

WellBeat: A Framework for Tracking Daily Well-being Using Smartwatches
Corresponding author: Marios Constantinides, marios.constantinides@nokia-bell-labs.com

WellBeat: A Framework for Tracking Daily Well-being Using Smartwatches

Sungkyu Park

KAIST, Daejeon (South Korea)

Marios Constantinides

Nokia Bell Labs, Cambridge (UK)

Luca Maria Aiello

Nokia Bell Labs, Cambridge (UK)

Daniele Quercia

Nokia Bell Labs, Cambridge (UK); King's College, London (UK)

Paul van Gent

Delft University of Technology, Delft, (Netherlands)

Abstract—Human physiology is a window to our physical, mental, and emotional states; our well-being. Today, a new wave of objective data derived from consumer-grade body sensors—like those equipped by smartwatches—paves the way towards a new approach in how well-being is being measured, continuously and unobtrusively. Here, we developed a framework for collecting and analyzing physiological data using smartwatches in-the-wild, and demonstrated its robustness in data obtained away from controlled laboratory settings. We found that changes in people's heart rate and heart rate variability are predictive not of momentary well-being (a scientific idea that continues to live on in the absence of in-the-wild evidence, aka, zombie theory) but of daily well-being.

Index Terms: O.1 Affect sensing and analysis; O.2 Modeling human emotion; O.6.1 Methods of data collection

Subjective Well-being (SWB) is a multifaceted construct. In literature, SWB and happiness are often used interchangeably, and refer to 'people's cognitive and affective evaluations of their life' [1]. Psychologists describe it as the aspect of happiness that can be empirically

■ INTRODUCTION

measured, or the presence of positive emotions and absence of negative ones. Measuring one's happiness is a challenging task. To grasp its constituents [2], scholars often resort to measure it using surveys (e.g., PANAS), usually administered to restricted groups of people, or occasionally, to larger populations over longer periods of time. Nevertheless, their high costs, and reliance on people's recollection to provide truthful reports [1] greatly limits their scale. The diffusion of mobile and wearable devices enables a scalable way of sampling people's experiences and feelings [3], thus facilitating well-being reports collection at scale, systematically, and objectively [4].

Often, these self-reports are used to train algorithms that predict well-being from behavioral traces. One line of work focused on predicting self-reports from contextual data like a person's current activity [4]. A second, more recent research stream used commercial-grade body sensors to predict well-being, stress, and affective states from biological signals [5], or motion [6]. The first approach can predict macro-scale well-being (e.g., people tend to be happier when outdoors), but fails to grasp the variation of momentary feelings when external conditions are fixed; this is a severe limitation, as people spend most of their time indoors performing a restricted number of activities. The latter approach has proven effective in predicting momentary feelings in controlled experimental settings (e.g., heart rate to identify emotional responses to videos), but has failed in estimating well-being *in-the-wild*, mainly because data collected by wearables used throughout the day is noisy, and affected by several confounders. To overcome these limitations, we propose to measure well-being in real-time and *in-the-wild* using consumer-grade smartwatches through systematic monitoring of people's physiological changes. In doing so, we make three contributions:

- We developed 'WellBeat', a framework for collecting physiological data, which consists of *i*) a Samsung Galaxy watch application that occasionally prompts users to report their momentary feelings, while continuously sampling their heart rate, and *ii*) a service for data storing and processing. To analyze the

collected data, we implemented a processing pipeline for conducting Heart Rate Variability (HRV) analysis on the raw Photoplethysmography (PPG) signal, and demonstrated its *robustness* (§FRAMEWORK).

- We conducted a three-week study with 12 subjects using an Experience Sampling Method (ESM). In total, we collected 1,121 hours of raw PPG signal, and a total of 1,032 self-reported labels related to happiness, awakenedness, and relaxedness levels (§ESM STUDY).
- We analyzed the extracted HR and HRV parameters, and the self-reported well-being measures in two aggregation levels: (a) momentary, and (b) daily (§ANALYSIS). The distinct patterns in people's heart rate variations explain their happiness, awakenedness, and relaxedness levels (§RESULTS) but, they do so to a greater extent when the aggregation is at daily level; a finding consistent with theoretical expectations. Higher variations in HR are linked to higher happiness levels, and higher variations in HRV are linked with higher levels of awakenedness and relaxedness. Compared to state-of-the-art HRV processing tools, the signal produced by the Wellbeat's data processing pipeline shows a stronger association with the user-reported labels.

