Journal of Biomedical Informatics 63 (2016) 141–149

Contents lists available at ScienceDirect

Journal of Biomedical Informatics

journal homepage: www.elsevier.com/locate/yjbin

Let's get Physiqual – An intuitive and generic method to combine sensor technology with ecological momentary assessments

F.J. Blaauw [a,b,*], H.M. Schenk [a], B.F. Jeronimus [a,c], L. van der Krieke [a], P. de Jonge [a,c], M. Aiello [b], A.C. Emerencia [a]

[a] University of Groningen, University Medical Center Groningen, University Center for Psychiatry, Interdisciplinary Center Psychopathology and Emotion Regulation (ICPE), The Netherlands

[b] University of Groningen, Johann Bernoulli Institute for Mathematics and Computer Science (JBI), Distributed Systems Group, The Netherlands

[c] University of Groningen, Department of Developmental Psychology, The Netherlands

ARTICLE INFO

Article history:
Received 8 March 2016
Revised 11 June 2016
Accepted 2 August 2016
Available online 4 August 2016

Keywords:
API
Platform
Interoperability
EMA
ESM
Wearables
Smartwatch
Fitbit
Android
Google Fit

ABSTRACT

The emergence of wearables and smartwatches is making sensors a ubiquitous technology to measure daily rhythms in physiological measures, such as movement and heart rate. An integration of sensor data from wearables and self-report questionnaire data about cognition, behaviors, and emotions can provide new insights into the interaction of mental and physiological processes in daily life. Hitherto no method existed that enables an easy-to-use integration of sensor and self-report data. To fill this gap, we present 'Physiqual', a platform for researchers that gathers and integrates data from commercially available sensors and service providers into one unified format for use in Ecological Momentary Assessments (EMA) or Experience Sampling Methods (ESM), and Quantified Self (QS). Physiqual currently supports sensor data provided by two well-known service providers and therewith a wide range of smartwatches and wearables. To demonstrate the features of Physiqual, we conducted a case study in which we assessed two subjects by means of data from an EMA study combined with sensor data as aggregated and exported by Physiqual. To the best of our knowledge, the Physiqual platform is the first platform that allows researchers to conveniently aggregate and integrate physiological sensor data with EMA studies.

1. Introduction

In Ecological Momentary Assessments (EMA) and other electronic diary methods, participants are repeatedly assessed for a certain period of time (usually days to weeks), by administering a single or a set of questionnaires on a relatively high frequency (e.g., daily or multiple times per day) [1,2]. The EMA approach has several advantages over traditional cross-sectional approaches in which an assessment is conducted from a large population sample at one or a few points in time [3,4]. With EMA, moment-to-moment fluctuations in physiological conditions and psychological states – such as cognition and affect – can be recorded in real-time, reducing recall bias. Additionally, personal daily dynamics can reveal the influences of time and setting on mental health [5].

Nowadays, many people measure various aspects of their lives using sensors in wearables including activity trackers and smartwatches [6,7]. Wearable sales have increased greatly over the past few years, which is an indication of their growing popularity [8]. Furthermore, with the recent introduction of the smartwatch, personal health monitoring gained widespread adoption. Personal health monitoring may include monitoring of activity or sleep patterns, calories used, and heart rate, depending on the type of sensors integrated in the wearable [9]. Also, in the medical field, the interest for – and prospects of – monitoring physiological parameters of patients using different types of sensors is increasing [10].

The combination of psychological data from EMA with physiological sensor data could provide new insights into the interaction of mental and physiological processes in daily life. When sensor data is proven to be sufficient, reliable, and accurate, sensors could be used complementary to or replace some of the self-report questionnaire items concerning physical activities. The combination of EMA data with sensor data has previously been explored by other

* Corresponding author at: Interdisciplinary Center Psychopathology and Emotion Regulation, University Medical Center Groningen, University of Groningen, Groningen 9700 RB, The Netherlands.

E-mail addresses: f.j.blaauw@rug.nl (F.J. Blaauw), h.m.schenk@umcg.nl (H.M. Schenk), b.f.jeronimus@rug.nl (B.F. Jeronimus), j.a.j.van.der.krieke@umcg.nl (L. van der Krieke), peter.de.jonge@rug.nl (P. de Jonge), m.aiello@rug.nl (M. Aiello), a.c.emerencia@rug.nl (A.C. Emerencia).

http://dx.doi.org/10.1016/j.jbi.2016.08.001
1532-0464/

researchers. For example, Booij et al. [11] performed an EMA study wherein the participants wore an accelerometer to measure physical activity whilst filling out EMA questionnaires. Such studies often include few participants and use expensive, single-purpose devices for the sensor measurements. The cost of these devices impedes large scale integration of sensor technology in EMA studies. However, with the increasing popularity and quality of smartwatches and other sensor-equipped wearables in recent years, it becomes possible to use sensor data from wearables that a participant is already wearing. Currently, integrating these commercially available wearables for use in large scale EMA studies is a nontrivial task due to the diversity of the different service providers and the incompatibility of the exported data formats with EMA data.

In this paper we present *Physiqual*, a novel approach to summarize data output by wearables in a unified format for use in EMA health research and Quantified Self. The novelty of Physiqual resides in the fact that to date and to the best of our knowledge no method exists that can automatically integrate data from commercially available wearable sensors with existing EMA studies with the potential to be used in large scale research. We provide a detailed description of the functionality of the Physiqual platform and also demonstrate its practical usefulness in a case study. For this case study, a trial was conducted in which two subjects wore a Fitbit (http://fitbit.com) or smartwatch compatible with Google Fit (http://fit.google.com) while participating in a 30-day longitudinal study using the "HowNutsAreTheDutch" project [12,5]. HowNutsAreTheDutch is a project that provides a cross-sectional study as well as a large scale EMA study [5]. Moreover, we provide an online demo of our implementation of Physiqual and released its source code as open-source software. Our implementation of Physiqual serves as a proof of concept and demonstrates its capabilities.

The paper is organized as follows: Section 2 gives an overview of the current state of the art with regard to the present work. In Section 3, the concept of Physiqual is elaborated. We describe the types of physiological data that are supported by Physiqual and how their different sampling rates are unified. We provide a concise overview of the implementation of Physiqual and outline its architecture. In Section 4, we describe the case study we performed using Physiqual in combination with an EMA study. We explain the steps taken to gather the data and shed light on the statistical analysis performed. Section 5 describes the validation of Physiqual, both in terms of effectiveness and accuracy. Section 6 shows the results of the case study and includes links to the source code and to a live demo of our implementation of Physiqual on an online platform. In Section 7, we discuss our findings and describe the limitations of our current approach. Section 8 concludes the work and contains recommendations for future research.

