people's daily lives in China. [40]

Then, we prepared six sets of characters for these six typefaces, same as the ones used in Chapters 3 and 4, which included all the 3755 characters, and we used the projection method to extract every character from texts.

- **Projection method**

This method is used to segment characters from a horizontal text line or a vertical text line. Character spaces are cut by projecting the character into the horizontal or vertical direction. The main idea of projection is to count the number of black pixels in each row as we calculate the horizontal projection, and count the number of black pixels in each column as we calculate the vertical projection. Thus, by this method, we can easily find the space between pair of characters.

After character segmentation, we processed every character and put them into a white box of the same size. This process allows them to keep their own character structures without destroying their original designs inside the box, because Chinese characters may not have the same heights and widths. Thus, we used the 128 *128 pixels empty canvas which was bigger than the largest width and height of all the 3755 characters as the white box and put the character into it, making the centre points of character and box overlapped. To realize this center points overlapping, we have to find out the white box and character centers which are both located at the position (height/2, width/2). Then, we move the character in the white box to make these center points matched.

### 5.1.2 Data preprocessing

After data preparation, to test whether or not the connection between human recognition and machine OCR exists and whether our hypothesis matches our expectations or not as well, we followed the character recognition processing steps by binarization, image smoothing, normalization, gradient feature, junction points and end points extraction, zooming and feature vector establishment methods.

#### 5.1.2.1 Binarization

Because our character data was selected from paper-based materials, and after scanning the materials, the black and white characters were changed to RGB colorized or grayscale version. For better processing, we cast them into binary colors. Actually, a binary array is a good choice for storing all the pixels of an image, where “0” and “1” are used instead of white and black pixels. This step makes the original image much easier to process, because it ignored all the colors or grey levels.

#### 5.1.2.2 Image Smoothing

There are two steps in smoothing images. The first one is called filling, while the second one is called removal. These two steps make the image much smoother than the original one because they remove the isolated pixels and noises, and fill the potential missing pixels back to their proper positions [41]. Below we introduce the algorithms of filling and removal processing.

- Filling:

Condition: Xo must be equal to zero (white).

| | | |
|---|---|---|
| X1 | X2 | X3 |
| X4 | X0 | X5 |
| X6 | X7 | X8 |

**Figure 21 The 3x3 mask used in image smoothing, where X0 is the centre pixel, and X1-X8 are the surrounding pixels.**

The algorithm for Filling processing described in Fig.21:

$$X_0 = X_0 \text{ U } (X_2 \cap X_7 \cap (X_4 \text{ U } X_5)) \text{ U } ((X_2 \text{ U } X_7) \cap X_4 \cap X_5)$$

Explanation: Filling makes the image much smoother by adding the possible missing pixels back to the image.

- Removal:

Condition: $X_0$ must be equal to 1 (black); (Same figure 20 )

The algorithm for removal is :

$$X_0 = X_0 \cap [((X_1 \text{ U } X_2 \text{ U } X_4) \cap (X_5 \text{ U } X_7 \text{ U } X_8)) \text{ U } ((X_4 \text{ U } X_6 \text{ U } X_7) \cap (X_2 \text{ U } X_3 \text{U } X_5))]$$

Explanation: Removal can delete the isolated pixels from the image to make it much easier to recognize. This is helpful in reducing noises.

Before processing the original image, it is necessary to add a new pad with value of “0” to avoid the exception of “out of image boundary”. The smoothing step should be done several times, to remove any the remnant pixels.

#### 5.1.2.3 Normalization

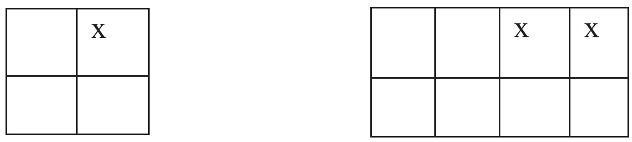
Normalization is one of the most important preprocessing steps for character recognition. Normally the character image is normalized to a standard size in order to be classified easily. The purpose of normalization is to reduce the noise and make the feature extraction easier. It also improves the accuracy of classification in order to get as close to perfect as possible [41].

