StressHacker: Towards Practical Stress Monitoring in the Wild with Smartwatches

Tian Hao, Ph.D.[1], Kimberly N. Walter, Ph.D.[1], Marion J. Ball, Ed.D.[1], Hung-Yang Chang, Ph.D.[1], Si Sun, Ph.D.[1], Xinxin Zhu, M.D., Ph.D.[1]
[1]IBM T.J. Watson Research Center, Yorktown Heights, NY 10598

Abstract

In modern life, the nonstop and pervasive stress tends to keep us on long-lasting high alert, which over time, could lead to a broad range of health problems from depression, metabolic disorders to heart diseases. However, there is a stunning lack of practical tools for effective stress management that can help people navigate through their daily stress. This paper presents the feasibility evaluation of StressHacker, a smartwatch-based system designed to continuously and passively monitor one's stress level using bio-signals obtained from the on-board sensors. With the proliferation of smartwatches, StressHacker is highly accessible and suited for daily use. Our preliminary evaluation is based on 300 hours of data collected in a real-life setting (12 subjects, 29 days). The result suggests that StressHacker is capable of reliably capturing daily stress dynamics (precision = 86.1%, recall = 91.2%), thus with great potential to enable seamless and personalized stress management.

Introduction

There is mounting evidence that psychological stress from work-related pressures, financial concerns, and family responsibilities is an important public health issue in the United States[1]. A majority of Americans report frequently feeling stressed and that stress negatively impacts their physical and mental well-being[1]. Further, psychological stress is associated with adverse health outcomes such as cardiovascular disease, diabetes, depression, and substance use disorders[2–4]. Despite the evidence that stress negatively affects health, there are few effective and practical stress management solutions that can be seamlessly integrated into an individual's daily life. In order to develop truly personalized, targeted stress management solutions, we need to be able to accurately and reliably identify occasions of stress in real-world settings.

Daily stress processes are typically assessed via self-report measures and/or biomarkers of stress system functioning, which are limited in their ability to capture the dynamic nature of daily stress. Self-reported stress cannot be assessed continuously and consequently may not capture all stressful events[5]. Further, it is subject to recall issues and self-report bias. Biomarkers of stress commonly used in real-world settings include cortisol, ambulatory blood pressure, and heart rate variability (HRV). Neither cortisol nor blood pressure can be measured continuously and they are also fairly invasive[6,7]. Although HRV can be assessed passively and continuously using commercially available wrist and chest sensors, these wearable sensors do not provide automated information about stress system functioning in real-time. Recently, numerous research efforts[8–10] have been made to advance the technologies for assessing stress in real-world scenarios using a variety of sensors including ECG (electrocardiography), RIP (respiratory inductance plethysmography), camera, microphone, etc. However, these technologies have limited practicality as they either require custom devices[8] or are cumbersome and therefore not suited for daily use[9,10].

The present study aims to address these limitations by evaluating the feasibility of using a novel HRV-based stress monitoring system to passively, continuously, and accurately capture occasions of stress in a person's daily life. HRV is a commonly used measure of autonomic nervous system (ANS) functioning, which is one of the two primary stress systems – the other being the hypothalamic-pituitary-adrenal axis[11]. The ANS consists of the sympathetic nervous system (SNS; "fight or flight") and the parasympathetic nervous system (PNS; "rest and digest"), which work in opposition of each other to control stress reactivity and recovery[12]. HRV measures the variation in time between two consecutive heart beats[13] and numerous studies have demonstrated that HRV is a valid measure of reactivity to acute psychological stressors[14–16]. Our HRV-based stress monitoring system, called StressHacker, uses photoplethysmography (PPG) sensor data that is available in commercially available wearable devices. StressHacker does not require the use of a custom biosensor devices. Rather, it can be easily integrated into commercially available smartwatches allowing for seamless integration into individuals' daily lives. Thus, StressHacker has the potential to make automated and personalized stress management a reality by detecting stress and offering real-time stress reduction recommendations in real-world settings.

