MouStress: Detecting Stress from Mouse Motion

David Sun[1,2], Pablo Paredes[1], and John Canny[1]

[1]EECS, University of California, Berkeley
{davidsun, paredes, jfc}@eecs.berkeley.edu
[2]Accenture Technology Labs, San Jose
david.q.sun@accenture.com

ABSTRACT
Stress causes and exacerbates many physiological and mental health problems. Routine and unobtrusive monitoring of stress would enable a variety of treatments, from break-taking to calming exercises. It may also be a valuable tool for assessing effects (frustration, difficulty) of using interfaces or applications. Custom sensing hardware is a poor option, because of the need to buy/wear/use it continuously, even before stress-related problems are evident. Here we explore stress measurement from common computer mouse operations. We use a simple model of arm-hand dynamics that captures muscle stiffness during mouse movement. We show that the within-subject mouse-derived stress measure is quite strong, even compared to concurrent physiological sensor measurements. While our study used fixed mouse tasks, the stress signal was still strong even when averaged across widely varying task geometries. We argue that mouse sensing "in the wild" may be feasible, by analyzing frequently-performed operations of particular geometries.

Author Keywords
Stress modeling, affective interfaces, mouse interaction

ACM Classification Keywords
H.5.2 User Interfaces (H.1.2, I.3.6): Theory and methods, Input devices and strategies

INTRODUCTION
Stress has a profound impact on the emotional, cognitive and physical well-being and the quality of life of individuals. It has been strongly linked to numerous chronic health risks, such as cardiovascular disease [25], diabetes, obesity [4], hypertension, and coronary artery disease [22]. Physiological reactions induced by stress are symptomatic of mental illnesses, such as anxiety disorder and depression which is a leading cause of suicides [19]. Chronic stress can induce mood swings, social isolation, and aggression which negatively affect interactions with peers, friends, and families, leading to a broader array of social issues.

CHI 2014, April 26 - May 01 2014, Toronto, ON, Canada

http://dx.doi.org/10.1145/2556288.2557243

Stress may result from and also cause difficulties in using computer interfaces and applications: users may become frustrated, and they may feel trapped and unable to do what they want to do, leading to poor user experience and overall reduced productivity. Being able to use "everyday" technologies to continuously monitor and detect stress on ordinary computers could open a new source of field data to gauge and understand user difficulties and to inform the design of better UIs or applications adaptive to user pains.

Self-report tools and specialized questionnaires have been successful in detecting and diagnosing acute distress disorder in clinical settings [5]. These methods, however, provide only a momentary snapshot of individual stress levels and are of limited use for in-situ, everyday monitoring. Routine unobtrusive monitoring can help individuals better understand stress patterns and can enable a variety of interventions, including break-taking, breathing, and visualization exercises.

A number of physiological markers have been linked with stress, including heart rate variability, muscle tension, pulse oximetry, and galvanic skin response. Custom wearable body sensors have been designed to continuously monitor these stress markers. However, the costs of acquisition and continuous wear, even before stress-related problems are evident, present barriers for wide-spread adoption.

This paper explores the use of common computer mouse operations for measuring stress. Studies by Lundberg [15] and Wahlstrom [28] have shown that increased arm muscle activity and muscular tension are prominent mental stress markers. We show that muscle stiffness of arm/hand movement can be directly captured from common mouse operations with a physiological model of hand-arm dynamics, the Mass-Spring-Damper system (MSD). We derive two novel mouse-derived stress metrics based on the parameters of this model and discuss a computational procedure that provide direct and accurate means to estimate these parameters. A controlled study was carried out using a within-subject, fully-balanced design, where mouse activity data, ECG, and subjective stress ratings were collected from 49 participants under both calm and induced stress conditions. The results suggest that within-subject mouse-derived stress measures are quite strong, and in fact stronger than parallel physiological methods. We also find that stress detection is feasible, and under controlled conditions, 10 samples (mouse motions) from a user in a fixed state of stress can infer their stress state at 70% accuracy. This requires 100-200 training movements to train a single-parameter model for that user.

BACKGROUND AND RELATED WORK

Stress and Muscle Activity

Increased muscle activity/tension is one of the most consistent physiological markers of stress and arousal [11]. [15] identified increased activity in the trapezius muscle from electromyography readings (EMG) while subjects performed cognitively demanding tasks. Independent studies by Wahlström [29] and Visser [27], showed (from EMG readings) that mentally demanding computer mouse tasks can also significantly increase co-contraction and tension in neck/shoulder muscles, upper arm (biceps/triceps) muscles, multiple forearm and finger muscles. Visser's study further indicates that the accumulation of muscle tension has a significant effect in the tempo of arm motion in pointing tasks.

Affective Mouse Devices

Prior work fitted mouse devices with custom sensors to measure and detect stress/arousal as well as affect during computer work. Kirsch and Picard used the SenticMouse [14] (a mouse with pressure sensor add-on) to collect finger pressure while subjects browsed affective images. Analysis suggested a correlation between finger pressure and valence states (positive versus negative). Wahlström et al. [29] collected grip and click forces with sensor addons during a computer stressor task. Click force was found to be significantly higher during the stressor compared to the control, but grip force had showed no significant differences.

A promising early exploration of the use of *mouse movements* to detect affect was the work of Maehr [16]. Maehr used several metrics of mouse movement, and emotions were induced in subjects by watching short videos. A few of the metrics showed some relationships to emotion, but many did not show significant relationships. Specifically, "motion breaks" (discontinuities in mouse movement) were significantly related to both arousal and disgust and close to significant for anger. While the features used were intuitive, they were not driven by physiological correlates of arousal or stress.

Stress and HRV

Heart rate variability (HRV), which captures the variation in the inter-beat interval of the heart, is a frequently-used physiological measure of arousal/stress. Numerous studies have demonstrated its use in detecting stress induced by cognitive and physical stressors, with varying degrees of success, e.g., mental math [2], interview [21], and computer work [28]. However, researchers have also consistently found HRV measurements to be frequently noisy and highly sensitive to motion artifacts when subjects are moving around [3] and inaccurate for short term measurements [18]. In fact, *automatic* HRV analysis almost always begin with *manual* data pre-processing of the ECG signal for artifact removal as per the recommendations by the Task Force European Society of Cardiology and the North American Society of Pacing and Electrophysiology [24].

