Stroke Break Analysis: a Practical Method to Study Timeout Value for Handwriting Recognition Input

Yanqing Cui
Nokia Research Center,
Nokia House 1, No. 11, He Ping Li Dong Jie
100013, Beijing, PRC
+86 10 65392828
yanqing.cui@nokia.com

Vuokko Lantz
Nokia Research Center,
P.O. Box 407,
FIN-00045 Nokia Group, Finland
+358 50 4821730
vuokko.lantz@nokia.com

ABSTRACT

Handwriting recognition (HWR) input method has been considered to be one of the most usable text entry methods for handheld devices, especially for languages with large and complicated character sets such as Chinese. The paper studies stroke break times within handwritten characters and presents a new method for setting HWR timeout by examining the break time distributions. For multi-stroke character HWR input, a timeout is widely used as a segmentation technique to initiate the recognition process. In this paper, we examine the largest stroke break time in each character and explore the relationship between break time distribution and optimal HWR timeout. The study used Chinese as test material and the test independent variables were writing condition (input box, full screen) and user's posture while they were writing (hold device in hand, keep device on table). The main findings are: (1) the stroke break times are similar in full screen and input box conditions, though the users tend to write larger characters in full screen condition. (2) The stroke break times fit into a tight distribution. It is feasible to estimate optimal HWR timeout by studying stoke break time distribution. A nonparametric histogram method was used to model the stroke break distributions and it showed that typical Chinese HWR default timeouts are around 99% percentile in the distribution. (3) Differences in HWR stroke break distributions are very significant between individual users. The stroke break time analysis can also be applied to design HWR timeout customization scale.

Categories and Subject Descriptors

H.5.2 User Interfaces: Input devices and strategies

General Terms:

Measurement, Design, Experimentation, Human Factors, Languages, Theory, Verification.

Keywords:

Handwriting Recognition, Stroke, Break, Timeout, Distribution.

1. INTRODUCTION

Handwriting recognition (HWR) input method has been considered to be one of the most usable text entry methods for handheld devices since the devices emerged in 1990s. It is especially fit for languages that use large and complicated character sets [1]. For example, the Thai alphabet has 88 letters, which are obviously not easy to input with a conventional keyboard.

An HWR input process includes several stages. The most important stages are writing and recognition. In single-stroke systems such as Graffiti, Unistrokes, and Jot, the device changes from writing to recognition immediately after each pen-up [1, 2]. But for multi-stroke characters, some segmentation method is needed for initiating recognition. A timeout is widely used for this purpose. The system waits for a new stroke after each pen-up. If a certain threshold time (timeout) is reached and the user has not started a new stroke, the user interface proceeds to the recognition stage. The previously input strokes are recognized as one character [5].

Another technique is to use a special key for skipping the timeout (timeout kill) and starting the pen trace recognition right away. Some devices use a combination of the two solutions. For example, HP iPAQ (HP Group, U.S.; http://www.hp.com/) supports multiple input boxes. It includes both a timeout and the provision for a timeout kill by writing into the other input box.

Setting the HWR timeout is a trade-off between the text entry speed and error rate. The timeout must be long enough to differentiate within-character stroke breaks from between-character stroke breaks reliably. Otherwise, errors will occur when the system starts recognition before user has finished all the strokes of the character. Meanwhile, the timeout must be as small as possible to avoid wasting the user's time. Too long a timeout both decreases user text entry speed and lowers the user's subjective experience. The handwriting input process is typically accompanied by many user mal-operations and system errors. All these errors may cause some irregular stroke breaks. These anomalies make it difficult to locate the optimal timeout value.

Although timeout is widely used as HWR segmentation tool, HCI research methods for setting it optimally are still very limited. Our research is motivated by the need for such methods to assist designers of pen-based computing systems.

