Further Investigating Pen Gesture Features Sensitive to Cognitive Load

Ling Luo, Ronnie Taib
National ICT Australia
13 Garden Street
Eveleigh, NSW 2015, Australia
{ling.luo, ronnie.taib}@nicta.com.au

Lisa Anthony, Jianwei Lai
UMBC Information Systems
1000 Hilltop Circle
Baltimore, MD 21250, USA
{lanthony, jianwei1}@umbc.edu

ABSTRACT

A person's cognitive state and capacity at a given moment strongly impact decision making and user experience, but are still very difficult to evaluate objectively, unobtrusively, and in real-time. Focusing on smart pen or stylus input, this paper explores features capable of detecting high cognitive load in a practical set-up. A user experiment was conducted in which participants were instructed to perform a vigilance-oriented, continuous attention, visual search task, controlled by handwriting single characters on an interactive tablet. Task difficulty was manipulated through the amount and pace of both target events and distractors being displayed. Statistical analysis results indicate that both the *gesture length* and *width over height ratio* decreased significantly during the high load periods of the task. Another feature, the symmetry of the letter '*m*', shows that participants tend to oversize the second arch under higher mental loads. Such features can be computed very efficiently, so these early results are encouraging towards the possibility of building smart pens or styluses that will be able to assess cognitive load unobtrusively and in real-time.

Author Keywords

Gesture features; cognitive load; user study; pen-based input; vigilance; attention; human-computer interaction.

ACM Classification Keywords

H.5.2. [Information interfaces and presentation]: User interfaces – Evaluation/methodology; Input devices and strategies.

INTRODUCTION

Cognitive load represents the mental effort imposed on a participant's cognitive system when performing a particular task [8]. When a task demands very high quality of performance, like air traffic control or machine operation, the degradation in quality caused by too-high cognitive load may lead to accidents or serious consequences. For example, it was found that the lack of "at least some cognitive availability and understanding of the situation" may have led pilots to ignore continuing alarms during the fatal accident [3] on the Rio to Paris flight AF447, which disappeared over the Atlantic. Being able to assess cognitive load in real-time can allow intervention when levels become too high and can prevent such accidents. Furthermore, in other less safety-critical settings, cognitive load assessment can still be useful. For example, in an educational setting, assessing students' cognitive states could help teachers to better control teaching content and pace, and thus improve learning effectiveness and efficiency. Therefore, research into ways to estimate human cognitive state and capability is critical to improving the quality of human computer interaction, increasing task performance, and developing optimized human-decision-support applications. Practically speaking, it is important to look for good methods of measurement and estimation, which will be not only accurate, but also unobtrusive and real-time, so they can reduce noise, unpredictable factors, and disruptions to the cognitive process.

There are four different ways to measure cognitive load explored in the literature [8]: (i) subjective assessment techniques; (ii) task and performance based techniques; (iii) behavioral measurements; and (iv) physiological measurements. Typically, more than one method is used, so their combination can improve accuracy. In our user study, both behavioral and physiological measures were used, but this paper focuses on behavioral measurement. Behavioral measurement is here defined as non-obtrusive data collection during natural multimodal interaction. In this paper, we report on a user study in which participants performed a vigilance-oriented, continuous attention, visual search task [2], controlled by handwriting single characters on an interactive tablet. We examine behavioral features of the pen gestures input in two cognitive load levels to identify efficient features that can be computed in real-time.

Prior work has looked at how gesture features are related to changes in cognitive load induced by task complexity [10] and task memory demands [11] for simple shapes (circles and crosses). That work found that features such as shape degeneration [10] and pen trajectory duration, speed, and length [11] are correlated with increases in cognitive load. Our study expands on that prior work by (a) using another way to induce cognitive load (i.e., task speed), (b) probing

IUI 2013 Workshop: Interacting with Smart Objects, March 18, 2013, Santa Monica, CA, USA

additional pen input features (i.e., pressure, bounding box size, and geometric features such as the symmetry of letter ‘m’), and (c) using a wider variety of pen input shapes (i.e., letters). Prior analysis based on the same experiment reported in this paper [2] found that *gesture duration*, *number of points* (correlated with *speed*), and *gesture length* were all significantly affected by cognitive load.