RELATED WORK AND BACKGROUND

Heart Rate Variability

In the human body, the Autonomic Nervous System (ANS) is responsible for controlling bodily functions that are not consciously directed [7]. HRV—the physiological phenomenon of variation in the time interval between heartbeats—is one of the most promising markers to assess ANS activity, particularly in psycho-physiological studies [8]. It is based on the measurement of the time elapsed between heartbeats, usually referred to RR intervals.

The most reliable tool for heart rate monitoring is the Electrocardiogram (ECG). It measures the electrical activity of the heart, and the distance between spikes on the ECG line is used to estimate the RR intervals [9]. ECG serves as the gold standard, but cannot be easily measured during daily activities.

A practical and non-invasive alternative to

assess heart's state is the PPG, which relies on optics. A LED sensor illuminates the skin beneath the sensor and a photodiode absorbs the amount of backscattered light, which corresponds to the discoloration of the skin as blood perfuses through it after each heartbeat. Each cardiac cycle appears as peak in the PPG signal. Often, PPG is easily distorted by artifacts introduced by motion, ambient light, or skin tones differences. However, the broad diffusion of PPG sensors in consumer-grade smartwatches, and the availability of signal processing tools to handle noisy signals [10], make it the best alternative to conduct HRV studies at scale [9]. Existing frameworks for HRV analysis like BioSPPy¹, pyHRV², and HRVAS³ toolbox use the Pan-Tompkins algorithm for HRV analysis.

Once RR intervals are extracted, they can be used to extract a number of parameters that model HRV. These parameters are defined in the time domain (e.g, RMSSD, SDNN), frequency domain (e.g., LF, HF), or as non-linear indices (e.g., SD1, SD2) [8].

FRAMEWORK

Next, we describe our data collection framework and the data processing pipeline for conducting HRV analysis.

Data Collection

To study the relationship between people's physiological changes and their well-being, we developed 'WellBeat', a three-tier framework (Figure 1a) composed by an application layer, a web services layer, and a data access layer.

Application layer

To access the watch application users register through a webpage to obtain a 4-digit unique number that anonymously identifies each device, and is used to activate the application on first use.

The application was developed using the Tizen platform; the Samsung's Operating System⁴. It has two components: *i*) a front-end that handles user-facing menus for self-reports collection, and *ii*) a native background service written in C

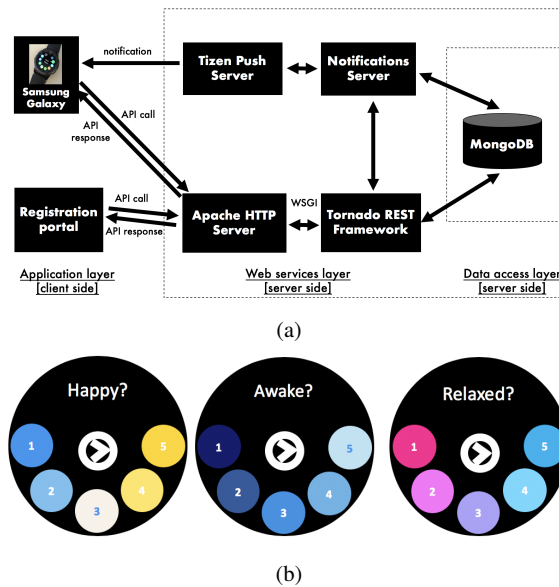


Figure 1: (a) 'WellBeat' framework overview diagram, and (b) the three questions were asked in a Likert-scale 1-5, with (1) indicating not at all, and (5) indicating extremely.

that continuously records the PPG signal, and communicates with the web services.