2. HANDWRITING TIMEOUT STUDY

2.1 Current research methods

Timeout has been studied intensively to be used as a segmentation tool in various user interface systems. These studies mainly try to verify the effectiveness of using timeout as segmentation tool or study how the users learn to use the timeout [3, 4, 6, 7].

Generally, users learn to use timeout segmentation rather quickly. The users also have different strategies in using the timeout [3]. But methods for optimal timeout setting are still necessary in product creation processes, especially in HWR design. People have already established a solid rhythm in natural writing. They certainly expect they can use the same rhythm in HWR input [5].

In applied research practices, researchers usually adopt classical experimental psychology methods to explore optimal timeout values. They ask users to test a HWR engine with several different timeouts and try to select the best one by analyzing user performance and subjective evaluation results. The research method is generally effective to find out acceptable timeouts after several rounds of tests. But it has some inherent shortcomings. Firstly, the timeout candidate selection may bias the result. Secondly, the test must be conducted using a working HWR engine, which may be not available in early product development stages. Thirdly, each user may demand different optimal timeout values. The research method does not cover the issue of how to customize the timeout for each user.

2.2 The study overview

We present a new method for finding optimal HWR timeout values by studying the stroke break time distribution. The distribution is formed from a database of natural handwriting samples and we try to explore its relationship with optimal HWR timeout.

There are plenty of HWR timeout study findings available, so we use them as a reference to identify the critical point in the stroke break distribution. The previous studies show that in a handwriting text entry task, the writers activate a series of cognitive processes before they start to write a character. For commonly used simple Chinese characters, the processes last for 439ms on average. The standard deviation is 48ms [5].

The current mainstream devices serve as reliable references of the optimal HWR timeout. These timeout settings reflect the current research results. Table 1 lists Chinese HWR timeout values for some mainstream products. These timeouts fall into an interval from 450ms to 600ms.

Table 1. Some typical HWR devices timeout settings.

Products	Default timeout
HP iPAQ 4150 (Win CE 2003)	450ms
Dopodo 686 (Win CE 2002)	500ms
Nokia 6108	500ms
Decoma	550ms
Nokia 7710	600ms
Motorola A780	600ms

In 2003, when we developed Nokia 6108, we had already confirmed that Chinese optimal HWR timeout is around 500ms through a series of experimental studies. By combining all these studies, we can propose that the optimal timeout for Chinese HWR is about 500 ms.

In this study, we map these previously proposed timeouts into the stroke break time distribution and try to locate the corresponding percentiles. In order to achieve this, we model break times with non-parametric and parametric distribution models. The paper also examines the individual differences in break time distributions. We use the critical percentile deduced form the results of previous studies to estimate optimal HWR timeout for each user. We also discuss how the stroke break analysis can be applied to creating HWR timeout customization options. This study explores the feasibility of a new research method to study HWR timeout. If it proves successful, it could be a practical method to be used in product creation processes. The research method could also be adopted for academic studies in the future.

3. THE EXPERIMENT

This study uses Chinese characters as test materials. The main reasons that we select Chinese are: (1) HWR has already been widely used in Chinese input. We can find many related studies and real products to verify our findings. (2) Chinese characters typically have several strokes, which is the present HWR development trend. As the recognized handwriting unit gets larger, the number of strokes in each unit will also grow.

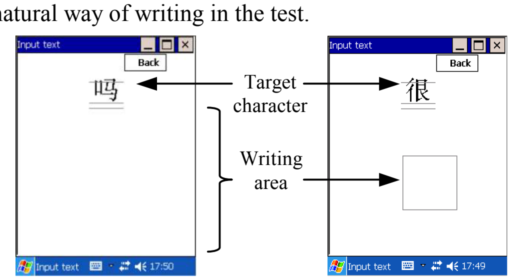
There are two common writing conditions in HWR systems. The user can either write in an input box or anywhere on the screen. There are two typical postures for using hand held devices - it is either held in hand (two-hand operation) or put on a table (one-hand operation).