In this paper we extend the list of features examined and find that features such as *normalized gesture length* and *width over height ratio* decreased significantly during the high load periods of the task. Also, after examining visualizations of all recorded gestures, we noticed that some letters seemed to exhibit different forms as cognitive load varied, for example, the symmetry of the letter ‘*m*’, showed that participants tend to oversize the second arch under higher mental loads, as illustrated in Figure 1. Such features can be computed very efficiently, so these early results are encouraging towards the possibility of building smart pens or styluses that will be able to assess cognitive load unobtrusively and in real-time.

Figure 1. Sample Input from two Participants.

MOTIVATION

Cognitive Load Impacts Performance

Cognitive load is closely related to the capacity of *working memory*, which, according to the cognitive load theory [8] refers to the brain system providing temporary storage of the input necessary to acquire information, process it, and prepare feedback actions when completing tasks. The capacity of working memory is limited; accepted estimates of the amount of information it can hold at a time is restricted to 7 ± 2 items [5]. When cognitive load exceeds working memory’s capacity limit, the participant’s performance starts to degrade [10]. This degradation could lead to longer reaction times, higher error rates, and decreased control of actions and movements [11]. Therefore, we are examining methods of unobtrusively detecting cognitive load spikes, such as using the gesture features discusses in this paper, to allow systems to intervene to reduce negative impacts of higher load.

Pen Gesture Input as an Unobtrusive Sensor

Pen gestures are input produced through a pen or stylus during a user’s interaction with a computer [9]. Previous research indicated that some gesture features can be used as indicators of cognitive load imposed by a task [10, 11, 12]. Compared to other modalities, pen gesture offers benefits such as naturalness for the user, low intrusiveness, and the possibility to automatically analyze data in real-time. It can capture variations in performance implicitly without interrupting the task, and the data is available for analysis once the current gesture is finished [2, 10]. In the market, there are already digital pen and paper systems, such as Anoto, which support gesture capture and online/offline analysis [1]. Prior research has shown that for specific tasks, for example, math problem solving by high school students, pen-based systems provide cognitive support and produce better learning results than traditional keyboard and mouse graphical user interfaces [6], so we believe there is similar potential in developing pen-based adaptive systems for both safety-critical and other tasks.

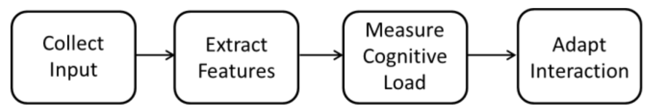

Figure 2. Smart Pen Workflow.

In our work, we use the workflow in Figure 2 as the accepted model of how smart pens process and react to user input. Based on the written content, the pen extracts geometric features and automatically classifies them according to pre-built models, possibly trained for specific users. Depending on the application, high cognitive load detection can be used to trigger alerts, e.g. when a mission-critical operator is experiencing high cognitive load, a manager may be alerted to provide additional resources or a break for the operator. In other contexts, the pace of the content can be adapted, e.g. when a student is learning online content using an interactive tablet.

Simulating Real-World Tasks

Previous research has used a variety of experiment designs to collect pen gesture input and correlate it to cognitive load. One example is the map tasks in [7, 10], in which participants are asked to look for routes and organize a green light corridor on a city map. There are also tasks instructing participants to compose sentences from three predefined words [12] or to solve mathematics problems [6], requiring participants to write down all the intermediate processes. In this paper, we use a continuous attention, visual search task, which simulates real-world vigilance tasks such as air traffic control or information analysis. Our task has two cognitive load levels, and our analysis focuses on specific geometric features of the single letter inputs.

