Experiences from a Wearable-Mobile Acquisition System for Ambulatory Assessment of Diet and Activity

Kristof Van Laerhoven
University of Siegen
Siegen, Germany

Mario Wenzel
Johannes Gutenberg University
Mainz, Germany

Anouk Geelen
Wageningen University & Research
Wageningen, The Netherlands

Christopher Hübel
Johannes Gutenberg University
Mainz, Germany

Maike Wolters
Leibniz Institute for Prevention Research and Epidemiology BIPS
Bremen, Germany

Antje Hebestreit
Leibniz Institute for Prevention Research and Epidemiology BIPS
Bremen, Germany

Lene Frost Andersen
University of Oslo
Oslo, Norway

Pieter van't Veer
Wageningen University & Research
Wageningen, The Netherlands

Thomas Kubiak
Johannes Gutenberg University
Mainz, Germany

ABSTRACT
Public health trends are currently monitored and diagnosed based on large studies that often rely on pen-and-paper data methods that tend to require a large collection campaign. With the pervasiveness of smart-phones and -watches throughout the general population, we argue in this paper that such devices and their built-in sensors can be used to capture such data more accurately with less of an effort. We present a system that targets a pan-European and harmonised architecture, using smartphones and wrist-worn activity loggers to enable the collection of data to estimate sedentary behavior and physical activity, plus the consumption of sugar-sweetened beverages. We report on a unified pilot study across three countries and four cities (with different languages, locale formats, and data security and privacy laws) in which 83 volunteers were asked to log beverages consumption along with a series of surveys and longitudinal accelerometer data. Our system is evaluated in terms of compliance, obtained data, and first analyses.

ACM Classification Keywords
H.5.m. Information Interfaces and Presentation (e.g. HCI): Miscellaneous;

Author Keywords
multi-modal data collection and presentation; barcode scanning; beverage consumption logging; activity recognition

iWOAR '17, September 21–22, 2017, Rostock, Germany

ACM ISBN 978-1-4503-5223-9/17/09. . . $15.00
http://dx.doi.org/10.1145/3134230.3134239

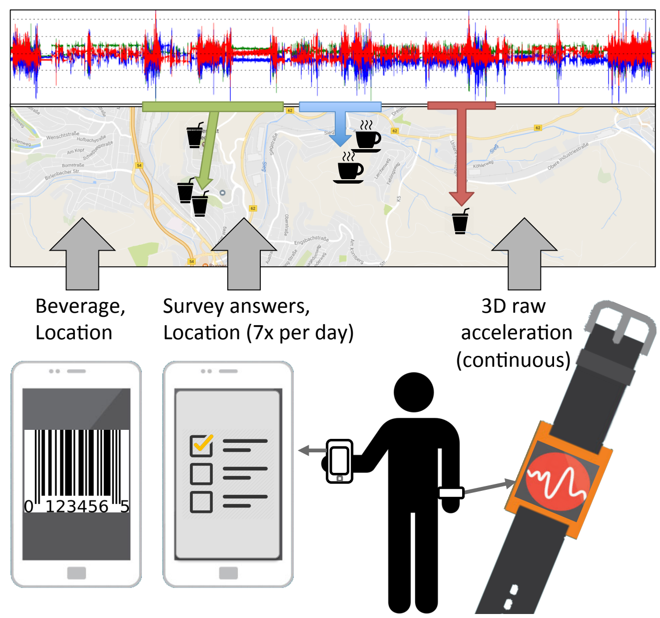

Figure 1. Our approach merges data collected from participants' smartphones and activity trackers to obtain all relevant information to soft-drink consumption and sedentary activity over the course of a week.

INTRODUCTION
Present-day data collection efforts to acquire specific information from across a large cross-section of a population tends to be a considerable undertaking. These surveys are highly needed, however, to gain a deeper and more accurate understanding into what factors influence diet and physical activity. By exploiting technology that is already present in the population, capturing these determinants across large populations then would enable us to more effectively promote healthy diet and physical activity. This paper will specifically highlight the challenges when harmonising the collected data in such studies across administrative and cultural borders, as individual

countries and states do not tend to use the same methodologies and collect inherently different multi-modal data sets.

We report in this paper on a pan-European research effort to design and study a system to gather data across borders, using participants' smartphones and custom wrist-worn activity logging devices that are able to record raw 3D acceleration. The presented research was part of the European Joint Programming Initiative Healthy Diet for a Healthy Life action Determinants of Diet and Physical Actitivy (DEDIPAC) knowledge hub (www.dedipac.eu, [13]). Through a combination of smartphone-based questionnaires and the collected location and activity data, we are able to capture the context (including the user's mood, location, activity, type of drink) of sugary drink intake. Important however is the scalability of such a system, so that collection efforts can be performed more frequently and at a large-enough scale. This paper's main contributions are therefore threefold:

- An information collection system is presented that operates through an App that is easy to deploy and run on the participant's smartphone, using its built-in sensors and interaction capabilities for questionnaires, with an activity logger.
- A first evaluation on the performance of the system is presented, detailing how the deployment was organized, as well as investigating the speed, reliability and robustness in the collection and fusion of all collected information.
- A pilot study across three countries with 83 participants forms a first evaluation of our approach, providing first insights into the challenges and pitfalls in starting and maintaining large-scale ambulatory assessment studies.

