Usability Comparison of Text CAPTCHAs Based on English and Chinese

Junnan Yu[(✉)], Xuna Ma, and Ting Han[(✉)]

School of Media and Design, Shanghai Jiao Tong University, Shanghai, China
{Junius,Hanting}@sjtu.edu.cn

Abstract. Text CAPTCHAs are widely deployed in nowadays for websites to defend malicious attacks. Although most text CAPTCHAs employ alphanumeric characters, there are emerging interests in designing CAPTCHAs based on regional languages. Here, we conducted experiments to compare the usability of CAPTCHAs based on English and Chinese. The results indicate that, comparing with CAPTCHAs that employ random English or Chinese characters, those based on frequently-used English or Chinese words provide the best usability in terms of efficiency, effectiveness and satisfactory for participants who are native Chinese speakers while familiar with English. CAPTCHAs based on random Chinese characters, however, is least user-friendly from a comprehensive perspective. The evaluation method and results presented here may shine a light for the design of CAPTCHAs that employ characters other than alphanumeric.

Keywords: CAPTCHAs · Usability · Human-computer interaction · Cross-culture design

1 Introduction

CAPTCHA (Completely Automated Public Turing test to tell Computers and Humans Apart), with its aim to distinguish human behavior from automatic scripts, is now widely used for online systems, particularly in registration and password verification scenes [1, 2]. For instance, Gmail employs it to filter out spammers; Facebook would benefit from preventing fake accounts and junk messages; PayPal utilizes it to enforce the financial security of its users and so on.

Principally, a well-designed CAPTCHA is expected to be easily recognized by humans while hard for bots to crack. Since its invention in 2002, CAPTCHAs in nowadays generally fall into three categories [3]: Text, Image and Voice. Given that Text form is the dominant one [4] and the focus of this paper, the word CAPTCHA mentioned afterwards represents only the text kind unless otherwise specified. Typically, a CAPTCHA includes several alphanumeric characters which are distorted and/or overlapped with each other, together with strikethrough lines and noise backgrounds [5, 6]. In this way, computer algorithms will have difficulty separating characters from one another and identifying them individually. With the increased complexity of those designs, it is more efficient to defend automatic scripts [5] but also at the cost of degraded

usability. Therefore, it's essential to study the usability of text-based CAPTCHAs with a variety of design complexities.

For instance, Chellapilla et al. [7] investigated the design factors that could balance between usability and security. Elie Bursztein et al. [8] identified a set of features of alphanumeric CAPTCHAs and classified them in to three categories—visual features (character sets and counts, font sizes, etc.), anti-segmentation features (character overlaps, random dot sizes, etc.), and anti-recognition features (rotated character counts and degrees, etc.), then further investigated their effects on the usability of alphanumeric CAPTCHAs. Lee [9] compared the usability of alphanumeric CAPTCHAs for native Chinese speakers of different ages and revealed that young group had better performance than the old group. Belk et al. [10] evaluated the effects of cognitive styles on people's performance of CAPTCHAs. They pointed out that, when designing a user-friendly CAPTCHA, not only should the intrinsic factors like noise, mask line, etc. be taken into account, but also some variables on a user's side such as his/her cognitive style, culture background, etc.

However, all those studies on the design and usability of CAPTCHAs are predominantly focused on those employing alphanumeric Characters. Under the background of globalization, there is also an increasing concern about designing localized CAPTCHAs that employ the regional languages. Shirali-Shahreza [11] designed a type of text CAPTCHA that employed Persian/Arabic characters with improved security and usability. Yang [12] explored the application of Korean characters in text CAPTCHAs, their results showed that the Korean CAPTCHAs could be easily understood by native Korean speakers while difficult to be defeated by OCR (Optical Characters Recognition) programs. Banday [13] investigated the usability of CAPTCHAs based on Urdu, one of the regional languages used in India. The results indicated that, for native speakers of Urdu who had few or no familiarity with English, they solved Urdu CAPTCHAs significantly faster and more accurately than those based on English. Shortly, localized CAPTCHAs are generally believed to provide better usability because people are intuitively more comfortable with their native languages.