A PHYSIOLOGICAL MODEL OF THE ARM

In this work we are interested in using motion produced by the user while operating a computer mouse to infer stress levels. We begin by considering a physiological model of the arm to establish a theoretical basis for the subsequent study and analysis. We argue that the model lends itself to directly capturing the physiological effect of stress and present a novel application of a computational technique to derive this model from data.

Mass Spring Damper Model

There is substantial empirical and theoretical evidence that the dynamics of human arm motion for two-dimensional tasks, such as handwriting and drawing, can be well approximated by a *mass-spring-damper* (MSD) system [10]. Variations of the MSD model have wide applications in science and engineering, such as in modeling arm motion for handwriting and drawing tasks [12] and in system design for robotics and haptic interfaces[17].

Consider the use of a single mass-spring-damper system to model arm motion along each axis of motion during computer mouse use. The *mass* is a lump representation of the arm/hand, and the mouse. The *spring* and *damper* components capture the interactions between the active and passive muscle elements of the arm. The system takes as input forces generated by the arm and produces as output an arm motion trajectory which is recorded as a mouse movement.

Mechanically, an MSD system consists of a mass (m) attached to a spring component (k) and a viscous damper (c), where k is the spring constant and c is the damping coefficient. The mass shall oscillate at a rate related to the *tension* of the spring but the oscillation will decay exponentially due to the drag/friction produced by the damper. This behavior is fully described by the two fundamental MSD parameters – the *damped frequency* (ω) and the *damping ratio* (ζ).

From basic physics, we know that the constant k determines *stiffness* of the spring component MSD model. Furthermore, prior research have demonstrated that increased arm muscle tension is a strong physiological correlate of stress during computer work [15, 28]. We thus postulate that an increase in stress will translate to an increase in the tension k of the spring component of the MSD model.

How is k related to the fundamental parameters ω and ζ? We know the damped frequency ω is proportional to the natural (undamped) frequency ω_0 and by definition:

$$\omega_0 = \sqrt{\frac{k}{m}} \;\; \zeta = \frac{c}{\sqrt{mk}}$$

We can see that, for a system with constant mass (e.g. arm/hand) and damping coefficient, the damped frequency will be proportional to the square-root of k, the spring constant: $\omega \propto \sqrt{k}$, while the damping ratio will be *inversely* proportional to the square-root of k: $\zeta \propto \frac{1}{\sqrt{k}}$.

By deriving the above relationship between stiffness and the MSD parameters, we have shown that the MSD model is well suited for directly modeling muscle tension, a strong stress indicator. Further, we may "capture" the effect of stress if we are able to obtain ω and ζ.

LPC Model

Direct identification of the MSD model will not be possible since only the system's output – mouse movement can be observed. Hence a major modeling challenge is how to "invert" the system, i.e. to infer the system's fundamental parameters when only the output signal is available.

Linear predictive coding (LPC) is a signal modeling technique which builds a predictive model for future samples based *only* on linear combinations of observed signals from the past [20]. It turns out that an ideal second-order system, such as the spring-damper system, has a simple second-order LPC model. That is, each sample can be predicted exactly from the previous two, given the MSD parameters[1]. Conversely if we build a second-order LPC model that best fits a series of samples, we can recover the MSD parameters.

Model Computation

We used a logger to record a sequence of raw mouse events describing x and y mouse motion. Each event is a tuple (dx, dy, t), where dx and dy are the displacements along x and y directions, and t is the timestamp. This event stream is segmented into individual trials, each corresponding to a particular instance of mouse task. The relative displacements are linearly interpolated, resampled and summed to give uniformly time-sampled absolute mouse displacements along x and y axes. This becomes the *observed* trajectory of the MSD system that we wish to identify.

To compute the MSD model, LPC is applied to a trajectory with an interpolation order p as input[2]. LPC produces a sequence of coefficients that define the characteristic polynomial of MSD system. The complex *roots* (r) of this polynomial characterize the MSD's damping behavior. Specifically, the damping frequency (ω) is the imaginary part of the complex root $(\omega = \mid \Im(r) \mid)$, and the damping ratio (ζ) is the ratio of the root's real part and its absolute value $(\zeta = \frac{|\Re(r)|}{||r||})$.

HYPOTHESES

Given the discussion in the previous section, we designed an experiment to evalute the following specific hypotheses:

Hypothesis 1a (H1a): Due to higher stress, the damped frequency ω will be higher during stressed mouse use compared to a baseline.

Hypothesis 1b (H1b): Due to higher stress, the damping ratio ζ will be lower during the stressed mouse use compared to a baseline.

That is we expect the MSD model's parameters to capture the within subject stress variation for some well defined set of mouse tasks.

We considered a third indicator variable: task completion time (t). Subjects under stress typically increase their speed, especially on familiar tasks [27]. Hence:

Hypothesis 2 (H2): The time to complete different mouse tasks will be shorter during stressed mouse use compared to an unstressed baseline

METHOD

Mouse Task Design

To decide which common mouse operations should be used in our study we first observe that the bulk of onscreen mouse interactions can be characterized by a small number of repetitive tasks. For example, we frequently move the mouse cursor over a button or icon to invoke an action, e.g. starting an application, opening a file, sending a message. Another common set of tasks involve moving or rearranging objects by dragging, e.g. positioning windows, copying files. Sometimes the mouse cursor must be moved through constrained "tunnels", such as in selecting an action from a drop-down menu or highlighting text in an editor. These and other common mouse interactions can be described by three abstract mouse operations: *point-and-click*, *drag-and-drop*, and *steering*. We designed the mouse tasks in this study based on these abstract operations for good generality and wide applicability, similar to literature on human psycho-motor modeling [8].

Point-and-click

The abstract point-and-click task shows two different color-filled targets, at equal distance from the center-point of the screen. The user moves and clicks the two targets in succession (left and right) as quickly and accurately as possible (Figure 1).

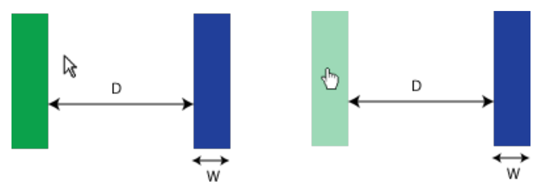


Figure 1: Point-and-click: Targets are dimmed when clicked-on to provide feedback.