BACKGROUND AND RELATED WORK

Diet and physical activity have been known to play a critical role in overall human health, and have as a consequence been studied extensively in the past decades [13]. Most of this research is based on data from pen-and-paper self-reports produced by volunteers, which tends to make the data gathering process slow, expensive, and prone to errors. The importance of performing such studies and analyzing their data appropriately is high, and is used for guiding policies worldwide. As an example, the WHO has recently argued for taxes on sugar sweetened beverages that would result in proportional reductions in consumption, especially if aimed at raising the retail price by 20% or more [18].

Traditional self-reporting is the gold standard approach for collecting and monitoring dietary behavior across larger populations. As with any self-reported data collection, it is likely to contain certain amounts of bias and erroneous data due to false recollections (a set of issues can be found in [12]). The efforts required to prepare such a study, to make sure that bias and errors are reduced to a minimum, and to collect and integrate all data afterwards is furthermore significant. Although there is little evidence on calculating sample sizes for such pilot studies, a recently published study [6] suggests a sample size of at least 9% of the population of the main study. Motivated by these types of problems, more objective measuring technologies have been suggested, which are able to monitor diet and physical activity with less effort from the study volunteers, and are easier to deploy and collect.

A vast range of technology-based dietary monitoring technologies have been reported on in the past decade. Some research has investigated the use of body-worn sensors to track the specific sounds [1], capacitive changes around the neck [**?**], or the gestures of arms, wrists, and the upper body while of eating or drinking [19]. Recent work has focused on using exclusively off-the-shelf smartwatches [23], obtaining fair precision (65%) and recall (78%) measures for such systems under free-living conditions, which still remain rather low for use in large data collection efforts.

Camera-based approaches such as presented in [20] or [2] have been evaluated for basic feasibility, but were found to be not yet ready for large-scale deployments, either. Systems presented vary between fully wearable systems that regularly take pictures of the user's field-of-view [17] and camera systems that need to be installed in the user's environment. The challenge of managing the thousands of first-person-view images of food scenes can be met, for instance with the help of machine learning techniques and crowdsourcing efforts, as for example presented in the Platemate project [16] or as proposed in [24]. However, the installing or wearing of a camera with a limited battery lifetime is mentioned as one of the main hurdles [2] for larger-scale *in the wild* studies.

Activity recognition has been an active research area since at least a decade and a half. Some of the previously-mentioned inertial approaches have been used to detect physical activity, sleep and sedentary segments, specific gestures, and routines [4]. Although first implementations on smartphones had considerable impact on the phone's battery lifetime due to the extra data collection tasks, recent methods have achieved improvements through down-sampling and prediction strategies [26]. An additional benefit of using smartphones is the ability of complementing the phone's sensor set with specific wireless sensors to collect physiological data of interest. Other work specifically targeted smartphone-based collection of data, allowing larger populations and longitudinal experiments. These studies have shown that deploying software on popular mobile platforms has become very effective and that the generation of such data can contain behavioral patterns [14]. The use of a mobile device does have its limitations, however. As for instance presented in [7], the assumption that most participants will have their smartphones always on and at-hand for collecting user data, and for triggering interactions with the user, is found to be generally not true.

This paper describes an alternative approach, in which we rely heavily on both wearable and mobile technology to capture events of interest with situated surveys. Two public health challenges were identified that should be studied as outcomes with the impact of selected determinants being examined: (1) sugar-sweetened beverages intake and (2) sedentary behavior. Our approach targets large deployments especially, by relying on the Android platform and Google's Play Store, enabling our smartphone App to be installed by the study participants on their personal smartphones. Activity loggers that can record accurate and raw acceleration data over extended periods of time, however, are not yet as ubiquitous and still need some installation and deployment effort from the study organizers.

SYSTEM DESIGN AND COMPONENTS

The presented system consists of both user-worn devices and a back-end server where all data is collected and unified. These components, as well as the distribution and collection processes, are described in more detail in this section.

Architecture and User Devices. Two mobile devices reside with the study participants during the data collection: A wearable activity logging device for capturing activity and a custom smartphone App for capturing questionnaire data and beverage consumption information. The information from the wearable unit is recorded continuously for the entire duration of the experiment, whereas the App's data is inherently event-driven. The data from both devices is recorded locally, and is merged and synchronized after the experiment on a secure server. Figure 1 gives a brief overview of the type of information that is collected during the study. Our system's software architecture is distributed over the smartphone's App and the server.

For the Android App, written in Java, we have used the myHealthHub event-based middleware [21] as a basis for collecting sensor information and managing questionnaire events, with the option of including bluetooth-based sensors in the future. This requires two processes to be active on the smartphone, the myHealthHub middleware and our data collection App, which are both launched on the participant's phone when the experiment starts. For the accelerometer devices, we used the software that comes with the units to start and configure them; This included setting their internal date and time for later synchronization across all data sources. The server-side software is a set of data preparation, fusing, and analysis routines written in Python, that collect and synchronize all data from the participants' devices, to store these in a database.

Capturing Sedentary Behavior and Activity. The past years have seen a large variety of wrist-worn fitness trackers that allow the users' activity levels and activities such as sleeping, walking, or jogging to be recorded. For our purposes, however, three requirements make the selection of fitting activity loggers considerable smaller. First, raw accelerometer data delivering 3D readings in a known unit (usually milli-g) is needed so that data can be compared and converted across different devices. Second, the devices need to be able to sample acceleration values at a rate of at least 25 times per second in order to be able to detect particular activities later on. Third, the devices need to be able to run continuously for at least a week so that participants do not need to take care of the device themselves on a regular basis. An additional aspect was the availability of exporting the raw acceleration data into a common readable format; Solutions from commercial devices tend to vary wildly in this regard and routinely require custom software and analysis algorithms.