Meanwhile, CAPTCHA designs that employ Chinese characters are also emerging and have already been deployed by leading internet companies, such as Baidu.com and Renren.com, the counterparts of Google and Facebook in China, respectively. Paralleling with those deployments, Wang [14] proposed a Chinese CAPTCHA design that added a semi-transparent layer of Chinese characters as the background of the main layer and further experimentally proved that it was an effective means against OCR. Shen et al. [15] explored a multiscale corner structure model that was capable of hacking Chinese CAPTCHAs, which was insightful to improve the security of Chinese CAPTCHAs. Studies of Chinese CAPTCHAs are mainly about their mechanism [16–18], the usability of such localized CAPTCHAs, however, has hardly been explored, particularly, its difference with respect to those based on English characters.

Here, we investigated and compared the usability of CAPTCHAs based on English and Chinese for Chinese users. This study focuses on the following questions: Would the subjects have better performance when interacting with CAPTCHAs that use their native language? What are the subjects' perceptions about those localized designs?

2   Method

2.1   Participants

Thirty participants (13 males and 17 females), who are native speakers of Chinese with English as a familiar second language, were recruited for current studies. Their average age was 21.6 with a standard deviation of 1.3. All participant were students from Shanghai Jiao Tong University, 9 of them were undergraduate students and the remaining were graduate students. All participants had passed the College English Test Band 6, a language proficiency test held by the Ministry of Education of China. Therefore, they were all familiar with the English words appeared in current experiments. In addition, each participant was an experienced computer user who spent at least 2 h per week on word processing with keyboard and mouse. During online activities, all subjects had encountered English CAPTCHAs, and 29 of them had experienced Chinese CAPTCHAs. None of them had trouble reading on the screen or operating the input devices of computer.

2.2   Apparatus

The experiments were conducted in a lab environment. All participants were instructed to solve CAPTCHAs on a same setup, which included a 20-inch liquid crystal display with a resolution of 1440 * 900, a computer running Windows 8.1 system, a set of regular QWERTY keyboard and mouse as the input devices. The input software for Chinese characters was Microsoft Pinyin, which was daily-used input method for all participants and also the pre-installed input method of Windows 8.1. The tilt angle, height and distance of the display and chair were adjusted by participants to comfort themselves. The CAPTCHAs were generated on a remote server and loaded in the form of a webpage to the local browser, which was Google Chrome in this study. After the CAPTCHA test, each participant was also required to finish an online questionnaire and interviewed to learn their subjective opinions regarding those CAPTCHA designs.

2.3   Tasks

All participants were required to finish three consecutive tasks: Firstly, each participant was required to get familiar with the experimental apparatuses through solving five CAPTCHAs prepared for testing purpose. After that, four types of CAPTCHAs were presented for participants to solve one by one and each type of design included 12 randomly generated CAPTCHAs. Finally, participants were asked to finish an online questionnaire and interviewed to learn their subjective perceptions about the CAPTCHA designs in the experiments.

2.4   Study Design

To compare the usability of English and Chinese CAPTCHAs for Chinese users, four types of CAPTCHAs, which were based on Random English Characters (REC),