Drag-and-drop

The abstract drag-and-drop task is similar to the pointing task. The user clicks on the left target and, while keeping the left mouse button depressed, moves and releases the object over the target on the right (Figure 2).

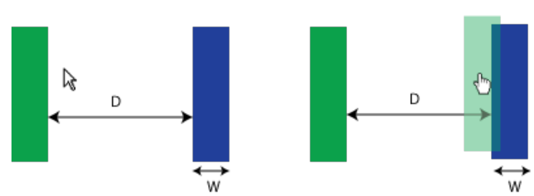


Figure 2: Drag-and-drop: A shadow copy of the dragged object moves along with the mouse to provide feedback.

Steering

The abstract steering task follows the design in[1]. where the participant draws a line through a tunnel, from left to right, as quickly and as horizontal as possible (Figure 3).

[1] Assuming there is no input, i.e. not external force on the system. When inputs are present, least squares estimates minimize the output error

[2] We used $p = 4$ in our analysis

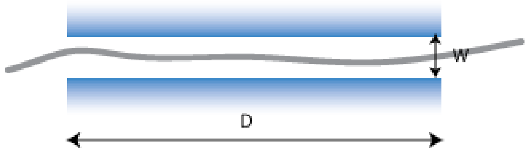


Figure 3: Steering: Line is visualized to provide feedback.

Task configurations

A task configuration corresponds to a particular assignment of values for two parameters: distance (D) and width (W). For point-and-click and drag-and-drop tasks, D describes the horizontal distance between the center point of the targets, while W describes the uniform width of the targets. For the steering task, D describes the horizontal length of the tunnel and W describes the vertical width of the tunnel. In our design, each mouse task has twenty D and W combinations, as summarized in Table 1. Each task configuration is presented 5 times to the subject for a total of 100 trials per mouse task.

	Distance (px)					Width (px)			
Pointing	64	128	256	512	1024	8	16	32	64
Dragging	64	128	256	512	1024	8	16	32	64
Steering	64	128	256	512	1024	16	32	64	128

Table 1: Different task configurations by varying the distance and width parameters

Conditioning Task Design

In this experiment we used two conditioning tasks, one to help induce stress in subjects and another to help destress the subjects. We note that the stressor and destressor tasks were not applied concurrently with the mouse tasks. We explain this design decision shortly.

Stressor Task

The basic stressor task requires participants to count down from a large prime number in decrements of 13 sequentially as quickly and accurately as possible. Recursive mental math calculations have been found to be effective in inducing cognitive stress with strong physiological markers, such as increased heart rate variation and reduced breathing rates [2].

Tasks containing the components of "uncontrollability" and "social-evaluative threat" are associated with some of the largest physiological stress responses and the longest recovery times [7]. Thus, we augment the basic stressor task the following additional components:

Social pressure: We wanted the participant to *feel* like his/her performance might be negatively judged by another person. To achieve this, during the mental math task, the participant was required to verbalize successive answers to the experimenter and the experimenter was seen to be recording all incorrect answers in front of the participant.

Timing pressure: The participant was given an upper limit of 5 seconds to verbalize the next number in the sequence. When a visible timer expired, the participant was immediately interrupted and asked to restart from the beginning.

Repetition: In addition to the timer, when an incorrect answer was verbalized at any point, the participant was interrupted and was required to restart the counting from the beginning.

Performance: The participant was informed at the start of the math stressor she can earn a bonus up to $10 depending on how well she performed. Specifically for every 5 correct answers in a row $2 will be added, and for every 2 incorrect answers, $1 will be deducted. This performance component simulates workplace stress where outcomes can impact compensation (e.g. year-end bonuses) or job standing.

Destress Task

The motivation for using a destress task was twofold: to provide a uniform control for pre-existing stressors that participants might carry into the experiment and to provide a recovery period at the end of study as required by the IRB for conducting stress-related experiments. The task is based on mindfulness meditation and visualization which has been found to be effective in alleviating stress under various situations [6].

Participants

Participants were recruited from a list of potential subjects maintained by the campus. No details regarding the experiment were disclosed to participants beyond that they will be asked to perform tasks similar to those encountered in academic settings. To minimize psychological and physiological confounds, subjects were screened to exclude those with significant psychiatric disorders, heavy smoker and drinkers, and those with shoulder, arm, wrist injuries in the past 6 months. Fifty-one, one-hour sessions were schedule over the course of two weeks. Forty-nine successfully completed the tasks, split into 26 female (53%) and 23 male subjects (47%). The average age of participants was 20 years old. The minimum compensation was guaranteed $25, even if the participant earned a negative amount on the math stressor task.

Procedures

The experiment was partitioned into four main phases:

- *Calm*-phase: a five-minutes *destressor* task designed to normalize for external factors and to help destress the participant.
- *mCalm*-phase: this phase is *coupled* with and takes place immediately after the *Calm* phase. The subject performs a collection of mouse tasks with a balanced number of pointing, drag-and-drop, and steering mouse operations. The mouse tasks and configurations are described in the **Mouse Task Design** section.
- *Stressor*-phase: a five-minutes stress-inducing task. Details of the task are described in the **Stressor Task** section.
- *mStress*-phase: this phase is *coupled* with and takes place immediate after the *Stressor*-phase. The subject performs the exact same set of mouse operations as the *mCalm* phase. The task orders are fully randomized.

Hence, each participant performed two sessions of mouse tasks, a control session (mCalm) and a stress session (mStress). To control for order effects in a repeated measure design, the ordering of the task couplings (*Calm-mCalm* and *Stressor-mStress*) were randomized between subjects, as shown in the two protocol schedules in Figures 4 and 5. A uniform three minutes of Recovery takes place at the end of the experiment as per study requirements. The participants gave subjective stress ratings (SSR) at the beginning and end of each task, as indicated in Figures 4 and 5. The data is roughly balanced for gender and for the two experimental protocols, with 25 (13 female, 12 male) subjects completing the normal protocol, and 24 subjects (13 female, 11 male) subjects completing the counter-balanced protocol.