For this reason, we selected three different activity logging devices that fit the above requirements (depicted in Figure 2), meaning they are able to log continuous accelerometer data at a relatively high sampling rate, over larger time spans: (1) The Actigraph GT3X[1] stores 3D acceleration and ambient light. Data from the GT3X were sampled at a frequency of 32Hz and after the study downloaded and stored as proprietary binary gt3x file or, after conversion, as comma-separated CSV text files. (2) The Movisens move II device[2] can record 3D acceleration, barometric air pressure and temperature. Unlike the two others, these units were worn at the waists and were generally taken off by the participants at night. The acceleration readings were selected to be sampled at 64 Hz and stored as binary data with the unisens[3] format. (3) The HedgeHog device [4] can record 3D acceleration, temperature, and ambient light. It is, unlike the previous two devices, a research prototype whose design specifications, both hardware and software, have been open-sourced. We used a sampling frequency of 25 Hz and an acceleration range from -4g to 4g with an accuracy of 0.003g. Data from these units is stored as compressed NPZ data files that cab be read using the widely-available *NumPy* package.

Figure 2. The particpants' activities were monitored with a variety of accelerometer-based activity loggers (from left to right): the Actigraph GT3X, the Movisens move II, and the Hedgehog. All capture 3D acceleration at a frequency of 30 to 100 Hz for at least a week, continuously.

Capturing Beverage Consumption Although several systems have been conceptualized that aim at automatically recognizing the consumption of beverages by means of wearable sensors, there are currently no devices that detect the type of beverage well enough *and* can be deployed at large scale. Instead, we opted for using the personal smartphone's built-in camera to allow the study participants to enter this information as quickly and accurate as possible. For this, three options have been explored and implemented: (1) Taking a picture of the beverage for later recognition, (2) Recognizing the barcode on the beverage, or (3) manual input. Since the first option was hardly used in pre-trials and needs significant post-processing, we kept only options (2) and (3) for the main study. For a fast selection of beverages, participants could choose from a top-tier list between a selection of the following options: Sugar-sweetened carbonated drink, other sugar-sweetened drink, sport drink, energy drink, diet carbonated drink, other diet drink, smoothie, fruit-, or vegetable juice.

The sugar-sweetened beverage consumption is characterized further at every occasion, by asking the participants for supplemental information by themselves. In addition to storing a description of the beverage, the time at which it was consumed and the particpant's location (when this service was

[1]The Actigraph GT3X: http://actigraphcorp.com/support/activity-monitors/gt3x/ [last access 06/2017]

[2]Movisens move II: https://www.movisens.com/en/products/activitysensor/ [last access 06/2017]

[3]The Unisens universal data format for multi sensor data: http://www.unisens.org/ [last access 06/2017]

[4]The HedgeHog [8] Activity Logger: http://kristofvl.github.io/HedgeHog/ [last access 06/2017]