In this paper we present preliminary data from an in-field study evaluating StressHacker's ability to monitor stress in people's daily lives. In summary, StressHacker is a system empowered by mobile sensing technologies that can translate continuous PPG signal or RR intervals into stress index – an indicator that is reflective of the change of stress level. At the current stage, we evaluate StressHacker using PPG data collected from wrist-worn wearables. However, the underlining algorithms of StressHacker can be tailored to take similar signals from other devices as input (e.g., electrocardiogram from chest band sensor). Ultimately, our hope is that through a set of data-driven analytical tools, StressHacker will be able to uncover powerful insights about daily stress processes by identifying sources of stress, recognizing patterns, and making personalized stress reduction recommendations. Possible real-world applications of StressHacker include corporate wellness programs, telemedicine in populations highly vulnerable to stress (e.g., chronic health conditions, substance use disorders, depression, anxiety), and automobiles to increase safety and enhance the driver's experience.

In-field Experiment

The primary goal of the in-field study is to assess the feasibility of continuously monitoring daily stress using smartwatches. To this end, we have recruited 12 subjects who are IBM employees, among which 8 are female and 4 are male. Most of the subjects are young adults. All the subjects voluntarily agreed to contribute to the data collection, and signed a consent form.

In order to simulate real-life scenarios, all of the experiments were un-controlled with the following procedure. Prior to data collection, the researchers briefly informed the subject of the purpose of the experiment, types of data to be collected, and their responsibilities. Then a wrist-worn wearable device (Empatica E4 wristband[17]) was provided to the subject. The device is designed to record bio-signals for research and development purposes, and equipped with the same sensors we intend to use in smartwatches. The device's extended battery life allows a continuous data collection for up to 2 days. The recorded data was stored locally in the flash memory of the device, and was later downloaded for further analysis. We demonstrated to each subject how to start and stop the recording. The subjects were allowed to start and stop the recording anytime, however, they were informed that a data collection that covers most of their daily activities is preferred (e.g., a recording starting in the morning before work and ending in the evening after work or before sleep).

After the recording was started, data including pulse wave and acceleration was continuously sampled from the PPG heart rate sensor and accelerometer, respectively, and stored locally for off-line examination. In the current stage of our evaluation, the subjects did not get any feedback from or have any interaction with the system during data collection. To simulate real-life scenarios, the subject was informed to conduct their daily activities as they normally do during data collection. After the data collection was completed, the subject was asked to provide a log containing a list of their daily activities covered by the recording. Each entry should include a time period and a brief description of the

Table 1: An example of the logs provided by the subjects reporting their daily activities and the associated stress information (optional). The stress information can be given in the form of a short descriptive sentence and/or perceived stress level (PSL) on a scale from 1 to 10. In the log shown in this table, both descriptions and PSLs are provided.

Time	Activity	Stress Information	PSL
6:30-7:30	Getting ready for work.	No particular emotions, routine	1
7:30-8:56	Driving, 1-hour traffic jam on the bridge	Anxious about being late.	2
8:56-11:06	Working	Attentive, stuck at a problem.	2
11:06-11:31	Lunch and responding to emails	Relaxed	1
11:31-12:53	Working	Attentive	2
12:53-14:15	At a seminar and working	Attentive. Stressed due to the multitasking.	4
14:15-15:00	Coffee break	Relaxed but still thinking about things.	2
15:00-17:00	Meeting	Attentive. The first half more stressed.	3

activity, for example, “1:00-2:30 group meeting.” In addition, to help us better evaluate StressHacker, we also asked the subjects to report stress information associated with each reported activity in the form of a short description and/or perceived stress level (PSL) on a scale from 1 to 10. In the evaluation, such stress information was considered as references of their daily stress dynamics, and compared with StressHacker’s result. Table 1 shows a log provided by one of the subjects as an example. Note that the stress information is optional, meaning the subjects can choose not to include stress information for certain or all of the activities in the log.

