Gesture Dynamics: Features Sensitive to Task Difficulty and Correlated with Physiological Sensors

Lisa Anthony, Patrick Carrington, Peng Chu, Christopher Kidd, Jianwei Lai
UMBC Information Systems
1000 Hilltop Circle
Baltimore MD 21250 USA
{lanthony, carpat1, pengchu1, kidd1, jianwei1}@umbc.edu

Andrew Sears
B. Thomas Golisano College of Computing and Information Sciences
Rochester Institute of Technology
20 Lomb Memorial Drive
Rochester, NY 14623
andrew.sears@rit.edu

ABSTRACT

This paper presents preliminary results regarding which features of pen-based gesture input are sensitive to cognitive stress when manipulated via changes in task difficulty. We conducted a laboratory study in which participants performed a vigilance-oriented continuous attention and visual search task. Responses to the search stimuli were entered via pen gestures (e.g., drawing a letter corresponding to the stimulus). Task difficulty was increased during predefined intervals. Participants' input behaviors were logged, allowing for analysis of gesture input patterns for features sensitive to changes in task difficulty. We also collected physiological sensor readings (e.g., skin temperature, pulse rate, and respiration rate). Input behavior features such as gesture size and pen pressure were not affected by task difficulty, but gesture duration and length were affected. Task difficulty also affected physiological sensors, notably pulse rate. Results indicate that both gesture dynamics and physiological sensors can be used to detect changes in difficulty-induced stress.

Categories and Subject Descriptors

H.5.m [**Information Systems**]: Information Interfaces and Presentation (e.g., HCI) – *miscellaneous*.

General Terms

Experimentation, Human Factors.

Keywords

Gesture dynamics, cognitive stress, user study, pen-based input, physiological sensors, affective computing, vigilance, attention, human-computer interaction.

1. INTRODUCTION

Understanding an individual's cognitive state is a key pursuit of affective computing. Related work has used a variety of signals to detect the user's cognitive state, including physiological measures such as electrodermal activity (EDA) / galvanic skin response (GSR), heart rate, eye movements, pupil dilation, or a combination of these sensors [3, 10]; behavioral measures such as posture [5], keystroke dynamics [11], acoustic features of speech [12], and mouse or touchpad motion or pressure [4, 9]; and self-report data such as cognitive load questionnaires [2]. Related work has also attempted to detect a variety of cognitive states, including cognitive load [6] and negative affect or emotion [4].

ICMI MMCogEmS 2011, November 17, 2011, Alicante, Spain.

We add to the existing body of work by presenting preliminary results from an empirical study that focuses on behavioral measures from pen-based gesture input and their sensitivity to cognitive stress manipulated via task difficulty. Pen input features have previously been examined as indicators of cognitive load caused by task complexity [6] and task memory demands [7] for simple shapes (circles and crosses). Our study augments prior work by (a) exploring another method to induce cognitive load (i.e., task speed), (b) probing additional pen input features (i.e., pressure and bounding box size), and (c) using a wider variety of pen input shapes (i.e., letters).

We conducted a laboratory study in which participants performed a vigilance-oriented continuous attention and visual search task, which we call the "target-finding task" (Figure 1). Participants performed the task using one of five input modalities: gesture, speech, typing, mouse click, and finger tap. This paper considers only the gesture modality data, focusing on identifying pen input features that are sensitive to changes in task difficulty. We call these features "gesture dynamics" after the analogous "keystroke dynamics" in prior related work [11]. We also collected physiological sensor readings (e.g., skin temperature, pulse rate, and respiration rate). In this paper, we provide preliminary results on the sensitivity of task performance, gesture dynamics, and physiological data to difficulty-induced stress, discuss implications, and describe future work motivated by these results.

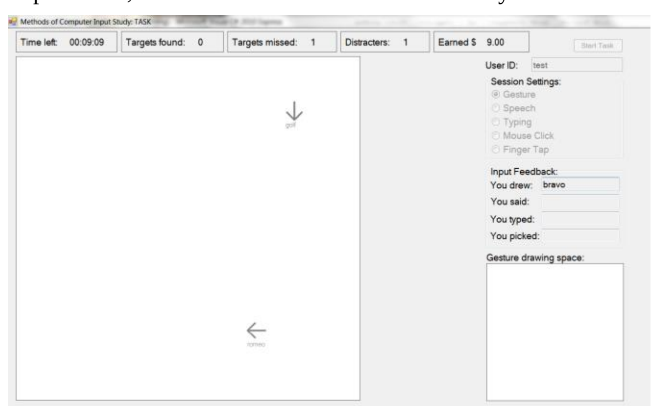

Figure 1. Snapshot of target-finding task, non-stress period.