2. Background

Advances in mobile technology have fostered the rise of EMA studies. Mobile technology allows for EMA studies to be conducted on a large scale, and participants can be measured more easily and more reliably than when using traditional methods (i.e., pencil and paper) [16]. The use of (mobile) technology allows for multimodal continuous data collection and automatic data entry at a high frequency [17,18].

Self-tracking and collecting longitudinal sensor data form the pillars of a movement called the Quantified Self (QS). In QS, an individual uses sensor-equipped wearable devices to quantify and gain insight into their day-to-day life, either out of personal interest or with the purpose of improving their quality of life [19]. The increasing popularity of QS has prompted the development of several tools and platforms for managing self-tracking data. The value of self-tracking data in health research has been demonstrated in previous studies. For example, the Experience Sampling for Total Hip Replacement (ESTHER) platform is developed to study experiences after surgery and to evaluate interventions which are developed to support patients during home recovery [14]. Nevertheless, this platform was specifically designed in the context of patients who underwent a total hip replacement.

The increased availability of sensors to assess physiological measures yields a substantial amount of data in the medical and social sciences [20,21]. The need for combining EMA data and sensor data is demonstrated by the development of several platforms specifically designed for this purpose. Gaggioli et al. [15] built the open-source platform *PsychLog* to collect data which can be used in psychophysiological research [15]. Unlike Physiqual, this platform does not support data collected from commercially available sensors and focuses on a specific set of sensors. That is, they only focus on electrocardiogram (ECG) and accelerometer data, whereas Physiqual is not tied to specific hardware and thus is compatible with any sensor that can interface with a supported service provider (e.g., Fitbit or Google Fit). Other researchers focus on the interpretation of psychological states or on deriving psychological states using sensors. For instance, Wagner et al. [22] show the possibilities to recognize emotions (such as anger and joy) in real time in multimodal online emotion recognition (OER) systems by fusing data from various sensors (e.g., data from audio and video). Technology can also be used for pattern identification and data analyses in automating EMA and ESM sensing. Shi et al. [23] showed that by using machine learning, information detected by sensors can be automatically classified to certain psychological states, such as stress.

An application similar to Physiqual is *mEMA* by Ilumivu [13]. mEMA is a complete EMA solution that uses a mobile application to perform measurements. Furthermore, Ilumivu provides options to enrich an EMA data set with physiological sensor data, as measured from the mobile phone sensors or wearable sensors. Although this functionality overlaps with some of the functions of Physiqual, there are several important differences. Firstly Physiqual focuses on sensors from external services and therefore supports a plethora of wearable sensor devices. Secondly, Physiqual can be used separately from an existing EMA solution and can be enabled after a study has been completed. Lastly, mEMA is a commercial proprietary solution, whereas Physiqual is freely available open-source software. A comparison between Physiqual and the three other platforms is presented in Table 1. The projects by Wagner et al. [22] and Shi et al. [23] are not included in this table as their main focus lies on data analysis instead of the EMA/sensor platform. This comparison addresses five properties: (i) the target group the platform focuses on, (ii) the sensor compatibility of the platform, (iii) the availability of the source code, (iv) the method of sensor data collection, and (v) the EMA system to be used with the platform.

Despite the increasing number of platforms and technologies that contribute to the collection of EMA and sensor data, to the best of our knowledge, an automated way to combine data from different sources in a functional data format is still missing. The goal of Physiqual is to fill this gap.

3. Physiqual

Physiqual is a novel means to collect, aggregate, and unify sensor data for use in EMA studies. With Physiqual we aim to offer a single point of access to gather sensor data from various service providers and to expose this data in such a way that it can be combined with EMA data. In order to offer this single point of access,

Table 1
Comparison between Physiqual and existing EMA and sensor platforms.

	Physiqual	mEMA [13]	ESTHER [14]	PsychLog [15]
Target group	General	General	Hip replacement patients	General
Compatibility	Wearables & smartphones	Wearables (beta) & smartphones	LiveView/ProMove-3D sensor	Specialized ECG and accelerometer data
Source availability	Open-source	Closed-source	Unknown/closed-source	Open-source
Sensor measurements	Continuous	Continuous	Continuous	Intermittent
Used EMA system	Variable	Specific	Specific	Specific

Physiqual gathers and processes data from the underlying service providers. One of its key features is the abstraction of any service provider-specific routines (e.g., connecting to the service provider or collecting the data from it), allowing for an approach that is unaware of the service provider being used. Hence, data exported by Physiqual always adheres to the same format. Fig. 1 gives an overview of the actors involved in the use of Physiqual and shows the main flow of information.

The steps in this flow (Fig. 1) are as follows. Physiqual ties into the EMA study platform managed by the researchers. Prior to the study, it requires the researcher to configure certain settings that are specific to the design of the EMA study (as shown in step 1) and identical for all participants (i.e., the duration of the study, the frequency of its measurements, and the type of imputation to be used). The researcher also needs to configure the credentials to access the service providers (step 2). For the entire duration of the EMA study, participants passively measure themselves using wearable devices supported by Physiqual (steps 3 and 4). In our envisioned scenario, Physiqual integrates seamlessly with the (web) application that hosts the EMA part of the study. Through this familiar front-end, participants are asked to provide the necessary authentication credentials for Physiqual to obtain their physiological measurements for use in the EMA study (step 5). The decision whether user permission should be requested prior, during, or after the study period lies with the researcher. The authorization credentials in Physiqual are stored persistently, allowing for data exports subsequent to study completion (unless access is explicitly revoked by the participant). Upon completion of the study for each participant that has granted permission, the researcher can call a routine in Physiqual to export all sensor data from a specified time interval (step 6). As Physiqual stores only the authorization information, the responsibility of scheduling exports and storing the retrieved data lies with the hosting platform managed by the researcher. Physiqual gathers the online sensor data from the service providers and the researcher merges the data with EMA data (step 7) to perform his or her research analysis.

3.1. Architecture

The architecture of Physiqual adheres to a layered approach as illustrated in Fig. 2. Each of the layers serves a specific purpose. The first layer, the service layer, gathers sensor data from the external service providers. The second layer, the aggregation and processing layer, performs several processing steps on the data. In this layer the data is summarized, aggregated, and unified to a format compatible with the EMA protocol. After this step, data flows to the third layer, the imputation layer, in which any missing values can be imputed using one of the supported data imputation algorithms (as outlined in Section 3.4). The final data set is then offered to the researcher through the top layer, the presentation layer, in various formats (i.e., JavaScript Object Notation (JSON), Comma Separated Values (CSV), or using a web page). Self-evidently, the "raw" data of the service providers is still available (also via Physiqual). Although Physiqual allows the researcher to use the sensor data, whilst unaware of the platform it originated from, the researcher can retrieve a list of participant codes in combination with the name of the connected service provider. The steps performed in each of these layers are described in more detail in the next sections.

