Received: 4 October 2022 | Revised: 3 February 2023 | Accepted: 28 February 2023

DOI: 10.1111/psyp.14293

PSYCHOPHYSIOLOGY SPR WILEY

ORIGINAL ARTICLE

The effect of socially evaluated multitasking stress on typing rhythms

Mark A. Wetherell[1] | **Shing-Hon Lau**[2] | **Roy A. Maxion**[3]

[1]Psychobiology of Stress & Wellbeing Group, Department of Psychology, Northumbria University Newcastle, Newcastle upon Tyne, UK

[2]Software Engineering Institute, Carnegie Mellon University, Pittsburgh, USA

[3]Computer Science & Machine Learning, Carnegie Mellon University, Pittsburgh, Pennsylvania, USA

Correspondence
Mark A. Wetherell, Psychobiology of Stress & Wellbeing Group, Department of Psychology, Northumbria University Newcastle, Newcastle upon Tyne, NE1 8ST, UK.
Email: mark.wetherell@northumbria.ac.uk

Funding information
National Science Foundation, Grant/Award Number: CNS-1319117

[Correction added on March 27, 2023 after first online publication: The Acknowledgment section is included in the article.]

Abstract

Individuals have unique typing rhythms characterized by specific keystroke dynamics. Changes in state and cardiovascular responding are well documented manifestations of the fight-flight response to stress. However, as stress also leads to changes in muscle tone and motor control, typing rhythms may also be impacted. We aim to determine which individuals are experiencing stress through their typing rhythms and identify universal keystroke markers of stress. Participants ($N = 116$) typed 80 repetitions of a 6-word, 30-character phrase before and after 15 min of critically evaluated multitasking stress. Cardiovascular, hemodynamic, and state variables were compared across baseline, stress, and recovery periods and measures of typing rhythm were derived for each period and classified using machine-learning algorithms. Critically evaluated multitasking led to significant changes in all stress measures, demonstrating highly robust stress reactivity. Machine learning algorithms accurately classified stressed typing for each individual based on their typing rhythms; however, no universal keystroke markers of stress were identified. Using typing rhythms. We were able to determine whether an individual was stressed or not, but the markers used for classification differed between individuals. These individual changes may provide opportunities for identifying stressful periods through keystroke monitoring, as well as the potential for early identification of disorders which may impact fine motor control. Typing rhythms could therefore be used to monitor health and well-being in individuals who use keyboards in various situations. This is the first rigorous assessment of stress and typing rhythms and has led to the development of a feasible and highly reproducible research protocol.

KEYWORDS

cardiovascular, heart rate variability, Individual differences, motor control

1 | INTRODUCTION

Biological responses are activated in situations interpreted as challenging or threatening, and where perceived demands exceed our perceived ability to respond appropriately. In the first instance, these responses are activated by the sympathetic-adrenal-medullary (SAM) axis, and the subsequent secretion of the hormonal endpoints adrenaline and noradrenaline, which are responsible for initiating the "fight or flight" response. These responses to challenges in daily life are our attempts to maintain allostasis and our ability to respond in this way is entirely adaptive. The fight-flight response is therefore characterized by short-term physiological changes that provide the necessary resources to deal with the presenting threat or challenge. These responses are well-established, and measures of cardiovascular reactivity including increased heart rate and blood pressure, are frequently assessed during acute stressor paradigms (Chida & Hamer, 2008). Other responses are less frequently measured in response to psychosocial stress, but nonetheless occur as a function of SAM activity, specifically, changes in muscle tone and subsequent increases in tremor. Tremors are involuntary muscle movements that result from the combination of mechanical and nervous activity (Tomczak et al., 2014). Given the role of the nervous system, tremor occurs in response to adrenaline, and has been observed following the induction of emotional states such as anger and fear (Lakie, 2010). Tremor can impact performance in settings where unaltered motor control is essential, such as in occupational settings and elite sports (Louis et al., 2011). Shooting ability, for example, is impacted by tremor (Lakie, 2010) and as further evidence of the role of adrenaline, β1-adrenergic blockade leads to reductions in tremor and improved shooting performance (Kruse et al., 1986). Other activities requiring fine motor control may also be subject to similar changes, for example, the control and movement of fingers during typing.

Keystroke dynamics, the term given to the procedure of measuring and assessing a user's typing rhythm, traces its roots to an 1897 Psychological Review article noting that individual telegraph operators could be distinguished by the differences in their rhythmic styles of sending the dots and dashes of Morse code (Bryan & Harter, 1897). Keystroke data are comprised of the timings of key-press and key-release events on a computer keyboard. These events are used to derive the principal keystroke features: durations of hold times (single keypress) and latencies (elapsed time between successive key presses) (Killourhy & Maxion, 2009). Variations in these features across individuals have been used to accurately differentiate users for typing strings of just 5 letters (Gaines et al., 1980), short, fixed texts such as usernames and passwords (Joyce & Gupta, 1990; Obaidat, 1995), and for longer-text paragraphs (Gunetti & Picardi, 2012). Most of the work in keystroke dynamics has focused on using the unique timings of typing to identify users wishing to log in to secure computer systems and have been included in several reviews of keystroke dynamics for computer security in recent years (cf. Banerjee & Woodard, 2012; Obaidat et al., 2019; Teh et al., 2013; Zhong & Deng, 2015).

Personal typing rhythms can also serve as proxies for various health conditions where changes in involuntary movement and muscle tone are features of symptomatology. Lakovakis et al. (2018) used touchscreen keystroke hold timings and pressures to detect fine-motor skill declination in Parkinson's patients in a step toward unobtrusive early detection of Parkinson's disease. Further, Lam, Meijer, et al. (2021) found keystrokes to reliably distinguish between multiple-sclerosis patients and controls, showing a strong association between higher keystroke latencies and clinical disability measures. They further suggest that keystroke dynamics demonstrates good responsiveness to changes in disease activity, fatigue, and clinical disability, and were better than some commonly used clinical measures (Lam, Twose, et al., 2021).

There is some evidence that keystroke dynamics may be related to psychological state; however, this evidence is limited and largely correlational. There have been some attempts to assess the effects of experimentally manipulated psychological state on typing – specifically attempts to induce states of acute stress. In a proof-of-concept study, Andren and Funk (2005) attempted to determine whether stress versus no-stress conditions could be correctly classified via keystroke dynamics. The authors report classification accuracies ranging from 20–100%; however, the study included only four participants and neither the method of stress induction nor the validation of any stressor effects was reported. Later studies have attempted to elicit stress either through manipulating the demands associated with the typing stimuli, or through exposure to additional stress. Lim et al. (2014a) induced demand through imposing time restraints while participants typed letter strings of varying lengths in either familiar or unfamiliar languages. Greater levels of self-reported stress were noted following the typing of longer, unfamiliar texts; however, there were no effects of time pressure. Measure of typing rhythm (keystroke speed per key, and latency between consecutive down presses), however, were affected by both time pressure and familiarity, but not text length. As the demand manipulations did not have consistent effects, it is difficult to attribute the changes in typing rhythm to stress; however, there is evidence that specific aspects of typing stimuli, in this case, familiarity of language, is related to increases in stress and modifications to typing rhythms. Using a similar approach, Kolakowska (2016) asked

computer science students to type lines of computer code in a neutral state while experiencing increased demand through reducing the support available during task completion and through increasing time pressure. Differences were observed across a range of typing measures; however, as levels of perceived stress were not assessed, it is not possible to attribute these changes to stress.