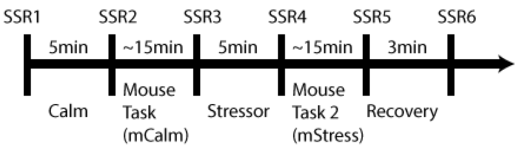


Figure 4: The normal (calm to stress) protocol.

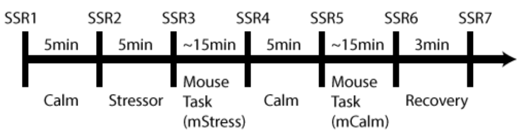


Figure 5: The counterbalanced (stress to calm) protocol.

mStress vs. Stressor Condition

In this experimental design mouse measurements were not made during the active application of a stressor or destressor, but during the next *m*-phase. This was a difficult experiment design choice but was made so that we could be sure that subjects were actually experiencing stress during the stress phases. We used stress-inducing procedures that have been shown to induce stress in many prior experiments. There was no concurrent mouse use during those stressors, and in many cases this would have been impossible. Prior attempts to use concurrent stressors in mouse experiments have had mixed success Garde [9], Wahlstrom [28], and Hoshikawa [13], although we feel that the right choice of stressors should work well, e.g. "social-evaluative threat" and "uncontrollability". Had we not done this we would have been at risk of not inducing enough stress, and of potentially measuring physiological phenomena other than stress. With the experiment as executed, we saw very strong self-report and HRV differences between *Stressor* and *Calm* and weaker (but significant under many measures) differences between *mStress* and *mCalm*.

The weakness of this approach is that the physiological stress induced during the *Stressor* phase will decay somewhat during the *mStress* phases. However, the time constants of stress decay (minutes to tens of minutes) should still afford good measurements well into the *mStress* phase. Similar arguments apply to *mCalm* which follows *Calm*. This is indeed what we observed.

Data Collection and Processing

We recorded three independent measures of stress for each subject: subjective stress ratings, continuous electrocardiogram(ECG), and mouse activity data.

Subjective Stress Rating

Multiple Subject Stress Ratings (SSRs) were recorded for each participant at different stages of the experiment. Each SSR was reported on a 11 point Likert scale, where 0 indicates no stress and 10 indicates extreme stress. Unlike ECG signals which are samples of a *continuous* signal, SSR are discrete, *point-in-time* values recorded at the start and at the end of an experimental phase in Figures 4 and 5. To estimate subjective stress *during* any experimental phase, we take the average of the two SSR values reported by the subject before and after the phase was completed.

Electrocardiogram data

Continuous ECG data was collected from a 3-lead ECG meter. The ECG electrodes were attached to a subject's chest in a triangular configuration, with 2 electrodes placed over the right and left coracoid processes, and one electrode over the xiphoid process. The ECG device was connected to a computer and streamed continuous ECG signals for the duration of the experiment.

We extracted the HRV indicators in Table 2 using the Kubios HRV analysis tool [26]. These measures have been shown in prior work as promising objective indicators of individual differences in emotional response, particularly as it relates to stress [23]. Kubios will automatically detect QRS complexes in the raw ECG signal, compute RR intervals, and produce all the most commonly used HRV parameters according to the guidelines given in [24]. We further removed incorrectly detected beats and added missing beats.

Mouse Data

The three mouse tasks and a mouse motion recorder were implemented using C++ and Microsoft Windows GDI+. The mouse motion recorder monitors and records device level *raw-input* events, which report mouse movements at the subpixel level. We obtain the timestamp of an event from a high precision hardware timer. A high resolution gaming mouse was used for the study. The mouse has a spatial resolution of 5700 CPI (counts per inch). While we did not collect data using a normal mouse, we did simulate normal mouse resolution by decimating the high-resolution mouse data to 400 CPI. The experiments reported here are for the high-resolution mouse. We repeated all tests with decimated data. This yielded the same detection accuracy (70%) as the high-resolution mouse. The t-scores for decimated data were all within 20% of the high-resolution values.

An MSD model is computed from the resulting data, as decribed in Section "Model Computation". Mouse motion parameters and interpretations are summarized in Table 3.

RESULTS AND ANALYSIS

Subjective Stress Rating

For SSR, a one-tailed, paired *t-test* was applied between *Calm* and *Stressor* phases, and between *mCalm* and *mStress* phases.

Measures	Description and Interpretation
MeanRR	Mean of the rhythm-to-rhythm (RR) interval series. Lower value indicates high stress
RMSSD	The root mean square of differences of successive RR intervals. RMSDD describes short term variation. Low value indicates high stress.
Power LF (nu)	The powers of LF in normalized units. LF demonstrates both sympathetic and vagal activation. High LF value indicates high stress.
Power HF (nu)	The powers of HF in normalized units. HF is modulated by the vagal (parasympathetic) tone. Low HF value indicates high stress.
LF/HF	Ratio of LF to HF. The ratio mirrors sympathetic and parasympathetic balance. High value indicates high stress.
SD1	The standard deviation of the Poincare plot orthogonal to the line-of-identity. SD1 describes short-term variation caused by sinus arrhythmia. Low value indicates high stress.
SD2	The standard deviation of the Poincare plot along the line-of-identity. SD2 describes long-term variability. Low value indicates high stress.

Table 2: Descriptions and interpretations of Heart Rate Variability Measures

Indicator	Description and Interpretation
ω_x	The damped frequency of the MSD system along the *x-axis* of motion. High value indicates high stress.
ω_y	The damped frequency of the MSD system along the *y-axis* of motion. High value indicates high stress.
ζ_x	The damping ratio of the MSD system along the *x-axis* of motion. Low value indicates high stress.
ζ_y	The damping ratio of the MSD system along the *y-axis* of motion. Low value indicates high stress.
t	Task completion time. Low value indicates high stress.

Table 3: Descriptions and interpretations of mouse motion based stress indicators.

All effects will be considered at 0.05 level of significance. On average the SSR indicator was significantly higher during Stressor phase ($M = 4.39, SE = .27$), compared to the Calm phase ($M = 2.42, SE = .24$)($t(48) = 12.63, p = 7e^{-17}$). The SSR indicator was also significantly higher during mStress ($M = 3.90, SE = .25$) compared to mCalm ($M = 2.67, SE = .26$) ($t(48) = 5.86, p = 4e^{-7}$).