Because the characters were selected from documents used in real life without decoration and processing, the original design of the fonts remained intact. Thus after character segmentation, different characters might have different heights and widths, even if they had the same structures. Thus, we measured the largest value of height and width among all the characters in same font, and then made a box of size maxheight+4 by maxwidth+4 pixels to contain characters. Then, we normalized this original image into smaller dimensions of size 64x 64 pixels, because if the image is too small, the feature information in each zone will not be clear enough.

Below, we introduced the normalization algorithm, followed by the example in Figure 22(a and b).

- Algorithm:
  - $X_{old} = X_{new} * row_{old} / row_{new}$
  - $Y_{old} = Y_{new} * column_{old} / column_{new}$

  Example:




**a** **b**

**Figure 22 In this example, an image of 2 by 2 pixels is normalized to 4 by 2 pixels.**

### 5.1.3 Feature Extraction

Each character or digit has many different kinds of features, and is just like a vector of features. Each feature helps people to identify the character to some extent. For feature extraction, the general purpose is to extract as few features as possible, while maximizing the accuracy of the recognition. In this section, we presented several techniques to extract useful character features, including gradient feature (Subsection 5.3.1), junction point and endpoint feature (Subsection 5.3.2), and the relative algorithms would be covered as well. The purpose of using gradient features is that gradient features are extracted from the character shapes, which is just like the human integral recognition of characters. Using endpoints and junction-points features is because these features all the local structural features, people also use this way to memorize characters [9].

#### 5.1.3.1 Gradient Features

The gradient feature is measured by the magnitude and direction of the greatest change in intensity in a small neighborhood of each pixel. Gradient features could be extracted according to several different models, and in this thesis, Sobel Operator [50] is applied. The Sobel templates compute the horizontal (X) and vertical (Y) components of the gradient as shown below.

- Sobel Masks:

| | | |
|---|---|---|
| -1 | 0 | 1 |
| -2 | 0 | 2 |
| -1 | 0 | 1 |

X

| | | |
|---|---|---|
| 1 | 2 | 1 |
| 0 | 0 | 0 |
| -1 | -2 | -1 |

Y

**Figure 23 Sobel operator templates used for convolution.**

Compared with the results of extracting several kinds of gradient feature, we used the original input image without skeletonization and contour processes, in order to preserve more information. Below, we present the algorithm of using Sobel masks to extract gradient features in detail [51].

Given an input image I of size $D_1 * D_2$, each pixel neighborhood is convolved with the templates in Figure 22 to determine the X and Y components, $S_x$ and $S_y$, respectively. That is:

$$S_x (i, j) = I (i-1, j+1) + 2*I (i, j+1) + I (i+1, j+1) – I (i-1, j-1) – 2*I (i, j-1) – I (i+1, j-1);$$

$$S_y (i, j) = I (i-1, j-1) + 2*I (i-1, j) + I (i-1, j+1) – I (i+1, j-1) -2*I(i+1, j) – I (i+1, j+1);$$

Here, i and j are the given range over the image rows ($D_1$) and columns ($D_2$), respectively. The gradient magnitude is then calculated as:

$$r(i,j) = \sqrt{S_x^2(i,j) + S_y^2(i,j)}$$

The gradient direction is calculated as:

$$\Theta(i,j) = tan^{-1}\frac{S_y(i,j)}{S_x(i,j)}$$

The adaptive gradient threshold is calculated as:

$$\bar{r} = \sum_{i,j}\frac{r(i,j)}{D1 * D2}$$

The complete gradient map is rich in information. However, some processing of the gradient map is necessary to highlight the important information for the purpose of character recognition. The threshold is computed as the average gradient magnitude over the whole character image gradient map, and is used to filter out the spurious responses to the Sobel operator.

This process made a threshold for gradient map of the image. In this process, a threshold was applied to nullify pixels whose gradient magnitude value lies below the computed threshold, as follows:

if r(i, j)≥ $\bar{r}(i,j)$, r'(i, j) = r(i, j), and $\Theta'(i,j) = \Theta(i,j)$. Else if r(i, j)≤ $\bar{r}(i,j)$, r'(i, j) = 0, and $\Theta'(i,j)$ =NaN.