Other studies have manipulated stress using paradigms that share attributes with the techniques more typically employed in the psychophysiology literature. Vizer et al. (2009) manipulated cognitive stress using a mental arithmetic task, and asked participants to type both free and fixed text samples. They observed changes in the timing of keystrokes (i.e., frequency of error-correction, navigation, and punctuation keys), and were able to correctly identify typing strings produced following the mental arithmetic task with a 75% accuracy using machine-learning algorithms. However, as with Kolakowska (2016), there was no check on whether the task induced stress.

Lim et al. (2014b) also used a mental arithmetic stressor to assess typed responses to questions of increasing difficulty. As question difficulty increased, participants reported greater levels of stress, and there were decreases in keystroke speed. Finally, Freihaut and Goritz (2021) asked participants to type a 6-digit number sequence following a socially evaluated mental arithmetic task. Stress led to changes intyping latency; however, typing strings produced following the stress manipulation could not be accurately classified using machine-learning algorithms.

Accordingly, there is some evidence that stress may alter typing rhythms; however, the literature is sparse and extremely limited in terms of its methodological rigor. Specifically, studies typically have small sample sizes; they vary in terms of their measures of typing rhythms and / or the validity of the stressor paradigms and the verification of stress manipulations; or the reporting of methodological details is insufficient to draw appropriate conclusions or allow for replication. Indeed, Lim et al. (2014a, 2014b) call for more rigorous experiments to verify the effects of psychological state on typing rhythms.

In the present study, we attempt to rectify the aforementioned shortcomings. Specifically, with an appropriately powered sample size, we use socially evaluated multitasking, an established technique, to elicit acute stress and assess its effects upon measures of typing rhythms. In line with previous research, this study aims to (1) further demonstrate the effectiveness of socially evaluated multitasking to elicit psychobiological stress reactivity, evidenced by increases in psychological, cardiovascular (Scholey et al., 2009; Wetherell et al., 2017; Wetherell & Carter, 2014), and hemodynamic (Allen et al., 2019) measures. Further, this study will (2) use this stressor paradigm to ascertain whether neutral and stressed typing from the same participant can be differentiated; and (3) identify underlying markers, that is, any feature of a participant's typing data that are responsible for significant differences between neutral and stressed typing.

2 | MATERIALS AND METHODS

The full experimental protocol is available at the Open Science Framework (osf.io/tb2ja).

2.1 | Participants

All recruitment and study procedures were granted ethical approval from the Institutional Ethics Committee. Participants ($N = 132$) were recruited via advertisements at Carnegie Mellon University. Participants had to be at least 18 years old; be a fluent English speaker; have at least 3 years of typing experience on a computer; and be able to type at least 30 words per minute. Participants were excluded if they had any history of cardiac, neurological, anxiety, stress or sleep disorders, stroke, or color-blindness (all ascertained through self-declaration). Following confirmation to participate, additional exclusion criteria were: hypertension (blood pressure exceeding 140/90); must not consume any psychoactive drugs or alcohol for 48 h prior to participation; must not consume more than 3 cups of coffee (or equivalent) for 24 h prior to participation; must not consume any caffeine or other stimulants for 2 h before the study. Fourteen participants failed to meet all inclusion criteria and a further two participants were excluded due to a lack of engagement in the tasks. Data were analyzed from 116 healthy participants with an age range of 18–60 ($N = 111$ ranging from 18–30) with 67 females and 49 males. Participants were compensated $60 for completion of the study.

2.2 | Materials

2.2.1 | Questionnaires

State anxiety was measured using a modified version (response scales converted to 100 mm visual analogue scales (VAS)) of the short-form State–Trait Anxiety Inventory (Marteau & Bekker, 1992). The scale is comprised of three negative items (tense, upset, worried) and three reverse-scored positive items (calm, relaxed, content). All item responses are summed to give a total score, with higher scores indicating greater levels of state anxiety. Perceived workload demands were assessed using the National Aeronautics and Space Administration task load index,

known as NASA-TLX (Hart & Staveland, 1988). The NASA-TLX consists of a set of six 100 mm visual analogue scales anchored with 'low' and 'high' at the extreme points. The scores for each scale are used to provide measures of six workload domains, three of which reflect the demand placed upon the respondent by the task (mental demand, physical demand and temporal demand) and three reflecting the interaction between the respondent and the task (effort, perceived performance, and frustration); higher scores reflect greater levels of each perceived workload domain. Both scales are validated and reliable measures of state anxiety ($\alpha = .82$) and perceived cognitive workload ($\alpha = .82$) respectively.

2.2.2 | Stressor task

Stress was induced using the Multitasking Framework (Purple Research Solutions, 2021), a platform for the presentation of performance-driven, cognitively demanding tasks that is analogous to working environments that require attendance and response to simultaneous stimuli (Wetherell & Sidgreaves, 2005). This study used four tasks: memory search, Stroop, visual monitoring, and high-number identification. All tasks are performance driven with points awarded for correct responses and points deducted for missed or incorrect responses. Participants are instructed to be as quick and accurate as possible on all of the tasks so as to achieve as high a score as they can; a cumulative total score is displayed in the middle of the screen as the tasks are running. The Framework has been used in previous studies to elicit psychobiological stress reactivity (Allen et al., 2019; Scholey et al., 2009; Wetherell & Carter, 2014). Additionally, participants also received critical social evaluation; this involved being video recorded and being critically evaluated by an assessor who delivers regular comments about the participant's performance and speed while engaging the Multitasking Framework. Critically Evaluated Multitasking leads to increases in cardiovascular and psychological markers of stress reactivity, providing an ecologically valid paradigm for eliciting acute stress within a controlled laboratory environment (Wetherell et al., 2017).