Heart Rate Variability

HRV analysis was done for each participant for the following four phases of the experiment: *Calm*, *mCalm*, *Stressor*, *mStress*. The descriptive statistics are summarized in Table 4.

Measures	Conditions			
	Calm	*Stressor*	*mCalm*	*mStress*
MeanRR	0.76 (0.02)	0.65 (0.02)	0.77 (0.02)	0.77 (0.02)
RMSSD	0.04 (0.003)	0.04 (0.004)	0.04 (0.003)	0.04 (0.003)
LF (nu)	60.51 (2.79)	66.44 (1.73)	59.89 (1.93)	61.28 (1.64)
HF (nu)	39.44 (2.78)	33.46 (1.73)	39.92 (1.92)	38.62 (1.63)
LF/HF	2.57 (0.41)	2.54 (0.26)	1.85 (0.17)	1.83 (0.13)
SD1	0.03 (0.002)	0.03 (0.003)	0.03 (0.002)	0.03 (0.002)
SD2	0.08 (0.01)	0.08 (0.01)	0.07 (0.00)	0.07 (0.00)

Table 4: Mean and Standard Error of various Heart Rate Variability measures in the four phases.

A paired, one-tailed t-test was applied to the HRV measures between *Calm* and Stress phases, and between *mCalm* and *mStress* phases. The direction of the tail is specified in accordance to the physiological interpretation of each measure according to Table 2. All effects are reported at .05 level of significance. We summarize the results in Table 5.

Measures	Conditions			
	Calm vs Stressor		*mCalm vs mStress*	
	t(48)	*p*	*t(48)*	*p*
MeanRR	10.95	$1.08e^{-14}$*	-0.23	.82
RMSSD	0.40	.35	-1.79	.96
LF (nu)	-2.03	.02*	-0.81	.21
HF (nu)	2.05	.02*	0.76	.23
LF/HF	0.05	.52	0.18	.57
SD1	0.40	.34	-1.78	.96
SD2	-0.56	.71	-1.11	.86

Table 5: Student t-test results for different Heart Rate Variability measures between *Calm* and Stessor phases and between *mCalm* and *mStress* phases. Star(*) indicates significance at .05.

Results show that on average the two measures: MeanRR ($t(48) = 10.95, p = 1.08e^{-14}$) and HF (nu) ($t(48) = 2.05, p = .02$), were significantly lower in the *Stressor* phase compared to the *Calm* phase. In addition, LF(nu) was significantly higher in the *Stressor* phase compared to the *Calm* phase ($t(48) = -2.03, p = .02$). These results are consistent with theoretically predicted direction of the parameters and indicate that the *Stressor* was effective in inducing stress. No HRV measure was found to be significantly different (with the appropriate direction) between *mCalm* and *mStress* phases.

Mouse Motion

In this section, we examine whether the motion derived parameters in Table 3 are good indicators of stress across several models.

Mixed Task Model

We start by considering the *Mixed Task* model. The Mixed Task model can be regarded as a kind of *omnibus* model that is independent of specific task types (clicking, dragging and steering) as well as task configurations (taget width by distance). The Mixed Task model is obtained by averaging across all task types, configurations, and repetitions (300 data points). For each subject a single average was obtained for each of the indicators: $\omega_x, \omega_y, \zeta_x, \zeta_y, t$, for the stressed mouse use phase (mStress) and for baseline phase (mCalm). Summary statistics and t-test results are given in Table 6.

All effects will be considered at .05 level of signficance. In accordance with Table 3, a single-tailed, paired *t-test* was applied between stressed mouse use *mStress* and the baseline *mCalm* for all the indicator variables. On average, the damped frequency for *both* x-axis (ω_x) ($p = .003$) and y-axis (ω_y) ($p = .002$) was significantly higher during mStress than during mCalm. The damping ratio for the x-axis of motion (ζ_x) was significantly lower during mStress compared to mCalm ($t(48) = 3.11, p = .002$), but the effect was not significant for the y-axis (ζ_y) ($t(48) = 1.32, p = .097$). Results indicate that t was not significantly lower during mStress compared mCalm ($t(48) = -.02, p = .49$).

Indicators	**Conditions**			
	mCalm	*mStress*	*t(48)*	*p-value*
ω_x	0.126(0.001)	0.130 (0.001)	2.83	.003*
ω_y	0.205(0.004)	0.215(0.004)	3.00	.002*
ζ_x	0.532(0.0004)	0.534(0.0003)	3.11	.002*
ζ_y	0.408(0.008)	0.414(0.008)	1.32	.097
t	1.03(0.05)	1.03(0.07)	-.02	.490

Table 6: Summary of descriptive statistics (Mean and Standard Error) and *t-test* results between *mCalm* and *mStress* phases of the experiment pertaining to mouse motion parameters under the *Mixed Task* model. Star ($*$) indicates significance at .05.

Task Specific Models

Next, we examine *Task Specific* models. We consider separate averages for each of the three mouse tasks: clicking, dragging, and steering. Task Specific models are obtained by averaging across all task configuration and repetitions. The motion indicator variables were tested between mStress and mCalm phases (along the same direction as *Mixed Task* model). To avoid α-inflation, we applied Bonferroni correction and set signficance to $\alpha = \frac{0.05}{3} = 0.0167$. This result is presented in Table 7.

For the Clicking task, the results indicate that the damped frequency was on average significantly higher during the *mStress* phase than the *mCalm* phase, along both x-axis (ω_x) ($t(48) = 4.54, p = 2e^{-5}$) and y-axis ($\omega_y$) ($t(48) = 3.94, p = 1e^{-4}$). In addition, the damping ratio was significantly lower for *mStress* compared to *mCalm* along the x-axis (ζ_x) ($t(48) = 4.54, p = 4e^{-5}$). No significant effect was observed for ζ_y.

For the Steering task, the analysis showed that the damped frequency was significantly higher during *mStress* compared to *mCalm* for ω_x ($t(48) = 2.40, p = .010$), and ω_y ($t(48) = 2.55, p = .007$). No signficant effect was observed for damping ratio along x-axis or y-axis. Furthermore, none of the damping parameters were signficant (under the discounted α) for the Dragging task.

