Cognitive Load Evaluation of Handwriting Using Stroke-level Features

Kun Yu^{1,2}Julien Epps^{1,2}Fang Chen²¹School of EE&T, University of New South Wales, Sydney, Australia
²ATP Laboratory, National ICT Australia, Sydney, Australia
kun.yu@nicta.com.auj.epps@unsw.edu.aufang.chen@nicta.com.aufang.chen@nicta.com.au

ABSTRACT

This paper examines several writing features for the evaluation of cognitive load. Our analysis is focused on writing features within and between written strokes, including writing pressure, writing velocity, stroke length and inter-stroke movements. Based on a study of 20 subjects performing a sentence composition task, the reported findings reveal that writing pressure and writing velocity information are very good indicators of cognitive load. A stroke selection threshold was investigated for constraining the feature extraction to long strokes, which resulted in a small further improvement. Differing from most previous research investigating cognitive load during writing based on task performance criteria, this work proposes a new approach to cognitive load measurement using writing dynamics, with the potential to allow new or improve existing handwriting interfaces.

Author Keywords

cognitive load, handwriting, stroke, interstroke

ACM Classification Keywords

H.5.2 Information Interfaces and Presentation: Evaluation/ methodology

INTRODUCTION

Writing is one of the most complex interaction skills humans have grasped, and is characterized by the intensive cooperation from the brain, the eyes, to the fingers. Most writing processes attract focused attention of a writer, and are supported by the intensive usage of brain resources.

Cognitive load is commonly used to describe the level of mental resource demand requested from the human brain for a specific task [2], and it is closely correlated with the performance of the person in concern. A task whose demands are optimally matched to a user will keep the cognitive load at an appropriate level, which leads to good user performance with high efficiency of learning [6]. Most writing tasks require the focused attention of the writers, and cognitive load evaluation will help us to know the mental effort they are experiencing, and further give us hints on how to improve their writing performance. Cognitively over-challenging tasks in writing could overload the writer, and result in significantly decreased writing quality or task performance. A low cognitive load sometimes is not desired as well, because users may not remain focused on the current task.

A series of methods have been proposed for cognitive load evaluation, analyzing behavior, performance, physiology and subjective rating measurements [6]. For handwriting, task performance based methods and subjective methods prevail because they are easy to fit into the framework of psychological experiments, and have less constraint on the writing process. However, task performance is the composite reflection of multiple factors, and we can not safely assume that it is equal to cognitive load. A commonly adopted technique for investigating cognitive load of handwriting is the dual-task approach [9, 10]. When the writer is engaged in the primary task, the test coordinator interrupts the subject now and then and asks the subject to finish some other task, i.e. the secondary task. Usually the secondary task is not correlated with the primary task, and it is believed that this method requires the release and reorganization of working memory when the subject attempts to allocate cognitive resources to access the secondary task, and the time for the subject to respond is taken as the measurement for the cognitive demands the writer is experiencing. Other performance methods [10] include the evaluation of the quality of written sentences in terms of grammar, word fitness, and sentence complexity. A questionnaire including subjective rating on a Likert scale usually follows the experiment, which is also used to examine the cognitive load on an expost facto basis.

Extensive research on the relationship between cognitive load and handwriting has been reported by psychologists. Most work is based on the tri-model framework of working memory [2, 1], where the central executive, the phonological loop and the visuo-spatial sketchpad are believed to interactively affect writing. Experts also separate the writing process into three phases, including planning, translating and reviewing [5, 12], and thus cognitive load is evaluated based on the demand and distribution of mental resources by the working memory components during the different phases of writing. For the planning process, the working memory is mainly occupied by the demands of the visuo-spatial sketchpad and the central executive component, and the writing quality will decrease if the required working memory is beyond the span of the writer [3, 10]. Translating is closely related with the writing behavior and dependent on the practices of the writers, and a well trained writer will have less cognitive demand during this phase; therefore the writing quality could be improved with the remnant working memory available [7]. Reviewing is considered an advanced and promotive skill to evaluate the quality of writing from the perspective of readers [3]. It will only occur when the expert writer has extra cognitive resources after meeting the demands of planning and translating, and writers capable of reviewing are good at cognitive load management during writing.