2.2.3 | Typing assessment

A target typing phrase must possess key attributes; specifically, it must be memorable, devoid of emotionally charged text, and easy to type in order to avoid practice effects. To satisfy these criteria, we developed a four-phase process. The desired attributes of the phrase were determined; candidate phrases were generated that held these attributes; the number of phrases was refined; the phrase that was the best fit with the original attributes was experimentally determined. This process generated 100 initial phrases; the 20 phrases which were most consistent with the desired attributes were assessed by 413 participants from Amazon's Mechanical Turk. Here, participants were asked to repeatedly type two of the twenty phrases (randomly drawn) and indicate which one they thought was easier to type. The resulting pairwise preferences were transformed into an ordinal ranking using a Thurstone model, as detailed in Critchlow and Fligner (1991), and the highest-ranked phrase was selected for use in the study. The phrase "great friends are good to have" was selected. Typing data were collected using custom software and a modified Apple USB-keyboard (model M9034LL/A) with a standard QWERTY layout. The keyboard was modified by removing the standard keyboard encoder and rerouting the output to a custom external timer that timestamps keystrokes with an accuracy of ±100 microseconds. The custom software displays instructions and stimulus prompts on the screen, in this case the all-lower-case phrase "great friends are good to have". Participants are presented with a blank text box in which the phrase must be typed, followed by the Enter key. All characters in the phrase must be typed correctly; if any typographical errors are made, the box is temporarily grayed out and the phrase must be retyped from the beginning, ensuring the capture of a perfectly typed set. During experimental phases participants are required to type the phrase 80 times, assuring a level of comfort and familiarity with the phrase.

Typing data are recorded as 'key-down' and 'key-up' events, denoting the press and release of a key respectively, from which hold times (how long a key is depressed) and latency times (the time between consecutive key processes) are derived. The target phrase has 31 characters (including the final 'enter' key); hence 31 hold times and 30 latency times are recorded for each typing iteration. Each participant therefore produces 61 typing features for each typing-phrase completion.

2.2.4 | Physiological assessment

In line with previous studies assessing multitasking, cardiovascular (Kelly-Hughes et al., 2014; Scholey et al., 2009; Wetherell et al., 2017; Wetherell & Carter, 2014) and hemodynamic (Allen et al., 2019) measures were assessed. Specifically, to assess cardiovascular reactivity, we measured heart rate (HR); systolic blood pressure (SBP); diastolic blood pressure (DBP); and mean arterial pressure (MAP). To assess heart-rate variability we measured median R-R interval (Median R-R) and standard deviation of R-R interval (SDRR). Median R-R, the time between

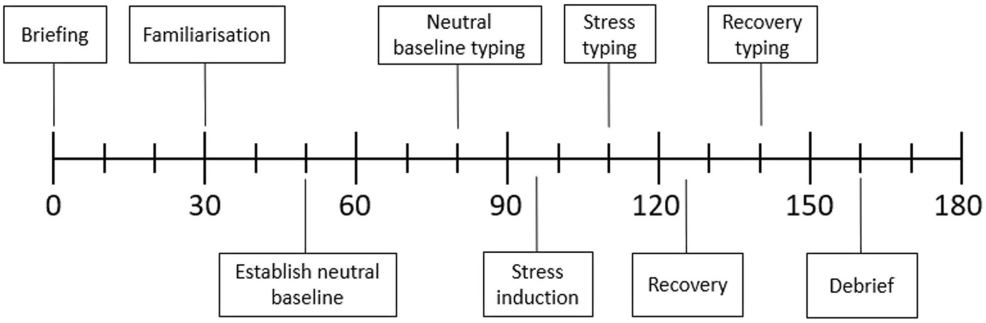

FIGURE 1 Protocol timeline (minutes).

heart beats, and SDRR, a measure of the consistency in heartbeat timings, are standard clinical parameters (Malik et al., 1996) that respond to psychological stress (Kim et al., 2018).

Heart rate and blood pressure were recorded using a GE Critikon Dinamap V100 vital signs monitor. Measurements were obtained at 5-min intervals via an inflatable cuff attached to the non-dominant arm. Heart-rate variability indices were assessed continuously at a sampling rate of 10,000 samples per second using an electrocardiogram (ECG) with three electrodes in a modified lead-II placement (Stern et al., 2001). Data were processed through a Dual Bio-Amp into a PowerLab 16/25 and analyzed with LabChart (AD Instruments, 2014).

2.3 | Procedure

Upon arrival at the lab, participants were greeted and briefed by the experimenter; informed consent was obtained. Participants were then familiarized with the procedures and materials. Specifically, participants completed a warm-up typing task where they were required to type the phrase "great friends are good to have" a total of 40 times without error; following brief demonstrations of the tasks to be performed, they completed a 2-min practice session with the Multitasking Framework. ECG electrodes were then attached, the integrity of the signal was verified, and the blood-pressure cuff was attached. All participants were tested individually. The protocol comprised three phases: baseline, stress, and recovery.

2.3.1 | Baseline

For a period of 30 min participants were asked to watch a video of an underwater scene, accompanied by soft orchestral music, while noting the categories of animals (e.g., fish, bird) that they saw. The task was designed to be calming and engaging, but not stimulating.

2.3.2 | Stress

Participants engaged the Multitasking Framework for a period of 15 min while receiving critical feedback and evaluation at regular intervals (Wetherell et al., 2017).

2.3.3 | Recovery

Participants then completed a 15-min recovery period where they were instructed to relax, breathe slowly, and continue to watch the underwater video used during baseline. At the end of each phase participants completed the VAS measures and the typing task. An overview of the protocol is presented in Figure 1.

2.4 | Treatment of data & statistical analyses

2.4.1 | Treatment of physiological data

Heart rate and blood pressure readings were taken every 5 min across the assessment periodand averaged for each phase (baseline, stress, recovery). Artifacts were mitigated automatically by the LabChart software. Data were computed for 5-min intervals and the values for all intervals were averaged for each phase.

2.4.2 | Psychobiological analyses

Cardiovascular (HR, SBP, DBP, MAP), heart-rate variability (Median R-R, SDRR); and psychological (state anxiety, perceived workload facets) measures were assessed across baseline, stress and recovery periods using repeated measures ANOVA, with Greenhouse–Geisser correction (Greenhouse & Geisser, 1959). Follow-up analyses were conducted as appropriate, using Bonferroni corrected ($p < \alpha = 0.00962$) *t*-tests. To verify that all participants demonstrated stress reactivity, these analyses were

repeated for those participants in the lowest responding quartile for each measure.

2.4.3 | Classifying typing as neutral or stressed

Whether a typing string was produced in a neutral or stressed state by each participant was established using machine-learning algorithms where the primary inputs to each algorithm are hold and latency times for each repetition of neutral and stressed typing from the same participant. These input data are referred to as the training data through which the algorithm learns a model for neutral and stressed typing from each participant. To evaluate the goodness of these models, we provide each learned model with testing data, consisting of new repetitions of neutral and stressed typing from the given participant. The goodness of the learned models is contingent on their ability to predict whether never-before-seen repetitions from the participant were produced in a neutral or stressed state. We report Random Forest algorithm results using the randomForest R package (version 4.6–2, Liaw & Wiener, 2002), where classification is based on the integration of multiple decision trees. Statistically significant classification of phrases typed during neutral or stressed phases requires an accuracy of 60% (based on a 95% confidence interval, with 80 total trials and a null hypothesis of 50% accuracy).

2.4.4 | Identifying universal typing markers

A marker was defined as any typing feature (of the 31 hold times and 30 latency times) with a mean change of more than 10% between baseline and stress typing sessions. For the purpose of identifying universal markers (one or more markers that are shared among all participants), we present the number of markers per participant, and provide heatmap visualizations (Figure 6) that depict marker patterns across participants.