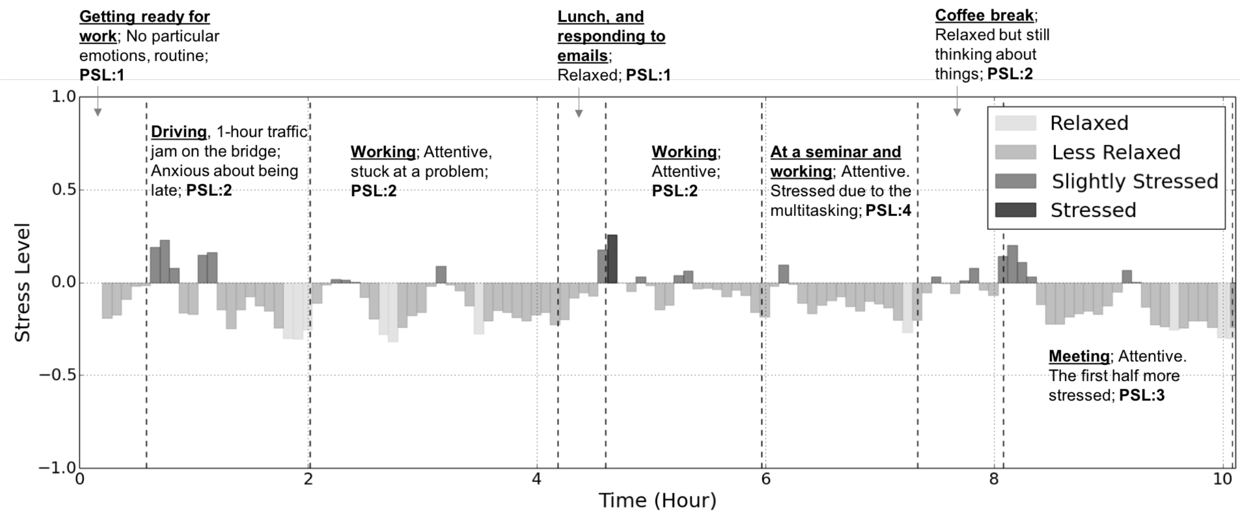

Figure 1: An example of StressHacker’s stress level output based on a 10-hour recording with annotated activities and stress information from the log shown in Table 1.

Dataset

We have conducted in-field experiments across 29 days, collecting recordings from 12 subjects. The durations of the recordings range from 3.78 hours to 46.6 hours (see Table 2 for details). The total duration of all recordings is 306.2 hours, containing a total of 237 reported activities, 130 of which were reported with associated stress information. Among the 130 activities with stress information, 102 of them were labeled as stressful.

After the recording device was started, it continuously recorded pulse wave signal (i.e., blood volume pulse) from PPG heart rate sensor (sampling rate = 64 Hz), and acceleration from accelerometer (sampling rate = 32 Hz), along with their timestamps. The pulse wave collected from the PPG sensor was used to calculate HRV-based features, which in turn were used to derive stress levels. The collected motion data was used to estimate the motion level, which was primarily used to gauge the quality of the PPG signal. Specifically, we first extracted RR intervals from pulse wave. Then the RR intervals contained within a 5-min window were used to calculate short-term HRV features that have been studied and suggested to be a good indicator of stress level[18,19]. We adopted the time-domain methods[20] to calculate two HRV features that reflect the total power and the high frequency power of HRV. Finally, we fed the resulted features into a model that maps the features to a stress level ranging from -1 to 1, where -1 indicates very relaxed, and 1 represents very stressed.

Figure 1 demonstrates StressHacker’s continuous stress level output using data from a typical in-field experiment as an example. The final output is a time series of stress levels ([-1, 1]) with a 5-minute interval. The stress level is categorized into four states: *Relaxed* [-1, -0.2], *Less Relaxed* (-0.2, 0], *Slightly Stressed* (0, 0.2] and *Stressed* (0.2, 1]. Note that in the current stage of evaluation, we use a fixed mapping from HRV features to stress level, and a pre-defined categorization of stress states. However, in practice, the system will be able to adjust the related parameters over time according to the user’s feedback.

Table 2: A list of detailed information for each individual experiment. N(Activity) is the number of activities reported by the subjects. N(StressInfo) indicates the number of reported activities with stress information. N(Stress) is the number of activities that have stress information and are labeled as stressful activity. The evaluation results are represented in TP (true positive), FN (false negative), TN (true negative) and FP (false positive). They are obtained by comparing StressHacker's stress level output against the stress information reported by the subject (precision=86.1%, recall=91.2%, F-measure=88.6%).