After filtering by threshold, we divided the 360 degrees of gradient directions into 16 bins, which meant that each direction contains 24.5 degrees. Then, we mapped the new gradient directions into the 16 bins for later processing. To better explain this process, we prepared an example of gradient directions mapping. Examples shown below in Fig. 24:

| | | |
|---|---|---|
| 32 | 43 | 305 |
| 180 | 55 | 2 |
| 358 | 280 | 200 |

**(a) Gradient direction value**

| | | |
|---|---|---|
| 2 | 2 | 14 |
| 8 | 3 | 1 |
| 16 | 12 | 9 |

**(b) Gradient direction bins**

**Figure 24 The example of gradient direction mapping: (a) the original gradient direction values, and (b) the direction bin values matching the direction value of the same position.**

- Zoning:

After finishing the gradient direction mapping processing, we separated the image into 4*4 zones for extracting the features in each zone. For each zone, we summed up the gradient magnitudes that were in the same direction bin as the gradient feature of this zone. Thus, for each zone, the dimension for the feature was 4*4*16 = 256. For comparison, we used the mean gradient magnitudes of each zone as well that were equal to (the gradient magnitude / number) of pixels in this zone.

#### 5.1.3.2 Endpoint and Junction-point features

To extract the endpoint features and junction-point features, the local structural features, we have to know the definite numbers of these points. Thus, the skeletonization processing plays an important role at this moment to solve this problem.

- Skeleton features:

Skeletonization is one of the most important steps for feature extraction, since a skeleton shows the general shape of a pattern, and many features can be extracted from it. For example, we can extract: the junction points, number of strokes, relative position and so on. The skeleton of a pattern is only one pixel wide, and yet it still keeps the structure of the original pattern.

The Zhang - Suen algorithm is used for shape thinning of the image. This algorithm has two steps that can be applied on the image, one by one, until there is no change in the image. This is a parallel algorithm which is more efficient than a single direction method. The algorithm is described below:

Firstly we define that:

A (p) = the number pertaining to times of transitions from 0 to 1 in the ordered sequence of $p_1$ to $p_8$.

B (p) = number pertaining to the sum of the black pixels that are the 8 neighbors of point p: $(p_1+p_2+p_3+p_4+p_5+p_6+p_7+p_8)$

The positions of P, $P_1$ to $P_8$ are shown in Figure 25.

Prerequisite: This algorithm is used on the black pixels position.

Then we can apply this algorithm in two steps, described follow:

- Step 1:

  Condition 1: 2<= A(p) <=6;

  Condition 2: A(p) = 1;

  Condition 3: $p_1 * p_3 * p_7 = 0$;

  Condition 4: $p_1 * p_5 * p_7 = 0$;

The point p that satisfies all the conditions above will be flagged, and after applying Step 1 on all the black pixels of the image, these flagged points will be removed from the image.

- Step 2:

  Condition 1: 2<= A(p) <=6;

  Condition 2: A(p) = 1;

  Condition 3: $p_1 * p_3 * p_5 = 0$;

  Condition 4: $p_3 * p_5 * p_7 = 0$;

After applying Step 1 to all the pixels of the image, Step 2 starts to be applied in the same way as Step 1 works. Step 1 and Step 2 are applied iteratively, until there are no pixel changes in the image. The result is the skeleton shape of the image.

| | | |
|---|---|---|
| p4 | p3 | p2 |
| p5 | p | p1 |
| p6 | p7 | p8 |

**Figure 25　　Example of Pixel positions.**

- Endpoint feature:

  If the central foreground pixel has only one pixel neighbor around it, then it is an end point.

| | | |
|---|---|---|
| | | x |
| | p | |
| | | |

**Figure 26　　Endpoint position.**

However, just using an endpoint number without knowing the endpoint position is not enough and clear for recognition. Thus, we separated the image into 3 zones horizontally and 3 zones vertically:

| | | |
|---|---|---|
| Left | Middle | right |

| |
|---|
| up |
| middle |
| down |

**(a) Horizontal** **(b) Vertical**

**Figure 27 Horizontal and vertical zones of image.**

➢ Equation:

Endpoint = {H_left, H_middle, H_right, V_up, V_middle, V_down };

Dimensions of the endpoint is 6.