3 | RESULTS

3.1 | Stress induction

Stress led to significant increases in cardiovascular reactivity: HR, $F(2,230) = 107.30$, $p < .001$, $\eta^2 = 0.48$; SBP, $F(2,230) = 300.98$, $p < .001$, $\eta^2 = 0.72$; DBP, $F(2,230) = 251.93$, $p < .001$, $\eta^2 = 0.69$; MAP, $F(2,230) = 358.65$, $p < .001$, $\eta^2 = 0.76$. Follow-up analyses demonstrated increased cardiac activity during stress compared with baseline and recovery periods ($p < .001$) for all measures. In contrast, stress led to significant decreases in heart-rate variability: Median R-R, $F(2,230) = 101.97$, $p < .001$, $\eta^2 = 0.47$ and SDRR, $F(2,230) = 152.85$, $p < .001$, $\eta^2 = 0.57$. Follow-up analyses demonstrated decreases during stress compared with baseline and recovery periods ($p < .001$) for both indices. Mean (and SE) data are presented in Figure 2.

Stress led to significant increases in state anxiety, $F(2,230) = 370.81$, $p < .001$, $\eta^2 = 0.76$ characterized by greater reported anxiety during stress compared with baseline and recovery ($p < .001$). Similarly, stress led to significant increases in facets of perceived workload: mental demand, $F(2,230) = 1205.45$, $p < .001$, $\eta^2 = 0.91$; physical demand, $F(2,230) = 162.25$, $p < .001$, $\eta^2 = 0.58$; temporal demand, $F(2,230) = 1523.55$, $p < .001$, $\eta^2 = 0.93$; effort, $F(2,230) = 837.49$, $p < .001$, $\eta^2 = 0.88$; and frustration, $F(2,230) = 162.83$, $p < .001$, $\eta^2 = 0.58$, where reported perceived workload was greater during stress compared with baseline and recovery periods ($p < .001$). In contrast to increased perceived demands, perceived performance was significantly reduced, $F(2,230) = 117.80$, $p < .001$, $\eta^2 = 0.50$, during stress compared with baseline and recovery periods ($p < .001$). Additional analyses also demonstrated significant cardiac reactivity ($p < .05$) for those participants in the lowest responding quartiles across all measures. Mean (and SE) data are presented in Figure 3.

3.2 | Stress & typing

3.2.1 | Classifying typing as neutral or stressed

The Random Forest classifier (version 4.6–2, Liaw & Wiener, 2002) was used to classify neutral and stressed typing strings. All participants had classification accuracies in excess of 60%, with over half (61/116) of the samples having classification accuracies above 90% (statistically significant identification requires an accuracy of 60%). The classification accuracy of all participants is presented in Figure 4.

3.2.2 | Identifying universal markers

A marker is defined as a mean shift of more than 10% between baseline and stress typing sessions. The number of markers was calculated for each participant. The minimum number of identified markers for any participant was 9; the maximum number of markers was 51. To search for patterns of markers across participants, heatmaps were developed to represent hold-time and latency

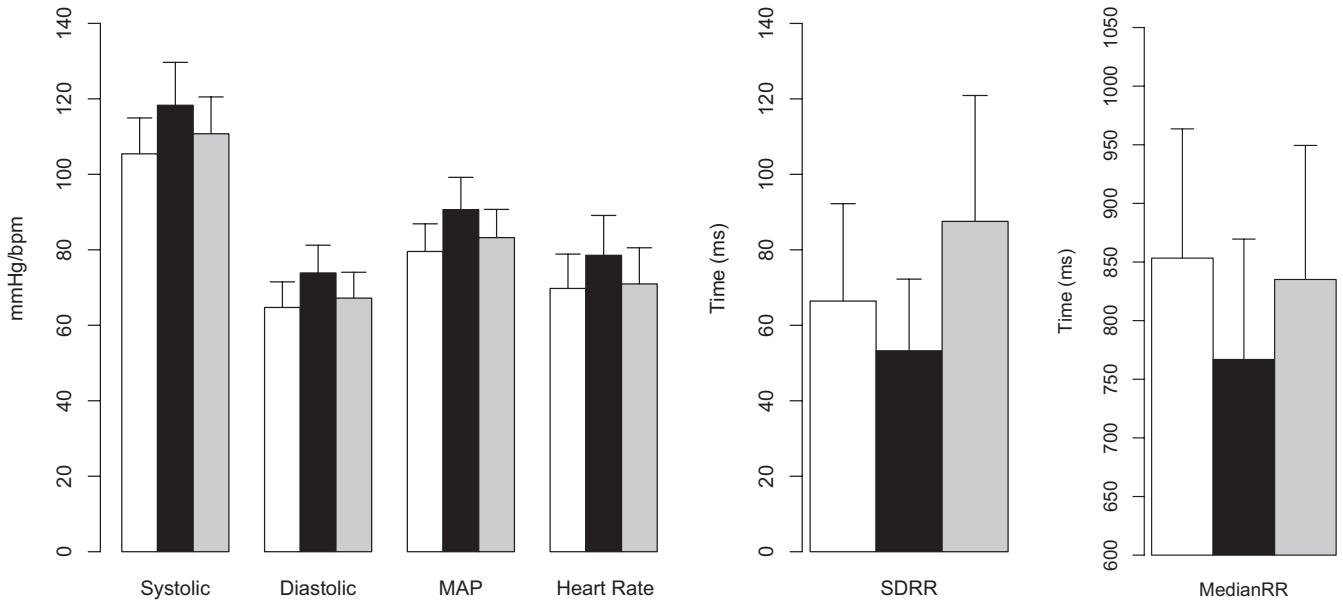

FIGURE 2 Mean (SE) values for cardiovascular and heart-rate variability measures during baseline (white), stress (black), and recovery (gray). Stress increased blood pressure and heart rate; heart-rate variability decreased.