Behavioral measures are correlated with the movements of the hand during writing. They are quite important for handwriting analysis, although not specifically for cognitive load purposes. For forensic investigations, temporal and spatial behavioral measures were taken into consideration, including the writing pressure, stroke length and range, hand movement time together with the number of peak velocities in one stroke [8]. However, these features have not been technically validated as reflective of cognitive load.

In this paper, we report our findings of behavioral indicators for different cognitive load experienced via handwriting. In previous research, velocity and shape information [11] were analyzed for limited pen gestures. Our work examined the relationship between comprehensive writing behaviors and cognitive load for normal writing, and provided hints for live monitoring and adjustment of cognitive load for pen users.

METHODS AND TASK DESIGN

Our task aims to vary the cognitive demands with tasks of different difficulties in a controlled environment, so that pressure and velocity changes in writing pattern resulting from cognitive load will be the main source of variation in observed responses. We also investigate whether stroke length affects cognitive load evaluation, as lower variance estimates of cognitive load might be expected from strokes that contain more measurements of position, pressure and velocity.

Task Description and Procedure

Starting from Baddeley's model [1], we would like to include all the three components of working memory in our written task design, while trying to engage the writers in their written tasks to the greatest extent. Specifically, we adapted Ransdell and Levy's experiments [10] to our tests. After seeing a set of randomly selected words, subjects were required to write down composed sentences based on the set of words. The interface for the test is shown in Figure 1 (a). Every time the subject pressed a key, a word list was displayed for a limited time before disappearing. The time for displaying the words was one second for one-word cases, two seconds for two words, and two seconds and a half for three words. The subject was required to remember the words, and write a sentence with the words given, as shown in the example response of Figure 1 (b). There was no time limit for writing, but subjects were not allowed to write the words down before writing the sentence. The subjects were also advised to use the given form of the words, but not necessarily in the given order.



Figure 1. The test interface (a), and corresponding writing sample (b).

We used the WACOM DTZ-1200W tablet to collect writing data, and subjects wrote the composed sentences in a $20 \text{cm} \times 10 \text{cm}$ space on the interface. The subjects could write in multiple lines, and make modifications in their preferred style. After a subject finished one sentence and proceeded to the next one, the writing space was cleared automatically. After all the tests, a questionnaire was completed by the subjects, rating the experienced difficulty of the tasks. The writing measurement recorded included the writing pressure, coordinates of the writing points and inter-stroke movement traces with the writing time, as depicted in Figure 2.

Twenty subjects including one left-hand writer, participated in the study, most of whom were research students. English was not the first language for nineteen out of the twenty subjects, but they all had learnt English for over ten years. Before the test, the test coordinator confirmed with them that they understood the meaning of the words. Each participant finished three blocks of ten tests. Each block corresponded to one induced cognitive load level, and the order of the blocks was randomized. All the subjects felt it easy to compose sentences with a single word, and challenging or especially challenging for the three-word tasks. This was also validated by both the increased time for thinking and the subjective rating scores collected; the averaged difficulty evaluations for the three test levels were 2.6, 4.5, 7.3 respectively, from a 9-point Likert scale. Most of the subjects finished all the tasks, and one subject did not write down a complete sentence for three tasks of the highest difficulty level.



Figure 2. Strokes (grey with red start point) and inter-stroke (green): the black points are the sampling points, with pressure depicted as blue arrows, velocity calculated as red arrows. The inter-stroke does not have pressure information.

EXPERIMENT ANALYSIS

Intra-stroke Feature Evaluation

Strokes are the major output of handwriting, and the local peak values and average for pressure, writing velocity for each individual stroke were calculated (Figure 2), together with the length of the written strokes. The selected features were intrinsic to the strokes, and accessible for popular pen tablets. A common understanding is that the pressure of the pen-tip is correlated with the writing velocity to some extent, and a high writing velocity often accompanies the low pressure of the pen, and vice versa. Figure 3 shows the pressure and velocity during strokes. The size of the sample point is proportional to the level of pressure or the writing velocity. If this relationship is reliable, it is feasible to examine either the pen pressure or the writing velocity for cognitive load evaluation instead of both.