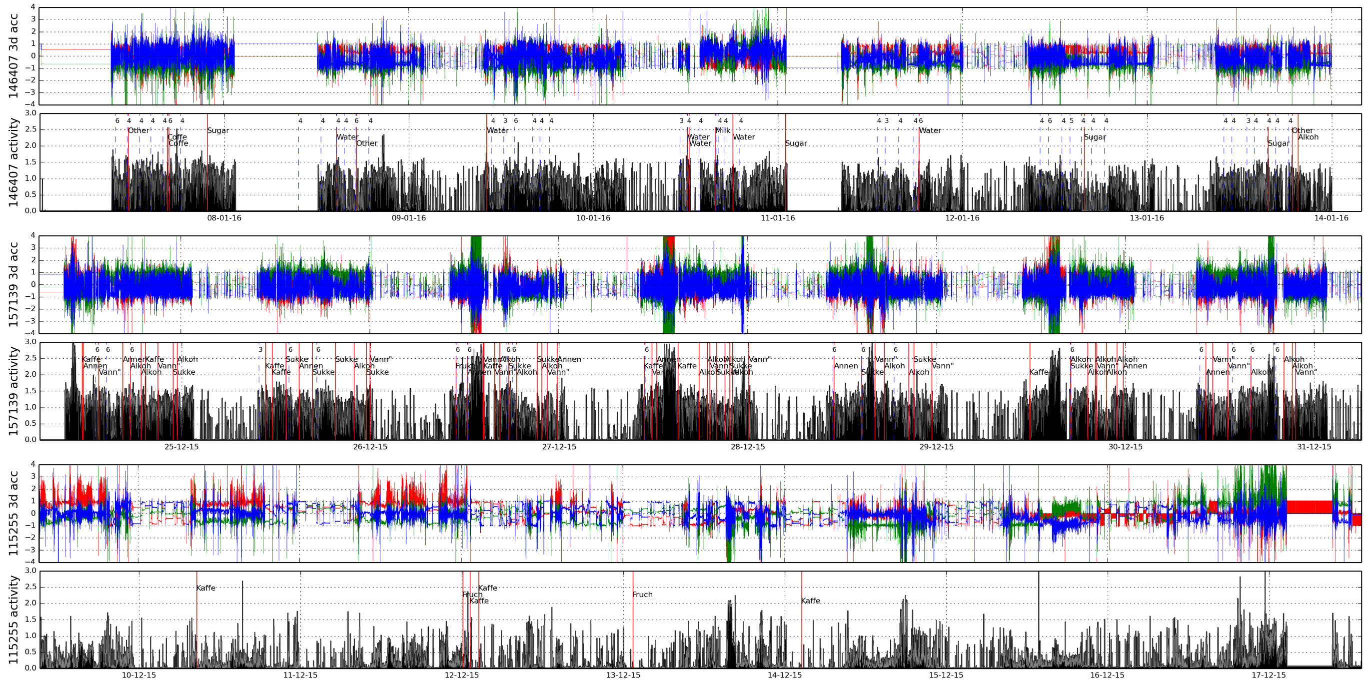

Figure 3. Some examples of raw 3D accelerometer (top plots) and accumulated activity intensity plots (bottom plots) gathered from the activity trackers from three participants. These show the data from one week, with times at which they filled in the questionnaires and at which they drank a beverage annotated with red markers and a description of the beverage. Note that participants sometimes took off their activity loggers at night (leading to no motion for several hours) and that beverage annotations were made in the local language (in Dutch, German, English, or Norwegian).

activated on the phone), also the amount (in glasses, or 250 ml) would be stored. Table 1 shows an excerpt of the information that was gathered for a set of beverage consumption instances. Whenever participants opted to use the beverages' barcodes, the obtained codes are decoded into beverage descriptions through a (paid) Open EAN/GTIN database[5] API, complemented with manual entries, on the back-end server.

ID	DATE	TIME	NAME	AMNT	BARCODE	BCN	SSB	LONG	LAT
142858	12-1-2016	09:28:14	Water	-	-	-	-	51.9850	5.6663
142858	13-1-2016	13:25:42	Water	-	-	-	-	48.0118	7.8343
142858	13-1-2016	17:13:47	Andere suikerh.	1.0	4890008100231	-	1	48.0118	7.8343
142858	14-1-2016	22:33:54	Energie drankjes	1.0	23001251	-	1	51.9728	5.6727
142858	15-1-2016	10:20:11	Water	-	-	-	-	-	-
142858	15-1-2016	12:45:42	Water	-	-	-	-	-	-
142858	15-1-2016	16:44:59	Andere suikerh.	0.5	23001244	-	1	51.9730	5.6730
142858	16-1-2016	20:37:58	Suiker+koolzuurh.	1.0	8713500008460	-	1	51.3997	5.3443
142858	16-1-2016	21:23:32	Water	-	-	-	-	51.3997	5.3443
142858	17-1-2016	17:21:11	Suiker+koolzuurh.	0.5	5000112544633	-	1	51.3997	5.3443
142858	18-1-2016	10:39:15	Water	-	-	-	-	-	-
142858	18-1-2016	12:31:17	Koffie, Thee	-	-	-	-	-	-
142858	18-1-2016	12:43:31	Andere suikerh.	1.0	87222548	-	1	51.9729	5.6726
142858	19-1-2016	10:48:49	Water	-	-	-	-	-	-

Table 1. Example of the beverage information gathered with the participants' smartphone: Participant ID, date and time, beverage description (name) and amount, barcode data, and position (when available).

[5]Open EAN/GTIN database: http://opengtindb.org/ with a 6-month fee of about $150 for API access[last access 01/02/2017].

Capturing Self-Report Data. Daily self-report surveys on the participants' mood, social vs. non-social situation, media use and availability, and social norms pertaining to sugar-sweetened beverages were triggered on the smartphone via the same App. These questionnaires were selected by the co-authors of this paper that have backgrounds in the domains of nutrition research and psychology. The goal was to assess, with surveys from key research studies presented in these fields [5, 15, 22, 9, 25], the state factors such as self-control, need for affiliation, social identity, social norm influence, exposure towards temptations, and mood.

Each day, starting at 9 a.m., seven signals are distributed throughout a time window of 14 hours, following recommendations of [10]. Within each 2-hours block, participants are randomly signalled by our smartphone App to complete the same questionnaires with the condition that two consecutive signals are at least 30 minutes apart. If the participants do not respond to the alarm within 30 minutes, the message on the smartphone disappears and the observation is listed as missing. The previously mentioned signal refers to the (semi-random) alarm scheduled by our smartphone App, which prompts participants to complete a particular questionnaires. The following event-triggered measures were thus assessed seven times per day, for the one week duration of the study:

1. State Self-Control Capacity Scale (10 items; [5])
2. Need for affiliation: Need to Belong Scale (1 item; [15])
3. Social identity: One item adapted from [22]: "I identify with/feel a connection to fellow university students."
4. Social norm influence: Three items on the presence of others or enactment models, respectively (see [9])

5. Environment: Exposure towards temptations, current availability of sugar-sweetened beverages and use of media equipment in the environment (7 items, self-report)
6. Multidimensional Mood Questionnaire (6 items; [25])
7. Objective measurement of the environment: Exposure towards temptations (e.g., availability of sugar-sweetened beverages) as well as availability of media equipment was assessed via GPS-based localization every 15 minutes.

PILOT STUDY DESIGN AND METHODOLOGY

The study was designed by a team of nutrition experts, psychologists, and engineers. The following provides the details in participant selection and study procedure.

Participants' data were collected in two waves in four study sites spread over three countries, with a different language each. The final sample consisted of 83 volunteers (56 females, age M = 23.0 years, SD = 3.3) with an average BMI of M = 21.5 kg/m^2 (SD = 2.5). Participants were recruited mainly on the study sites' campuses. For participating in the trial, participants from Wageningen have received a renumeration of $50 after study completion, those from Oslo received a gift card of about $25, while participants from Mainz and Bremen did not receive monetary compensation, but instead received partial course credit. Table 2 summarizes the inclusion and exclusion criteria used, with the assessment of mental and physical disorders solely being used for the purpose of assessing the inclusion- and exclusion criteria and discarded afterwards.