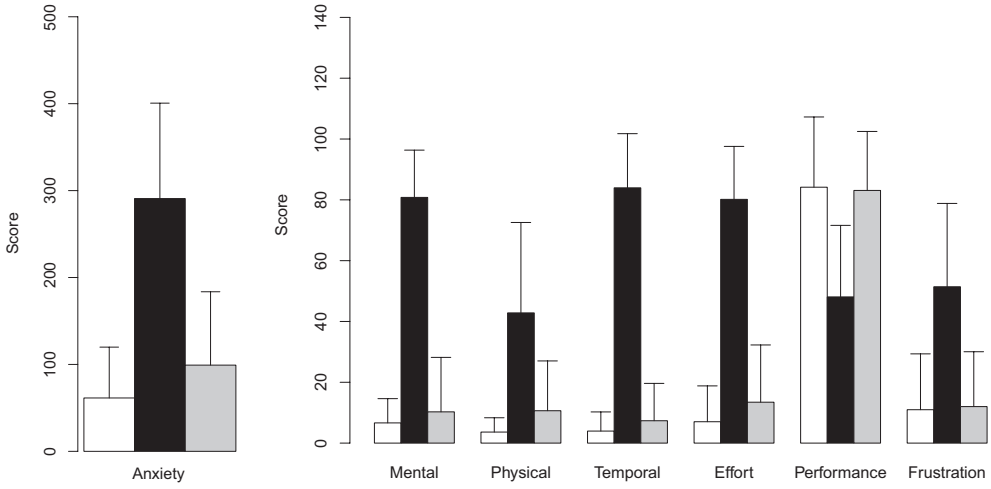

FIGURE 3 Mean (SE) values for state anxiety and perceived workload, left panel; and indices during baseline (white), stress (black), and recovery (gray), right panel. Stress increased anxiety on all task-related elements except performance, which decreased.

features that were either unchanged (black), shorter (blue), or longer (red) during stress compared with neutral typing; each column represents a feature, and each row represents a participant. A universal marker would be characterized by a column that is entirely, or mostly, non-black, representing a feature that was universally altered across all participants during stress. More specifically, a universal marker would be either entirely blue or red, indicating that this feature is shorter, or longer, across all participants during stress. The heat maps do not demonstrate any universal markers for either hold time or latency. The number of markers per participant is presented in Figure 5. Heatmaps for hold time and latency-time markers are presented in Figure 6.

4 | DISCUSSION

This study used an established stress paradigm to induce acute psychosocial stress and to explore its effects on typing rhythms. Specifically, our aims were to (1) ascertain whether stress induces any changes in typing rhythms, and (2) whether we could identify any underlying universal typing features that led to this classification.

The stressor paradigm performed as required, eliciting significant increases in key measures of psychobiological stress reactivity. Specifically, consistent with previous studies, the stressor led to increases in cardiovascular, psychological (Scholey et al., 2009; Wetherell et al., 2017; Wetherell & Carter, 2014), and

hemodynamic (Allen et al., 2019) indicators of stress reactivity. Although the aggregated analyses demonstrate highly significant changes across all our stress reactivity measures, to verify that the stressor was efficacious in all participants, we also demonstrated significant reactivity in those participants who were the lowest responders. We can, therefore, be confident that the stressor elicited stress in the whole sample, and we can explore the aims relating to the classification of typing rhythms under stress.

We were able to identify specific markers of stress within each participant. Specifically, each participant demonstrated at least 9 markers that differentiated their typing between neutral and stressed conditions. The existence of these markers led to the successful classification of typed phrases using machine learning (Random Forest algorithm (Breiman, 2001)) with classification accuracies exceeding 60% in all participants. Moreover, classification accuracies of 90% were achieved in more than half of the participants. The existence of markers that are impacted by stress supports previous related work demonstrating changes in keystroke dynamics related to increases in perceived stress (Lim et al., 2014b) and in increased demands Kolakowska (2016). Furthermore, the use of machine-learning algorithms to classify stressed from neutral typing, is reminiscent of the work of Vizer et al. (2009) albeit, with the current classifier achieving greater accuracy. Furthermore, unlike Kolakowska (2016) and Vizer et al. (2009) we verified stress reactivity in our participants, allowing us to attribute these changes to the stressor with greater confidence.

Despite obtaining high accuracies on within-participant classification, in response to the second aim we were unable to find *universal* markers for stress that were common to all participants. Direct examination of marker patterns, represented through heat maps, demonstrated that the most common marker is only shared by 69 out of 116 subjects. As indicated by earlier research, individuals have unique typing rhythms that allow for highly accurate classification (Obaidat et al., 2019). It is perhaps, therefore, not surprising that individuals demonstrate similarly unique responses to stress. That is, every individual has their own set of markers that allow for accurate stress classification; however, these are not broadly shared across other individuals.

These findings have implications for the potential wider applications of stress identification through typing rhythms. The absence of universal stress markers in typing rhythms limits the effectiveness of systems that utilize keystroke dynamics for stress detection in open-world environments where potential users are unknown to the system. Instead, the unique changes in typing rhythms mean that any identification system must be personalized for a user. This does, nevertheless, lead to a number of interesting and useful applications in closed-world environments where the users and their typing rhythms are known to the system. In its broadest application, if it were possible to detect stress by analyzing the way operators interact with their computer systems, individuals who are experiencing stress could be identified and measures could be taken to mitigate the impact of stress on performance and wellbeing. This may occur in situations where

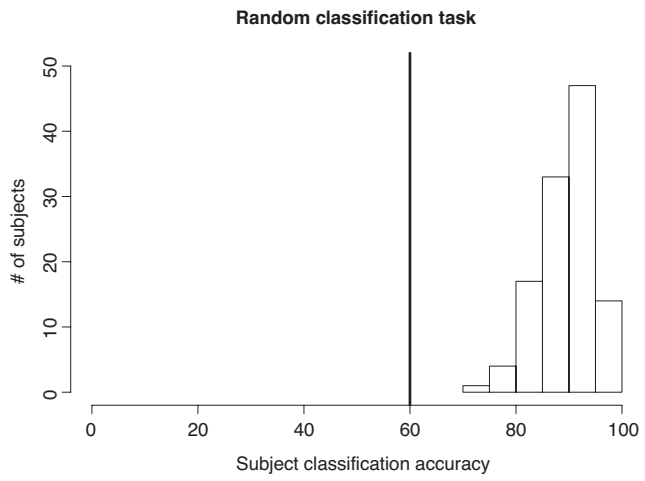

FIGURE 4 Classification accuracy for all participants using a Random Forest classifier to distinguish between neutral versus stressed typing; any classification accuracy greater than 60% (vertical line) is statistically significant.

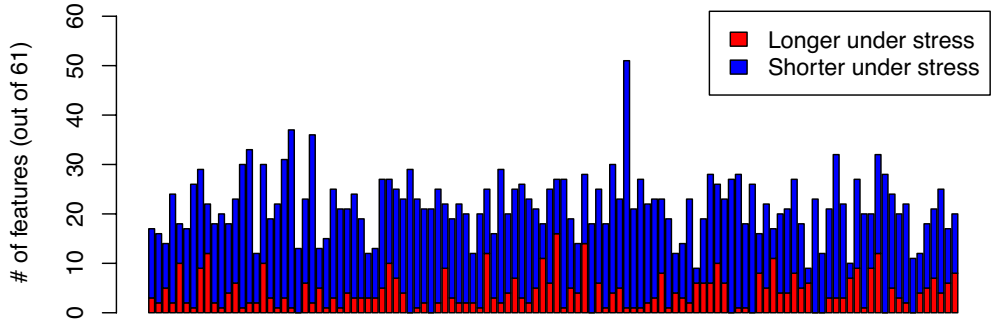

FIGURE 5 Number of identified markers per participant (each column represents an individual participant).