- Junction-point feature:

  If there are more than three neighbors around the central point and these neighbors are not connected, then this central point is the junction point. See Fig. 25.

| | | |
|---|---|---|
| | | x |
| x | p | |
| | x | |

**Figure 28 Position of the junction points.**

After feature extraction, we built a feature vector that included all the features collected from gradient feature extraction, endpoint features and junction point feature extraction.

The fomula for this vector is: Vec = {gradient_f, endpoint_f, junction_f}. The dimension of this feature vector is 263 elements.

### 5.1.4 Similarity Comparison by Euclidean Distance

Although we cannot get the core algorithms of the classifiers of HanWang OCR and Wintone OCR, we can use the base of classification algorithms that is named as the Euclidean distance to measure the similarities roughly. Euclidean distance is used to measure the distance of two points spatially. In the theory of classification, if the distance between character A and character B is smaller than that between character A and character C, then character A should be assigned to the same side with character B. In this way, we can use every character to be compared with all the other 3754 characters of the same font to measure the distance. Then we may find out how many character pairs fall below the distance threshold. These pairs that are below the threshold would create troubles for the machines to recognize them.

- The formula for calculating the Euclidean distance between two characters is described below:

  ρ(**A**，**B**) = sqrt [ ∑( a[i] - b[i] )^2 ] (i = 1，2，…，n)

Wintone OCR and HanWang OCR are the most popular and best sold OCR tools in mainland China. The companies have claimed that their products all have a precise recognition result above 99%. However, through our pangram recognition test, the best recognition accuracy was not as good as they stated. (Their recognition result is shown below) Thus, we used the best actual result of pangram recognition as a threshold to compare the similarity of character pairs (refer the pangram result in Table 22).

According to the recognition result, the misrecognition rate of 2% of current OCR software can be used as a threshold to filter the data approximately. We chose the threshold value in the FZHT result because this was consistent with best recognition result both by Wintone OCR and HanWang OCR. After filtering the first 2%, we rounded the value to the nearest integer larger than 10. Then, we set this value as the threshold to count the pair numbers which had the smaller distance than the threshold for all the six fonts. The results are shown below (Table 21).

There are 3755 characters in each set, and we need to measure the distance between every pair of characters. Thus, we have to measure (3755 * 3755 -3755) / 2 = 7048135 distinct pairs. Because the distance between character A and character B is equal to the distance between character B and character A, and the distance between character A and character A is equal to 0.

Euclidean distance measurement results:

| Typeface | Number of pairs under the threshold of 10 | Total pairs |
|---|---|---|
| FZHT | 154444 | 7048135 |
| FZSS | 255209 | |
| FZWB | 193612 | |
| HWZS | 164732 | |
| WRJKT | 320103 | |
| WRYH | 713468 | |

**Table 21 Results of these six different typefaces for the Euclidean distance measurement under threshold ten.**

From this result, we can see that the number of pairs under the threshold of 10 of FZHT is less than the other five typefaces. Theoretically speaking, FZHT has less potential for mismatched character pairs than the rest of the five typefaces. This means documents printed in FZHT could be recognized with a higher accuracy than that in the other five typefaces. A detailed analysis will be provided in the following discussion part.

## 5.2 Sentence Recognition by Machine OCR

To test whether the fonts we have filtered out were fit for current OCR tool processing and can be recognized with high accuracy or not, we had to simulate the human real-time recognition progress by using OCR tools other than the single character theoretical simulation described in the section above. As we all know, human recognition is to recognize characters in the sentences then process the characters' meaning in their minds rather than to recognize the single characters separately. The OCR tools are used to process text recognition as the real people's reading way as well. Thus it is a good way to test the performance of our recommended typefaces.