EXPERIMENT DESIGN

Participants performed a vigilance-oriented continuous attention and visual search task [2]. During the experiment, arrows facing one of four directions (↑, ↓, ← and →) were displayed sequentially (with some overlap) on the screen, and each of them was companied by a text identifier underneath. There were 12 possible identifiers: {**a**lpha,

bravo, **d**elta, **e**cho, **g**olf, **h**otel, **i**ndia, **l**ima, **m**ike, **o**scar, **r**omeo, **z**ulu}. At any moment, all the identifiers visible on the screen were unique. The participants were instructed to detect any arrow facing down ↓ while ignoring all the other objects (distractors) on the screen, and to write down the first letter (the highlighted character in the above list) in a "gesture drawing space" located at the bottom right of the screen. The user interface is shown in Figure 3.

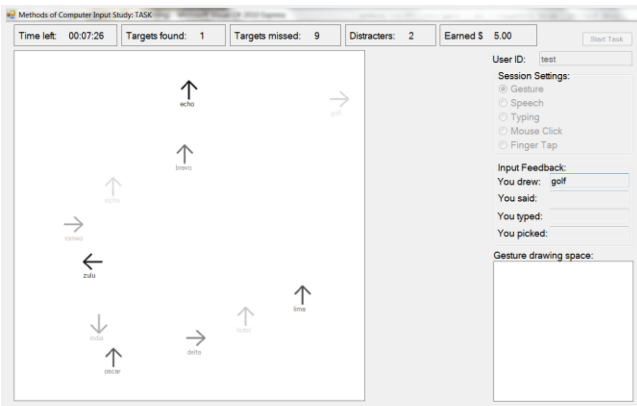

Figure 3. Experiment User Interface.

There were two levels of cognitive load in the task, labeled *Normal* and *High*, and the level was manipulated by controlling the time interval between arrows and the frequency of occurrence of target objects. During High periods, the higher frequency of actions required increased intrinsic cognitive load, and the higher number of distractors increased extraneous load, so this condition is labeled *high load* in our analysis.

There were 12 participants (7 males and 5 females) who used the pen modality to perform the task. Two participants (both female) were excluded from data analysis due to post-hoc analysis showing that in-task recognition of their gestures had had very low accuracy (i.e., less than two standard deviations below the mean), leaving an N of 10.[1] The equipment used to collect gestures during the experiment was a Tablet PC (Fujitsu Lifebook T-900). The system collected all task information, including the target and distractor objects and their order of appearance (same sequence for every user), task performance (recognized result and the system response: True Hit, Miss, etc.), and pen input trajectories. The analysis is based on these trajectories, which store each sampled gesture point (timestamp, character written, coordinates, pressure), and whole gesture information (start and end indicators).

[1] This low accuracy could have caused an additional load on the user and, for this investigation, we wanted to isolate the load caused by the task difficulty manipulation only.

DATA ANALYSIS RESULTS

Bounding Box

The term *bounding box* refers to the smallest box that can contain a gesture entirely. The bounding box has several geometrical features, including *height*, *width*, *area* and the *width to height ratio* (width/height, which is also $\cot\ \alpha$ in Figure 4).

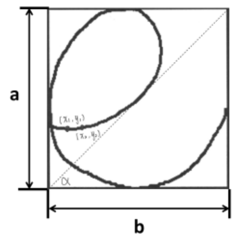

Figure 4. Defining a bounding box for the gesture.

The mean *width over height ratio* across all gestures is 0.851 (σ = 0.203, N = 10) in the Normal condition and 0.768 (σ = 0.169, N = 10) in the High condition. The decreasing trend of mean width over height ratio between Normal and High is consistent for all but one participant. The result of a two-tailed t-test showed the width over height ratios varied significantly between Normal and High ($t(9) = 3.05$, $p < 0.05$).

However, when comparing all letters together, we must take into account that the generic letter shapes exhibit different width over height ratios. For example, the width over height ratio for **l**ima and **i**ndia are quite small compared to the ratio for **m**ike. Moreover, the frequency of occurrence of different letters was not completely balanced between the two conditions because there were many more targets in the High condition. For example, **l**ima or **i**ndia appear 9 times as the targets in the High condition, but only once in the Normal condition. Therefore, the significant result above may be partly due to a bias linked to targets (letters) with smaller width over height ratio occurring more frequently in the High condition.