Exp ID	**Duration (Hr)**	**N(Activity)**	**N(StressInfo)**	**N(Stress)**	**TP**	**FN**	**TN**	**FP**
A1	30.85	17	10	8	8	0	1	1
B1	11.77	7	6	5	3	2	1	0
C1	7.57	11	9	6	6	0	0	3
C2	7.41	7	6	6	6	0	0	1
D1	7.42	18	10	10	9	1	0	0
E1	5.74	5	0	0	0	0	0	0
E2	8.11	10	0	0	0	0	0	0
E3	6.37	6	0	0	0	0	0	0
F1	14.29	9	8	5	4	1	2	1
F2	7.5	5	5	4	3	1	1	0
F3	46.6	20	16	13	11	2	3	0
G1	8	9	6	4	4	0	1	1
G2	8.13	5	4	2	2	0	0	2
H1	22.31	9	6	4	3	1	1	1
H2	21.79	9	6	5	4	1	0	1
I1	15.49	18	12	6	6	0	4	2
I2	10.11	8	7	6	6	0	0	1
J1	8.88	5	0	0	0	0	0	0
J2	3.78	5	3	3	3	0	0	0
J3	9.54	6	3	3	3	0	0	0
K1	6.68	4	0	0	0	0	0	0
K2	7.15	6	3	3	3	0	0	0
K3	7.4	6	2	2	2	0	0	0
K4	8.1	8	4	3	3	0	1	0
L1	7.46	12	4	4	4	0	0	0
L2	7.74	12	0	0	0	0	0	0
total	306.19	237	132	102	93	9	15	15

Result

In this section, we evaluate the feasibility of continuously monitoring daily stress using smartwatches using quantitative metrics that reflect the correlation between StressHacker's output and subject-reported stress information. Therefore, the evaluation is based on the 20 in-field experiments in which the subject provided stress information (i.e., experiments in Table 2 with N(StressInfo)> 0).

First, we identify "stressful" activities based on the reported stress information. Specifically, we classify an activity as "stressful" when its associated stress information satisfies one of the following two conditions: (1) it contains word(s) such as "stressful", "stressed", "nervous" or "anxious"; (2) its associated perceived stress level (PSL) is larger than the baseline 1 (in a few cases, subjects reported PSL=2 as relaxed, thus we used 2 as baseline in such cases).

Next, for the purpose of calculating the correlation, we determine a similar binary label ("stressful" or "not stressful") for each activity using StressHacker's output within the activity's associated time period. Specifically, an activity will be labeled as "stressful" if the stress levels within its time period satisfy one of the following two conditions: (1) at least one of the stress levels are categorized as "Stressed" (0.2, 1]; (2) at least 10% of the stress levels are categorized as “Slightly Stressed” (0, 0.2].

Lastly, we compare the binary labels determined by the self-report stress information and StressHacker’s output. The detailed result of the comparison is shown in Table 2. We can see StressHacker labels 108 activities (93TP+15FP) as “stressful,” with 93 classifications being true positives, leading to a precision of 86.1%. Out of 102 “stressful” activities reported in the subjects’ logs, 93 activities are correctly labeled by StressHacker, resulting in a recall (sensitivity) of 91.2%. Overall, StressHacker achieves an accuracy of 81.8% in the binary classification of “stressful” activity. We believe that such evaluation results are sufficient to suggest that the stress level information provided by StressHacker can be used to reliably monitor the dynamics of one’s daily stress.

Discussion

In this section, we look beyond the scope of quantitative evaluation and discuss StressHacker’s potential for offering valuable and timely insights that could fuel a broad range of decision makings in not only stress management at an individual level but also mental wellness at an organization or community level.