3.1 Participants

Twenty, 10 males and 10 females, native Chinese speakers participated in the study. Their ages ranged from 22 to 35, the average being 28. All the users were right handed and have only had limited previous experience with touch screen devices.

3.2 Test task and device

The test task was to write 299 common Chinese characters one by one into a handheld device. All these characters have multiple strokes, 6.69 strokes on average. The subjects were asked to use their natural way of writing in the test.



(a) (b)

Figure 1. Experiment screenshots illustrating full screen (a) and input box (b) writing conditions.

Typical screenshots of the experiment are shown in Figure 1. The users were asked to write the target character either anywhere under the target text or in a dedicated input box. The test software did not support recognition. It dismissed the written pen trace and proceeded to the next target 1200ms after the user had finished the last stroke. We used an HP iPAQ 4150 as a test device. The display resolution is 240*320 pixels, 1 pixel is 0.24mm. A Java application was built for running the test, i.e. showing the target characters and saving all the pen movements.

3.3 Test design

We used a two-variable mixed experimental design in this test. The independent variables were user posture (between-subjects) and writing condition (within-subjects). The writing condition was either full screen or input box. In the full screen condition, the user wrote the character anywhere on the display, see Figure 1(a). In the input box condition, the user only wrote in the fixed input box, see Figure 1(b). The input box size was 1.6*1.6 cm^2.

While writing, the users held the device either in their hand or on a table. To keep the test environment similar to real use situations, users conducted the test on their own, without any moderator present, at their office desks. The test users were allowed to pause to do any other tasks during the test sessions.

4. RESULTS

The largest stroke break time in each character determines the minimum value for the timeout. In this paper, we measure the largest stroke break time from each input character and use it in the following analyses. Each user has up to 598 records.

4.1 The effect of writing condition and posture

Figure 2 presents the mean of stroke break times under each test condition. As we can see, in the input box condition, holding the device in hand yields shorter stroke break times than holding it on a table. Under full screen condition, the two writing postures cause similar break times.

The ANOVA test reveals that neither of the main factor effects is significant. User posture: $F_{(1, 18)}= 0.98$, $p>0.05$; writing condition: $F_{(1, 18)}= 0.11$, $p>0.05$. The interaction effect between two factors is nearly significant in statistics: $F_{(1, 18)}=4.94$, $p=0.07$. As the input box and full screen writing conditions have similar stroke break times we intermix the data from these test conditions in following data analyses.

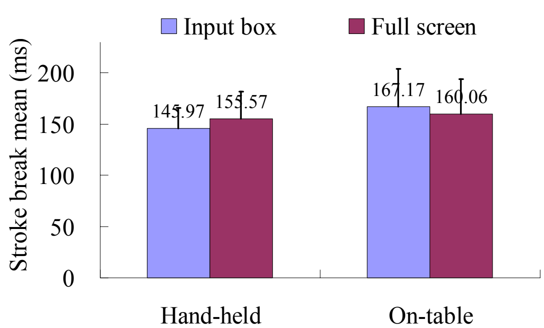


Figure 2. Mean of the largest stroke break time within a character.

4.2 Stroke break time distribution and individual differences

In Figures 3 and 4, cumulative histograms illustrate how the largest stroke break times within characters are distributed in the case of individual test subjects (U1-20 data) and on average (WI data). The writer-dependent (WI) distributions, with the exception of U20, are relatively tight and the steep cumulative histograms start to saturate between 150ms and 350ms. The optimal HWR timeout can be set by identifying the beginning of this saturated area. According to our analysis, the HWR timeouts of the current mainstream devices correspond to the 98.5-99.4% percentiles of the writer-independent stroke break time distribution and will lead to erroneous segmentation of 0.6-1.5% of input characters.

