Modelling arousal phases in daily living using wearable sensors

Martin Kusserow, *Member, IEEE,* Oliver Amft, *Member, IEEE,* Gerhard Tröster, *Senior Member, IEEE*

Abstract—In this work, we introduce methods for studying psychological arousal in naturalistic daily living. We present an activity-aware arousal phase modelling approach that incorporates the additional heart rate (AHR) algorithm to estimate arousal onsets (activations) in the presence of physical activity (PA). In particular, our method filters spurious PA-induced activations from AHR activations, e.g., caused by changes in body posture, using *activity primitive patterns* and their distributions. Furthermore, our approach includes algorithms for estimating arousal duration and intensity, which are key to arousal assessment. We analysed the modelling procedure in a participant study with 180 h of unconstrained daily life recordings using a multi-modal wearable system comprising two acceleration sensors, a heart rate monitor, and a belt computer. We show how participants' sensor-based arousal phase estimations can be evaluated in relation to daily activity and self-report information. For example, participant-specific arousal was frequently estimated during conversations and yielded highest intensities during office work. We believe that our activity-aware arousal modelling can be used to investigate personal arousal characteristics and introduce novel options for studying human behaviour in in daily living.

Index Terms—Arousal estimation, affect monitoring, daily hassles, wearable system, heart rate, body motion, daily activity.

1 Introduction

Daily life is rich of different situations in which our body triggers affective responses involving physiological activation, termed *arousal* [1]. Positively and negatively valenced arousal involves the activation of the autonomic nervous system (ANS) [2], and is a key aspect of analysis in affective computing (AC) [3]. Short-term affective responses can be considered as spikes in arousal during the day [4] and there is evidence that accumulating negatively valenced arousal, often denoted as stress, e.g., interpersonal tension or arguments with others, can result in detrimental effects for well-being [5], [4]. To better understand the influences of daily life arousal, an automatic and unobtrusive monitoring is needed that can objectively quantify arousal in daily life. A quantification would comprise when, how long, and how intense the experienced arousal was. In addition, a relation to the situational context would be included to enable an understanding of the arousal trigger (negatively or positively valenced) and provide possible feedback mechanisms for support. As part of an assistant system, such a monitoring could assess individual arousal phases with respect to specific daily activities and provide support for effective coping with negatively valenced arousal in known arousal contexts, e.g., support during public talks [6].

However, towards this vision objective assessment strategies to continuously monitor and characterise arousal situations during daily living are lacking. While current ubiquitous technology enables us to design unobtrusive wearable monitoring systems, quantifying arousal responses in daily living remains a challenging problem. Most importantly, because physical activity (PA) elicits similar bodily changes as psychological arousal, changes in physiological signals cannot be uniquely attributed to psychological causes [7]. Towards quantifying arousal, physiological responses to psychological stressors must be (1) robustly disentangled from PA effects and (2) described regarding onset, duration, and intensity. The relevance of the latter for affect estimation in continuous manner has been recognised by other researchers in AC, e.g. [8]. Arousal monitoring using sensors is an appealing option but methods are needed that quantify individual daily arousal exposure from sensor signals.

In this work, we explore a modelling and estimation approach for analysing human behaviour in daily life using on-body monitoring. To identify activations, we incorporated an algorithm from Psychophysiology [9] and extended it by an activity-aware filtering. More importantly, this work introduces algorithms that estimate arousal duration and intensity in addition, to provide a complete arousal phase estimation.

1.1 Arousal estimation challenges in daily living

Research of arousal analysis from physiological signals, including cardiac responses, has focussed on lab-based investigation, controlling for PA and setting artificial arousal condition, e.g. [10], [11] (see [12] and [13] for a comprehensive overview). However, transfer and applicability of lab findings to naturalistic environments is

- *M. Kusserow and G. Tröster are with the Wearable Computing Lab, ETH Zurich, 8092 Zurich, Switzerland. E-mail: kusserow@ife.ee.etz.ch, troester@ife.ee.ethz.ch*
- *O. Amft is with the ACTLab, Signal Processsing Systems, TU Eindhoven, P.O.Box 513, 5600 MB Eindhoven, The Netherlands. E-mail: o.amft@tue.nl*

limited. In the context of this work, *naturalistic* refers to monitoring individuals' real life situations where stressors naturally occurred and were not artificially generated, and physical activity and behaviour was not restricted. In contrast to the lab, monitoring affect in natural environments imposes higher demands on sensing systems and analysis methodology to collect and meaningfully interpret sensor data. Therefore, literature advocated to develop novel analysis methods for natural daily living conditions [3], [14], [15], [16], which is the aim of our work.

Towards estimating arousal in daily living, the following challenges have been addressed in this work: (1) continuous monitoring of activity and arousal using a wearable system, (2) modelling of arousal phases to determine arousal onset, duration, and intensity, and (3) using evaluation strategies to interpret estimated arousal episodes in daily living studies.