Frequent English Words (FEW), Random Chinese Characters (RCC) and Frequent Chinese Words (FCW) and illustrated in Fig. 1, respectively. REC and FEW designs utilized English characters RCC and FCW employed Chinese characters. For each language, the characters were presented as either random characters (REC, RCC) or words (FEW, FCW) that are frequently used in daily life. All other design factors were kept the same. For instance, each CAPTHCA was 230 pixel in width and 70 pixel in height. The font size was the same for all designs and the font family employed was Microsoft Yahei, which supports both English and Chinese Characters. The characters displayed on each CAPTCHA had a transparency of 25 % while were surrounded by 3 random lines and the same background noise levels. The distortion of each character was also kept the same by setting the same parameter. Furthermore, although each English CAPTCHA included 8 letters while the Chinese one included 3 or 4 characters, the average keystrokes [19] required for their inputs were the same under current experimental setting. Therefore, it maintained a similar workload to input different CAPTCHA types and was expected to provide a similar condition to evaluate the solving time of different designs.
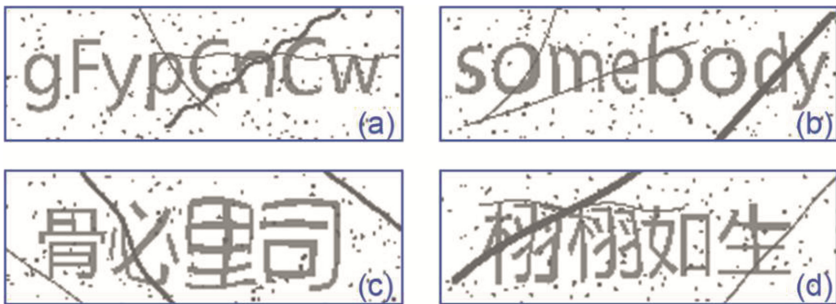

Fig. 1. Illustration of Text CAPTCHA styles explored in current study: (a) Random English Characters (REC); (b) Frequent English Word (FEW); (c) Random Chinese Characters (RCC); (d) Frequent Chinese Word (FCW). These CAPTCHAs were generated through a re-developing of the widely-used Securimage code [20].

During the experiment, only one CAPTCHA was presented on the web interface each time. Each participant was instructed to recognize, input and submit the characters shown on that CAPTCHA, which simulated the general CAPTCHA verification scene used by most websites in nowadays. After submitting his/her recognition result, a record will be generated on the remote server, indexing the solving time, the user input and whether the CAPTCHA was correctly input. Meanwhile, the webpage refreshed automatically and the participant was directed to solve the next CAPTCHA till the end of the task cycle, which included 48 CAPTCHAs in total, 12 for each kind. The collected data were further analyzed to obtain the average solving time and correction rate for each type of CAPTCHA design.

The usability of each CAPTCHA design was evaluated by three independent variables of usability [21]: effectiveness, efficiency and satisfaction. The effectiveness and efficiency were measured by the average solving time and correction rate for each type of CAPTCHA, respectively. The satisfaction was obtained through an online questionnaire and a face-to-face interview with each participant.

2.5   Procedure

The experiment was carried out in three stages—experiment preparation, testing and interview. During the preparation stage, we reset the testing apparatuses and described the purpose and tasks of the experiment to each participant, who was also informed that this test was anonymous and any data collected would be restricted for the use of current study only. After that, a participant was instructed to get familiar with the experiment apparatuses through solving five CAPTCHAs prepared for testing purpose. In the testing stage, a participant was left alone in the lab to solve four consecutive CAPTCHA sections and one online questionnaire without any disturbances. However, the experiment instructor would wait outside the lab in case the participant would need any tech support. For the final stage, participants were interviewed to learn their additional comments about the different CAPTCHA designs as well as their emotional feelings. After that, each subject was given a small gift to appreciate his/her cooperation.

3   Results and Discussion

3.1   Comparison of Efficiency and Effectiveness Between English and Chinese CAPTCHAs

The average solving time for all four kinds of CAPTCHA design, which based on Random English Characters (REC), Frequent English Words (FEW), Random Chinese Characters (RCC) or Frequent Chinese Words (FCW), were illustrated in Fig. 2.