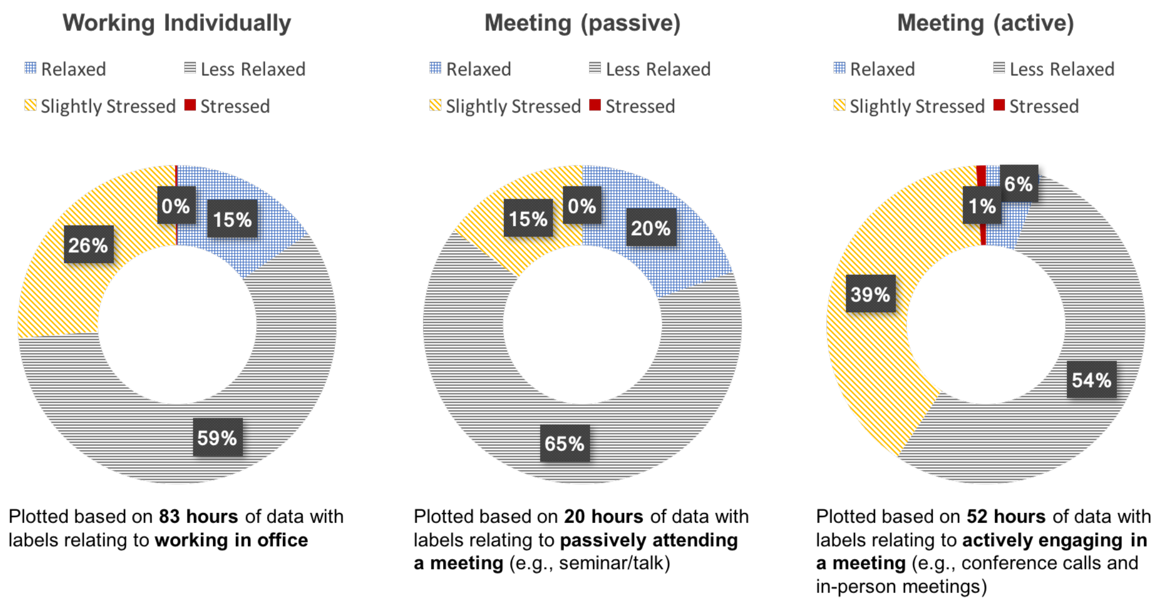

Figure 2: Comparing StressHacker’s stress level output across three typical work scenarios.

We first look at StressHacker’s stress level output in three typical work scenarios reported in the in-field experiment, which are working individually (83 hours), passively attending a meeting (20 hours) and actively engaging in a meeting (52 hours). As all of the in-field experiments were conducted during week days, the combination of these three scenarios accounts for more than half of the overall data collection. Figure 2 shows the comparison of stress levels. A key observation is that, overall, the subjects were most stressed while actively engaging in a meeting (40% of the time), compared with while working individually (26% of the time) and during a passive meeting (15%). Such insight can also be supported by the percentages of the “relaxed” state detected by StressHacker in each scenario, which suggests that the subjects were less relaxed during active meetings (only 6% of the time), compared with 15% of the time when working individually and 20% of the time during passive meetings. As all of the 12 subjects are employees at IBM Research, we speculate that the result above could be reflective, to some extent, of the overall “stress profile” of the organization under different working scenarios.

In addition to the work scenarios, we have also captured a typical non-work scenario that was labeled as “getting ready in the morning” (reported by 4 subjects in 6 days). Figure 3 shows the percentages of different stress levels. We can see that, although it has higher percentage of time being relaxed, the overall percentage of time being stressed (41%) is similar to that in the scenario of actively engaging in a meeting. Such insights offered by StressHacker could potentially be used to identify major stressors and provide personalized recommendations to help users navigate through daily stress. For instance, when significant stress has been detected on every Monday morning, the system could recommend trying to get up half an hour earlier.