The User Interface consists of three screens (Figure 1b). Users are nudged into submitting a self-report via notifications triggered by the Tizen Push Server thrice a day, at random times (Figure 1a). Random scheduling notification policies capture well the spontaneous nature of happiness and feelings [4], while being less prone to the risk of cognitive biases than regular scheduling policies [11]. In line with previous work on *experience sampling* [4], a notification consists of three optional questions prompted on separate screens (Figure 1b). The timestamp of responses is recorded.

The PPG signal is sampled every 100ms (10Hz); we refer to this signal as '*ppg signal*'. Instantaneous heart rate, post-processed by Tizen is also sampled; we refer to this signal as '*hrm signal*'. Sampling at 10Hz allowed us continuous recordings throughout the day, and ensuring battery life up to 14 hours Both signals and self-reports are temporarily stored on the watch, and they are periodically transmitted to the server whenever Internet is available, ensuring negligible effects on battery life due to data transmission.

¹<https://biosppy.readthedocs.io/en/stable/>.

²<https://pyhrv.readthedocs.io/en/latest/>.

³<https://github.com/jramshur/HRVAS>.

⁴<https://www.tizen.org/>.

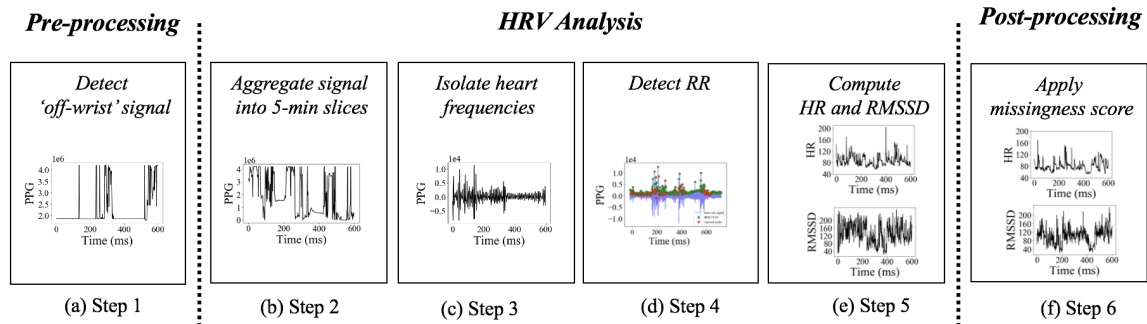


Figure 2: Schematic data flow in our six-step data processing pipeline for conducting HRV analysis. (a) Step 1: detect and filter out off-wrist signal; (b) Step 2: signal slicing; (c) Step 3: pass aggregated signal slices through a bandpass filter to filter out heart frequencies (i.e., 0.8-3.67Hz); (d) Step 4: RR detection using a peak detection algorithm; (e) Step 5: compute HR and HRV estimates using the extracted RR; (f) Step 6: apply the missingness score to the extracted HR and HRV estimates for additional motion artifacts correction.

Web services layer

Using the Tornado framework, we developed a RESTful API that handles layers' communication through three endpoints (Figure 1a). The first endpoint stores data from the registration portal, and generates unique user identifiers. The second stores sensor data and self-reports received from the watch. The third implements an observer pattern that allows the watch to register in the Notifications Server. In turn, it submits POST requests to the Tizen's Push Server, which notifies the watch.

Data access layer

The raw data is stored in a MongoDB instance, in two database collections; the *Sensor* and the *Self-reports* collections.

- **Sensor:** [userID, sensor_type, sensor_value, ts], where *userID* is the 4-digit user identifier, *sensor_type* hrm or ppg, *sensor_value* is the sensor reading, and *ts* is the time that the sensor reading was sampled.
- **Self-reports:** [userID, type, value, ts], where *userID* is the 4-digit user identifier, *type* is the category of the self-reports (i.e., Happy/Awake/Relaxed), *value* is the Likert-scale value of the self-report, and *ts* is the time that the label was reported.

Data processing pipeline and HRV analysis

We built a six-step data processing pipeline, which we describe next

Pre-processing

Step 1: Detect 'off-wrist' signal: The watch application measures PPG continuously, even when users were not wearing the watch. To filter out noise introduced by 'off the wrist' signal, we matched the timestamps of the 'ppg' and 'hrm' signals. When the watch is off the wrist, the instantaneous heart rate provided by the Tizen (hrm signal) is marked as 0 or -3. Therefore, we cleaned the ppg signal using these thresholds (Figure 2a).