Study Procedure. The study comprises the core ambulatory assessment study, a pre-monitoring baseline laboratory session and a post-monitoring session. Participants received information about the study procedures via email or during a meeting and had the opportunity to ask questions regarding the project. They also received the inclusion and exclusion criteria, and upon confirming inclusion criteria, they participated in a first session where they signed informed consent and where the use of the smartphone questionnaires via the App and the wearable accelerometer was explained. Furthermore, participants completed trait questionnaires regarding their level of self-control, need for affiliation, social identity, and social norm influence. Starting the day after this session, participants wore the accelerometer unit continuously for one week, while also recording beverage consumption information and completing the triggered questionnaires on their smartphones.

inclusion criteria	exclusion criteria
Age: 18 to 30 years old	Pregnant or breast feeding (self-report)
University student	Students of nutrition, food, or sports science studies
Provision of informed consent	Mental disorder including eating disorder or substance abuse
Sufficient knowledge of the language of the respective country	BMI ¡18.5 or ¿35.0 (self-report)
Sugar-sweetened beverage consumption of at least one per week	Diabetes mellitus or other relevant disease (self-reported). Intake of drugs affecting metabolism like cortisol or diuretics
Possession of a compatible smartphone (Android 4.0 +; 100 MB available)	

Table 2. A short overview of inclusion and exclusion criteria for our pilot study across 4 cities (and three nations).

After one week, participants returned for a second laboratory session where they were debriefed. During this debriefing session, the participants were asked whether any complications using the accelerometer device or the myHealth Assistant framework occurred. After that, they were asked to complete a questionnaire to judge study feasibility. To underpin these subjective measures, the number of missing observations and responses were counted. After data collection and analysis was completed, participants received a document containing standardised feedback information on sugar-sweetened beverage consumption, physical activity and self-control.

Study Measures. Besides participants' compliance, sugar-sweetened beverage consumption and physical activity were the primary outcomes of the pilot study. The following measures were assessed both between-subjects at baseline and within-subjects seven times per day for one week: Sugar-sweetened beverage consumption, self-control, need for affiliation, social identity, and social norm influence. Physical activity and mood exposure towards media equipment were only assessed within-subjects. All acceleration data was converted to a common format and processed in two subsequent passes. In a first pass, the raw acceleration data was converted to actigraphy-like readings where the standard deviation was taken per axis for each minute of data, and summed up over all axes (the bottom plots in Figure 3 depict the results of this transformation). These were used to facilitate the estimation and quantification of sedentary activities throughout the data. In a second pass, a set of activity recognition classifiers was used on the raw 3D acceleration data that was determined to be non-sedentary in the first pass, to detect the more specific classes of walking and running (using the large-scale accelerometer-based dataset corpus and the random forest classifier suggested by [11]).

DISCUSSION OF RESULTS

The first results obtained from the collected data, in particular the participants' compliance and first analyses on the beverage and activity data, show the following:

Compliance to signal-contingent self-reports. To study compliance to the study protocol, we first investigated the proportion of completed signals to the total signals, by conducting a multilevel logistic regression with signal as the outcome (1=completed, 0=not completed). As predictors, we included the running signal number as well as study site and wave, controlled for gender and age, since these variables have been found to be associated with compliance to ambulatory assessment. This model revealed an acceptable compliance of 65.7%. Moreover, we found a significant effect of study site, $\chi^2(3)$ = 17.04, p ¡ .001. Participants in Oslo (M = 45.7%, SEM= 8.0) and Bremen (M = 52.6%, SE_M = 7.1) completed fewer signals on average than participants in Mainz (M = 72.7%, SE_M = 3.6) or Wageningen (M = 77.4, SE_M = 4.5), with p ¡ .017 and no difference between the latter, p = 0.447. Furthermore, although wave only approached statistical significance, $\chi^2(1)$ = 2.55, p = .111, the simple slopes showed that compliance improved from wave 1 (M = 60.0%, SE_M = 4.2) to wave 2 (M = 70.6%, SE_M = 3.5).
Regarding the effect of time on compliance, the model revealed a significant but small effect of the running signal

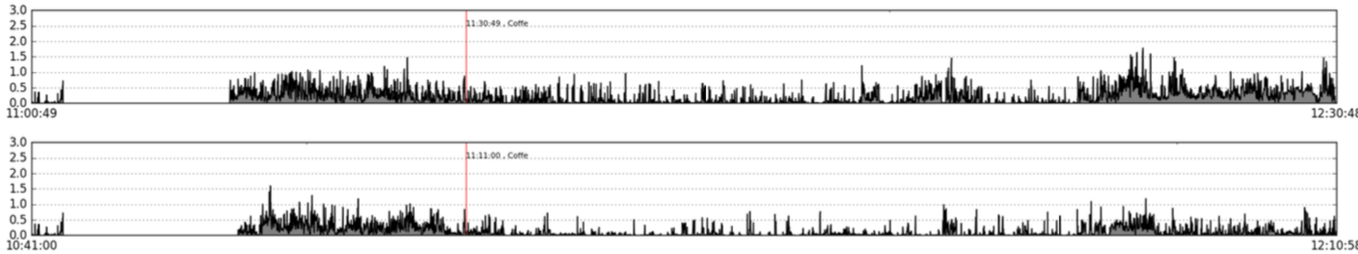

Figure 4. An example where windowed activity intensity for a participant around different beverage consumption instances exhibits strong correlation.