In order to mitigate the effect of the shape of the letter, the ratio of each gesture was normalized by a "standard" ratio for each specific letter, reflecting the shape of the letter. The standard ratio is calculated as the average width over height ratio across all occurrences of that letter from all participants, across both conditions. (The limited number of gestures from each participant did not permit us to establish a standard ratio per letter per participant.) Each individual occurrence of a letter is normalized by dividing its ratio by the standard ratio for that letter:

$$normalized\_ratio = original\_ratio / standard\_ratio$$

The standard ratios in Table 1 validate that different ratios apply to different letters. For example, **i**ndia and **l**ima have relatively small values, whereas **m**ike and **z**ulu are larger.

After that processing, a two-tailed t-test was used once more on the normalized width over height ratios, but found no significant differences between Normal and High conditions when controlling for properties of the letter entered ($t(9) = -0.32$, *n.s.*).

Letter	Standard Ratio	Letter	Standard Ratio
alpha	1.13	**i**ndia	0.19
bravo	0.62	**l**ima	0.40
delta	0.60	**m**ike	1.48
echo	1.03	**o**scar	0.79
golf	0.53	**r**omeo	0.97
hotel	0.64	**z**ulu	1.87

Table 1. Standard Letter Ratio for each letter in the study.

Gesture Pressure

Every point of a gesture has a pressure value as sensed by the hardware during input, and we define *gesture pressure* as the mean pressure value across all points of that gesture.

The pressure sensing capability malfunctioned during the study, so another two participants had to be excluded from just this analysis (leaving an N of 8). The mean values of *gesture pressure* for the remaining 8 participants were 25,743 screen dots in the Normal condition and 26,164 dots in the High condition (the TabletPC range was [0~32,767 screen dots]). A similar normalization process to the one described for the bounding box was used here:

$$normalized\_pressure = current\_pressure / standard\_pressure$$

where the standard pressure for a specific letter is calculated as the average pressure across all occurrences of that letter from all participants, across both conditions. The mean values after normalization were 0.936 ($\sigma = 0.141$, N = 8) and 0.938 ($\sigma = 0.126$, N = 8) for Normal and High conditions, respectively. These values indicate that participants tended to press slightly harder in the High condition than in the Normal condition. However, a two-tailed t-test indicated that this trend was not significant ($t(7) = -0.26$, *n.s.*), reducing the utility of this feature for detecting changes in cognitive load.

Gesture Length

Gesture length is the sum of the Euclidean distances between every two consecutive points in a single gesture, which is computed by the following formula:

$$\sum_{i=0}^{n}\left(\sqrt{(x_{i+1}-x_i)^2+(y_{i+1}-y_i)^2}\right)$$

where (x_i, y_i), (x_{i+1}, y_{i+1}) are the coordinates for two consecutive points, and the gesture length is the sum of Euclidean distances between every two consecutive points.

Extending prior results [2], here we normalize length using a standard length (defined as the mean length across all occurrences of each letter from all participants, across both conditions). After such normalization, the mean values of *gesture length* were 1.08 ($\sigma = 0.335$, N = 10) for Normal and 0.93 ($\sigma = 0.29$, N = 10) for High, indicating a shorter gesture length in the High condition which was significant by a two-tailed t-test ($t(9) = 3.79$, $p < 0.05$), further supporting this feature's relationship to cognitive load.

Symmetry of letter 'm'

To examine whether our anecdotal observations of differences in the symmetry of the letter '*m*' were supported by the data, we compared the widths of the left arch and right arch in High and Normal load conditions.

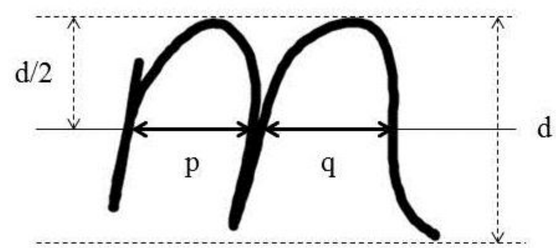

Figure 5. Symmetry of the letter 'm'.