3.2. Service layer and service providers

Physiqual applies a Service-Oriented Architecture (SOA) to retrieve the sensor data from the service providers [24], enabled using the Open Authorization protocol v2 (OAuth2). The OAuth2 protocol allows users to give certain applications permission to access their data. With OAuth2, the credentials of the user remain at the service provider and are never transferred to a third-party service. Moreover, the participant can revoke the permission at any time, without needing to change credentials.

Physiqual is designed to be compatible with certain service providers rather than with specific sensor hardware. This is because the service providers themselves already support many different sensor types. Sensors, including the ones used in his study, have some limitations, as the level of accuracy of these sensors might vary [25]. The development and validation of sensors for measuring physiological data is outside of the scope of this paper. Physiqual is currently compatible with two service providers for accessing sensor data, viz. Google Fit and Fitbit.

Google Fit is a platform to capture, manage, and aggregate data from a variety of (third-party) devices. Data for Google Fit can be collected using a Google Fit enabled device. Android, a mobile operating system by Google designed for smartphones and tablets, and Android Wear, an operating system specially designed for smartwatches and other wearables, have applications that are compatible with Google Fit. For example, when using the Google Fit application one can collect steps using a smartphone and heart rate using a smartwatch. Furthermore, data can be collected by a third-party application and/or device. Retrieving the data from Google Fit is possible using specific libraries or by using the application programming interface (API) directly.

Fitbit is a company specialized in developing consumer software and hardware for measuring activity and health-related data. They currently offer eight different wearable sensors, with functionality from basic step counting to heart rate monitoring and location tracking. The data can be stored on the device, from which it is synced to the Fitbit platform. Furthermore, both companies offer an elaborate API to gather daily data from a user. Gathering intraday data from Fitbit, however, requires access to the so-called partner API, to which access is granted on a per-project basis.

Loose coupling with these service providers by means of an API allows Physiqual to bind at runtime. That is, the internals of the service providers can be changed without affecting Physiqual.

3.3. Aggregation and processing layer

Data sources offered by various service providers can be in a different format or granularity. For example, one service provider may list steps per second, while another lists steps per minute. Additionally, it is unlikely that the sampling rate exactly corresponds to the sampling rate of the EMA data. Physiqual therefore resamples the data in a way that renders it useful and intuitive to the researcher.

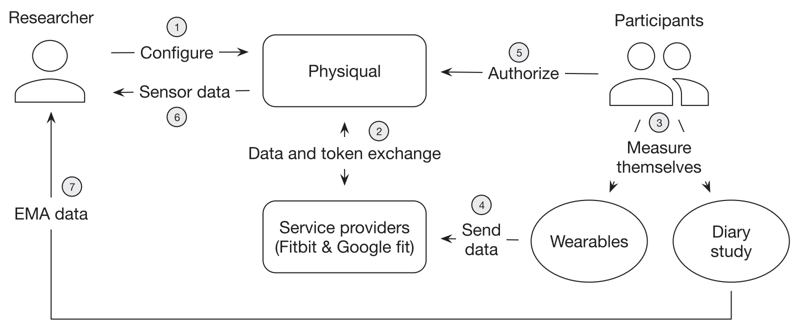

Fig. 1. Overview of actors and flow of information in Physiqual.

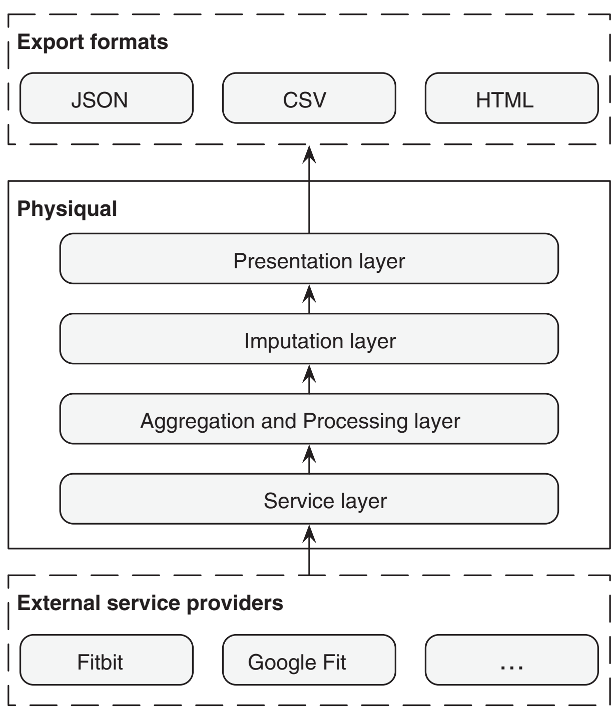

Fig. 2. Overview of the layers in the Physiqual architecture.

EMA studies administer questionnaires using a certain schedule or protocol. For Physiqual, we currently support studies which use equidistant measurement protocols. In such protocols, the measurements are conducted at equidistant time intervals (e.g., every six hours) for a certain number of measurements per day. For example, in the HowNutsAreTheDutch study, each participant was measured three times per day, with the first measurement approximately 12.5 h before a user-specified bedtime, the second measurement approximately 6.5 h before bedtime, and the last measurement approximately half an hour before bedtime [5,26]. To adhere to the measurement schedule of the EMA study, the sensor data requires a resampling step. Physiqual combines all sensor data from the time of the measurement moment, including the first measurement time, up-to the next measurement time. For example, in the aforementioned schedule (a measurement every six hours) when having the first measurement at 10:00:00 AM, the last sensor reading included will be the one at 3:59:59 PM. Depending on the type of variable, this resampling step takes one of three forms.

3.3.1. Steps, distance, and calories

A meaningful way for researchers to summarize steps, distance, or calorie expenditure over a certain time-span is by calculating their respective sums. This approach is incorporated in Physiqual. In order to down-sample the measurements, Physiqual sums the values (per category) to derive a value that best represents the interval between subsequent measurements. For the first measurement of the day it might not be desirable to include all preceding measurements, as some analysis methods omit the period of night. Therefore, the previous interval for the first measurement can be configured to a fixed number of hours. Thus, the decision of whether or not to include the night lies with the researchers.

3.3.2. Sleep

Sleep is measured slightly different from steps, distance, or calories. Several EMA studies adjust their schedule in such a way that no questionnaires are administered during the night in order to reduce the impact of the study on its participants. However, if Physiqual were to comply exactly to the EMA study schedule for the sleep metric, chances are that large parts of sleep during the night are not measured by Physiqual. Therefore the sleep metric is provided for each measurement as the time spent sleeping since the previous measurement, in minutes.