number, OR = 0.98, z = -6.65, p ¡ .001. Thus, with each signal, the chance of completing a signal decreased by approximately 2%. Although this might seem small, effect sizes in analyses with repeated measurements sum up; In this case, compliance is reduced by each trial, leading to a drop of compliance of approximately 16.6 percentage points from the first signal (M = 75.5%, SE_M = 2.2) to the last of study (M = 58.9%, SE_M = 2.7). Interestingly, a three-level regression model, with signals nested within days nested within participants, on the completed signals with study day (e.g., the second day of the study) and the current signal of the study (e.g., the third signal of the second day) as predictors revealed that participants completed fewer signals with each subsequent day, OR = 0.82, z = -7.96, p ¡ 0.001, but more with increasing time during the day, OR = 1.06, z = -2.71, p = 0.007.

Compliance to event-contingent self-reports. We repeated the same analyses as for the signal-contingent self-reports by conducting a mixed regression model with the number of reported beverages within two signals and the predictors outlined above. While study site was not significantly associated with the number of reported drinks when wave 1 was combined with wave 2, the latter exhibited a significant effect on beverages reporting: Participants in the second wave (M = 0.44, SEM = 0.04) reported significantly fewer drinks within two signals than participants in the first wave (M = 0.15, SEM = 0.04), b = 0.29, z = 4.77, p ¡ 0.001.

As in the signal-contingent self-reports, participants reported less drinks over the course of the study, b = -0.003, z = -4.32, p ¡ .001, leading to a drop of compliance on beverages reports of approximately 0.13 drinks within two signals, from the first two signals (M = 0.36, SEM = 0.02) to the last two of study (M = 0.23, SEM = 0.02). We followed up on this by a three-level mixed regression on the reported number of drinks within two signals, including study day and signal of the study as predictors. This model showed that both study day, b = -0.02, z = -5.39, p ¡ .001, and signal of the day, b = -0.01, z = -3.11, p = .002, were significantly negatively associated with beverages reporting, in that participants reported fewer drinks over the course of a study day and the study. // Regarding beverage consumption, participants registered a total of 2613 drinks, which amounts to about four and a half drinks per day (2613 drinks / (83 participants * 7 days) = 4.50). Out of these, 178 drinks have been registered as sugar-sweetened, with participants reporting to consume on average 2.22 glasses of sugar-sweetened beverages (SD = 3.22). An ANCOVA on total reported sugar-sweetened beverages, controlled for gender and age, yielded significant main effects for study site, $F(3, 73)$ = 4.44, p = .007, η_p^2 = .16, and wave, $F(1, 73)$ = 9.18, p = .004, η_p^2 = .12. The main effect of study site was driven by lower sugar-sweetened beverage consumption on average in Bremen (M = 0.70, SD = 1.35). The other sites Mainz (M = 2.76, SD = 3.54), Oslo (M = 2.75, SD = 3.22), and Wageningen (M = 1.83, SD = 3.63) had no significant differences. For wave as a factor, participants in wave 2 (M = 2.68, SD = 3.41) reported higher sugar-sweetened beverage consumptions on average than participants in wave 1 (M = 1.80, SD = 3.03). However, the total number of reported drinks was also higher in wave 2 and the effect of the wave was reduced but still significant, $F(1, 73)$ = 5.32, p = .024, η_p^2 = .07, when total number of reported drinks was included in the ANCOVA as a control variable. Out of 2613 reported drinks, location could be retrieved for 1235 drinks (47.3%). We conducted a multilevel logistic regression which yielded a significant effect of site, $\chi^2(3)$ = 19.37, p ¡ .001, indicating that localization worked significantly better in site Mainz (M = 63.1%, SD = 4.8) and in Oslo (M = 65.9%, SD = 4.8), compared to site Bremen (M = 27.5%, SD = 4.3) and Wageningen (M = 27.5%, SD = 4.5). However, the low proportion of location coordinates may be due to the fact that some smartphones were initially provided to the participants without SIM cards, or that the smartphones owned and used by the participants did not always have mobile internet available. We supplied the smartphones used in the second wave of studies with 4G SIM cards, which improved location coverage substantially: The proportion increased from under 34.3% to 72.7%.

Activity, beverage consumption, and routines. Beverage consumption and activity intensity occur within routines of an individual and can be found in the data through correlation analysis using the beverage report times and the sedentary behavior data (as obtained through axes-summed standard deviation over the window of a minute). We used time windows varying from 10 up to 60 minutes, for which we measured the correlation in acceleration intensity between different drinking instances for the same person. The effect of correlation window size, intensity of activity before and after the beverage consumption, and subsequent timestamps for beverage consumption were explored as possible variable parameters across all 83 participants. Potential weekday routines (i.e., similar time-of-day events in the data on different days) were used to estimate a lower bound on the hit rate of finding a potential user routine. Of all participants, 47 participants showed to have at least one such routine in their data to verify against, with 20 participants showing at least one such recurring routine for all weekdays. After exploring all parameters, it was found that using one timestamp, proceeded by a window of 30 minutes for activity intensity and followed by one of 60 minutes, provided the best results. As a more qualitative result from the first analyses of the data, we argue that the recognition of specific activities, such as in our case walking and running, within the acceleration signals can be used to identify these routines more accurately, though these were found to be far less numerous in the pilot study's data. Several of such instances were found through visual inspection of the data, which could be automated so that participants can be interviewed for these routines in further studies. It is important to note that such routines are likely to be different in studies that would recruit participants from other age categories or environments; As our pilot study recruited mostly university students, other populations might see higher (e.g., in school children or office workers) or lower (e.g., persons with more flexible daily schedules) amounts of such routines.