HRV analysis

Step 2: Signal slicing: HRV parameters are defined on temporal intervals of the PPG signal. Experts suggest an absolute minimum of 5 minutes to conduct HRV analysis [7]. Therefore, we partitioned the PPG signal in consecutive 5-minute *slices*, and we refer to the i^{th} slice of participant u as $slice_i^u$ (Figure 2b).

Step 3: Isolate heart frequencies: According to the Nyquist-Shannon theorem, when sampling a continuous signal the sampling rate should be at least twice that of the highest frequency to be captured. The heart of most humans beats in a range between 40-220 bpm, which translates to a minimum of 0.67Hz and a maximum of 3.67Hz. Therefore, we filtered out *slices* whose sampling rate is less than the 7.34Hz ($2 \times 3.67Hz$) needed to capture the maximum heart frequency. Despite configuring the watch application recordings at even higher sampling rate (i.e., 10Hz), the sam-

pling rate might fall below this threshold. This could be due to CPU balancing or instantaneous sensor glitch. Then, using Python's SciPy signal package, we applied a band-pass filter to a *slice* to suppress all the frequencies outside the range of the human heart beat (Figure 2c).

Step 4: Detect RR: We first upsampled the PPG signal to 250Hz using the Fourier method [10], which is considered an acceptable frequency range to perform HRV analysis [12]. We then developed an adaptive threshold peak detection algorithm to detect the RR intervals. The algorithm computes a moving average over a temporal window of 1.5 seconds centered on each data point, discovers regions of interest between pairs of points in which the signal amplitude is larger than the moving average, and finally identifies peaks within each region (Figure 2d). The moving average is incremented stepwise, and the analysis is iterated until an optimal peak detection fit has been determined [10].

Step 5: HRV parameters: From RR intervals, we computed the HRV parameters and the μ_{HR} for each signal $slice_i^u$. The output is a vector that corresponds to the HRV parameters (§RELATED WORK AND BACKGROUND) and the instantaneous heart rate (HR) (Figure 2e).

Post-processing

Step 6: Missingness score: To ensure reliable HRV parameters we computed a 'missingness score' [13]; a metric for adjusting error tolerance in HRV analysis by removing motion artifacts. It is calculated using Formula (1), and provides a confidence level of the number of RRs that should have been recorded to ensure reliable estimates; the higher the value, the less reliable the estimates are. As prior work suggests [13], we filtered out any HR and HRV parameters whose missingness score is $> 35\%$ (see Figure 2f).

$$missingness_score = 1 - \frac{observedRRs + 1}{\mu_{HR} \times minutes}, \quad (1)$$

where observedRRs is the number of RR intervals observed in the *slice*.

ESM STUDY

We conducted a three-week ESM study to demonstrate our framework's effectiveness. We

recruited 12 individuals (3 Female) with no medical conditions. They all agreed to share their heart rate data, and were consented in writing for their participation. We anonymized any user identification and made our dataset publicly available⁵.

In total, we collected 1,032 self-reported momentary labels across the three label categories, and sampled 1,121 hours of raw PPG signal. Forty percent of the initial PPG signal (about 445 hours) was eliminated due to participants not wearing their watches all the time (Step 1). About 40 hours were filtered out during HRV analysis (Steps 2-5). A total of 291 hours was left after correcting for motion artifacts (Step 6). This data loss is a general problem often observed in studies where controls over participation are absent. However, we ensured data reliability by following the six-step procedure. For example, the final step (Figure 2f) of the data processing pipeline yields HR and HRV parameters which are aligned with the normative values of healthy adults [14].

ANALYSIS

With HR and HRV parameters at hand, we investigated the extent to which heart rate changes are associated with well-being. We did so by exploring two aggregation levels: (a) momentary, and (b) daily. Some previous studies showed that HR and HRV cannot approximate momentary self-reported labels [15], while others report contrasting results [16], [17], or even question the use of various HRV indicators through the idea of zombie theories [18]. In both aggregation levels, we focused on the HR and the RMSSD parameter as the HRV measure due to its wide use across studies [8].