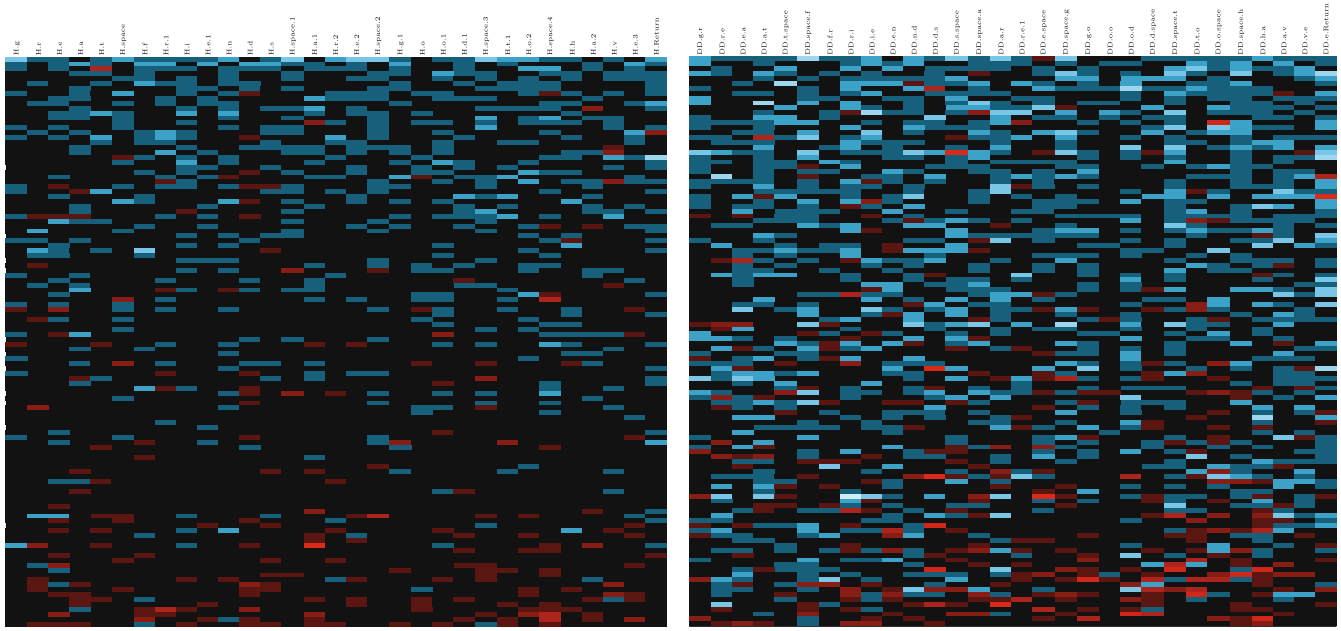

FIGURE 6 Heatmap visualizations for hold-time markers (left panel) and latency-time markers (right panel). Each column represents a typing feature, and each row represents a participant. Black rectangles indicate that a feature is not a marker for a participant; blue rectangles indicate markers where the feature was shorter during stress; red rectangles indicate markers where the feature was longer during stress. More vibrant color indicates a stronger marker (i.e., a larger difference between neutral and stress typing).

the potential risk of a stress-related errors is catastrophic, for example, in air-traffic control or critical infrastructure settings, or could assist with the maintenance of wellbeing in users in less-critical settings. The ability to detect changes in stress could also be applied to the detection of other affective states, for example, frustration, boredom or excitement; and to these ends there is the potential for application in situations that strive for computer systems to respond to human emotion. Furthermore, in line with previous studies that have shown that keystrokes can distinguish some disease states (Lakovakis et al., 2018; Lam, Meijer, et al., 2021; Lam, Twose, et al., 2021), the ability to identify deviations from an individual's typical typing rhythm could be utilized to detect various neurological conditions where changes in fine motor control may be detectable through typing rhythms.

We set out to address several significant limitations in the previous literature and, as such, this study has a number of clear strengths, notably related to experimental rigor and ensuring the validity of the methods employed. In contrast to previous work in this area, where typing stimuli were variable or lacking in a clear rationale for selection, our typing material was carefully crafted to cause minimal intrusion to either typing or affective state. We followed a clearly documented procedure to choose a 30-character phrase that was objectively judged to be easy to type, encouraged a smooth flow of typing uninterrupted by pauses, was easily held in memory without the need for a reminder, and not susceptible to practice effects. We used a stressor that has been shown to elicit change in our chosen indices of psychobiological reactivity and recovery (e.g., Allen et al., 2019; Scholey et al., 2009; Wetherell & Carter, 2014), and followed a paradigm that has high levels of ecological validity (Craw et al., 2021). Our method of stress manipulation and its measurements are therefore well-validated and is representative of everyday working environments that are characterized by high levels of stress induced through high levels of cognitive demand, effort, pressure, and frustration (Wetherell & Sidgreaves, 2005).

Given the lack of reporting and rigor in previous studies, this study has led to the development of a highly reproducible protocol with a focus on rigorous attention to detail. We suggest that these details have contributed to the highly significant changes we observed in cardiovascular and psychological stress reactivity and recovery. These effects were in line with prediction and previous work using similar stressor paradigms; however, the effects in the present study are more pronounced, with high levels of consistency across the samples. It is noteworthy that the study was run by researchers with no prior experience of stress manipulations or sampling of psychobiological indices. To ensure consistency, a highly detailed protocol was developed which covered every aspect of the study and data collection. Although a working protocol is standard practice, this protocol contained some details (e.g., standardized greetings and interactions) that are perhaps assumed, or at least not made explicit, in other reported studies. Interactions

between the experimenter, environment, and participant can have subtle, but meaningful, impact on psychobiological markers of affect. To these ends, the execution of this study has highlighted the importance of highly detailed protocols that are strictly adhered to, to ensure consistency and to facilitate reproducibility.

Our findings should, nonetheless, be considered in light of some limitations. Although the sample significantly exceeds that of any previous study in the area, the sample was drawn largely from a university population comprising healthy, young, and skilled participants. The generalizability of these findings can only be ascertained with replication in different populations. Specifically, with regard to the potential for using typing rhythms as early identification of health status and disease-progression, replication in clinical populations is recommended. This study involved the typing of one phrase. The phrase was carefully selected, and it is therefore likely to be representative of other phrases; however, without replication with other stimulus items, the true representativeness of our phrase remains unknown. It would be interesting to allow participants to type free text, as this would more closely resemble typing in a natural environment; however, such an increase in ecological validity would come at the cost of consistency in the typing stimuli, which has been an issue with some prior work in the area. Similarly, participants completed their sample typing once in each of the baseline, stress, and recovery periods. Previous studies have demonstrated that stress responses are consistent across repeated applications of the Multitasking Framework (Wetherell & Carter, 2014); however, the current results represent one exposure to the stressor only. Despite highly accurate classification of phrases that were typed following stress, there may exist some keystroke markers that were undetectable using the current methods. That is, more sensitive methods involving motion capture and pressure-sensitive keyboards may afford greater identification accuracy which may be better able to detect individual and/or universal markers. Additionally, it may be the case that non-standard measures of keystroke dynamics or incorporating other sources of related information may also facilitate the search for universal markers. For example, in addition to changes in keystrokes, Lim et al. (2014b) also observed reductions in mouse speed with increases in perceived stress. More recently, studies using smartphones have reported associations between more variable typing speed and lower accuracy and bipolar-disorder mood states (Zulueta et al., 2018) and depression severity (Vesel et al., 2020). Given that smartphones represent an increasingly common communication format, they may provide future avenues for research in this area; however, such research would be hampered by the differences in typing styles across a range of devices and the variability in conditions of use that would impact upon typing rhythms.