The cumulative histograms show that individual differences between users are significant. The timeout value corresponding to 99% critical point of the cumulative writer-dependent (WD) histograms varies between 159ms and 709ms, 505ms being the average estimated with WI data. Obviously, there is no general timeout value that would be optimal for all users. These findings highlight the need for timeout customization. In this test user group, the optimal timeout options could be 350ms, 500ms, and 700ms for fast, ordinary, and slow writers, respectively.

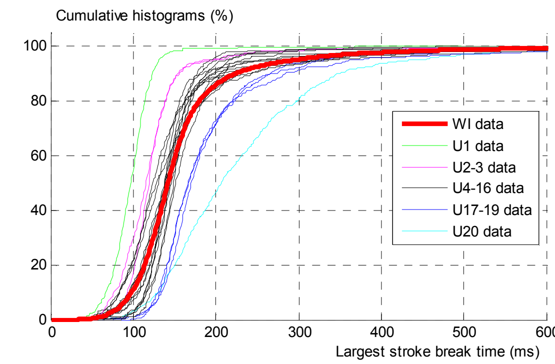


Figure 3. Cumulative histograms of stroke break times.

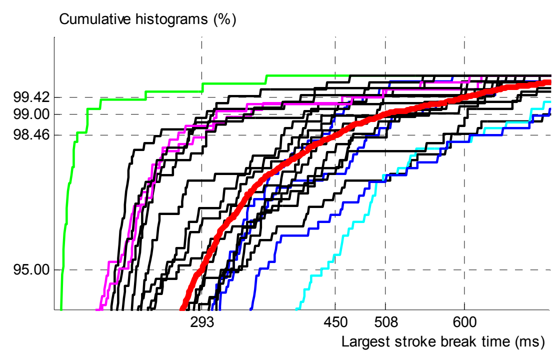


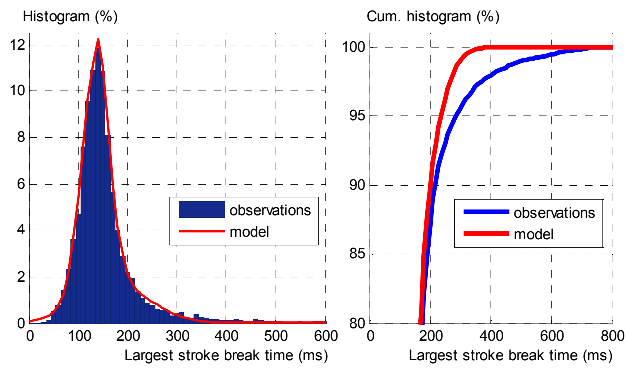
Figure 4. Close-up of cumulative histograms of stroke break times.

4.3 Model for stroke break distribution

Histograms are nonparametric distribution models fine for describing observations but they do not help understanding the underlying process generating the data. Model accuracy and memory consumption depend directly on time resolution. In order to adapt the HWR timeout for individual users, the system needs to keep a record of the observed stroke break times.

As an alternative approach to histograms we tried to model stroke break time distributions with various parametric models. The main advantage of parametric models is that they are fully defined with only a few parameters and are computationally straightforward to use. A writer-dependent model can be adapted into a writer-dependent one by adjusting the model parameters on the basis of new stroke break time observations.

We experimented with various different parametric distribution models, e.g. single Gaussian, Weibull, and Gamma. The best fit between the model and observed data was obtained with a mixture of three Gaussian distributions. Figure 5(a) shows how the model captures the basic shape of the distribution well, the RMSE fitting error being as low as 0.001. However, the model underestimates the right tail of the distribution, the area with only a few observations, see Figure 5(b), and thus cannot predict the optimal HWR timeout accurately. The accuracy of parametric model depends on how well it describes the data generation and on the number of data samples used for estimating the model parameters. On the basis of these results, we can conclude that in practice the histograms are better suited for setting optimal HWR timeout values than parametric models.



(a) (b)

Figure 5. (a) Gaussian mixture model fitted to the histogram of the WI stroke break times, (b) close-up of the observed and modeled cumulative histogram of WI stroke break times.