Figure 5 illustrates how to extract the feature. The first step is to find the bounding box of the letter, then compute the distance *d* between the top and bottom lines, and then draw a straight horizontal line in the middle of the box, *d* / 2 away from top and bottom lines. Typically, there will be about 5 crossing points between the line and the gesture, but the actual number of crossing points may vary according to the way the letter was written. After that, the ratio between the two longest segments of the horizontal line, *q* and *p* is used to check the symmetry of this letter:

$$Symmetry = q / p$$

The closer to 1 this value is, the more symmetrically the letter was formed; otherwise, the left arch may be much larger or smaller than the right one. Although this method is only an estimation of symmetry, it is easy to implement and can be computed very efficiently.

The mean *symmetry* value in the Normal condition was 0.88 ($\sigma = 0.751$, N = 10), which was smaller than 1.28 ($\sigma = 0.418$, N = 10) observed in the High condition, but a two-tailed t-test ($t(9) = -1.59$, *n.s.*) showed that this difference was not significant. We also assessed the same feature at different horizontal cross sections of the letter, for example moving the crossing line up or down by 10% away from the middle height, but there were still no significant differences in the t-test results. Hence, while there was a trend for participants to write the right arch wider than the left arch under higher mental load, the fluctuation was not significant from a statistical viewpoint.

DISCUSSION

As mentioned, prior work has examined what gesture features are related to changes in cognitive load [10, 11, 12]. Our study expands on prior work by (a) exploring another way to induce cognitive load (i.e., task speed), (b) probing additional pen input features (i.e., pressure, bounding box size, and geometric features such as the symmetry of letter 'm'), and (c) using a wider variety of pen input shapes (i.e., letters). *Shape degeneration* [10] and pen trajectory *duration*, *speed* and *length* [11] have been found in other experiments to correlate to cognitive load. Previous analysis based on the same experiment reported in this paper [2] also found that *gesture duration*, *number of points* (correlated with *speed*), and *gesture length* were significantly affected by difficulty-induced cognitive load. We extended this past work by looking at even more gesture input features and explored how they responded to changes in cognitive load, all in search of a canonical set of gesture features that can be efficiently computed in real-time systems and are responsive to all types of cognitive load.

In the feature analysis, normalization over the standard value of the features per letter played a critical role, by effectively decreasing the impact of factors that vary from letter to letter. These factors included, among others, bounding box size and gesture length. Normalization also compensated for the unbalanced letter distribution of the task design. During the analysis, features like *width over height ratio* and *gesture length* showed statistically significant relationships to cognitive load originally. However, after normalization, differences in the *width over height ratios* between load conditions were not statistically significant, indicating that previous positive results may be affected by the letters that were input.

In summary:

- Both *gesture length* and *normalized gesture length* exhibited significant relationships with cognitive load.
- The bounding box *width over height ratio* showed significant differences as the load increased, but after normalization (over standard letter ratio), it was not significantly affected.
- Neither *gesture pressure* nor *normalized gesture pressure* were significantly affected by increasing cognitive load.
- The symmetry of the letter 'm' exhibits an increased trend in the high load condition, but it is not significant.
- From a practical perspective, a simple feature like *gesture length* can estimate cognitive state unobtrusively and can be computed very efficiently, making it a good candidate for a smart pen or stylus.

The dimensions of the gesture bounding box are important features. The results showed declining trends during High load (although not significant), indicating that cognitive load may impact fine-grained handwritten production, although degeneration in gesture shapes were observed in past research [10]. However, it has also been postulated that handwriting skills are based on primary communication systems (*folk psychology*) and hence should not get taxed by variations in cognitive load [4]. Further experimentation is required to determine which explanation prevails. In particular, in future explorations, we would change the experiment design in order to balance written input and collect more samples of each type of letter from more participants. Such changes would also allow us to group participants based on *post hoc* analysis of their individual inputs, for example, users who tend to write larger or smaller or experience larger impacts due to cognitive load.