2. EXPERIMENT DESIGN

The study is a two-factor mixed design in which modality is a between-subjects factor and "stress" (caused by task difficulty) is a within-subjects factor. The analyses in this paper focus exclusively on the gesture modality data. We developed a fixed-attention target-finding task modeled after single-stream vigilance tasks from the literature [1]. These vigilance tasks tend to involve one stream of input in which only one target or distracter is visible (or audible) at any given time, and responses involve a simple key or button press. In order to increase the richness of the features we could extract from each response, we increased the complexity of the task by adding an element of visual search. Objects (arrows facing the four cardinal directions) faded in and out gradually onscreen. Users were instructed to select downward-facing arrows (targets), and ignore all other arrows (distracters). Objects could appear anywhere on the blank field, for a random duration of between two and four seconds, with a random interval of between one and three seconds between objects. As a result, multiple objects (including targets) could be onscreen at once. Every participant experienced the exact same sequence of objects.

Task difficulty was manipulated by the regular occurrence of "stress periods," defined as follows. First, the interval between subsequent objects was decreased by 80%, causing many more objects to be onscreen at once. Second, the interval between targets was decreased from approximately 45 seconds during non-stress periods (with other distracter objects appearing in between these intervals) to approximately 4.5 seconds during stress periods. The range of possible durations that objects were onscreen remained unchanged for both the stress and non-stress periods. In each 10-minute task, three 1-minute stress periods occurred at regular intervals. A snapshot in time from a non-stress and stress period is shown in Figures 1 and 2, respectively.

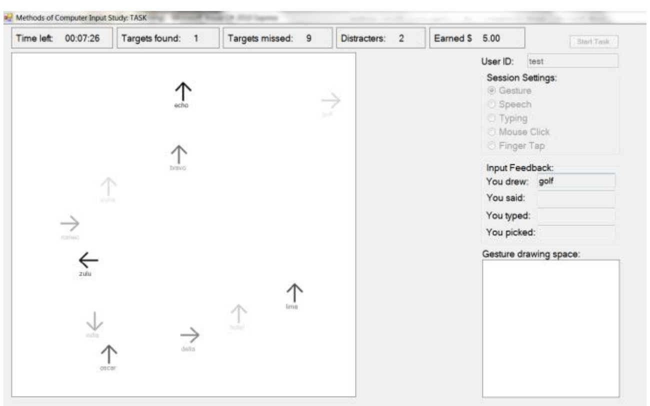

Figure 2. Snapshot of target-finding task, stress period.

All objects were accompanied by a short one-word identifier (ID) displayed just below the object. This ID was chosen from a set of IDs selected from the NATO phonetic alphabet. Here are the 12 IDs used: {"alpha", "bravo", "delta", "echo", "golf", "hotel", "india", "lima", "mike", "oscar", "romeo", "zulu"}. The IDs for all objects onscreen at any one point in time were unique, but object IDs were repeated during the session and were used for both targets and distracters. In the gesture modality, users drew the first letter of the ID in a space at the bottom right of the screen.

3. DATA ANALYSIS AND RESULTS

Twelve participants (7 male) completed the task using the gesture modality. All were members of the university community who answered advertisements via email, flyers, or website announcements for the study. Two participants (both female) were excluded from data analysis because of low recognition accuracy for their gestures because recognition errors could cause additional stress and, for this preliminary investigation, we wanted to isolate the stress caused by task difficulty. The remaining 10 participants had an average age of 23.5 years (min=20, max=29, σ=3.3). All were right-handed and had a minimum of 7 years of computer experience. Most (8) indicated they used pen and stylus input on computers "sometimes," while the others indicated they had tried it only once or twice. Most (9) had played games similar to the target-finding task before, and four of those people indicated they played such games "often."

3.1 Task Performance

We analyzed task performance to understand the overall pattern of responses and whether any task performance measures were sensitive to task difficulty. See Table 1 for descriptive statistics. In all tables, shaded dependent variables (also marked with an *) show significant contrasts at the $p<0.05$ level. In all cases, means were computed across all data points across all participants, unless otherwise specified.

Table 1. Descriptive statistics for task performance measures.