LESSONS LEARNED

Although it is foreseeable that a study distributed across several nations is far from trivial, we here summarize the more prominent technical challenges in organizing the pilot study.

Ethics committee approval for studies like our pilot study is mandatory, and consists of a different process in each of the countries where a deployment of the pilot study was planned. For the system to be deployable in all study sites, it was required to implement a more stringent set of requirements over all three countries in which the study was held. The following technical requirements were made across all sites by one of the ethics committees: (1) Location entries from GPS (as logged by the smartphone App) needed to be partially filtered so that residents' addresses were, after data collection, blurred using a method implemented by [3]. (2) Our acceleration-logging devices were accepted by the respective ethics review committees, largely on the premise that data was kept locally on the devices. (3) All information on diet, media use, mood, and self control were stored under a pseudonym in a secured data base at the local organizer's site.

Automatic updates via the app store for the smartphone App during pre-trials and the main trial at every site has proven to be a big advantage. Originally, we moved to the app store to facilitate deployment, but this feature was also used during the study. This meant that errors occurring at the data gathering phase could be investigated and fixed immediately, and a new version of the App could be published via Google's Play Store, triggering automatic updates on the participants' phones, at any time. This feature has been used extensively and saved time in redeployment for updates, as the App has seen 14 major revisions over the course of 7.5 months.

Diversity in platforms. Although our App was designed to solely run on Android version 4 and above, the platforms used in our study were, through the inclusion of participants' phones, diverse. Android OS versions 4 were installed on smartphones from 7 participants, version 5.0 and 5.1 on smartphones from 16 and 19 participants, respectively, and version 6.0 was installed on 41 smartphones. If participants did not own a compatible Android smartphone, a Motorola Moto G (2nd gen) was supplied, which was the model that was most often used by the participants (45 times). Other popular smartphones used by the participants include the Samsung Galaxy S3 to S7 (15 times) and Sony Xperia Z3 (12 participants).

Unification of data was an unexpected challenge, both on the level of fusing data from different devices, as well as for the merging of all survey data over the different languages and region-specific settings of the study sites. The setup of the study at all sites contained a slight overhead because of different languages, renumeration setup, and region-specific software errors, but designing the system and unifying the data for all nations involved became particularly challenging. Furthermore, as the App was designed at one site, software updates regularly contained the fixing of spelling mistakes for the other languages. Many of the App's reported bugs were very specific and often could be tracked down to software interfaces with Android libraries: One instance for example occurred only on phones with a Norwegian locale, while scanning a barcode in landscape mode. Establishing data unification was initially hampered by the fact that the data from the acceleration units tended to be large (up to around 1.5 GB per participant) and took long to be processed on the server.

CONCLUSIONS

Self-reports, which are completed by study participants, are currently the standard method when investigating diet, sedentary behavior, and physical activity across large populations. Mobile and wearable technologies can complement and enhance such collection methods with more accurate and real-time data. This paper presents a study with such as system, in which a collection system with smartphones and wearable accelerometer units was developed to monitor the sugar-sweetened beverage consumption and physical activity of 83 participants, across four cities in three different nations.

The results of our pilot study have been promising: Participant compliance did vary across the study sites but was found to be acceptable during the one-week deployment. The activity data has shown that coarse routines can be automatically extracted around beverage consumption events and that basic activity recognition approaches help in identifying and verifying these.

There were several lessons learned from a technology perspective. Studies that aim at surveying large populations, even with the help of mobile and wearable technologies, will always require a significant effort, though we have found several unexpected pitfalls and "lessons learned" throughout our specific pilot study. Cross-nation studies will need to comply with more firm ethics requirements, as they need to answer to multiple committees at the same time. Making smartphone Apps available via app stores not only accelerates the deployment phase, but is particularly helpful in spreading regular updates. This is all the more helpful as a highly diverse ecosystem of smartphones is likely to lead to problems during the study. Finally, it is easy to underestimate the unification of all data, with language barriers and a current lack of standards between the wearable accelerometers being especially challenging.

ACKNOWLEDGMENTS

This work was supported by the Joint Programming Initiative Healthy Diet for a Healthy Life as part of the Determinants of Diet and Physical Activity action. The funding agencies supporting the research reported in this article were (in alphabetical order of participating Member States): Germany: Federal Ministry of Education and Research; The Netherlands:

The Netherlands Organisation for Health Research and Development (ZonMw); Norway: The Research Council of Norway, Division for Society and Health.