5 | CONCLUSION

Notwithstanding the aforementioned limitations, we have developed a highly reproducible protocol that has led to the accurate classification of individuals experiencing stress through their typing rhythms. This study is the largest, most rigorous assessment of stress and typing rhythms to date, and the findings have implications for the long-term monitoring of individuals and their stress-induced performance and wellbeing, in a range of settings that involve human-computer interaction.

AUTHOR CONTRIBUTIONS

Mark A. Wetherell: Conceptualization; methodology; writing – original draft. **Shing-Hon Lau:** Formal analysis; writing – review and editing. **Roy A. Maxion:** Conceptualization; funding acquisition; methodology; supervision; writing – original draft.

ACKNOWLEDGMENTS

The authors would like to acknowledge the contributions of David Banks, Huayun Huang, and Patricia Loring for statistical guidance, programming, and data collection.

FUNDING INFORMATION

This work was supported by the US National Science Foundation, Award: CNS-1319117.

CONFLICT OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

The data that support the findings of this study are available from the corresponding author upon reasonable request.

ORCID

Mark A. Wetherell https://orcid.org/0000-0002-1026-2798
Roy A. Maxion https://orcid.org/0000-0002-2833-7276

REFERENCES

AD Instruments. (2014). http://www.adinstruments.com/

Allen, S., Wetherell, M. A., & Smith, M. A. (2019). An experimental investigation into cardiovascular, haemodynamic and salivary alpha amylase reactivity to acute stress in type D individuals. *Stress*, *22*(4), 428–435. https://doi.org/10.1080/10253890.2019.1583741

Andren, J., & Funk, P. (2005). A case-based approach using behavioral biometrics to determine a user's stress level. In *Proceedings of the workshop on CBR in the health sciences, the sixth international conference on case-based reasoning (ICCBR-05)* (pp. 9–17). Springer.

Banerjee, S. P., & Woodard, D. L. (2012). Biometric authentication and identification using keystroke dynamics: A survey. *Journal of Pattern Recognition Research*, *7*(1), 116–139. https://doi.org/10.13176/11.427

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Bryan, W. L., & Harter, N. (1897). Studies in the physiology and psychology of the telegraphic language. *Psychological Review*, *4*(1), 27–53. https://doi.org/10.1037/h0073806

Chida, Y., & Hamer, M. (2008). Chronic psychosocial factors and acute physiological responses to laboratory-induced stress in healthy populations: A quantitative review of 30 years of investigations. *Psychological Bulletin*, *134*(6), 829–885. https://doi.org/10.1037/a0013342

Craw, O. A., Smith, M. A., & Wetherell, M. A. (2021). Manipulating levels of socially evaluative threat and the impact on anticipatory stress reactivity. *Frontiers in Psychology*, *12*, 622030. https://doi.org/10.3389/fpsyg.2021.622030

Critchlow, D., & Fligner, M. (1991). Paired comparison, triple comparison, and ranking experiments as generalized linear models, and their implementation in GLIM. *Psychometrika*, *56*(3), 517–533. https://doi.org/10.1007/BF02294488

Freihaut, P., & Goritz, A. S. (2021). Does peoples' keyboard typing reflect their stress level? An exploratory study. *Zeitschrift fur Psychologie*, *229*(4), 245–250. https://doi.org/10.1027/2151-2604/a000468

Gaines, R. S., Lisowski, W., Press, S. J., & Shapiro, N. (1980). *Authentication by keystroke timing: Some preliminary results*. RAND Corporation. https://www.rand.org/pubs/reports/R2526.html

Greenhouse, S. W., & Geisser, S. (1959). On methods in the analysis of profile data. *Psychometrika*, *24*, 95–112.

Gunetti, D., & Picardi, C. (2012). Keystroke analysis as a tool for intrusion detection. In *Continuous authentication using biometrics: Data, models, and metrics* (pp. 193–211). IGI Global.

Hart, S. G., & Staveland, L. E. (1988). Development of NASA-TLX (task load index): Results of empirical and theoretical research. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (Vol. *52*, pp. 139–183). Elsevier.

Joyce, R., & Gupta, G. (1990). Identity authentication based on keystroke latencies. *Communications of the ACM*, *33*(2), 168–176.

Kelly-Hughes, D., Wetherell, M. A., & Smith, M. A. (2014). Type D personality and cardiovascular reactivity to an ecologically valid multitasking stressor. *Psychology and Health*, *29*(10), 1156–1175. https://doi.org/10.1080/08870446.2014.915970

Killourhy, K. S., & Maxion, R. A. (2009). Comparing anomaly-detection algorithms for keystroke dynamics. In *IEEE/IFIP international conference on dependable systems and networks (DSN-09)* (pp. 125–134). IEEE Computer Society Press. https://doi.org/10.1109/DSN.2009.5270346

Kim, H.-G., Cheon, E.-J., Bai, D.-S., Lee, Y. H., & Koo, B.-H. (2018). Stress and heart rate variability: A meta-analysis and review of literature. *Psychiatry Investigation*, *15*(3), 235–245. https://doi.org/10.30773/pi.2017.08.17

Kolakowska, A. (2016). Towards detecting programmers' stress on the basis of keystroke dynamics. In *2016 federated conference on computer science and information systems (FedCSIS)* (pp. 1621–1626). https://doi.org/10.15439/2016F263

Kruse, P., Ladefoged, J., Nielsen, U., Paulev, P.-E., & Sørensen, J.-P. (1986). β-Blockade used in precision sports: Effect on pistol shooting performance. *Journal of Applied Physiology*, *61*(2), 417–420.