3.3.3. Heart rate

For heart rate, summation of the data does not always provide EMA studies with a measure that is intuitive or useful. Simply taking the average does not suffice either as questions in EMA studies are often formulated to ask for current feelings or for feelings that best describe the time since the previous measurement [5]. We assume researchers are more interested in knowing the heart rate that was measured most frequently during a time interval, instead of a mean or cumulative score.

Fig. 3 gives a hypothetical example of why a normal histogram or mode may not suffice. The gray bars show the histogram values (with the corresponding mean, median, and mode). In this example, we have detected 28 occurrences of heart rate 110, while we detected 24, 20, and 20 occurrences of respectively heart rate 71, 72, and 73. Using the mode selecting the most occurring heart rate estimates the heart rate of 110 as occurred the most frequent one. Although this is true, this is probably not what the researcher is interested in.

To solve this issue, Physiqual implements a top hat kernel density estimation (KDE) method to determine the heart rate that best represents the time interval [27, pp. 1–31]. Fig. 3 shows how the top hat kernel density estimation method would select a bin. This method effectively collects the heart rate measurements in a histogram where each measurement not only increases its own bin, but also the k surrounding bins. For example, if $k = 2$, and we detect a heart rate of 80, we do not only increase the frequency of the 80-heart rate bin, but also of the 78, 79, 81, and 82 bins. After performing the top hat kernel density estimation, we select the mode from the new data set. The top hat kernel density estimation method reduces the effect of inaccuracies in and small fluctuations

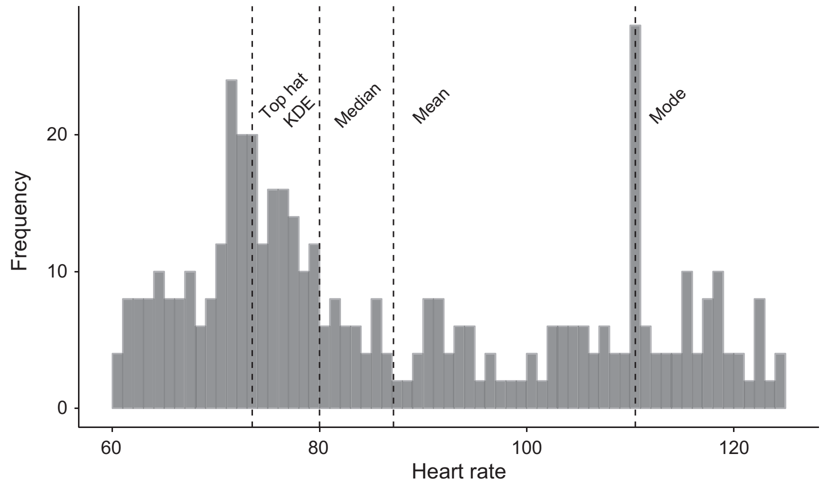

Fig. 3. The gray bars correspond to the bins of a regular histogram. The dashed lines point out respectively the bin selected by top hat KDE, median, mean, and mode. Using the mode of the data would yield a heart rate of 110, while using the mode after top hat kernel density estimation is 73.

of heart rate. When there are multiple bins with the maximum number of occurrences, we choose the bin that lies closest to their mean. In case of a tie, we return the average of the tied values.

3.3.4. Unifying data

Different service providers may use different formats for their exported data sources. For example, Google Fit lists the timestamp for steps in nanoseconds, while Fitbit uses a more conventional date time notation, and one service provider might use the metric system to export its data, whereas another service provider exports data in the imperial system.

To make sure that the format of the exported data is not affected by a specific service provider, Physiqual unifies the output format of the variables across different service providers. This unification maintains the abstraction of service providers as interchangeable parts and allows the hosting application to remain unaware of which service provider is used. Researchers can use this single datafile without being bothered by the details of each service provider that the participants use, or all required transformations, and use the data as-is.

3.4. Imputation layer

Physiqual can resolve missing values through imputation. To prevent information loss, Physiqual imputes the data at one of the top layers in the architecture, thus after the data has been aggregated. Consequently, Physiqual only imputes aggregated values so that imputation is only needed when all values considered for the aggregate are missing. This is a rare occurrence because in a typical EMA measurement interval sensor data is measured many times.

The default imputation method is Catmull-Rom interpolation, a cubic spline interpolation technique [28, pp. 317–326]. The researcher can also select a different method. The selected imputation method will be used to impute each of the aggregated variables. Physiqual currently supports the following imputation methods:

- **Mean imputation:** missing values are imputed with the mean of the observed values.
- **Last observation carried forward:** missing values are imputed with the last observed value.
- **K-Nearest Neighbors:** missing values are imputed with the mean of the values of the K-surrounding neighbors (i.e., the K-Nearest Neighbors algorithm).
- **Spline inter/extrapolation:** missing values are imputed with re-sampled data points that have been derived with a spline function fitted on the available data.
- **Catmull-Rom**: missing values are imputed with a spline interpolation technique that uses cubic interpolation splines.
- **No imputation**: it remains possible to refrain from imputation.

3.5. Presentation layer

Data from Physiqual is, depending on the needs of the researcher, presented in one of three formats: (i) JSON, (ii) CSV, and (iii) HTML. These export formats each comprise the same set of variables. In Table 2 we provide an overview of the data sources per platform. For a more elaborate overview of the sensor data provided by the service providers, we refer to the API documentation of these service providers.[1]

4. Case study

We designed and executed a case study with two subjects that participated in an EMA study while using a wearable device with sensor readings over a period of 30 days. This case study illustrates how integrating physiological data into an EMA study can provide new insights into the relations and interactions between physiological and mental processes, further demonstrating the utility of Physiqual in a practical setting. In contrast to cross-sectional studies, which provide average values, the main aim of EMA is to identify relationships within individuals and to find associations at the individual level [29]. Multiple repeated measurements can be linked to physiological data collected with wearables, revealing meaningful information for that specific individual. We do not aim to generalize the results, because what holds for one individual, is not necessarily true for another. Separate analysis are conducted for each individual to elucidate individual patterns.

4.1. EMA and sensors

An overview of the case study design is provided in Fig. 4. The EMA data in this case study was collected using a Dutch national mental health measurement project, known as HowNutsAreThe-

[1] Fitbit: https://dev.fitbit.com and Google Fit: https://developers.google.com/fit.

Table 2
Supported variables in Physiqual.