Besides the pangram we used in Chapter 2 to search for the typefaces people preferred, we prepared another test materials that were made with 800 characters, selected from "Hanzi Frequency statistics"[48], including 500 high frequency characters, 200 normal frequency characters, and 100 low frequency characters. When combined with the pangram recognition result, we finally analyzed the sentence recognition result.

We used the two best famous Chinese OCR tools as previously mentioned, Wintone OCR and HanWang OCR, to complete this step. The recognition results are shown below.

| Typeface | HanWang OCR's Error Rate (%) | | | Wintone TH-OCR's Error Rate (%) | | |
|---|---|---|---|---|---|---|
| | Pangram | 700 characters | 100 characters | Pangram | 700 characters | 100 characters |
| WRJKT | 7.02% | 5% | 15% | 1.754% | 2.14% | 73% |
| HWZS | 2.32% | 1.86% | 5% | 1.754% | 0.71% | 73% |
| WRYH | 0.877% | 0.43% | 2% | 1.754% | 0.57% | 75% |
| FZSS | 18.42% | 11.57% | 14% | 1.754% | 0.57% | 72% |
| FZWB | 0.526% | 0.71% | 1% | 13.16% | 13.29% | 82% |
| FZHT | 0.877% | 0.43% | 3% | 1.754% | 0.43% | 75% |

**Table 22 The two different machine OCR recognition results. The Wintone OCR only included a few low frequency characters in its training data. Thus it is not fair to calculate the error rate for all the characters together to compare the recognition results made by the two different OCR tools. Thus, we use a different way to display the results of the 800 characters test. The number displayed in the column of 700 characters is the result of the first 700 characters' recognition result, including the high frequency and normal frequency characters. The number displayed in the column of 100 characters is the recognition result of the 100 least frequently used characters.**

From Table 22, we found that FZHT and WRYH were all fit for current OCR tools, with the lowest error rates, and these fonts could be used directly without any trouble of selecting OCR tools. From these results, we can also see that the training database of Wintone OCR did not contain all the least frequently used Chinese characters (see the recognition result of the last 100 characters). Moreover, character segmentation would affect the recognition results seriously. For example, as Fig. 29 provided, almost all the characters in the red square were misrecognized, because the area segmentation was incorrect, which would directly affect line segmentation and single character segmentation.

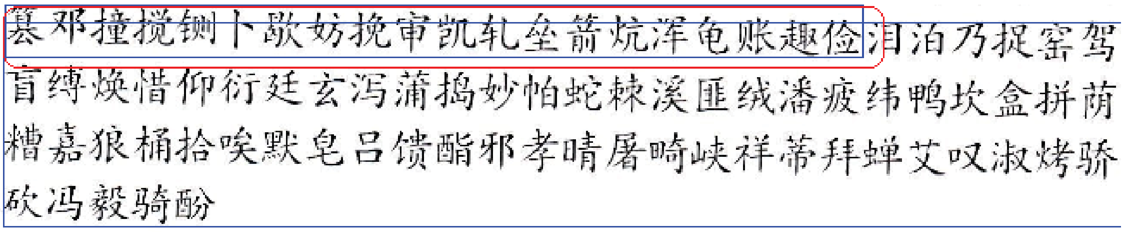



**Figure 29 Segmentation example of the character recognition processed by Wintone OCR. The red square covers the missegmented characters.**

## 5.3 Discussion

The recognition result of FZHT in Section 5.1 is same as the result in single character legibility experiment in Section 5.4. In the legibility experiment, SongTi (HWZS and FZSS) performed not as good as in the similarity comparison because the vertical and horizontal stroke contrast of SongTi did not affect machines as it affected humans. FZHT still performed best among these six typefaces. However, the performances of WRYH in the two different tests described in this chapter were not consistent. It is reasonable to have differences between theoretical deduction and practical products' verification, because as a mature OCR software, it may has extracted many other features and used more complicated methods to improve the accuracy of classification.

Moreover, according to the practical recognition experiment result, we can see that the space between pairs of characters is very important, not only for human recognition, but also for machine OCR segmentation.

# Chapter 6. Conclusion

In this chapter, we summarize the contributions of this thesis and present some future research ideas.