Hypotheses

We formulated three hypotheses:

- **H₁:** Happiness is a multi-faceted construct [1]. It is often linked to activity [5] or explains a state of relaxation [19]. Due to its complex structure, we investigated both HR and HRV, as follows: (a) The higher or the lower the HR, and (b) the lower the HRV, people are, on average, *happier* [19].

⁵<https://social-dynamics.net/wellbeat/dataset>

Department Head

- **H₂**: Awakeness is characterized by higher alertness levels and reactivity to cognitive tasks. Higher HR is often linked to higher alertness levels [20]. Therefore, we hypothesize that the higher the HR, the more *awake* people are.
- **H₃**: Relaxedness is often linked with the absence of stress. Prior works showed that HRV is a good predictor of physiological arousal and, in turn, how well the body copes with stress [16], [17]. Therefore, we formulate a hypothesis as follows: The higher the HRV, people are, on average, more *relaxed*.

Momentary aggregation

Similar to [15], [16], we matched the extracted HRV parameters at a window of ± 10 minutes from the time a momentary self-reported label was reported. This translates into 2 slices before and after the reported label.

Daily aggregation

We computed a metric that captures the *daily* variations of people's HR and HRV parameters compared to their baseline. We generalized the procedure such that it applies for each HRV parameter f described in (§RELATED WORK AND BACKGROUND).

First, we computed the user's baseline for each feature f as μ_B , where B is the set of a user's values for feature f across all days. Then, we performed an hourly aggregation by computing the $f@hour = \mu_{hour} - \mu_B$, which is the difference of the hourly value of feature f compared to the user's baseline for that feature. Finally, we aggregated the hourly changes into the daily metric using Formula (2).

$$\Delta_f@day = \frac{1}{n} \sum_{i=1}^n f@hour_i, \quad (2)$$

where n is the number of total hours per day, and i is the i^{th} hour of a day.

Next, we matched each user's daily metric with each self-report category's probabilities. To do so, we computed the probability, $P(W)@day$, for each $W \in \{\text{Happy, Awake, Relaxed}\}$, as $P(W) = \frac{|WL|}{|L|}$, where L is the set of all W self-reports across all days, WL is the set of W self-reports computed as $WL = \{l \in L \mid l \geq \bar{M}L\}$,

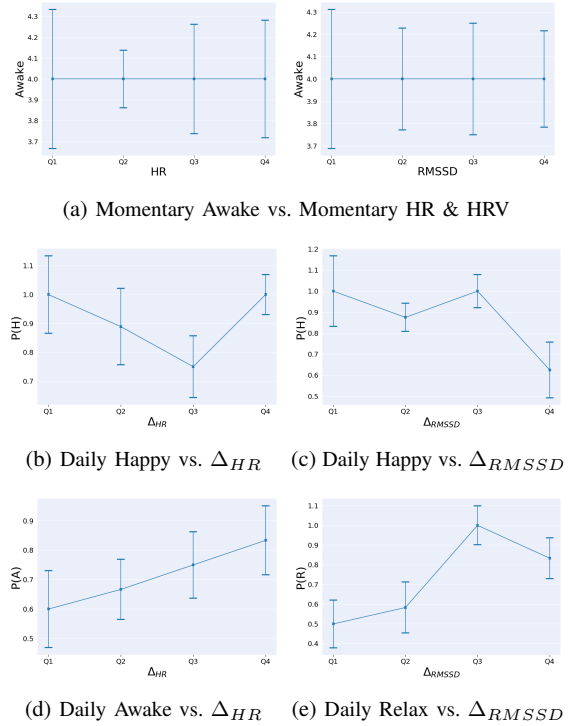


Figure 3: HR and RMSSD *momentary* (a) and *daily* (b-e) variations for Happiness, Awakeness, and Relaxedness (x-axis: quartiles; y-axis: (a) 5-point Likert scale, and (b-e) daily probability for each self-report category).

where l is a self-report, and $\bar{M}L$ the median value of all W self-reports. The median value represents each user's most frequently self-report value in each label category. Given each user's own self-assessment, the choice of median served as a representative way to discriminate the positive and negative classes in each label category.