	Fitbit	Google Fit (with smartwatch)
Steps	Supported	Supported
Heart rate (bpm)	Supported	Supported
Sleep (minutes slept)	Supported	Supported (using 3rd party app)
Distance (km)	Supported	Supported
Calories (expended)	Supported	Supported

Dutch [5,12]. HowNutsAreTheDutch offers an EMA study with a predefined protocol, viz. three questionnaires per day, for thirty consecutive days. Each questionnaire has a total of 43 items, of which 42 items are predefined, and one question can be selected from a list of possible items (or be defined by the participant). The participant is prompted to fill out a questionnaire at fixed times: every six hours, with the last questionnaire approximately half an hour before the bedtime of the participant. This bedtime has to be specified by the participant before the start of the study.

The case study included two subjects; a 26-year old male and a 32-year old female. The former collected data using the Google Fit service, wearing a Motorola Moto 360 (1st generation) in combination with a Motorola Moto G (2013) for collecting heart rate and steps, and an application called Cinch.[2] Cinch is a fitness application which was used to automatically measure heart rate every five minutes. Participant two collected data using the Fitbit Charge HR in combination with a Samsung Galaxy S3 Mini. Both participants gave consent for using their data for this case study.

4.2. Statistical analyses

Statistical analyses were performed on the combined data sets. The data sets contained the psychological variables as described by Van der Krieke et al. [26], combined with some of the physiological variables exposed by Physiqual (viz., steps, calories, heart rate, and distance). For the top hat kernel density estimation method we used a k of 2, and we configured Physiqual to include the measurements of six hours prior to the first measurement of the day.

To investigate the relations between the variables in the combined data set, we fitted vector autoregression (VAR) models [30]. VAR is a statistical method that can be used to fit a regression model on a time-series data set while accounting for the contemporaneous relations between variables (relations between variables at the same moment in time) and the time-lagged relations between variables (relations in which a variable is related to itself or a different variable at a previous moment in time). Here, the contemporaneous relations were defined as the residual Pearson correlations, and the time-lagged relations were defined as the significant Granger causalities at the $p \leqslant 0.05$ level [31]. Granger causality is a notion of causality describing that the variance of one variable (x) can better be explained using time lagged values of both x and of another variable (y) instead of merely using lagged values of x. In such cases, y is said to Granger cause x. For a detailed description, we refer to [30,31]. Fitting the VAR model was performed using Autovar, a program that automates the process of fitting VAR models for time series data [32]. For this analysis, we selected for each participant five variables from their data set that were reasonably normally distributed and had high variance. Furthermore, we included at least one physiological variable (as collected using Physiqual) in the model.

[2] Website: https://bit.ly/cinch-app.

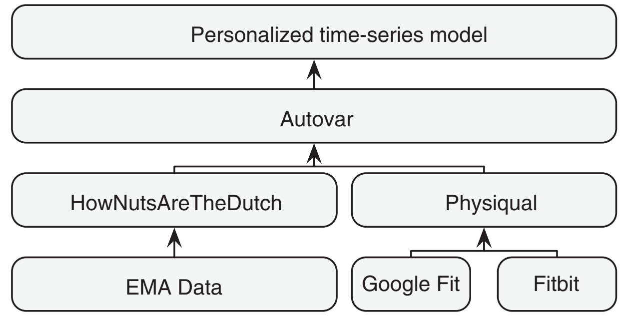

Fig. 4. Overview of the experimental setup for the Physiqual case study. Autovar refers to automated vector autoregression (VAR) analysis, see Section 4.2.

5. Validation

We performed a first validation of Physiqual in terms of effectiveness and accuracy. Firstly, we determined the effectiveness of Physiqual by comparing it with the manual analysis of a domain expert, in terms of results, time spent, and ease of use. Secondly, we validate Physiqual in terms of accuracy. In this validation, we illustrate how our proposed techniques for summarizing measurements to a single data point are in line with the design of EMA research, and how the results are equivalent to those used in EMA practice.

5.1. Effectiveness

To validate the effectiveness of Physiqual, our automated procedure was compared to a previously used manual procedure to collect and process data from sensors applied in research. The research used for this comparison has been published in a Dutch magazine [33]. Information about the manual procedure was collected by interviewing researchers who applied this procedure.

The procedure was described as follows. Sensors were read out and a raw data file was created. For the manual study, it was necessary to complete missing data about length and weight, which was completed manually. The raw file was converted to a Microsoft Excel-file. If more than one wearable was used over time, files were merged manually. The Excel-file was opened, and data labels about the start of the study and questionnaire intervals of the EMA-study for the duration of the study (e.g. thirty days) were inserted manually in the data file. Next, data was copied into another pre-programmed Excel-template, and descriptive statistics were computed using Excel. Due to a small error in the template, equations had to be adjusted manually. After this procedure, data was ready for statistical analyses.

Everything considered, it took an experienced researcher around 20–30 min to process the data of a single participant. Besides the time effort, this process is prone to mistakes due to the number of manual steps involved. After the initial one-time setup (that is, updating the EMA platform to use the Physiqual plugin and to manage the communication between Physiqual and the EMA application), Physiqual can be used to perform the process automatically. Generating the aforementioned data file using the Physiqual procedure would take several seconds (depending on the service providers used), which is negligible compared to the 20–30 min in manual analysis. We tested the response time of Physiqual for both service providers by exporting twenty 30-day data sets for all supported variables. The average response time for the Google Fit platform was 3.71 s ($sd = 0.34$, $range = 3.19–4.41$, $n = 20$). For Fitbit the response time was considerably higher, with an average response time of 57.73 s ($sd = 1.76$, $range = 55.83–62.39$, $n = 20$). This difference is caused by the number of API calls Physiqual makes. Google Fit allows Physiqual to retrieve a longitudinal data set per variable using a single request (i.e., with 5 variables this makes 5 requests in total). For Fitbit however, Physiqual needs to perform a request per day, for each variable for which to retrieve data (i.e., $5 \times 30 = 150$ requests in total). Nevertheless, compared to the manual analysis, Physiqual saves more than 95% of the time (over 19 min per participant). Importantly, as sensor data can be retrieved online, no physical contact between researcher and sensor is required, that is the sensors do not need to be physically available to the researcher. This enables for a large scale implementation of sensors in an EMA study, which would have been impossible with expensive, single-purpose sensor devices.

Self-evidently, saving time by replacing a manual procedure with an automated procedure like Physiqual is only interesting when the time savings outweigh the set up time of Physiqual. To estimate the length of the initial setup time of Physiqual in an existing EMA platform, we made an existing (large scale) EMA platform (HowNutsAreTheDutch [5]) compatible with Physiqual. Although most of the authors of the present work are involved in creating HowNutsAreTheDutch, the development of this EMA platform was completed prior to the inception of Physiqual, and as such, can be considered to be an arbitrary EMA platform choice. By following the description as provided on the open-source software repository of Physiqual, it took a single experienced software engineer less than one hour to enable Physiqual support in this platform. Comparing this estimate to the previously mentioned lower bound of 20 min for the manual procedure indicates that the implementation of Physiqual could already be beneficial in a study with more than five participants.