## 6.1 Summary

In this thesis, we have made an attempt to discover the legibility and readability of Chinese digital fonts used for reading digital documents and information from tablets and e-book readers. We used one acceptable and prevailing survey to reduce the amount of the potential research typefaces from 18 to 6. Then we applied two systematic, experiments to analyze, compare and select the best typeface. We also made a practical comparison between machine OCR and human recognition from two different aspects: single character recognition and sentence recognition. We concentrated on helping current tablet producers to find the most suitable display typeface for their products from the existing Chinese fonts, and then these producers might get more market share in the competition for Chinese digital tablets and e-book readers by following our research. Moreover, we hope that this study will incite more research in this area so that one day researchers will be able to identify the most legible digital typefaces for both human and machine reading.

By applying the Chinese typeface preference survey that was designed by a professional psychology method, we filtered out the unacceptable, less legible, less comfortable and less formal fonts, and kept the six remaining popular and well accepted fonts: FZHT, WRYH, HWZS, FZSS, WRJKT, and FZWB.

After conducting the legibility and readability experiments, designed by the combination of pattern recognition, data mining, and psychology, we found that FZHT performed best over the other typefaces, without any doubts. Moreover, we discussed the difference between HeiTi and SongTi fonts, the representations of Sans serif typeface and Serif typeface separately. Then we found that the paper based advantage of SongTi font, a contrast between the Heng (horizontal stroke) and the Shu (vertical stroke), played a role of excess baggage on screen display performance. We also made a comparison between FZHT and WRYH, all belonging to HeiTi category. We found that although the HeiTi performed better than other categories pertaining to Chinese typefaces, the structure and design details can affect the legibility and readability of any particular font in the same category, for example, the unsatisfied performance of WRYH comparing with FZHT. Besides, through the black and white contrast test, the WRJKT font was filtered out, because of its relative lower contrast between black and white. When people are reading quickly on tablets or e-book devices, strokes could be overlooked, leading to character misrecognition, causing more eye fatigue and increasing rereading frequency.

Based on the above aspects, we deduced the following rules for selecting a good font to display on tablet and e-book readers:

- To maintain human reading and writing habits as much as possible. (See the difference between FZHT and WRYH)
- Serif typefaces in Chinese are not fit for screen displays. The paper-based advantage of serif fonts are the biggest burden in reading on screen. (See the analysis of SongTi in Chapter 3)

- A good threshold of black and white contrast could reduce the recognition mistakes when reading fast. To choose this threshold, we can use the SCFRT method to find a good parameter.
- The space between pairs of characters is very important for human reading. Unsuitable space would affect the readability of the paragraphs. (See the comparison between FZHT and FZWB)

## 6.2 Future work

We have finished the research on the legibility and readability of Chinese fonts displayed on tablet and e-book readers. We retain one Chinese typeface (FZHT) from the currently existing Chinese fonts to be fit for screen displays with good performance. The results of our thesis could be helpful for these tablet and e-book device manufacturers to improve the legibility and readability of their new products. Also our study offers a systematic method of selecting and analyzing fonts to be displayed on a screen. However, this is just an initial step and is far from finished.

Currently, filtering out one or several typefaces from current existing Chinese fonts for screen displays can cause expediency but may not be a permanent solution for the tablet and e-book reader manufacturers. Moreover, due to the limits of our experimental methods in our study, there were some unaddressed issues that may have influenced the participants' responses. For example, the participants' age ranges, the participantal knowledge, their familiarity with the studied typefaces, etc. In the future, we can look at how to design and create a new typeface that is specifically fit for tablet and e-book devices reading, and make it as the main target and mission. Meanwhile, all the factors mentioned above should be taken into consideration and investigated.

Thus, firstly the possible features which may or may not improve the legibility and readability of Chinese characters should be extracted separately from different font categories and typefaces. One or more detailed and feasible comparisons and analyses have to be done in order to find what can improve or attenuate the legibility and readability of typefaces. Certainly, if it is possible, in the future we wish to cover as many Chinese typefaces as possible, including all the mainstream categories and non-mainstream calligraphic fonts. This would make a more complete comparison by keeping any important, useful and valuable information hidden in the gaps of Chinese characters.