Dependent Variable	Stress / Normal	Mean	Standard Deviation (σ)	N
Response Latency (ms)	Normal	1352.9	361.7	72
	Stress	1418.6	360.4	312
Response Latency Proportion	Normal	0.47	0.11	72
	Stress	0.47	0.11	312
* Response Duration (ms)	Normal	652.6	193.4	72
	Stress	552.7	192.8	312

3.1.1 Targets and Distracters Hit

Overall, there were 47 targets and 609 distracters. On average, participants found 38.4 targets (min=32, max=46, σ=4.67), missed 8.6 targets (min=1, max=15, σ=4.67), and hit 3.6 distracters (min=1, max=8, σ=2.27). Although some hit distracters were caused by misrecognitions, many (42%) were intentional hits caused by visual search challenges of the task. A total of 486 gestures were recorded (mean=48.6, min=42, max=58, σ=5.85).

3.1.2 Response Latency and Duration

We compared whether stress had an effect on the absolute latency for correct responses (e.g., the time between the onset of the target and the onset of the response), but found no significant difference between stress and normal periods ($F_{1,382}=1.94$, *n.s.*). (All tests are one-way ANOVAs on the within-subjects factor of stress unless otherwise stated.) There was also no significant difference in response latency as a proportion of the object's duration on screen ($F_{1,382}=0.01$, *n.s.*). For response latency, there was an interesting trend suggesting that users may begin their response more quickly during non-stress periods, but validating this trend would require additional data. We also analyzed whether there was an effect of stress on the duration of a response (e.g., the time between the onset and completion of the response) also, and found that the duration was significantly shorter during stress periods ($F_{1,382}=15.70$, $p<0.0001$).

These results imply that users may be more relaxed during non-stress periods, entering gestures more slowly (e.g., with longer durations), but beginning their responses more quickly after they appeared (e.g., with shorter latencies). To explain: during normal periods, fewer objects were onscreen, allowing users to focus on each object more quickly, decreasing the latency for responses. When many objects were onscreen, users took longer to identify which objects were targets, increasing the latency for responses. Then, since users knew another target may appear soon in stress periods, they entered their responses more quickly.

3.2 Gesture Dynamics

We analyzed several features of the gestures input by participants to identify those that were sensitive to changes in task difficulty. Table 2 includes descriptive statistics for this section.

Table 2. Descriptive statistics for gesture dynamics measures.

Dependent Variable	Stress / Normal	Mean	Standard Deviation (σ)	N
* Gesture Duration	Normal	624.5	236.1	95
	Stress	535.0	216.8	391
Gesture Speed	Normal	3.46	1.18	95
	Stress	3.72	1.21	391
* Number of Points	Normal	65.1	22.5	95
	Stress	53.2	20.9	391
* Gesture Length	Normal	224.8	97.8	95
	Stress	199.0	98.3	391
Gesture Height	Normal	79.3	35.8	95
	Stress	80.5	34.7	391
Gesture Width	Normal	56.0	29.9	95
	Stress	52.4	31.7	391
Gesture Area	Normal	4632.5	3474.8	95
	Stress	4409.5	3646.0	391
Gesture Average Pressure	Normal	25965.3	4378.3	74
	Stress	26301.3	4059.2	300
Gesture Pressure Per Point	Normal	20769.1	11908.3	6639
	Stress	20849.0	12337.9	20358

3.2.1 Gesture Duration and Speed

Not surprisingly, we found a significant difference in the gesture duration (computed by time from the first gesture point to the last gesture point) between stress and normal periods ($F_{1,484}$=12.58, p<0.0001). We also found a marginally significant difference in gesture speed (computed by dividing the length of a gesture by the number of points collected within the gesture) between stress and normal periods ($F_{1,484}$=3.74, p<0.055). In addition, the number of points collected per gesture is related to speed because the faster the user writes, the fewer points will be collected due to sampling rate. We found that number of points per gesture was significantly fewer in stress periods ($F_{1,484}$=24.34, p<0.0001). All of these results indicate that users drew gestures faster during stress.

3.2.2 Gesture Length and Size

We analyzed whether there was an effect of stress on the length of a gesture (e.g., the path length from the first point of the gesture to the last point), finding that the length was significantly shorter during stress periods ($F_{1,484}$=5.30, p<0.05). We further compared whether there was an effect of stress on the size of a gesture as determined by its bounding box (e.g., the height and width of the smallest box that entirely contains the gesture) but found no significant differences for height ($F_{1,484}$=0.09, *n.s.*), width ($F_{1,484}$=1.06, *n.s.*), or area ($F_{1,484}$=0.29, *n.s.*).