The analysis further showed that task completion time (t) was significantly lower during *mStress* compared to *mCalm* under the Clicking task ($t(48) = -2.65, p = .005$) but not significant for the Dragging and Steering tasks.

DISCUSSION

On the basis of the above findings, we revisit the hypotheses set forth earlier and discuss the implications.

Hypothesis 1a (H1a) is well supported by the inferential statistics. Results for the *Mixed Task* model suggests that the effects of stress was strong and consistent for damped frequency for *both* x-axis (ω_x) and y-axis (ω_y) of motion. Further analysis of the Clicking task model showed consistent results (significant effect) for both ω_x and ω_y, while the Steering task model showed signicant effect for ω_x. The ω parameters were not consistently signficant under every task specific model, but this should not taken as an absence of effect. In fact, it is clear that the effect of stress is present for the Dragging task ($\omega_x : p = .056, \omega_y : p = .087$), only that this effect is somewhat weaker compared to the Steering and Clicking tasks.

From our analysis for the *Mixed Task* model, we see that the damping ratio was significantly lower along the x-axis direction of motion (ζ_x), which supports *Hypothesis 1b (H1b)*. Analysis of the Clicking and Steering tasks show significant effect of stress on ζ_x, which further supports this hypothesis. The effect of stress on the damping ratio appears to be weaker along the y-axis of motion, as none of the models showed significant differences for ζ_y. Intuitively this is not very surprising since the measureable effect of stress should reflect more prominantly in the dominant direction of movement, which is in the left-to-right x-axis direction in the design our mouse tasks. Further studies are needed to verify this hypothesis.

Hypothesis 2 (H2) was supported for the Clicking task only (at $p = .005$), but not for the Dragging and Steering tasks, or for the Mixed task.

From the practical standpoint, the significance of the MSD parameters under the *Mixed task* model suggests that these parameters are strong indicators of stress and good candidates for stress detection during daily computer mouse use, which naturally contains a blend of different mouse operations.

Results from the Subject Stress Ration (SSR) was consistent with the mouse motion parameters. Analysis showed a signicantly higher self-report stress level in the *mStress* phase compared to the baseline *mCalm*

ECG analysis yielded several HRV measures (meanRR, LF (nu) and HF (nu)) that were significantly different between the *Stressor* phase and the *Calm* phases. This is consistent with the self-reported stress indicator, which showed very strong differences in stress level. However, no HRV measure was found to be significantly different between *mCalm* and *mStress* phases. SSR reports for *mStress/mCalm* showed that significant but weaker differences were present in stress between *mStress* and *mCalm*. On the other hand, many mouse stress metrics (damped frequency being the best) yielded strong, significant differences in stress between *mStress* and *mCalm*. It appears then that mouse measurements are more sensitive, or at least less prone to artifacts, than classical stress measures such as HRV.

Indicators	Tasks and Conditions											
	Click				Drag				Steer			
	mCalm	*mStress*	*t(48)*	*p-value*	*mCalm*	*mStress*	*t(48)*	*p-value*	*mCalm*	*mStress*	*t(48)*	*p-value*
ω_x	0.13(.001)	0.14(.001)	4.54	.00002*	0.126(.002)	0.129(.001)	1.62	.056	0.123(.002)	0.128(.003)	2.40	.010*
ω_y	0.20(.005)	0.21(.005)	3.94	.0001*	0.24(.009)	0.25(.008)	1.38	.087	0.177(.005)	0.185(.007)	1.53	.066
ζ_x	0.53(.0003)	0.53(.0002)	4.54	.00004*	0.53(.0004)	0.53(.0003)	1.68	.049	0.531(.001)	0.533(.001)	2.55	.007*
ζ_y	0.46(.007)	0.46(.007)	0.82	.194	0.38(.009)	0.39(.01)	1.02	.160	0.39(.01)	0.40(.01)	0.96	.171
t	0.89(0.02)	0.84(0.02)	-2.65	.005*	1.23(.07)	1.30(.11)	.66	.257	0.98(.08)	0.96(.11)	-0.16	.435

Table 7: Summary of *t-test* results which show the significance of the stress indicator variables: damped frequency, damping ratio, and completion time for task specific models: clicking, dragging, and steering. Star ($*$) indicates significance at .0167.

Subjects performing mouse tasks during mCalm and mStress phases naturally had bigger and more frequent upper-body/arm movement compared to Calm and Stressor phases, where they were predominantly sitting still (mindful meditation or speaking). We believe this is a plausible explanation for the noisier ECG signals we found for mStress and mCalm.

STRESS DETECTION

We studied the accuracy of within-subjects stress classification, i.e. given some labeled samples for one subject as training data, we studied the accuracy of stress classification on some unseen samples. We did this by taking a random sample of k of the data points derived during the study, training a classifier on the remaining $n-k$ points, and using this to classify the initial k samples. We used a very simple model-based classifier, relying on the structure that is evident in Figure 6. The model has a staircase structure, i.e. we model canonical stress behavior as having a simple step-wise dependence on target distance, and a separate (step slope) dependence on target size. The step for target distance is proportional to the log of the distance: the distances in our experiment were powers of two, so their logs are evenly spaced, matching Figure 6(a). We use the same step magnitude for all subjects (-0.01 for a 2x distance step), and zero step slope for the click and drag tasks. The step slope for the steering task was 0.005 for a 2x increase in the target size. There remains a single model parameter: the vertical offset of the entire staircase, which needs to be learned for each user.

An advantage of this model for "real-world" mouse stress analysis is that it requires only knowledge of the distance of a mouse motion, not the target size. Thus a low-level logging application can be used, which does not need to be aware of what task or application the user is using, which would entail increased privacy risks.

Given a user sample p which comprises a distance d and a feature value (e.g. the natural frequency), we normalize the value by computing a normalized feature values $\hat{p} = d - S(d)$ where $S(d)$ is the value of the stair model at distance d (it doesnt matter what height offset $S(d)$ has). Now the effects of target distance have been accounted for, and we have a simple one-dimensional label classification problem on the samples $\hat{p}$. We used a very simple one-dimensional classifier: namely we choose the real threshold which gave the best classification accuracy on the training set. This one-dimensional value sets the height of the stair model for that user. This is equivalent to an SVM with uniform penalties, but our implementation was simple and direct (i.e. we enumerated all data values and midpoints between them as potential thresholds, scored each for accuracy, and took the best).