RESULTS

At daily level, people are happier when their HR varies little from their baseline, or in extreme ranges. High deviations from one's baseline (Figure 3b) might be linked to activity, as previous work established the link between activity and happiness [5], while small deviations might explain a state of relaxation. Consistent with prior emotion research and ANS activity [19], we observed that lower HRV deviations (Figure 3c) are linked to higher happiness levels. Awakeness was higher when HR deviated significantly from one's baseline, in line with prior work [20]. Higher relaxation was observed when people's

HRV varied in higher ranges than their baseline. Previous work established the link of HRV as a proxy to assess one's physiological arousal [16], [17]; higher values suggest that the body copes better with stress.

For momentary aggregations, however, no specific patterns emerged (Figure 3a). This result confirms the difficulty in tracking people's momentary affective states documented in previous work [15]. Contrarily, when aggregating at a coarser-grained granularity, the patterns in people's HR and HRV matched their happiness, awakenedness, and relaxedness levels as expected from our hypotheses.

Additionally, we compared our WellBeat six-step pipeline (a) with a different pipeline that replaces steps 4 and 5 with a combination of the state-of-the-art tools BioSPPy and pyHRV, and (b) with a BioSPPy and pyHRV configuration without step 6 (c). We computed the daily metrics using these three configurations. We then set up a classification task to predict the three daily labels from the metrics computed for the three configurations separately. We implemented Logistic Regression classifiers and cross-validated them in a 90/10 train-test split setting. We observed, on average, a relative improvement of 9.06% in AUC among all three labels for configurations (a) over (b), corroborating our framework's robustness against state-of-the-art. We observed the highest relative gain of 12.7% when predicting relaxedness levels (overall AUCs of 0.592 and 0.57 in configurations (a) and (b) respectively), followed by awakenedness with 9.3%, and happiness levels with 5.2%. Such finding is in line with previous HRV literature [8], but also corroborates the difficult nature of this task under free-living conditions [15]. We also observed, an average improvement of 5.2% of configuration (b) over (c), which highlights the importance of additional checks when operating with noisy PPG signals in free-living conditions.

CONCLUSION

We examined the feasibility of measuring people's SWB using physiological data in-the-wild. In doing so, we developed 'WellBeat', a framework for collecting heart rate data from consumer-grade smartwatches and conducting HRV analysis. Unlike controlled laboratory stud-

ies, we deployed our framework in the wild.

We found distinct patterns in people's heart rate variations which are linked with their happiness, awakenedness, and relaxedness, but only when aggregating the signal at daily level. This supports the idea that tracking momentary constituents of well-being might not be feasible, particularly in-the-wild [15]. The daily level aggregation instead revealed distinct patterns which are consistent with theoretical expectations, either conducted under controlled settings [6] or utilized post-processed data from wearable devices such as Empatica E4 [16].

The results corroborate the validity of our framework, and pave the way towards a new approach of systematic monitoring people's physiological states unobtrusively and continuously.

REFERENCES

1. Diener, E., Lucas, R. & Oishi, S. Subjective well-being: The science of happiness and life satisfaction. *Handbook Of Positive Psychology*. 2, 63–73 (2002)
2. Csikszentmihalyi, M. Flow: The psychology of happiness. *Random House*. (2013)
3. Hänsel, K., Alomainy, A. & Haddadi, H. Large scale mood and stress self-assessments on a smartwatch. *Proceedings of the ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct*. 1180–1184 (2016)
4. Mackerron, G. & Mourato, S. Happiness is greater in natural environments. *Global Environmental Change*. 23, 992–1000 (2013)
5. Gloor, P., Colladon, A., Grippa, F., Budner, P. & Eirich, J. Aristotle said "happiness is a state of activity": Predicting mood through body sensing with Smartwatches. *Journal Of Systems Science And Systems Engineering*. 27, 586–612 (2018)
6. Quiroz, J., Geangu, E. & Yong, M. Emotion Recognition Using Smart Watch Sensor Data: Mixed-Design Study. *Jmir Mental Health*. 5, e10153 (2018)
7. Camm, A., Malik, M., Bigger, J., Breithardt, G., Cerutti, S., Cohen, R., Coumel, P., Fallen, E., Kennedy, H. & Kleiger, R. Heart rate variability: standards of measurement, physiological interpretation and clinical use. Task Force of the European Society of Cardiology and the North American Society of Pacing and Electrophysiology. (1996)
8. Laborde, S., Mosley, E. & Thayer, J. Heart rate variability and cardiac vagal tone in psychophysiological research—recommendations for experiment planning,