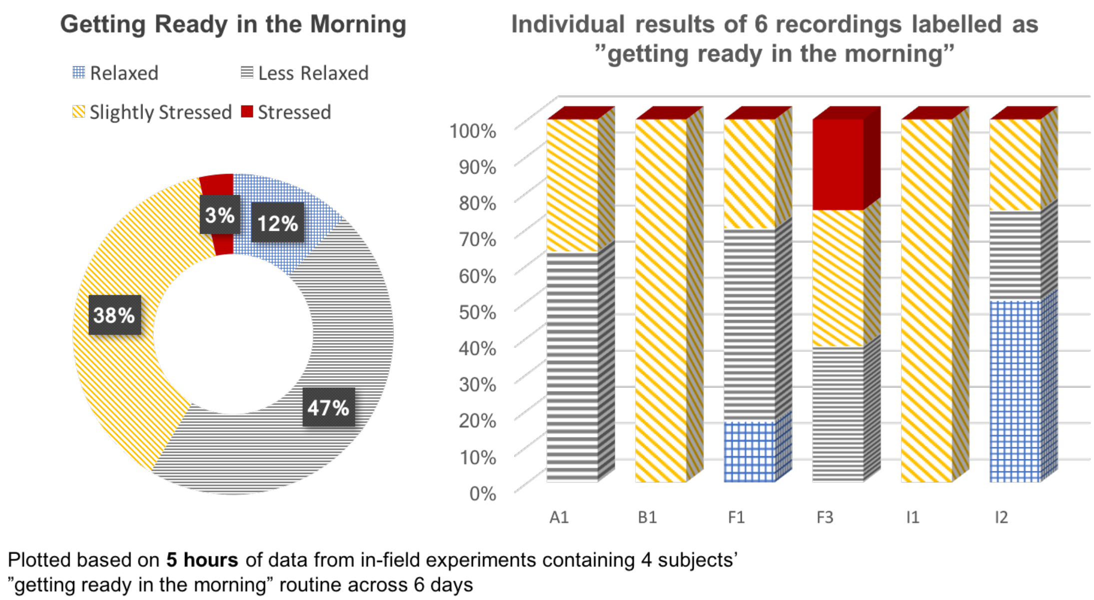

Figure 3: Percentages of stress levels detected by StressHacker in a typical non-work scenario labeled as “getting ready in the morning.”

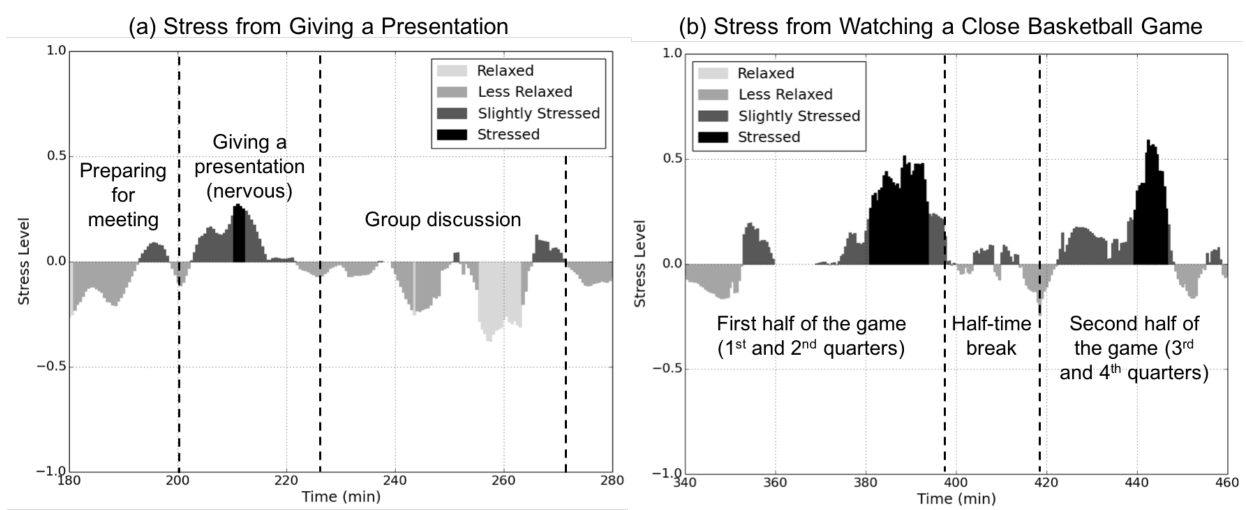

Figure 4: Demonstrating StressHacker’s ability to reveal fine-grained and insightful stress dynamics using two daily events captured in the in-field experiments as examples. The continuous stress level detected by StressHacker and the self-report activity/stress information are plotted over time.

Next, we zoom in on StressHacker’s continuous stress level output, and use two examples captured in the in-field study to showcase that StressHacker’s result allows us to take a closer look at the dynamics of stress associated with an event of interest, which in turn could potentially help us make more informed decision to minimize the negative impacts of stress. Figure 4 (a) shows that StressHacker captures the subject’s stress level changes before, during and after her presentation. During the time when she was preparing for the meeting, we can observe the transition from being relaxed to slightly stressed, which could be the effect of working under a close deadline and/or the anticipation of the coming presentation. Interestingly, around the 200th minute, StressHacker's result shows that she recovers from being slightly stressed to less relaxed, which could be because that she had finished the preparation resulting in a reduction in her stress from working under deadline. Unsurprisingly, StressHacker's result shows a substantial increase in her stress level especially in the first half of her presentation, which is confirmed by the self-reported stress information where the subject reported as being "nervous". In the course of the following group discussion, we can see the subject stayed at being slightly stressed for about 10 minutes before she started recovering and became relaxed at around the 260th minute.