These results taken together with those for gesture duration and speed indicate that the users did not draw smaller or larger gestures during stress periods, but they did draw them faster. During the stress periods, more objects were onscreen at once and so users may have felt rushed to enter responses more quickly so they could prepare for the next target. Users felt this pressure even though the duration of the objects was the same in both periods.

3.2.3 Gesture Pressure

We also compared whether there was an effect of stress on the pen pressure exerted by the user during the gesture. We collected pen pressure using the real-time stylus methods from the Windows Tablet PC SDK. For this analysis, we excluded two participants for whom the pressure sensing capability malfunctioned, yielding a constant pressure. For the remainder of data, a simple main-effects ANOVA did not find a significant difference between stress and normal for the average pressure during a gesture ($F_{1,372}$=0.39, *n.s.*) or for per-point pressure ($F_{1,26995}$=0.21, *n.s.*).

However, we hypothesized that gesture pressure might be user-dependent. For average gesture pressure, including the user identifier as a factor does not show a significant effect for stress level ($F_{1,7.2}$=2.6, *n.s.*). For per-point pressure, both the user *and* the gesture to which the point belongs might be important. It is expected that the pressure of points *within* a gesture will be more related than the pressure of points *between* gestures. An ANOVA of per-point pressure considering stress, user ID, and a unique gesture identifier as factors indicated a trend for increased pressure during stress ($F_{1,92}$=4.0, p<0.08), but additional data would be required to validate this observation. These results indicate that, in general, gesture pressure does not tend to change with the introduction of difficulty-induced cognitive stress, suggesting that we must focus on real-time computation of gesture features such as length and duration for unobtrusive measures.

3.3 Physiological Sensors

We collected physiological sensor readings using a BioPac MP150 data acquisition system. The sensors used included skin temperature (SKT100C), pulse (PPG100C with TSD200), and respiration rate (RSP100C with TSD201). Sensors automatically recorded the participants' readings every 4.8 milliseconds (ms). To make data analysis more tractable, we averaged every 10 sensor readings, yielding a data granularity of 48 ms.

The skin temperature sensor was mounted on the participant's forehead roughly between the eyebrows, a location previously determined to show temperature changes correlated to cognitive load [8]. Each participant wore two respiration sensors, one just under the chest and one around the waist, as recommended by BioPac, since different people tend to breathe more strongly from one area or the other. The pulse sensor was fixed to the participant's index finger on the non-dominant hand.

Biopac's AcqKnowledge software was used to collect the sensor readings and calculate on-the-fly rates for the cyclical sensor data: pulse rate and respiration rate (abdomen and chest). See Table 3 for descriptive statistics. For these data, means were first computed for each participant, and then grand means and standard deviations were estimated across all participants. Preliminary analysis comparing rates during normal periods and stress periods (repeated measures, within subjects, estimated grand means) yielded the following results:

- Pulse Rate: $F_{1,9}$=8.04, p<0.05
- Skin Temperature: $F_{1,9}$=3.08, p<0.115
- Respiration Rate (Chest): $F_{1,9}$=0.24, *n.s.*
- Respiration Rate (Abdomen): $F_{1,9}$=0.01, *n.s.*

Table 3. Descriptive statistics for physiological sensors.

Sensor Reading	Stress / Normal	Mean	Standard Deviation (σ)	N
* Pulse Rate (BPM)	Normal	80.85	10.36	10
	Stress	85.00	10.75	10
Skin Temperature (°F)	Normal	92.86	1.24	10
	Stress	92.88	1.26	10
Respiration Rate (Chest) (BPM)	Normal	14.46	1.35	10
	Stress	14.72	1.43	10
Respiration Rate (Abdomen) (BPM)	Normal	15.24	0.58	10
	Stress	15.20	1.42	10

Pulse rate was the most reliable indicator of stress in our data, followed by a marginal trend for skin temperature. Respiration rates did not show significance overall. However, during the study, experiment staff observed anecdotally that participants often sighed visibly at the onset of stress periods or when a target was just missed. We intend to analyze these data in more detail at timestamps near events of interest (e.g., targets appearing, stress periods onset, etc.), to determine if these anecdotal observations are supported by the quantitative data.

3.4 Other Data Collected

We also used pressure and distance sensors mounted to a chair in such a way that we could measure posture changes. After the session, participants filled out a self-report questionnaire regarding their experiences during the study. Due to space limitations and the preliminary nature of the current analysis, results for these data are not included here.