Figure 3. Velocity and pressure of strokes from subject 5. Left column: low cognitive load, right coloum: high cognitive load; top row: pressure, bottom row: velocity.

An ANOVA test conducted on the averaged pressure, peak pressure, averaged velocity, peak velocity and stroke length were employed to measure the strength of discrimination of each with respect to cognitive load. Results are shown in Table 1. Here, 'max P' indicates the local peaks of the pressure, 'min P' indicates the local valleys of the pressure, and 'avg P' is the averaged pressure for a single stroke. Similar abbreviations are applied to the velocity features. 'Len' refers to the length of the stroke. An immediate finding is that local peaks of pressure, and local minima of velocity are good indicators (F>7, p<0.005) of the different cognitive load levels, which is consistent with the common understanding mentioned above. However, local minima of pressure and the local maxima of velocity are not as effective. The averaged pressure is also promising for cognitive load discrimination (F>6, p<0.005). According to the experiments, the stroke length information is not suitable for cognitive load discrimination.

Inter-strokes are movements of the pen-tip when it is not on the interface. Compared with strokes, inter-strokes are more complex, and may result from hand movements between characters, lines or various writing pauses. Similar to the analysis of strokes, inter-strokes are also evaluated in terms of writing velocity and length (excepting pressure). Results of the ANOVA test for inter-strokes (not shown) were far less promising compared with the stroke-based methods.

Table 1. ANOVA test for stroke features. (Second row for F-ratio and last row for p-value).

max P	min P	avg P	max V	min V	avg V	Len
7.7	4.8	6.8	1.8	10.1	4.5	0.2
0.002	0.016	0.004	0.191	0.001	0.020	0.786



Subject-specific Stroke Selection

The ANOVA test did help us to find some good features for cognitive load evaluation, but it can be noted that no writerspecific information or language-dependent factors were considered during the analysis. Some writers may prefer cursive writing with long strokes, but others may not. For English writing, short strokes may inevitably occur now and then, e.g. the horizontal stroke of 't' and 'f', and some strokes for capital letters. A histogram of the length of strokes by all the subjects at different load levels is shown in Figure 4, from which it can be observed that a significant number of strokes are shorter than 4 mm. This distribution is also observed on an individual basis. If short strokes are not as reflective of the cognitive load as the long ones, our analysis result may be biased by the accumulation of short strokes.

To examine the influence of the short strokes, we introduced a factor, the α coefficient, to identify short strokes. Practically, the α factor is applied to the average stroke length of the respective writer L_{ave} , and such that a stroke is considered short if its length is less than the threshold αL_{ave} . The α value is clearly both feature dependent and data dependent. Figure 5 shows how the F value and p value for local maximum pressure and minimum velocity changes when α varies. Here we can find that the optimal α value is close to 0.08 for the maximum F value and minimum p value. It can also be noted that for a wide range of choices of α , statistical significance is maintained, so precise estimation of α is beneficial but not necessary.

Table 2 shows the F values and p values based on the selected long strokes when α is equal to 0.08. It is intuitive that the adoption of α enhanced the statistical significance with larger F value and smaller p value for most of the selected features, showing that selective removal of the short strokes improves the cognitive load measurement performance.

Table 2. ANOVA test after α selection. (Second row for F-ratio and last row for p-value)

max P	min P	avg P	max V	min V	avg V	Len
8.2	4.9	7.7	2.4	11.6	6.7	0.4
0.002	0.016	0.002	0.115	0.0002	0.004	0.702

DISCUSSIONS

In our evaluation of handwritten strokes, results showed that local maximum writing pressure, and local minimum writing velocity for strokes in particular are sensitive to the cog-



Figure 5. Relationship between F value, p value and α .

nitive load of the writer. The relationship between writing velocity and writing pressure is also observed when tracking the writing process. Maximum writing pressure usually appears at the beginning, corners and end of strokes, where the minimum writing velocity is also observed. This could be explained in terms of the importance of the stroke parts for English [8], and writers paid more cognitive-related attention when writing these parts than others. On the contrary, when writing the straight parts of a stroke, there is no change in the direction of the pen trace, so the writer might spare extra cognitive resources in other writing correlated issues, e.g. reviewing the previous stroke. If this is true, the cognitive load during writing should be fluctuating during the process of a stroke, correlated with the tempo of stroke construction.