The second case shown in Figure 4 (b) captures the subject's stress dynamics while watching a close basketball game. From StressHacker's result, we can see that overall, the subject was stressed through out the game, with a brief recovery during the half-time break. Interestingly, StressHacker's result reveals the the subject's transition from being slightly stressed to stressed in both the first and the second half of the game, which might reflect the effect of stress accumulation. Another interesting observation is that, in the second half of the game, the subject only stayed being slightly stressed for about 15 minutes (from around the 425th to the 440th minute) until his stress elevated to a higher level (shown in dark bars), whereas the duration of which in the first half of the game was about twice as longer (from around the 350th to the 380th minute).

We believe that StressHacker's ability to offer such fine-grained and real-time stress monitoring can significantly advance the automation of personalized stress management. To put it into perspective, we compare StressHacker with cStress[10], a stress monitoring system proposed in a recent research study. Compared with StressHacker, cStress relies on data collected from a larger set of sensors including inductive plethysmography (RIP), a two-lead electrocardiograph (ECG) and 3-axis accelerometers, which are integrated into a chest band that users need to wear all the time. An in-field study had been conducted to evaluate cStress in real-life scenarios based on a similar task to the one used in our study (identifying "stressful" and "not stressful" activities and comparing with subjects' self-report data). The evaluation result suggests that cStress achieves a median accuracy of 72%, whereas StressHacker's overall accuracy is 81.8%. Therefore, although only relying on data from PPG sensor, a less reliable heart rate sensor compared with ECG, StressHacker is still able to provide reliable results that are sufficient for daily stress monitoring.

Limitations and Future Work

In the paper, we focus on evaluating the feasibility of using bio-signals obtained from off-the-shelf wearables to infer daily stress. A major cause of the limited 81% accuracy is that data collected from PPG sensor is susceptible to various interferences. We have examined the collected data in attempts to identify factors that might affect the robustness of the system. Such factors include daily activities that involve relatively high level of motion (e.g., walking up/down stairs). As PPG sensor is susceptible to motion artifacts, such activities can lead to unreliable or even corrupted sensor data, which in turn, result in inaccurate stress detection results. As part of our future work, we aim to resolve this limitation by developing algorithms that can identify and mitigate the impacts of motion artifacts. To further improve the robustness, we plan to leverage the user's historical data to better understand the user's response to a certain context. For instance, the system might be able to learn over time that the user rarely gets stressed when s/he is driving home, and therefore classifies an isolated stress state with a short duration incorrectly detected in that scenario as false alarm.

As a next step of this study, we will investigate how users can address their stress experiences in real time and how not to distract users during high concentration activities. We are in the process of conducting a study aimed at providing users with real time data on their stress levels via a smartwatch. We intend to make these devices adaptive and personalized to deliver context aware notifications and stress management recommendations. This means that the notification and recommendations should not be delivered when users are performing high concentration tasks or activities. The printout of stress level output shown in Figure 1 is the first step toward such a real-time stress feedback and management system. In future studies we will investigate if the notifications help individuals become more mindful of stress from common daily activities and if increased mindfulness is associated with improved well-being. Further, we will evaluate if completion of real-time stress management recommendations (e.g., meditation, deep breathing) is associated with improved well-being.

Conclusion

In this paper, we present the result of our uncontrolled in-field study that aims to evaluate the feasibility of continuous daily stress monitoring using smartwatches. Based on more than 300 hours of data collected in real-life setting, the quantitative results of the evaluation suggest that the proposed system, StressHacker, can be used as a practical tool for effective and personalized stress management. Moreover, we have also presented analyses at both individual and community levels to showcase StressHacker's potential to offer valuable data and insights that can be used to advance the automation and personalization of stress management and promote mental wellness in general.