Once we have a model for a user, we can make an estimate of their stress by using single or (an average of) multple samples for the user in the same stress state.

Figure 7 shows the results. The vertical axis is the classification accuracy (fraction of correctly labeled samples). The horizontal axis is the number of samples used for *measurement* with the remainder used for training. This accuracy value is an average of classifaction accuracy for positive and negative samples. That is, at 5 on the x-axis, 5 measurements of natural frequency were randomly selected in the stressed state and 5 were taken in the unstressed state (for each user). The remaining 190 (95 for each stress condition), were used to train the model above for each user. The 5 measurements in each condition were averaged and compared against the threshold, and assigned as correct or incorrect based on the actual label of those 5 measurements. This process was repeated 10000 times to get an overall accuracy measure.

There are three plots in the figure. The lowest is a baseline model (not the model above) using a simple average of the p as the threshold (no stair model). The middle plot shows a simple average of the $\hat{p}$ as threshold. The top figure shows the max-accuracy threshold of the $\hat{p}$ as the model. Note that the stair model with max-accuracy threshold provides a useful $>$ 6 percentage-point improvement in accuracy at 10 samples. We do not show the max-accuracy classifier applied to the raw samples p - it gave very close to random (50%) accuracy.

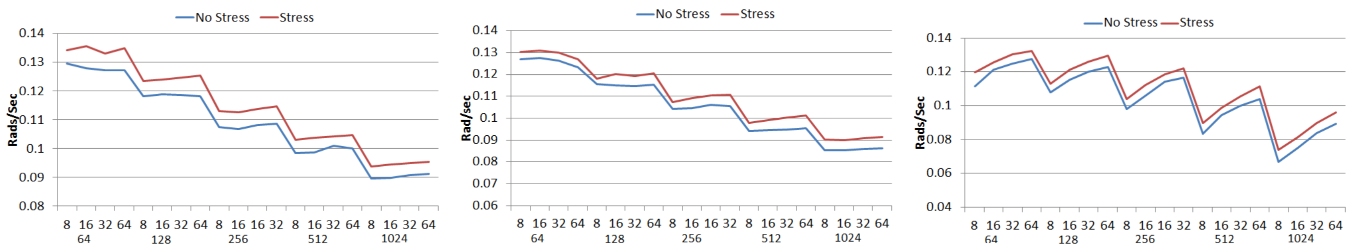
Model accuracy improves with the number of samples at measurement time, peaking at about 71% accuracy with 30 samples. As the number of measurement samples increases beyond 30, accuracy starts to fall because there are not enough remaining samples (100-k) to build an accurate model.

The upshot is with around 10 mouse movements with the user in a fixed stress state (and stress state usually changes over many minutes) should yield around 70% accuracy in classifying stress.

CONCLUSION AND FUTURE WORK

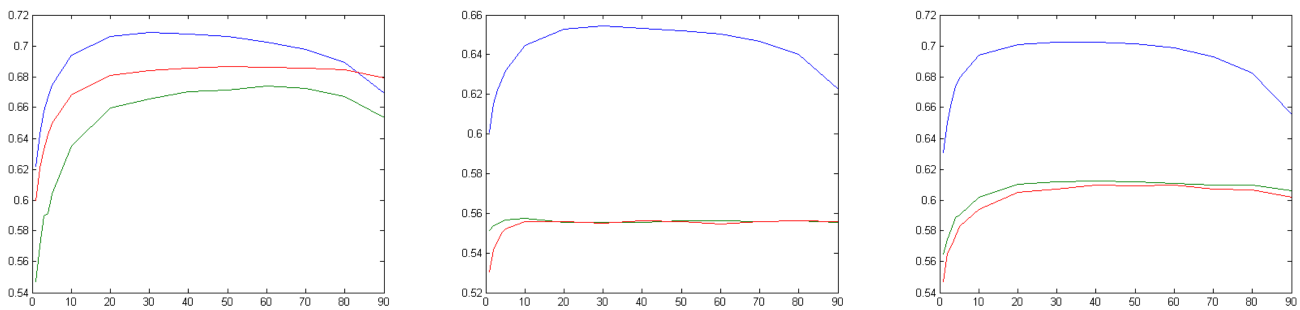
This paper makes several contributions to affective computing and human computer interaction.

We have shown a new physiologically motivated measures of stress that can be computed from common mouse operations.



(a) Clicking (b) Dragging (c) Steering

Figure 6: Damping frequency feature (y-axis) vs. task configuration (x-axis) for (a) Clicking, (b) Dragging and (c) Steering tasks. Note the strong staircase/sawtooth behavior.



(a) Clicking (b) Dragging (c) Steering

Figure 7: Accuracy (y-axis) vs. number of samples (x-axis) of the damping frequency feature for the user in a given stress state. The top curves are for the stair/sawtooth model and best accuracy threshold. The middle curves are stair model with simple mean threshold, and bottom curves are no model and simple mean threshold.

This is the first work that establishes a theoretical connection between a simple mechanical model of the arm (MSD) and muscle stiffness – a strong physiological correlate of stress. Further, we demonstrated the use of LPC as a computationally efficient and accurate method to estimate two fundamental parameters of this model – damping ratio and damping frequency. Our statistical tests suggested that both measures are sensitive to stress states in mix-task and task-specific conditions; and when stress is present, they are much stronger than HRV measures derived from concurrently recorded ECG signals.

This work makes novel use of the computer mouse – a widely available input device as a stress "sensor", which in practice has three advantages compared to existing sensing solutions: it is universally *accessible* to computer users; it is unobtrusiveness and requires no wiring; and it is suitable for long-term and in-situ monitoring when people are engaged in stressful tasks.

This work examined a broad range of target sizes and distances, and showed that an accurate detector can be agnostic to the target size – which suggests that a stress sensing process can run on the user's computer without knowledge of the underlying application. These are encouraging evidences that this work can be used as a basis for stress sensing solutions in the wild.