The α coefficient takes effect in classifying the strokes in to long and short ones, and in our experiments the long strokes have been shown to carry more load-related information than the short ones. Varying the α coefficient could help us to find the optimal set of strokes for cognitive load evaluation, but the α value should not be too large, otherwise a large proportion of strokes will be removed, which increases the p-value. As a consequence, we limit the value of α in the range from 0.01 to 0.2.

IMPLICATIONS FOR INTERFACE DESIGN

Research into cognitive load during handwriting is important for improving the performance and experience of users in pen-based interactions [4, 6]. As a non-intrusive supporting component, a cognitive load measurement module can provide a useful reference to control the difficulty level of writing tasks, to ensure users can write with high efficiency without a heavy cognitive burden. It may also be helpful when teaching young kids to write, e.g. for diagnosing problems with spelling or stroke construction.

CONCLUSION

Strokes and inter-strokes provide a comprehensive record of writing behavior, which have been found herein to convey rich information reflective of the cognitive load of a writer. In this paper we statistically examined the velocity, length and pressure information as stroke-level features against different cognitive demands, and discovered that the local maximum pressure with the local minimum velocity information for strokes are relatively reliable indicators of cognitive load compared with other features. The introduction of a simple stroke length threshold was effective in reducing the feature variability inherent in short strokes, further improving discrimination. As an extension of the current work, in future we will apply probability models to the selected features for cognitive load classification, and if this can be used to monitor and thus fine-tune the cognitive load during writing, it will not only enhance the writing experience on individual basis, but also boost the development of cognitive aware devices in the broader application area.

REFERENCES

- A. Baddeley. Working memory and language: an overview. *Journal of Communication Disorders*, 36:189–208, 2003.
- 2. A. Baddeley, N. Thomson, and M. Buchanan. Word length and the structure of short-term memory. *Journal of Verbal Learning and Verbal Behavior*, 14:575–589, 1975.
- 3. C. Bereiter and M. Scardamalia. *The psychology of written composition*. Lawrence Erlbaum, 1987.
- G. Frankish, R. Hull, and P. Morgan. Recognition accuracy and user acceptance of pen interfaces. In *Proc. CHI*, pages 503–510, 1995.
- 5. J. R. Hayes and L. S. Flower. Identifying the organization of writing process. In *Cognitive Processes in Writing*, pages 3–30. 1980.
- S. John, J. J. Merrienboer., and F. G. Paas. Cognitive architecture and instructional design. *Educational Psychology Review*, 10(3):251–296, 1998.
- R. T. Kellogg. Training writing skills: A cognitive developmental perspective. *Journal of Writing Research*, 1(1):1–26, 2008.
- 8. G. Luria and S. Rosenblum. Comparing the handwriting behaviors of true and false writing with computerized handwriting measures. *J. of Applied Cognitive Psychology*, (published online), 2009.
- 9. T. Olive. Working memory in writing: Empirical evidence from the dual-task technique. *European Psychologist*, 9(1):32–42, 2004.
- 10. S. Ransdell and C. M. Levy. Writing, reading and speaking memory spans and the importance of resource flexibility. In *The Cognitive Demands of Writing*, pages 99–113. 1999.
- 11. N. Ruiz, R. Taib, Y. Shi, E. Choi, and F. Chen. Using pen input features as indices of cognitive load. In *Proc.* 9th ICMI, pages 315–318, 2007.
- 12. R. Vanderberg and H. L. Swanson. Which components of working memory are important in the writing process? *Reading and Writing: An Interdisciplinary Journal*, 20(7), 2007.