5.2. Accuracy

We validated the accuracy using the data from the manual analysis described in Section 5.1. We checked whether the aggregated data from Physiqual was the same as the output from the manual analysis. In order to do so, we compared the results of the manual analysis described in Section 5.1 with the analysis performed by Physiqual. For this comparison, Physiqual's data retrieval procedure was slightly adapted as the data in that study was collected using an Actical device,[3] instead of a supported Fitbit or Google Fit device. The layered architecture of Physiqual allowed these changes to remain isolated to the service layer, leaving the rest of the program/code unaffected. The results of the manual analysis were equivalent to the output as retrieved from Physiqual. Note that for this analysis, imputation was done beforehand, so both Physiqual and the manual analysis received the same imputed data set.

6. Results

The results comprise a fully working, open-source implementation of the proposed platform and a case study illustrating the interaction between EMA data and sensor data. While the implementation establishes the feasibility of our architecture, the case study serves to demonstrate the practical utility of how adding sensor data can provide new insights.

6.1. Software implementation

Our implementation of Physiqual is available as open-source software and can be downloaded from https://github.com/roqua/physiqual. We implemented Physiqual in the Ruby on Rails framework[4] as a plugin (or Engine, in Ruby on Rails parlance) so that it can be easily integrated in third-party projects. Physiqual persists the data regarding the authentication of the participants to the external service providers in a database (i.e., the tokens that allow access to a participant's account). Our current implementation of Physiqual exposes the data in three formats, a JavaScript Object Notation (JSON) format, a Comma Separated Values (CSV) format, or as a web page on which a dashboard is shown presenting a general overview of the data.

A live demo of our Physiqual implementation can be found at http://www.physiqual.com. This simple web application facilitates account creation and supports data exports in predefined formats. It also shows a dashboard overviewing the measured activities, steps, heart rate, distance, and calories. For those without Fitbit or Google Fit data, example data can be shown instead.

6.2. Case study

Network representations of the case study analysis results are shown in Figs. 5 and 6. These network images illustrate the relations between variables as determined using VAR analysis. In these network images, the nodes depict the measured EMA variables (in this case psychological variables) and physiological variables (from Physiqual). Green nodes depict positive variables, red nodes depict negative variables, and blue nodes depict neutral variables. The edges (arrows) between the nodes depict the Granger causal relations between two nodes. That is, a directed edge from node A to node B shows that changes in node A precede changes in node B at the $p \leqslant 0.05$ level, or to put it differently, *A Granger causes B*. If the edge is undirected, the effect is contemporaneous, meaning that the variables affect each other at the same moment in time. See the work of Van der Krieke et al. for more information on these network images [5].

The results of the analyses for participant one (Fig. 5) showed a positive time-lagged association from the number of steps to humor and vice versa (Fig. 5a). Moreover, there was a negative time-lagged association from the number of steps to feeling down. These relations can be interpreted as follows: if this person reported more laughter at time $t = 0$, the person tends to have an increase in the number of steps at the next measurement moment ($t = 1$). Furthermore, when this person does more steps at time $t = 0$, he is expected to report more laughter – and to feel less down – at time $t = 1$.

In the contemporaneous model (Fig. 5b), steps were negatively associated with a personal question (i.e., a question determined by the participant). This relationship denotes that whenever this participant took more steps, he would have a decrease in this personal question at the same time. Due to technical issues, the Motorola Moto 360 smartwatch worn by participant one did not collect data for two weeks. Nevertheless, a valid model involving steps was found because steps were still collected by the Google Fit application on the smart phone.

For participant two (Fig. 6), the time-lagged model showed a positive influence of cheerfulness on the calories expended (Fig. 6a). Moreover, the calorie expenditure has a negative association with the feeling of falling short of something, which in turn had a negative association with cheerfulness. That is, when this person felt more cheerful, she would have an increase in the amount of calories expended, which in turn caused a decrease in concentration and a decrease in the feeling of falling short of something. In the contemporaneous model, no significant association

[3] Website: http://actigraphy.com/devices/actical.

[4] Website: http://rubyonrails.com.

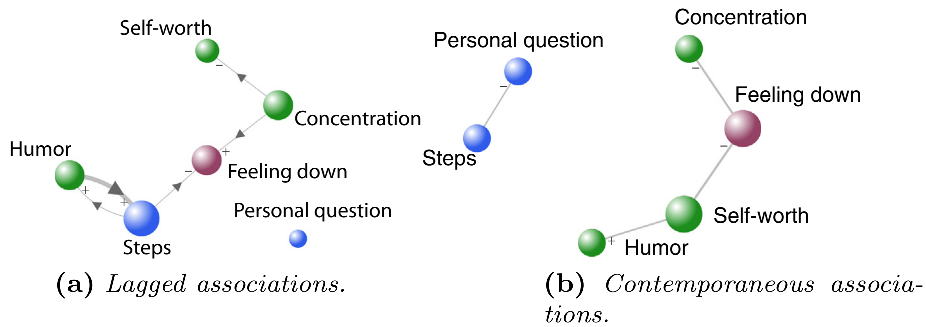

Fig. 5. Lagged and contemporaneous associations determined from the case study of EMA data and sensor data for participant one.

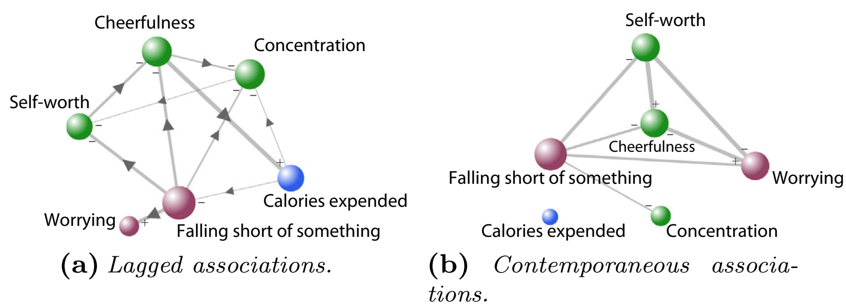

Fig. 6. Lagged and contemporaneous associations determined from the case study of EMA data and sensor data for participant two.

was found between calorie expenditure and any of the other variables in the model (Fig. 6b).

7. Discussion

The present study introduced Physiqual, a novel approach for processing sensor data for use in EMA studies. Physiqual enables the use of sensor data from commercially available wearable devices in EMA mental health research by interfacing with service providers to export data in applicable formats. The case study showed how Physiqual can be useful in adding physiological data to EMA data, potentially enabling new insights in psychophysiological research at the individual level. Currently Physiqual supports two service providers, but the platform can be easily extended to interact with other service providers in the future.