References

1. American Psychological Association. Stress in America [Internet]. 2015 [cited 2017 Feb 28]. Available from: https://www.apa.org/news/press/releases/stress/2014/stress-report.pdf
2. Mommersteeg PMC, Herr R, Zijlstra WP, Schneider S, Pouwer F. Higher levels of psychological distress are associated with a higher risk of incident diabetes during 18 year follow-up: results from the British household panel survey. BMC Public Health. 2012 Dec 23;12:1109.
3. Schneiderman N, Ironson G, Siegel SD. STRESS AND HEALTH: Psychological, Behavioral, and Biological Determinants. Annu Rev Clin Psychol. 2005;1:607-28.
4. Sin NL, Sloan RP, McKinley PS, Almeida DM. Linking Daily Stress Processes and Laboratory-Based Heart Rate Variability in a National Sample of Midlife and Older Adults. Psychosom Med. 2016 Jun;78(5):573-82.
5. Althubaiti A. Information bias in health research: definition, pitfalls, and adjustment methods. J Multidiscip Healthc. 2016 May 4;9:211-7.
6. Hoyt LT, Ehrlich KB, Cham H, Adam EK. Balancing scientific accuracy and participant burden: testing the impact of sampling intensity on diurnal cortisol indices. Stress Amst Neth. 2016 Sep;19(5):476-85.
7. Cohen JB, Cohen DL. Integrating Out-of-Office Blood Pressure in the Diagnosis and Management of Hypertension. Curr Cardiol Rep. 2016 Nov 1;18(11):112.
8. Healey JA, Picard RW. Detecting stress during real-world driving tasks using physiological sensors. IEEE Transactions on intelligent transportation systems. 2005 Jun;6(2):156-66.
9. Lu H, Frauendorfer D, Rabbi M, Mast MS, Chittaranjan GT, Campbell AT, Gatica-Perez D, Choudhury T. Stresssense: Detecting stress in unconstrained acoustic environments using smartphones. InProceedings of the 2012 ACM Conference on Ubiquitous Computing 2012 Sep 5 (pp. 351-360). ACM.
10. Hovsepian K, al'Absi M, Ertin E, Kamarck T, Nakajima M, Kumar S. cStress: towards a gold standard for continuous stress assessment in the mobile environment. InProceedings of the 2015 ACM international joint conference on pervasive and ubiquitous computing 2015 Sep 7 (pp. 493-504). ACM.
11. McEwen BS. Stress, adaptation, and disease. Allostasis and allostatic load. Ann N Y Acad Sci. 1998 May 1;840:33-44.
12. McCorry LK. Physiology of the Autonomic Nervous System. Am J Pharm Educ [Internet]. 2007 Aug 15 [cited 2017 Feb 25];71(4). Available from: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC1959222/
13. Shaffer F, McCraty R, Zerr CL. A healthy heart is not a metronome: an integrative review of the heart's anatomy and heart rate variability. Front Psychol. 2014;5:1040.
14. Hughes JW, Stoney CM. Depressed mood is related to high-frequency heart rate variability during stressors. Psychosom Med. 2000 Dec;62(6):796-803.
15. Mazurak N, Sauer H, Weimer K, Dammann D, Zipfel S, Horing B, et al. Effect of a weight reduction program on baseline and stress-induced heart rate variability in children with obesity. Obes Silver Spring Md. 2016 Feb;24(2):439-45.
16. Ruediger H, Seibt R, Scheuch K, Krause M, Alam S. Sympathetic and parasympathetic activation in heart rate variability in male hypertensive patients under mental stress. J Hum Hypertens. 2004 May;18(5):307-15.
17. Garbarino M, Lai M, Bender D, Picard RW, Tognetti S. Empatica E3 – A wearable wireless multi-sensor device for real-time computerized biofeedback and data acquisition. InWireless Mobile Communication and Healthcare (Mobihealth), 2014 EAI 4th International Conference on 2014 Nov 3 (pp. 39-42). IEEE.

18. Esco MR, Flatt AA. Ultra-short-term heart rate variability indexes at rest and post-exercise in athletes: evaluating the agreement with accepted recommendations. Journal of sports science & medicine. 2014 Sep;13(3):535.

19. Hjortskov N, Rissn D, Blangsted AK, Fallentin N, Lundberg U, Sgaard K. The effect of mental stress on heart rate variability and blood pressure during computer work. European journal of applied physiology. 2004 Jun 1;92(1-2):84-9.

20. Camm AJ, Malik M, Bigger JT, Breithardt G, Cerutti S, Cohen RJ, Coumel P, Fallen EL, Kennedy HL, Kleiger RE, Lombardi F. Heart rate variability. Standards of measurement, physiological interpretation, and clinical use. European heart journal. 1996 Mar 1;17(3):354-81.