Department Head

- data analysis, and data reporting. *Frontiers In Psychology*. **8**, 213 (2017)
9. Schmidt, P., Reiss, A., Duerichen, R. & Vanlaerhoven, K. Wearable affect and stress recognition: A review. *Arxiv Preprint Arxiv:1811.08854*. (2018)
 10. van Gent, P., Farah, H., van Nes, N. and van Arem, B. Analysing noisy driver physiology real-time using off-the-shelf sensors: heart rate analysis software from the taking the fast lane project *Journal of Open Research Software*. **7** (2019)
 11. Consolvo, S. & Walker, M. Using the experience sampling method to evaluate ubicomp applications. *Ieee Pervasive Computing*. **2**, 24–31 (2003)
 12. Quintana, DS and Alvares, Gail A and Heathers, JAJ. Guidelines for Reporting Articles on Psychiatry and Heart rate variability (GRAPH): recommendations to advance research communication. *Translational psychiatry*. **6**, e803 (2016)
 13. Ryan, W., Conigrave, J., Basarkod, G., Ciarrochi, J. and Sahdra, B.K. When is it good to use wristband devices to measure HRV?: Introducing a new method for evaluating the quality of data from photoplethysmography-based HRV devices. *PsyArXiv*. (2019)
 14. Nunan, D., Sandercock, G.R. and Brodie, D.A. A quantitative systematic review of normal values for short-term heart rate variability in healthy adults. *Pacing and Clinical Electrophysiology*. **33**, 1407–1417 (2010)
 15. Schmidt, P., Dürichen, R., Reiss, A., Van Laerhoven, K. and Plötz, T. Multi-target affect detection in the wild: an exploratory study. *Proceedings of the 23rd ACM International Symposium on Wearable Computers*. 211–219 (2019)
 16. Gjoreski, M., Luštrek, M., Gams, M. & Gjoreski, H. Monitoring stress with a wrist device using context. *Journal Of Biomedical Informatics*. **73** pp. 159–170 (2017)
 17. Kim, D., Seo, Y., Cho, J. and Cho, C.H. Detection of subjects with higher self-reporting stress scores using heart rate variability patterns during the day. *30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. 682–685 (2008)
 18. Heathers, J and Goodwin, M. LF/HF HRV: the 'Life After Death' of a refuted theory. *PsyArXiv*. (2017)
 19. Kreibig, Sylvia D. Autonomic nervous system activity in emotion: A review. *Biological psychology*. **3**, 394–421 (2010)
 20. Hansen, A.L., Johnsen, B.H. and Thayer, J.F. Vagal influence on working memory and attention. *International Journal of Psychophysiology*. **48**, 263–274 (2003)

Sungkyu Park is a Ph.D. candidate in the Graduate School of Culture Technology at KAIST (South Korea). He is interested in understanding human behavior and psychiatric disorders through large-scale data. Contact him at shaun01.park@gmail.com.

Marios Constantinides is a Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge (UK). He is interested in human-computer interaction and affective computing. Contact him at marios.constantinides@nokia-bell-labs.com.

Luca Maria Aiello is a Senior Research Scientist in the Social Dynamics team at Nokia Bell Labs Cambridge (UK). He conducts interdisciplinary computational social science research. Contact him at luca.aiello@nokia-bell-labs.com.

Daniele Quercia is the Department Head at Nokia Bell Labs in Cambridge (UK) and Professor of Urban Informatics at King's College London. He works in the areas of computational social science and urban informatics. Contact him at quercia@cantab.net.

Paul van Gent is a Postdoctoral Researcher at Delft University of Technology (Netherlands). His work focuses on driver physiology, embedded systems, sensing hardware, and software. Contact him at p.vangent@tudelft.nl.