Physiqual is a way to manage data and fill a niche with the rising interest in QS fueled by the increasing popularity of wearable devices. Physiqual embodies the recognition of the value of personalized medicine and the search for cheaper alternatives for collecting patient data to bring the rising costs of health care to a standstill [34].

As with every new development, Physiqual has its limitations. While sensors sample data at a high frequency, EMA data is collected over longer intervals. To compensate for this discrepancy, heart rate data is downsampled, adhering to the low frequency of the EMA data. By downsampling, the most frequently occurring heart rate is presented, which we consider to be most in line with an EMA study. However, due to this downsampling we lose information about short but possibly intense shifts in heart rate. These intense changes could conceal short physiological (stressful or pleasant) events which might have a considerable influence on mental phenomena [35].

In addition, the downsampled data, extracted from a sensor, is a summary of the measurements within a predefined period of time. This summary can be based on a varying number of measurements. Hence, the reliability of the exported measurements can differ. Currently, the format in which the data is exported does not accommodate a representation for the notion of reliability. However, note that EMA self-reported sleep duration or physical activity have been shown to be notoriously unreliable [36]. From this perspective, sensor data shall often be more 'reliable' and objective than corresponding EMA questions (albeit both approaches may also capture different information).

Practical limitations of Physiqual include the type of access allowed and the data exported by the service providers. For example, Fitbit permits intraday access to measurements only on a per-project basis, and the number of requests allowed has an hourly limit. Self-evidently, Physiqual can only process data of a wearable sensor when this data is accessible. Some wearable platforms currently have limited options for data extraction by third party applications such as Physiqual. One of the most popular smartwatches at the time of writing is the Apple Watch [37]. Although the Apple Watch provides several sensors useful for EMA research, Physiqual currently is not able to support it. At present, the Apple Watch does not provide an API accessible via the Internet, nor does the Apple HealthKit platform. These platforms currently only provide a mobile iOS – the mobile operating system by Apple Inc. – API to retrieve data from the watch. No method for exposing this data directly to Physiqual is therefore available. To support the Apple Watch in an external platform like Physiqual, a third party mobile application must be developed which is capable of uploading the Apple Watch its data to either one of the existing supported service providers or to a new platform.

8. Conclusions and future work

Physiqual is the first approach for combining data from commercially available wearable sensors and EMA studies. An important contribution is that we provide a generic, open-source platform to serve as a means to aggregate and unify these data. By automating the time-consuming task of data retrieval and aggregation, Physiqual potentially enables the usage of sensor data with EMA data on a large scale. With Physiqual, existing wearable devices can be used in EMA, instead of acquiring a specialized device for each participant. Our study provides a base for future developments, and invites researchers and corporations to cooperate on solutions for challenging social and mental health questions. Apart from the Fitbit and Google Fit platforms, Physiqual could support other platforms. For instance, Jawbone,[5] NikeFuel,[6] and Misfit[7] all provide a developer API that could be consumed by Physiqual.

Data exported by Physiqual may be used to complement or replace certain EMA data. Comparing physiological data with EMA data could provide new insight regarding the correlation between perceived and measured physical activity. The most appropriate EMA questions for this purpose would be those regarding activity or sleep. For those sensors that are validated in scientific studies, and where physiological data is significantly correlated with existing questions in EMA studies, replacing those questions with data exported by Physiqual could form a first step in alleviating some of the burden of EMA studies through the use of passive monitoring from sensors.

Acknowledgments

Physiqual is developed by researchers involved in the development of both the HowNutsAreTheDutch project and the Leefplezier project. The HowNutsAreTheDutch project is partly funded by a VICI grant (No. 91812607) awarded to Prof. Peter de Jonge by the Netherlands Organization for Scientific Research (NWO-ZonMW). We thank all people working on the HowNutsAreTheDutch project. The Leefplezier project is funded by Espria. Espria is a healthcare group in the Netherlands consisting of multiple companies targeted mainly at the elderly population.