We could also attempt to make characters more decorative by combining the features that may be extracted as mentioned above, basing on the most suitable, original unchanged FZHT font, including the positive and negative improved features as well. After that, we could filter the candidates with the characteristic features of FZHTs, and eliminate the unreasonable and unfeasible combinations by comparing these fonts with the original FZHT through a public survey displayed on screens. This survey would verify the validity, correctness and feasibility of improving character legibility.

After conducting the survey, we would use the remaining fonts to make a readability experiment similar to the one described in Chapter 4. Additionally, we have to find a way to conduct an experiment to find the best suitable space between pairs of characters of the specific typefaces for digital reading.

If all the steps run well, there would be another typeface created, based on FZHT. If we get enough financial support and time, we could design and create this potential font transformed from the original HeiTi concept and theory. This would be different from recommending the continued use of a specific existing typeface.

# References


[1] http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/about.asp

[2] http://www.displaysearch.com/cps/rde/xchg/displaysearch/hs.xsl/110314_touch_screens_in_tablet_pcs_forecast_to_reach_60m_in_2011.asp

[3] http://www.chinatoday.com/data/data.htm

[4] http://www.seekglasses.com/news_2010-1-4/31733.html

[5] type cast: http://www.apa.org/monitor/2010/11/read-onscreen.aspx

[6] K. Fukunaga, *Introduction to Statistical Pattern Recognition*, 2nd ed., Boston, USA: *Academic Press*, 1990.

[7] C.-L. Liu and C. Y. Suen, "A new benchmark on the recognition of handwritten Bangla and Farsi numeral characters," *Pattern Recognition*, vol. 42, no. 12, 2009, pp. 3287-3295.

[8] F. Tang and H. Tao, "Fast linear discriminant analysis using binary bases," *Pattern Recognition Letters*, vol. 28, no.16, 2007, pp. 2209–2218.

[9] Aibao Zhou, Xuemin Zhang, Hua Shu, Liguo He, (2005). "The Influence to the Recognition of Chinese Characters Caused by the Font Categories, Font Sizes and Parts of Speech," *Chinese Journal of Applied Psychology*. Vol.11. No. 2, pp. 128-132, 2005.

[10] C. M. Bishop, *Pattern Recognition and Machine Learning*, Germany: *Springer-Verlag*, Berlin/Heidelberg, 2006.

[11] M. Cheriet, N. Kharma, C.-L. Liu, and C.Y. Suen, *Character Recognition Systems*, USA: John Wiley & Sons, Inc., New York, 2007.

[12] E. Kavallieratou, N. Fakotakis, and G. Kokkinakis, "Slant estimation algorithm for OCR systems," *Pattern Recognition*, vol. 34, no. 12, 2001, pp. 2515 – 2522.

[13] Chunlei He, "Reliable Pattern Recognition System with Novel Semi-Supervised Learning Approach," Computer Science Department, Concordia University, 2010

[14] C. Y. Suen and M. Komoda, "Legibility of digital type-fonts and comprehension in reading," in J. C. van Vliet (ed.), *Text Processing and Document Manipulation,* Cambridge University Press, pp. 178~181, 1986.

[15] R. J. Woods, K. Davis, and L. F. V. Scharff, "Effects of typeface and font size on legibility for children," *Am J of Psychological Research*, vol. 1, 86-102, 2005.

[16] Eva R. Brumberger, "The Rhetoric of Typography: The Persona of Typeface and Text," *Technical Communications*, Volume 50, Number 2, pp. 206-223(18), May 2003

[17] Dawn Shaikh, "Know Your Typefaces! Semantic Differential Presentation of 40 Onscreen typefaces," *Usability News*, October 2009, Vol.11 Issue 2

[18] Ying Li and Ching Y. Suen, 2010, "Typefaces personality traits and their design characteristics," *In proceedings of the 9th IAPR International Workshop on Document Analysis Systems*, pp. 231~238, June 9th -11th, 2010, Boston, MA.