We believe the ability to use everyday technologies on ordinary computers to continuously monitor and detect stress opens a new source of field data to gauge and understand user difficulties. Such affective information can also be usefully presented to help people interpret changes based on exposures to stressors.

In the sequel to this work, we plan to use stressors that are both concurrent with mouse use and more natural (e.g. email tasks). Preliminary model results computed from a normal resolution mouse are quite positive which strongly suggests that the model can be deployed widely in work and office settings, and we plan to do so in future work.

REFERENCES

1. Accot, J., and Zhai, S. Beyond fitts' law: models for trajectory-based hci tasks. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, ACM (1997), 295–302.
2. Bernardi, L., Wdowczyk-Szulc, J., Valenti, C., Castoldi, S., Passino, C., Spadacini, G., and Sleight, P. Effects of controlled breathing, mental activity and mental stress with or without verbalization on heart rate variability.

Journal of the American College of Cardiology 35, 6 (2000), 1462–1469.

3. Berntson, G. G. Heart rate variability: Origins, methods and interpretive caveats. *Psychophysiology 34* (1997), 623–648.
4. Brindley, D., Rolland, Y., et al. Possible connections between stress, diabetes, obesity, hypertension and altered lipoprotein metabolism that may result in atherosclerosis. *Clinical science (London, England: 1979) 77*, 5 (1989), 453.
5. Bryant, R., Moulds, M., and Guthrie, R. Acute stress disorder scale: a self-report measure of acute stress disorder. *Psychological Assessment; Psychological Assessment 12*, 1 (2000), 61.
6. Davis, M., Eshelman, E., and McKay, M. *The relaxation and stress reduction workbook*. New Harbinger Publications, 2008.
7. Dickerson, S., and Kemeny, M. Acute stressors and cortisol responses: a theoretical integration and synthesis of laboratory research. *Psychological bulletin 130*, 3 (2004), 355.
8. Fitts, P. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of experimental psychology 47*, 6 (1954), 381.
9. Garde, A., Laursen, B., Jørgensen, A., and Jensen, B. Effects of mental and physical demands on heart rate variability during computer work. *European journal of applied physiology 87*, 4 (2002), 456–461.
10. Hill, A. The heat of shortening and the dynamic constants of muscle. *Proceedings of the Royal Society of London. Series B, Biological Sciences 126*, 843 (1938), 136–195.
11. Hoehn-Saric, R., and McLeod, D. Somatic manifestations of normal and pathological anxiety. *Biology of anxiety disorders. Progress in Psychiatry 39*, 36 (1993), 177–222.
12. Hollerbach, J. An oscillation theory of handwriting. *Biological Cybernetics 39*, 2 (1981), 139–156.
13. Hoshikawa, Y., and Yamamoto, Y. Effects of stroop color-word conflict test on the autonomic nervous system responses. *American Journal of Physiology-Heart and Circulatory Physiology 272*, 3 (1997), H1113–H1121.
14. Kirsch, D. The sentic mouse: A tool for measuring emotional valence. *retrieved Oct 2* (1998), 1–2.
15. Lundberg, U., Kadefors, R., Melin, B., Palmerud, G., Hassmén, P., Engström, M., and Elfsberg Dohns, I. Psychophysiological stress and emg activity of the trapezius muscle. *International Journal of behavioral medicine 1*, 4 (1994), 354–370.
16. Maehr, W. *eMotion: Estimation of User's Emotional State by Mouse Motions*. VDM Verlag, 2008.
17. Massie, T. H., and Salisbury, J. K. The phantom haptic interface: A device for probing virtual objects. In *Proceedings of the ASME winter annual meeting, symposium on haptic interfaces for virtual environment and teleoperator systems*, vol. 55, IOS Press (1994), 295–300.
18. McNames, J., and Aboy, M. Reliability and accuracy of heart rate variability metrics versus ecg segment duration. *Medical and Biological Engineering and Computing 44*, 9 (2006), 747–756.
19. Murray, C., Lopez, A., et al. The global burden of disease. *Geneva: WHO 270* (1996).
20. O'Shaughnessy, D. Linear predictive coding. *Potentials, IEEE 7*, 1 (1988), 29–32.
21. Pagani, M., Mazzuero, G., Ferrari, A., Liberati, D., Cerutti, S., Vaitl, D., Tavazzi, L., and Malliani, A. Sympathovagal interaction during mental stress. a study using spectral analysis of heart rate variability in healthy control subjects and patients with a prior myocardial infarction. *Circulation 83*, 4 Suppl (1991), II43.
22. Pickering, T. Mental stress as a causal factor in the development of hypertension and cardiovascular disease. *Current Hypertension Reports 3*, 3 (2001), 249–254.
23. Rajendra Acharya, U., Paul Joseph, K., Kannathal, N., Lim, C. M., and Suri, J. S. Heart rate variability: a review. *Medical and Biological Engineering and Computing 44*, 12 (2006), 1031–1051.
24. Rawenwaaij-Arts, C., Kallee, L., Hopman, J., et al. Task force of the european society of cardiology and the north american society of pacing and electrophysiology. heart rate variability. standards of measurement, physiologic interpretation, and clinical use. circulation 1996; 93: 1043-1065. *Intern. Med 118* (1993), 436–447.
25. Schnall, P., Landsbergis, P., and Baker, D. Job strain and cardiovascular disease. *Annual review of public health 15*, 1 (1994), 381–411.
26. Tarvainen, M. P., Niskanen, J.-P., Lipponen, J., Ranta-Aho, P., and Karjalainen, P. Kubios hrva software for advanced heart rate variability analysis. In *4th European Conference of the International Federation for Medical and Biological Engineering*, Springer (2009), 1022–1025.
27. Visser, B., De Looze, M. P., De Graaff, M. P., and Van Dieën, J. H. Effects of precision demands and mental pressure on muscle activation and hand forces in computer mouse tasks. *Ergonomics 47*, 2 (2004), 202–217.
28. Wahlström, J. Ergonomics, musculoskeletal disorders and computer work. *Occupational Medicine 55*, 3 (2005), 168–176.
29. Wahlström, J., Hagberg, M., Johnson, P., Svensson, J., and Rempel, D. Influence of time pressure and verbal provocation on physiological and psychological reactions during work with a computer mouse. *European journal of applied physiology 87*, 3 (2002), 257–263.