The solving time of FEW (M = 4.68 s, SD = 1.4 s) is essentially the same as that of the FCW (M = 4.46 s, SD = 2.7 s). This same solving time can be explained by the fact that all those participants were familiar with both the English and Chinese words appeared in this study. Therefore, participants had a similar response to both kinds of CAPTCHA design. It is also indicated in Fig. 2 that, solving RCC designs (M = 9.38 s, SD = 4 s) takes the longest time, followed by REC designs (M = 7.75 s, SD = 2.3 s). The results of both RCC and REC are much longer than those of FEC and FCC results. The longer solving time for CAPTCHAs based on both random English and Chinese characters reveals that, it took more time for participants to recognize each characters individually and then type them into the test interface. The similar solving time for both FEC and FCC further shows that it took basically the same effort for participants to response to their native language and a familiar second language. In general, CAPTCHAs based on frequently-used English and Chinese words have better efficiency than those employ random characters while there are no significant difference for the solving time of frequent English and Chinese words.
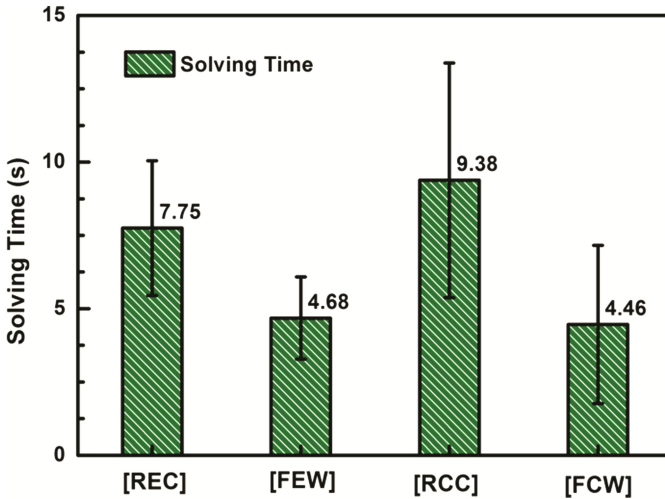
Fig. 2. Average solving time for all four kinds of CAPTCHA design: Random English Characters (REC), Frequent English Words (FEW), Random Chinese Characters (RCC), Frequent Chinese Words (FCW)

The effectiveness of those four CAPTCHA designs are represented by the percentage of CAPTCHAs that were correctly solved. As shown in Fig. 3, solving accuracy for FEW (99.22 %), RCC (97.66 %) and FCW (98.44 %) are almost the same, while REC (74.68 %) gave a significantly lower correction rate. The high correction rate for CAPTCHAs based on Chinese language and English words demonstrated that there is no intrinsic difference for participants to recognize those kinds of English and Chinese CAPTCHAs. To understand why the correction rate is much lower for CAPTCHAs based on REC, we further analyzed the user inputs for such kind of CAPTCHA. It turned out that, a majority of incorrect inputs were due to the confusion of similar English letters, such as "I" and "L". Therefore, we removed CAPTCHA inputs that contained any confusion letters and reanalyzed the correction rate of REC, which is illustrated in Fig. 4. It is clear that, without those confusion letters, the correction rate of CAPTCHAs based on REC has been improved by more than 10 %. Even without those confusion letters, however, the correction rate of REC is still at least 10 % lower than the other three designs. This is because the random lines and back ground noises, which were integrated for an improved security, sometimes would partially merge with the English characters, making them difficult for participants to identify correctly. While for English words, even though one or two letters of a word were masked, it was still possible for participants to recognize that word as a whole and correctly solve it. Therefore, the effect of random lines and background noise is more pronounced on REC design than FEW one. Furthermore, due to its complexity, even if most part of a Chinese character was blurred by the random line and background noise, participants had no difficultly recognizing it as a whole and therefore it maintained a high correction rate. Briefly, the correction rate is lowest for REC while quite good for FEW, RCC and FCW CAPTCHAs.
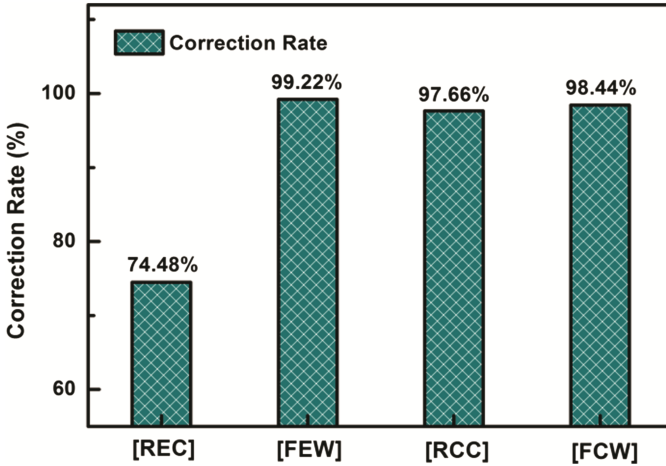
Fig. 3. Average correction rate for all four kinds of CAPTCHA design: Random English Characters (REC), Frequent English Words (FEW), Random Chinese Characters (RCC), Frequent Chinese Words (FCW)
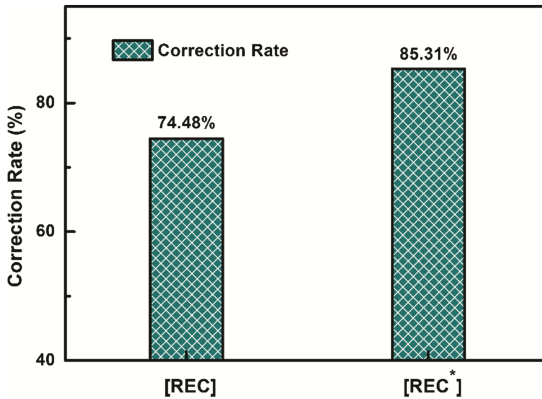