5. DISCUSSIONS

5.1 The effect of writing condition and user posture

In the full screen condition, the users tended to write larger characters (average size: 1.36cm *1.16cm) than in the input box condition (average size: 1.02cm *0.82cm). However, the users kept similar breaks between strokes in both conditions.

The result reveals that each user has a stable rhythm in writing characters. Their optimal timeout is invariant across changes in writing conditions. The finding is consistent with English studies. Wright C. E. (1993) study shows writing time only changes slightly when the size of the characters grows from the normal size preferred by a given writer [8].

5.2 HWR stroke break and timeout threshold

This study verifies our hypothesis that stroke break distribution can be used to locate optimal HWR timeout design. By mapping the timeout values used in the current devices to the stroke break time distribution, we can deduce that the cut point 99% percentile of the distribution can be used for optimal HWR timeout.

In the paper, the data analysis is based on the biggest stroke break time in Chinese characters. Therefore, in principle the percentile findings can be valid for other languages that are widely using multi-stroke characters, for example, upper case Latin letters. However, due to the language differences, further study is necessary to confirm the generality of the results.

5.3 Individual difference in HWR timeout

The test users show significant differences in their stroke break time distributions. Thus, each user has a different optimal HWR timeout. The cumulative histogram-based method presented here can be used to create a HWR timeout customization scale. By collecting handwriting samples from several users with different backgrounds and writing styles, we can decide the optimal timeout separately for each of them, or for some predefined user categories. Users can be grouped on the basis of their stroke break time histograms. This can be done by a human expert, like in Figure 4, or by an automatic clustering algorithm and using e.g. Kullback-Leibler divergence as a similarity measure between the histograms.

6. ACKNOWLEDGMENTS

We owe a special thanks to Sachi Mizobuchi, Liu Ying and Miika Silfverberg, who facilitated our experiments and gave valuable comments on our paper draft, and to Risto Suitiala and Chen Shuo, who provided the technical support for this study.

7. REFERENCES

[1] MacKenzie I. S., and Soukoreff R. W. (2002). Text entry for mobile computing: models and methods, theory and practice. *Human-Computer Interaction*, 17, 147-198

[2] Goldberg D., and Richardson C. (1993). Touch-typing with a stylus, In *Proceedings of the SIGCHI conference on Human factors in computing systems (INTERCHI' 93)* (Amsterdam, The Netherlands, April 1993). ACM Press, New York, 80-87

[3] Marila J., and Ronkainen S. (2004). Time-out in user interface: the case of mobile text input. *Personal and Ubiquitous Computing,* 8, 2, 110-116

[4] Silfverberg M., MacKenzie I. S., and Korhonen P. (2000). Predicting text entry speed on mobile phones. In *Proceedings of the ACM conference on Human factors in computing systems (CHI'00)* (The Hague, Amsterdam, April 2000). ACM Press, New York, 9-16

[5] Wu C., Yang Q., and Zhang K. (2003). Human performance modeling in temporary segmentation Chinese handwriting recognizer. *Acta Psychologica Sinica,* 35, 4, 483-491

[6] Hinckley K., Baudisch P., Ramos G., and Guimbretiere F.(2005). Design and analysis of delimiters for selection-action pen gesture phrases in Scriboli. In *Proceeding of the SIGCHI conference on Human factors in computing systems (CHI'05)* (Portland, Oregon, USA, April 2005) ACM Press, New York, 451-460

[7] Li Y., Hinckley K. Guan Z., and Landay J. A. (2005). Experimental Analysis of Mode Switching Techniques in Pen-based User Interfaces. In *Proceeding of the SIGCHI conference on Human factors in computing systems (CHI'05)* (Portland, Oregon, USA, April 2005) ACM Press, New York, 461-470

[8] Wright C.E. (1993). Evaluating the special role of time in the control of handwriting. *Acta Psychologica,* 82, 5-52