The symmetry of the letter '*m*' is an early attempt at exploring specific geometric features of individual pen gesture shapes. The results highlighted an increasing trend during High load (although not significant), which means higher cognitive load may have an effect on the way people form their letters, especially towards the end of the gesture: the right arch tended to increase in width compared to the left one under higher load. Again, a possible reason for the non-significant trend might be that there are individual differences among participants which could be explored by capturing more data in the future.

CONCLUSIONS AND FUTURE WORK

Focusing on smart pen or stylus input, this paper explores features capable of detecting high cognitive load in a practical set-up. Participants performed a vigilance-oriented, continuous attention, visual search task, controlled by handwriting single characters on an interactive tablet. Task difficulty was manipulated through the amount and pace of both target events and distractors being displayed. Both *gesture length* and *width over height ratio* decreased significantly in the high load session. Another feature, the symmetry of the letter '*m*', showed that participants tend to write the right arch wider than the left one under higher mental load. *Gesture pressure* and *bounding box size* were not significantly affected by cognitive load, though. Features such as *gesture length* can be computed very efficiently, making them good candidates for a smart pen or stylus to assess cognitive load unobtrusively in real-time.

In the future, more research will be needed to validate these results and to explore more gesture features to detect changes in cognitive load robustly. For example, other geometric features will be explored, such as angles or curvature of segments composing letters.

In order to ensure the high load we impose on the participants is actually high enough, we are planning to modify the experiment design through different timings and distracters, and by adding other sources of load, for example, using a dual task methodology. We will also balance the number of individual letters collected under each condition, and increase the number of inputs elicited so we can analyze the gesture data on a per-user basis.

We believe that a combination of features will be required to estimate cognitive load from handwritten input. Once identified, this work can lead to the construction of smart pens and styluses that will be able to monitor a user's performance and adapt the task at hand implicitly to moment-to-moment fluctuations in cognitive load.

ACKNOWLEDGEMENTS

The authors acknowledge Andrew Sears, Patrick Carrington, Peng Chu, and Christopher Kidd for contributions to the original study. We also thank Northrup Grumman for partial support of the original study.

REFERENCES

[1] Anoto website. http://www.anoto.com/about-anoto-4.aspx.

[2] Anthony, L., Carrington, P., Chu, P. and Sears, A., Gesture Dynamics: Features Sensitive to Task Difficulty and Correlated with Physiological Sensors, *ICMI 2011, MMCogEmS*.

[3] BEA, Final Report On the accident on 1st June 2009 to the Airbus A330-203 registered F-GZCP operated by Air France flight AF 447 Rio de Janeiro - Paris (2012).

[4] Geary, D. C., Principles of evolutionary educational psychology. *Learning and Individual Differences*, *12*, 317-345.

[5] Miller, G., The Magic Number Seven, Plus or Minus Two: Some Limits on Our Capacity for Processing Information, *The Psychological Review, vol. 63* (1956), 81-97.

[6] Oviatt, S., Human-centred Design Meets Cognitive Load Theory: Designing Interfaces that Help People Think, *MM 2006*.

[7] Oviatt, S., Coulston, R. and Lunsford, R., When Do We Interact Multimodally? Cognitive Load and Multimodal Communication Patterns, *ICMI 2004.*

[8] Paas, F., Tuovinen, J.E., Tabbers, H. and Van Gerven, P., Cognitive Load Measurement as a Means to Advance Cognitive Load Theory, *Educational Psychologist, 38(1)* (2003), 63-71

[9] Rubine, D., Specifying Gestures by Example, *Computer Graphics*, *Vol. 25*, *No. 4* (1991).

[10] Ruiz, N., Taib, R., Shi, Y., Choi, E., and Chen, F., Using Pen Input Features as Indices of Cognitive Load, *ICMI 2007*.

[11] Ruiz, N., Feng, Q.Q., Taib, R., Handke, T. and Chen, F. Cognitive skills learning: pen input patterns in computer-based athlete training. *Proc. ICMI-MLMI, 2010*, ACM Press (2010).

[12] Yu, K., Epps, J. and Chen, F., Cognitive Load Evaluation of Handwriting Using Stroke-level Features, *IUI 2011*.