Fig. 4. Correction rate of CAPTCHAs based on Random English Characters. REC or REC* represents the correction rate with or without confusion letters appeared in a CAPTCHA, respectively.

3.2 Satisfactory Questionnaire and Interview

In addition to the efficiency and effectiveness studies, each participant was also required to finish a questionnaire and interviewed to acquire their subjective opinions toward those four types of CAPTCHA designs. The results reveal that more than 97.3 % of the participants preferred to solve CAPTCHAs based on frequently used words rather than random characters. They believed that those CAPTCHAs could be easily recognized with just a single glance. On the contrary, for CAPTCHAs based on random characters, they would have to recognize each character individually and therefore it took more

efforts to solve them. When asked which of the four kinds of CAPTCHAs they prefer to solve the most, 56.07 % of the subjects were in favor of CAPTCHAs based on English words while the remaining 43.3 % were in favor of Chinese. The subjects who supported English words felt it was more natural and straightforward to type English words because they do not need to switch the input method between English and Chinese. For those who preferred CAPTCHAs based on Chinese words, they felt more comfortable with native language and the Pinyin input methods in nowadays are smart enough to make it fast to type Chinese. Although more than 78 % of the participants believed that CAPTCHAs based in random Chinese characters provided the most security, there were hardly any participant who was willing to encounter such type of CAPTCHAs.

4   Conclusion

The usability of CAPTCHAs based on English and Chinese were compared through a usability study conducted with participants who were familiar with both languages. Within the framework of similar design factors such as font size, font family, amount of distortion, random lines, background noise level and typing workload, it was found that, the effectiveness and efficiency of CAPTCHAs based on frequently-used English or Chinese words are similar while better than those based on random English or Chinese characters. CAPTCHAs based in random Chinese characters, however, turned out to provide the least overall usability. And the satisfactory questionnaire and interview showed that participants also preferred to encounter CAPTCHAs based on frequently-used words. In a word, comparing with English CAPTCHAs, Chinese also boasts the potential of serving a user-friendly CAPTCHA design. Therefore, the study presented here supports the application of Chinese CAPTCHAs to a large extent.