REFERENCES

1. O Amft, M Staeger, P Lukowicz, and G Troester. 2005. Analysis of Chewing Sounds for Dietary Monitoring. In *LNCS*. Springer Nature, 56–72.
2. L Arab, D Estrin, D H Kim, J Burke, and J Goldman. 2011. Feasibility testing of an automated image-capture method to aid dietary recall. *European Journal of Clinical Nutrition* 65, 10 (may 2011), 1156–1162.
3. C. Buck, S. Dreger, and I. Pigeot. 2015. Anonymisation of address coordinates for microlevel analyses of the built environment: a simulation study. *BMJ Open* 5, 3 (2015).
4. Andreas Bulling, Ulf Blanke, and Bernt Schiele. 2014. A tutorial on human activity recognition using body-worn inertial sensors. *Comput. Surveys* 46, 3 (jan 2014), 1–33.
5. N. Ciarocco, J. Twenge, M. Muraven, and D. Tice. 2004. The state self-control capacity scale: Reliability, validity, and correlations with physical and psychological stress.
6. Kim Cocks and David J Torgerson. 2013. Sample size calculations for pilot randomized trials: a confidence interval approach. 66, 2 (2 2013), 197–201.
7. A K Dey, K Wac, D Ferreira, K Tassini, J-H Hong, and J Ramos. 2011. Getting Closer: An Empirical Investigation of the Proximity of User to Their Smart Phones. In *Proc. of UbiComp '11*. ACM, 163–172.
8. E Berlin et al. 2015. Low-power lessons from designing a wearable logger for long-term deployments. In *2015 IEEE Sensors Applications Symposium (SAS)*. IEEE.
9. Wilhelm Hofmann et al. 2012. Everyday temptations: An experience sampling study of desire, conflict, and self-control. *Journal of Personality and Social Psychology* 102, 6 (2012), 1318–1335.
10. J M Hektner, J A Schmidt, and M Csikszentmihalyi. 2006. *Experience Sampling Method: Measuring the Quality of Everyday Life*. SAGE Publications.
11. H Ichino, K Kaji, K Sakurada, K Hiroi, and N Kawaguchi. 2016. HASC-PAC2016: Large Scale Human Pedestrian Activity Corpus and Its Baseline Recognition. In *Adj. Proc. of UbiComp '16*. ACM, 705–714.
12. David R. Jacobs. 2012. *Challenges in Research in Nutritional Epidemiology*. Humana Press, 29–42.
13. J Lakerveld, H P van der Ploeg, W Kroeze, W Ahrens, O Allais, L Frost Andersen, G Cardon, L Capranica, S Chastin, A Donnelly, U Ekelund, P Finglas, M Flechtner-Mors, A Hebestreit, I Hendriksen, T Kubiak, M Lanza, A Loyen, C MacDonncha, M Mazzocchi, P Monsivais, M Murphy, U Noethlings, D J O'Gorman, B Renner, G Roos, A J Schuit, M Schulze, J Steinacker, K Stronks, D Volkert, P van't Veer, N Lien, I De Bourdeaudhuij, and J Brug. 2014. Towards the integration and development of a cross-European research network and infrastructure: the DEterminants of DIet and Physical ACtivity (DEDIPAC) Knowledge Hub. *International Journal of Behavioral Nutrition and Physical Activity* 11, 1 (nov 2014).
14. Abhinav Mehrotra, Robert Hendley, and Mirco Musolesi. 2016. Towards Multi-modal Anticipatory Monitoring of Depressive States Through the Analysis of Human-smartphone Interaction *(UbiComp '16)*. ACM, 1132–1138.
15. Austin Lee Nichols and Gregory D. Webster. 2013. The single-item need to belong scale. *Personality and Individual Differences* 55, 2 (jul 2013), 189–192.
16. J Noronha, E Hysen, H Zhang, and K Z. Gajos. 2011. Platemate: Crowdsourcing Nutritional Analysis from Food Photographs. In *24th Annual ACM Symposium on User Interface Software and Technology (UIST '11)*. ACM, New York, NY, USA, 1–12.
17. G O'Loughlin, S J Cullen, A McGoldrick, S O'Connor, R Blain, S O'Malley, and G D Warrington. 2013. Using a Wearable Camera to Increase the Accuracy of Dietary Analysis. *AJPM* 44, 3 (mar 2013), 297–301.
18. World Health Organization. 2016. Fiscal policies for diet and the prevention of noncommunicable diseases. (2016).
19. T Rahman, M Czerwinski, R Gilad-Bachrach, and P Johns. 2016. Predicting "About-to-Eat" Moments for Just-in-Time Eating Intervention. In *Proc. of DH '16*. ACM, 141–150.
20. Sasank Reddy, Andrew Parker, Josh Hyman, Jeff Burke, Deborah Estrin, and Mark Hansen. 2007. Image Browsing, Processing, and Clustering for Participatory Sensing: Lessons from a DietSense Prototype *(EmNets '07)*. ACM, 13–17.
21. C Seeger, K Van Laerhoven, and A Buchmann. 2015. MyHealthAssistant: An Event-driven Middleware for Multiple Medical Applications on a Smartphone-Mediated Body Sensor Network. *IEEE JBHI* 19, 2 (2015).
22. F. Marijn Stok, Denise T.D. de Ridder, Emely de Vet, and John B.F. de Wit. 2012. Minority talks: The influence of descriptive social norms on fruit intake. *Psychology & Health* 27, 8 (aug 2012), 956–970.
23. E Thomaz, I Essa, and G D Abowd. 2015. A Practical Approach for Recognizing Eating Moments with Wrist-mounted Inertial Sensing. In *Proc. of UbiComp '15*. 1029–1040.
24. E Thomaz, A Parnami, I Essa, and G D Abowd. Feasibility of Identifying Eating Moments from First-person Images Leveraging Human Computation. In *Proc. of SenseCam '13*. ACM, 26–33.
25. Peter Wilhelm and Dominik Schoebi. 2007. Assessing Mood in Daily Life. *European Journal of Psychological Assessment* 23, 4 (jan 2007), 258–267.
26. Zhixian Yan, Vigneshwaran Subbaraju, Dipanjan Chakraborty, Archan Misra, and Karl Aberer. 2012. Energy-Efficient Continuous Activity Recognition on Mobile Phones: An Activity-Adaptive Approach. In *Proc. of ISWC '12*. IEEE, 17–24.