Lakie, M. (2010). The influence of muscle tremor on shooting performance. *Experimental Physiology*, *95*(3), 441–450. https://doi.org/10.1113/expphysiol.2009.047555

Lakovakis, D., Hadjidimitriou, S., Charisis, V., Bostant-zopoulou, S., Katsarou, Z., & Hadjileontiadis, L. J. (2018). Touchscreen typing-pattern analysis for detecting fine motor skills decline in early-stage Parkinson's disease. *Scientific Reports*, *8*(1), 1–13. https://doi.org/10.1038/s41598-018-25999-0

Lam, K. H., Meijer, K. A., Loonstra, F., Coerver, E. M. E., Twose, J., Redeman, E., Moraal, B., Barkhof, F., de Groot, V., Uitdehaag, B. M. J., & Killestein, J. (2021). Real- world keystroke dynamics are a potentially valid biomarker for clinical disability in multiple sclerosis. *Multiple Sclerosis Journal*, *27*(9), 1421–1431. https://doi.org/10.1177/1352458520968797

Lam, K.-H., Twose, J., McConchie, H., Licitra, G., Meijer, K., de Ruiter, L., van Lierop, Z., Moraal, B., Barkhof, F., Uitdehaag, B., de Groot, V., & Killestein, J. (2021). Smartphone-derived keystroke dynamics are sensitive to relevant changes in multiple sclerosis. *European Journal of Neurology*, *29*(2), 522–534. https://doi.org/10.1111/ene.15162

Liaw, A., & Wiener, M. (2002). Classification and regression by random forest. *R News*, *2*(3), 18–22.

Lim, Y. M., Ayesh, A., & Stacey, M. (2014a). Detecting emotional stress during typing task with time pressure. In *2014 Science and information conference* (pp. 329–338). https://doi.org/10.1109/SAI.2014.6918207

Lim, Y. M., Ayesh, A., & Stacey, M. (2014b). Detecting cognitive stress from keyboard and mouse dynamics during mental arithmetic. In *2014 Science and information conference* (pp. 146–152). https://doi.org/10.1109/SAI.2014.6918183

Louis, E. D., Hafeman, D., Parvez, F., Liu, X., Alcalay, R. N., Islam, T., Ahmed, A., Siddique, A. B., Patwary, T. I., Melkonian, S., Argos, M., Levy, D., & Ahsan, H. (2011). Tremor severity and age: A cross-sectional, population- based study of 2,524 young and midlife normal adults. *Movement Disorders*, *26*(8), 1515–1520. https://doi.org/10.1002/mds.23674

Malik, M., Bigger, J. T., Camm, J., Kleiger, R. E., Malliani, A., Moss, A. J., & Scwartz, P. J. (1996). Heart rate variability: Standards of measurement, physiological interpretation, and clinical use. *European Heart Journal*, *17*(3), 354–381. https://doi.org/10.1093/oxfordjournals.eurheartj.a014868

Marteau, T. M., & Bekker, H. (1992). The development of a six–item short–form of the state scale of the Spielberger state-trait anxiety inventory (STAI). *British Journal of Clinical Psychology*, *31*(3), 301–306. https://doi.org/10.1111/j.2044-8260.1992.tb00997.x

Obaidat, M. S. (1995). A verification methodology for 995 computer systems users. In *ACM Symposium on Applied 996 Computing (SAC)* (pp. 258–262). 997 ACM Press.

Obaidat, M. S., Venkata Krishna, P., Saritha, V., & Agarwal, S. (2019). Advances in key stroke dynamics-based security schemes. In M. S. Obaidat, I. Traore, & I. Woungang (Eds.), *Biometric-based physical and cybersecurity systems* (pp. 165–187). Springer International Publishing. https://doi.org/10.1007/978-3-319-98734-7

Purple Research Solutions. (2021). http:// www.purple-research.co.uk/framework.html

Scholey, A. B., Haskell, C., Robertson, B., Milne, A., Kennedy, D., & Wetherell, M. A. (2009). Chewing gum alleviates negative

mood and reduces cortisol during acute laboratory psychological stress. *Physiology & Behavior*, *97*(3–4), 304–312. https://doi.org/10.1016/j.physbeh.2009.02.028

Stern, R. M., Ray, W. J., & Quigley, K. S. (2001). *Psychophysiological recording* (2nd ed.). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195113594.003.0004

Teh, P. S., Teoh, A. B. J., & Yue, S. (2013). A survey of keystroke dynamics biometrics. *The Scientific World Journal*, *2013*. Article ID 408280, 24. https://doi.org/10.1155/2013/408280

Tomczak, A., Gajewski, J., & Mazur-Różycka, J. (2014). Changes in physiological tremor resulting from sleep deprivation under conditions of increasing fatigue during prolonged military training. *Biology of Sport*, *31*(4), 303–308. https://doi.org/10.5604/20831862.1127343

Vesel, C., Rashidisabet, H., Zulueta, J., Stange, J. P., Duffecy, J., Hussain, F., Piscitello, A., Bark, J., Langenecker, S. A., Young, S., Mounts, E., Omberg, L., Nelson, P. C., Moore, R. C., Koziol, D., Bourne, K., Bennett, C. C., Ajilore, O., Demos, A. P., & Leow, A. (2020). Effects of mood and aging on keystroke dynamics metadata and their diurnal patterns in a large open-science sample: A BiAffect iOS study. *Journal of the American Medical Informatics Association*, *27*(7), 1007–1018. https://doi.org/10.1093/jamia/ocaa057

Vizer, L. M., Zhou, L., & Sears, A. (2009). Automated stress detection using keystroke and linguistic features: An exploratory study. *International Journal of Human- Computer Studies*, *67*(10), 870–886. https://doi.org/10.1016/j.ijhcs.2009.07.005

Wetherell, M. A., & Sidgreaves, M. C. (2005). Secretory immunoglobulin-A reactivity following increases in workload intensity using the Defined Intensity Stressor Simulation (DISS). *Stress and Health: Journal of the International Society for the Investigation of Stress*, *21*(2), 99–106. https://doi.org/10.1002/smi.1038

Wetherell, M. A., & Carter, K. (2014). The multitasking framework: The effects of increasing workload on acute psychobiological stress reactivity. *Stress & Health*, *30*(2), 103–109. https://doi.org/10.1002/smi.2496

Wetherell, M. A., Craw, O., Smith, K., & Smith, M. A. (2017). Psychobiological responses to critically evaluated multitasking. *Neurobiology of Stress*, *7*, 68–73. https://doi.org/10.1016/j.ynstr.2017.05.002

Zhong, Y., & Deng, Y. (2015). A survey on keystroke dynamics biometrics: Approaches, advances, and evaluations. In Y. Zhong & Y. Deng (Eds.), *Recent advances in user authentication using keystroke dynamics biometrics* (pp. 1–22). Science Gate Publishing. https://doi.org/10.15579/gcsr.vol2.ch1

Zulueta, J., Piscitello, A., Rasic, M., Easter, R., Babu, P., Langenecker, S. A., McInnis, M., Ajilore, O., Nelson, P. C., Ryan, K., & Leow, A. (2018). Predicting mood disturbance severity with mobile phone keystroke metadata: A BiAffect digital phenotyping study. *Journal of Medical Internet Research*, *20*(7), e241. https://doi.org/10.2196/jmir.9775

How to cite this article: Wetherell, M. A., Lau, S.-H., & Maxion, R. A. (2023). The effect of socially evaluated multitasking stress on typing rhythms. *Psychophysiology*, *60*, e14293. https://doi.org/10.1111/psyp.14293