3.5 Gesture Recognition Accuracy

Participants performed the target-finding task on a Windows Tablet PC; recognition was done using Microsoft's Tablet PC SDK. We were concerned about recognition errors as an additional source of stress, so results for two participants with very low recognition accuracy were excluded from the analysis. (Recognition accuracy for them was less than two standard deviations below the mean.) Recognition accuracy was coded by hand during post-hoc data analysis by the first author based on an inspection of the gesture input. Accuracy was approximately 88% across all responses (min=81%, max=95%, σ=5.6%). Recognition accuracy could be improved to ensure minimal impact on task performance and physiological sensor readings.

4. CONCLUSIONS AND FUTURE WORK

The results presented in this paper represent preliminary investigations into effects of cognitive stress, induced by increases in task difficulty, on a wide range of gesture dynamics and physiological sensor readings. Participants performed a vigilance-oriented continuous attention and visual search task in which difficulty increased at certain intervals. Input behavior features such as gesture size and pressure were not significantly affected by difficulty-induced stress, but gesture duration (e.g., speed) and length were. Task difficulty also affected physiological sensors, notably pulse rate. Results indicate that both gesture dynamics and physiological sensors can be used to detect changes in difficulty-induced stress. We plan to continue this work with more detailed analyses focusing on the correlation of the various sensor channels with task performance, and cross-modality comparisons to determine which modalities are more or less sensitive to changes in stress. Recognition accuracy was a small but important factor, so we intend to explore the effects of per-user recognition accuracy on reactions to stress and task difficulty. We also have other data (self-report questionnaires and posture data) to analyze. Data collected during this study can be made available to interested researchers; please contact the first author.

5. ACKNOWLEDGMENTS

We thank Isaac D. Simmons for database assistance, Lisa Vizer for pointers to related work, and the study participants. We also thank Northrup Grumman for partial support of this research.

6. REFERENCES

[1] Auburn, T.C., Jones, D.M. and Chapman, A.J. 1987. Arousal and the Bakan vigilance task: The effects of noise intensity and the presence of others. *Current Psychology* 6, 3 (Sep. 1987), 196-206.

[2] Hart, S. and Staveland, L. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, P. A. Hancock and N. Meshkati, Eds. North Holland Press.

[3] Ikehara, C.S. and Crosby, M.E. 2005. Assessing Cognitive Load with Physiological Sensors. *Proc. HICSS 2005*, IEEE Press, 295a.

[4] Mentis, H.M. and Gay, G.K. 2002. Using TouchPad pressure to detect negative affect. *Proc. ICMI 2002*, ACM Press, 406-410.

[5] Mota, S. and Picard, R.W. 2003. Automated Posture Analysis for Detecting Learner's Interest Level. *Proc. CVPRW 2003*, IEEE Press, 49-55.

[6] Ruiz, N., Taib, R., Shi, Y., Choi, E. and Chen, F. 2007. Using pen input features as indices of cognitive load. *Proc. ICMI 2007*, ACM Press, 315.

[7] Ruiz, N., Feng, Q.Q., Taib, R., Handke, T. and Chen, F. 2010. Cognitive skills learning: pen input patterns in computer-based athlete training. *Proc. ICMI-MLMI 2010*, ACM Press, Article 41, 4 pgs.

[8] Shastri, D., Merla, A., Tsiamyrtzis, P. and Pavlidis, I. 2009. Imaging Facial Signs of Neurophysiological Responses. *IEEE Transactions on Biomedical Engineering* 56, 2 (Feb. 2009), 477-484.

[9] Schuller, B., Rigoll, G. and Lang, M. 2004. Emotion recognition in the manual interaction with graphical user interfaces. *Proc. ICME 2004*, IEEE Press, 1215-1218.

[10] Strauss, M., Reynolds, C., Hughes, S., Park, K., McDarby, G. and Picard, R.W. 2005. The HandWave Bluetooth Skin Conductance Sensor. In *Affective Computing and Intelligent Interaction*, J. Tao, T. Tan, and R.W. Picard, Eds. Springer Berlin Heidelberg. 699-706.

[11] Vizer, L.M., Zhou, L. and Sears, A. 2009. Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human-Computer Studies* 67, 10 (Oct. 2009), 870-886.

[12] Yin, B. and Chen, F. 2007. Towards Automatic Cognitive Load Measurement from Speech Analysis. In *Human-Computer Interaction: Interaction Design and Usability*, J.A. Jacko, Ed. Springer Berlin Heidelberg. 1011-1020.