References

1. von Ahn, L., Blum, M., Hopper, N., Langford, J.: CAPTCHA: using hard AI problems for security. In: Biham, E. (ed.) Advances in Cryptology — EUROCRYPT 2003, vol. 2656, pp. 294–311. Springer, Heidelberg (2003)
2. Von Ahn, L., Blum, M., Langford, J.: Telling humans and computers apart automatically. Commun. ACM **47**, 56–60 (2004)
3. Yan, J., Ahmad, A.S.E.: Usability of CAPTCHAs or usability issues in CAPTCHA design. In: Proceedings of the 4th Symposium on Usable Privacy and Security, pp. 44–52. ACM, Pittsburgh (2008)
4. Jeng, A.B., Tseng, C.-C., Tseng, D.-F., Wang, J.-C.: A study of CAPTCHA and its application to user authentication. In: Pan, J.-S., Chen, S.-M., Nguyen, N.T. (eds.) ICCCI 2010, Part II. LNCS, vol. 6422, pp. 433–440. Springer, Heidelberg (2010)

5. Bursztein, E., Martin, M., Mitchell, J.: Text-based CAPTCHA strengths and weaknesses. In: Proceedings of the 18th ACM Conference on Computer and Communications Security, pp. 125–138. ACM (2011)

6. Moradi, M., Keyvanpour, M.: CAPTCHA and its Alternatives: A Review. Secur. Commun. Netw. **8**(12), 2135–2156 (2014)

7. Chellapilla, K., Larson, K., Simard, P., Czerwinski, M.: Designing human friendly human interaction proofs (HIPs). In: Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, pp. 711–720. ACM (2005)

8. Bursztein, E., Moscicki, A., Fabry, C., Bethard, S., Mitchell, J.C., Jurafsky, D.: Easy does it: more usable captchas. In: Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems, pp. 2637–2646. ACM (2014)

9. Lee, Y.-L., Hsu, C.-H.: Usability study of text-based CAPTCHAs. Displays **32**, 81–86 (2011)

10. Belk, M., Fidas, C., Germanakos, P., Samaras, G.: Do human cognitive differences in information processing affect preference and performance of CAPTCHA? Int. J. Hum Comput Stud. **84**, 1–18 (2015)

11. Shirali-Shahreza, M.H., Shirali-Shahreza, M.: Persian/Arabic baffletext CAPTCHA. J. UCS **12**, 1783–1796 (2006)

12. Yang, T.-C., Ince, I.F., Salman, Y.B.: A korean CAPTCHA study: defeating OCRs in a new CAPTCHA context by using korean syllables. Int. J. Contents **5**, 50–56 (2009)

13. Banday, M.T., Shah, N.A.: Challenges of CAPTCHA in the accessibility of Indian regional websites. In: Proceedings of the Fourth Annual ACM Bangalore Conference, pp. 1–4. ACM, Bangalore (2011)

14. Wang, T., Bøegh, J.: Multi-layer CAPTCHA based on Chinese character deformation. In: Yuan, Y., Wu, X., Lu, Y. (eds.) ISCTCS 2013. CCIS, vol. 426, pp. 205–211. Springer, Heidelberg (2014)

15. Shen, Y., Ji, R., Cao, D., Wang, M.: Hacking Chinese touclick CAPTCHA by multi-scale corner structure model with fast pattern matching. In: MM 2014 – Proceedings of the 2014 ACM Conference on Multimedia, pp. 853–856 (2014)

16. Chen, D.: Research of the Chinese CAPTCHA system based on AJAX. WSEAS Trans. Circ. Syst. **8**, 53–62 (2009)

17. Hai-kun, J., Wen-jie, D., Li-min, S.: Research on security model with Chinese CAPTCHA. Comput. Eng. Des. **6**, 023 (2006)

18. Xu, S., Lau, F., Cheung, W.K., Pan, Y.: Automatic generation of artistic Chinese calligraphy. Intell. Syst. IEEE **20**, 32–39 (2005)

19. MacKenzie, I.: KSPC (Keystrokes per Character) as a characteristic of text entry techniques. In: Paternó, F. (ed.) Mobile HCI 2002. LNCS, vol. 2411, p. 195. Springer, Heidelberg (2002)

20. Securimage. https://www.phpcaptcha.org/

21. Standard, I.: Ergonomic requirements for office work with visual display terminals (vdts)– part 11: guidance on usability. ISO standard 9241-11: 1998. International Organization for Standardization (1998)