References

[1] S. Shiffman, A.A. Stone, Ecological momentary assessment: a new tool for behavioral medicine research, Technol. Methods Behav. Med. (July) (1998) 117–131.
[2] M. Aan het Rot, K. Hogenelst, R.A. Schoevers, Mood disorders in everyday life: a systematic review of experience sampling and ecological momentary assessment studies, Clin. Psychol. Rev. 32 (6) (2012) 510–523.
[3] P.C.M. Molenaar, A manifesto on psychology as idiographic science: bringing the person back into scientific psychology, this time forever, Meas.: Interdiscip. Res. Perspect. 2 (4) (2004) 201–218.
[4] E.L. Hamaker, Why researchers should think 'within-person': a paradigmatic rationale, in: Handbook of Research Methods for Studying Daily Life, Guilford Publications, New York, NY, 2012, pp. 43–61 (Chapter 3).
[5] L. Van der Krieke, B.F. Jeronimus, F.J. Blaauw, R.B. Wanders, A.C. Emerencia, H. M. Schenk, S. De Vos, E. Snippe, M. Wichers, J.T. Wigman, E.H. Bos, K.J. Wardenaar, P. De Jonge, HowNutsAreTheDutch (HoeGekIsNL): a crowdsourcing study of mental symptoms and strengths, Int. J. Methods Psychiatr. Res. 25 (2) (2016) 123–144.
[6] M. Almalki, K. Gray, F. Sanchez, The use of self-quantification systems for personal health information: big data management activities and prospects, Health Inform. Sci. Syst. 3 (Suppl. 1) (2015).
[7] J. Andreu-Perez, D.R. Leff, H.M.D. Ip, G.-Z. Yang, From wearable sensors to smart implants – toward pervasive and personalized healthcare, IEEE Trans. Biomed. Eng. 62 (12) (2015) 2750–2762.
[8] International Data Corporation (IDC), Worldwide Quarterly Wearable Device Tracker, Jun 18th 2015. <http://www.idc.com/getdoc.jsp?containerId=prUS41100116> (accessed October 15, 2015).
[9] T. Ferguson, A.V. Rowlands, T. Olds, C. Maher, The validity of consumer-level, activity monitors in healthy adults worn in free-living conditions: a cross-sectional study, Int. J. Behav. Nutr. Phys. Act. 12 (1) (2015) 1–9.
[10] J.-H. Park, D.-G. Jang, J. Park, S.-K. Youm, Wearable sensing of in-ear pressure for heart rate monitoring with a piezoelectric sensor, Sensors 15 (9) (2015) 23 402–23 417.
[11] S.H. Booij, E.H. Bos, M.E.J. Bouwmans, M. Van Faassen, I.P. Kema, A.J. Oldehinkel, P. De Jonge, Cortisol and α-amylase secretion patterns between and within depressed and non-depressed individuals, Plos One 10 (7) (2015) 1–15.
[12] F. Blaauw, L. Van der Krieke, E. Bos, A. Emerencia, B.F. Jeronimus, M. Schenk, S. De Vos, R. Wanders, K. Wardenaar, J.T.W. Wigman, M. Aiello, P. De Jonge, HowNutsAreTheDutch: personalized feedback on a national scale, in: AAAI Fall Symposium on Expanding the Boundaries of Health Informatics Using AI (HIAI'14): Making Personalized and Participatory Medicine A Reality, November 2014, 2014, pp. 6–10.
[13] Ilumivu, Mobile EMA, 2016. <https://ilumivu.com/solutions/ecological-momentary-assessment-app> (accessed February 21, 2016).
[14] J. Jimenez Garcia, N.A. Romero, S.T. Boerema, D. Keyson, P. Havinga, ESTHER, in: Proceedings of the 3rd ACM MobiHoc Workshop on Pervasive Wireless Healthcare – MobileHealth '13, July, 2013, ACM Press, New York, New York, USA, 2013, p. 7.
[15] A. Gaggioli, P. Cipresso, S. Serino, G. Pioggia, G. Tartarisco, G. Baldus, D. Corda, G. Riva, An open source mobile platform for psychophysiological self tracking, Stud. Health Technol. Inform. 173 (January) (2012) 136–138.
[16] T.J. Trull, U.W. Ebner-Priemer, Using experience sampling methods/ecological momentary assessment (ESM/EMA) in clinical assessment and clinical research: introduction to the special section, Psychol. Assess. 21 (4) (2009) 457–462.
[17] S.S. Intille, Technological innovations enabling automatic, context-sensitive ecological momentary assessment, in: S. Shiffman, A.A. Stone, A.A. Atienza, L. Nebeling (Eds.), The Science of Real-time Data Capture, Oxford University Press, New York, NY, 2007, pp. 308–337.
[18] S. Kumar, W.J. Nilsen, A. Abernethy, A. Atienza, K. Patrick, M. Pavel, W.T. Riley, A. Shar, B. Spring, D. Spruijt-Metz, D. Hedeker, V. Honavar, R. Kravitz, R. Craig Lefebvre, D.C. Mohr, S.A. Murphy, C. Quinn, V. Shusterman, D. Swendeman, Mobile health technology evaluation: the mHealth evidence workshop, Am. J. Preventive Med. 45 (2) (2013) 228–236.
[19] M. Swan, Health 2050: the realization of personalized medicine through crowdsourcing, the quantified self, and the participatory biocitizen, J. Personalized Med. 2 (3) (2012) 93–118.
[20] R.M. Chang, R.J. Kauffman, Y. Kwon, Understanding the paradigm shift to computational social science in the presence of big data, Decis. Support Syst. 63 (1) (2014) 67–80.
[21] A. Markowetz, K. Błaszkiewicz, C. Montag, C. Switala, T.E. Schlaepfer, Psycho-informatics: big data shaping modern psychometrics, Med. Hypotheses 82 (4) (2014) 405–411.
[22] J. Wagner, E. Andre, F. Jung, Smart sensor integration: a framework for multimodal emotion recognition in real-time, in: 2009 3rd International Conference on Affective Computing and Intelligent Interaction and Workshops, September 2009, IEEE, Amsterdam, 2009, pp. 1–8.
[23] Y. Shi, M. Nguyen, P. Blitz, B. French, Personalized stress detection from physiological measurements, in: International Symposium on Quality of Life Technology, Las Vegas, NV, September 2015, 2015.
[24] K.B. Laskey, K. Laskey, Service oriented architecture, Wiley Interdiscip. Rev.: Comput. Stat. 1 (1) (2009) 101–105.
[25] T.J.M. Kooiman, M.L. Dontje, S.R. Sprenger, W.P. Krijnen, C.P. Van der Schans, M. De Groot, Reliability and validity of ten consumer activity trackers, BMC Sports Sci., Med. Rehab. 7 (1) (2015) 24.
[26] L. Van der Krieke, F.J. Blaauw, A.C. Emerencia, M.H. Schenk, J. Slaets, E.H. Bos, P. De Jonge, B.F. Jeronimus, Temporal dynamics of health and well-being: a crowdsourcing approach to momentary assessments and automated generation of personalized feedback, Psychosom. Med. (in press).
[27] B.W. Silverman, first ed., Density Estimation for Statistics and Data Analysis, vol. 26, no. 1, Chapman and Hall/CRC, London, 1986.
[28] R.E. Barnhill, R.F. Riesenfeld, in: R.E. Barnhill, R.F. Riesenfeld (Eds.), Computer Aided Geometric Design, vol. 6, Academic Press, Inc., 1974.
[29] S.L. Van Ockenburg, S.H. Booij, H. Riese, J.G. Rosmalen, K.A. Janssens, How to assess stress biomarkers for idiographic research?, Psychoneuroendocrinology 62 (December) (2015) 189–199
[30] C.A. Sims, Macroeconomics and reality, Modelling Economic Series, vol. 48, no. 1, Clarendon Press, Oxford, 1980.
[31] C. Granger, Investigating causal relations by econometric models and cross-spectral methods, Economet.: J. Economet. Soc. 37 (3) (1969) 424–438.
[32] A.C. Emerencia, L. Van der Krieke, E.H. Bos, P. De Jonge, N. Petkov, M. Aiello, Automating vector autoregression on electronic patient diary data, IEEE J. Biomed. Health Inform. 20 (2) (2016) 631–643.
[33] D. Van der Neut, Meten is weten, Psychol. Mag. (33) (2014).
[34] M.D. Majmudar, L.A. Colucci, A.B. Landman, The quantified patient of the future: opportunities and challenges, Healthcare 3 (3) (2015) 153–156.
[35] M. Myrtek, Heart and Emotion: Ambulatory Monitoring Studies in Everyday Life, first ed., Hogrefe & Huber Pub, 2004.
[36] D.S. Lauderdale, K.L. Knutson, L.L. Yan, K. Liu, P.J. Rathouz, Self-reported and measured sleep duration, Epidemiology 19 (6) (2008) 838–845.
[37] Statista, Market share of smart wristwear shipments worldwide by operating system in 2015 and 2019, November 2016. <http://www.statista.com/statistics/466563/share-of-smart-wristwear-shipments-by-operating-system-worldwide/> (accessed February 20, 2016).

[5] Website: https://jawbone.com.
[6] Website: http://nikeplus.nike.com.
[7] Website: https://misfit.com.