[19] Walter Tracy, 2003, Letter of Credit: A View of Type Design, Dr. Godine. Chapter 4, Legibility and readability, 31.

[20] http://en.wikipedia.org/wiki/Chinese_character

[21] Tinker, M.A. *Legibility of Print, 3rd edition*. Iowa: Iowa State University Press, 1963

[22] Zachrisson, B., *Studies in the Legibility of Printed Text*. Stockholm: Almqvist & Wiksell, 1965

[23] Lund, O., Knowledge Construction in Typography: The case of legibility research and the legibility of sans serif typefaces. Thesis submitted for the degree of Doctor of Philosophy. Reading: The University of Reading, Department of Typography & Graphic Communication, 1999.

[24] Reynolds, L., "Legibility studies: Their relevance to present-day documentation methods," *Journal of Documentation*, 35(4), 307-340, 1979

[25] http://alexpoole.info/which-are-more-legible-serif-or-sans-serif-typefaces

[26] De Lange, R. W., Esterhuizen, H. L., Beatty, D., "Performance differences between Times and Helvetica in a reading task," *Electronic Publishing*, 6(3), 241-248, 1993

[27] Guiqing Bei, Xuetao Zhang, *Hanzi Frequency statistics*, Electronic Industry Press, Beijing, 1985

[28] Hyunjin Song and Norbert Schwarz, “If It’s Hard to Read, It’s Hard to Do - Processing Fluency Affects Effort Prediction and Motivation,” *Psychological Science*, vol. 19, issue 10, SAGE Publications, pp. 986-988, 2008

[29] Convince with Simple Fonts: http://www.neurosciencemarketing.com/blog/articles/simple-fonts.htm

[30] Walter Schneider, Amy Eschman & Anthony Zuccolotto, 2002, *E-Prime: User's Guide*. Pittsburgh, Psychology Software Tools Inc.

[31] http://en.wikipedia.org/wiki/Analysis_of_variance

[32] http://www.statsoft.com/textbook/basic-statistics/

[33] http://en.wikipedia.org/wiki/Pairwise_comparison

[34] Gregory Carey, 1998, Multivariate Analysis of Variance (MANOVA), http://ibgwww.colorado.edu/~carey/p7291dir/handouts/manova1.pdf

[35] http://www.01.ibm.com/software/analytics/spss/?pgel=ibmhzn&cm_re=masthead-_-products-_-sw-spss

[36] http://en.wikipedia.org/wiki/Normal_distribution

[37] http://www.mzandee.net/~zandee/statistiek/syllabus/LSD.pdf

[38] http://en.wikipedia.org/wiki/Post-hoc_analysis

[39] http://www.graphpad.com/faq/viewfaq.cfm?faq=176

[40] http://wenku.baidu.com/view/f697390c6c85ec3a87c2c5ad.html

[41] Cheriet, M., Kharma, N. Liu, C.L., and Suen, C., *Character recognition systems: A guide for students and practitioners,* Wiley-Interscience, 2007

[42] http://rafflesmedicalgroup.com.sg/health-articles/health-topics/myopia-causes-and-prevention.aspx

[43] Jacobi FK, Pusch CM. “A decade in search of myopia genes,” *Front Biosci.*,15:359-72, Jan 1, 2010

[44] http://www.about-vision.com/myopia/93-causes

[45] http://www.dahlsys.com/misc/antialias/index.html

[46] http://www.surl.org/usabilitynews/22/font.asp

[47] http://www.surl.org/usabilitynews/31/fontSR.asp

[48] http://wenku.baidu.com/view/0902c8cf05087632311212b3.html

[49] http://en.wikipedia.org/wiki/Walter_Tracy

[50] http://en.wikipedia.org/wiki/Sobel_operator

[51] Geetha, Srikantan, Stephen W.Lam and Sargun N. Srihari, "Gradient-based Contour Encoding for Character Recognition," *Pattern Recognition*, Vol 29, No.7, pp.1147-1160, 1996

[52] http://www.68design.net/Graphic/Theory-Graphic/49430-1.html

[53] http://en.wikipedia.org/wiki/Font_rasterization