1.1.1 Continuous wearable sensing & activity annotation

Due to the interaction of arousal and PA, multi-modal sensing should include modalities to estimate both. Besides sensing modalities that were used to record physiological signals, e.g., electrodermal activity (EDA) [17], the Electrocardiogram (ECG) has been shown to provide robust arousal information. The relation to arousal was maintained even when recorded in everyday life [9] and during sports competition situations [18]. Frequently, investigations made use of heart rate (HR) as one feature of the ECG, which can be extracted with a high signal-to-noise ratio (SNR). To capture PA during daily living, motion-sensitive sensors can be attached to the body. Body-attached sensors must not interfere with the wearer's daily activities and can be used for arousal monitoring, e.g., BodyANT [19] used in this work.

Supervised information on behaviour and arousal is difficult to collect in daily living studies, as individuals follow their personal routines. Usually, type and frequency of arousal triggers are different and uncontrollable for individuals, and their appraisal highly individual calling for participant-specific analysis approaches [4]. Consequently, arousal estimation in naturalistic settings cannot be considered the same recognition problem as for controlled environments [20]. Instead, alternative approaches are needed that explore plausible behaviour information to interpret arousal estimations in naturalistic environments and that can be combined with sensor-based measurements, e.g., daily activities [15]. Thus, annotations of daily activity could be obtained to analyse individual arousal estimations. We consider daily activities as contextual classes typically having durations of minutes to hours. For these time frames, individuals can maintain diaries of their daily activities, including, e.g., eating and working times, which would allow interpreting plausible individual stressors.

1.1.2 Evaluation of arousal estimations

Even in lab environments no gold standard exists to evaluate the performance of affective computing systems [12]. To evaluate the coherence of sensor signals and psychological states, self-reports are often considered imperative. While in the lab, individuals can explicitly focus on reporting tasks, e.g., when watching music videos [21], self-reporting in daily living studies are impaired by several aspects [22], [23]. In particular, individuals often lack abilities to perceive and report arousal sensations in a concise and reproducible manner. To some extend, standardised questionnaires can alleviate this issue. Nonetheless, in daily living it is often not possible to report instantaneously. This issue remains, even when using mobile phones that facilitate timely and comfortable user interfaces [24]. Consequently, the reliability of self-reports is restricted in natural settings.

Saliva samples were used to measure arousal responses as concentration of biochemical markers [10], e.g., cortisol. Saliva samples can be taken and stored by an experimenter in a lab environment. However, the sampling procedure is not practical for study participants during daily living, as this would require obtaining and storing saliva samples during the day.

Albeit the limited transferability of lab-based evaluation instruments, a combination of available assessments (daily activities, self-reports, questionnaires) could provide means for interpreting plausibility of arousal estimations from sensor data, as considered in this work.

1.1.3 Arousal estimation approach

In natural daily living, PA is a substantial source of noise constraining the options for arousal estimation in any physiological signal, including HR used in this work. Previous approaches often used body acceleration to identify and exclude PA-intensive data segments from analyses, e.g. [24], [25]. Because of the tight relation of PA and HR, rather than excluding all segments containing PA a-priori, information extracted from PA could be used to estimate arousal, as introduced below.

1.2 Arousal estimation by additional heart rate

From seminal investigations in Psychophysiology [26], the additional heart rate (AHR) algorithm [9] emerged (see Section 3.2 for its specifics). AHR capitalises on the finding that PA and HR are tightly related [27] and measured HR can be regarded as the additive combination of HR responses to physical and psychological stressors [28], [29]. The disproportion between HR increases without an accompanying increase in PA was attributed to psychological origin (*additional* referring to the HR component due to psychological arousal). The validity of the AHR algorithm to identify arousal onsets in the presence of PA was investigated on more than 1300 individuals [22].

Limitations of AHR-based arousal estimation. However, the AHR algorithm is sensitive to some PA-induced

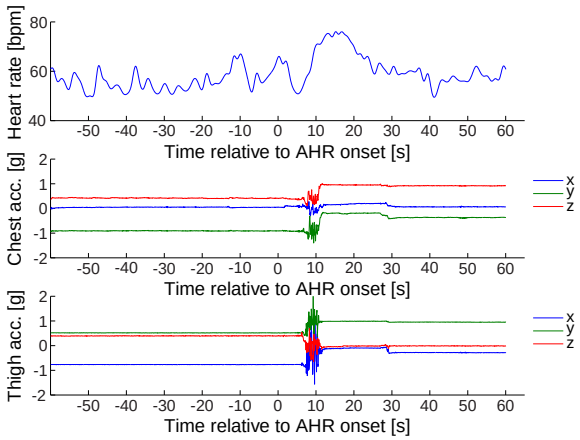

Fig. 1. PA-induced arousal estimation of the AHR algorithm during a sit-stand transition: HR (top) and 3D acceleration signals of chest (middle) and thigh (bottom). HR increased from ∼55 to ∼75 bpm for a period of 10 s after the transition at $t = 10$ s, suggesting that posture information can help identifying PA-induced arousal.

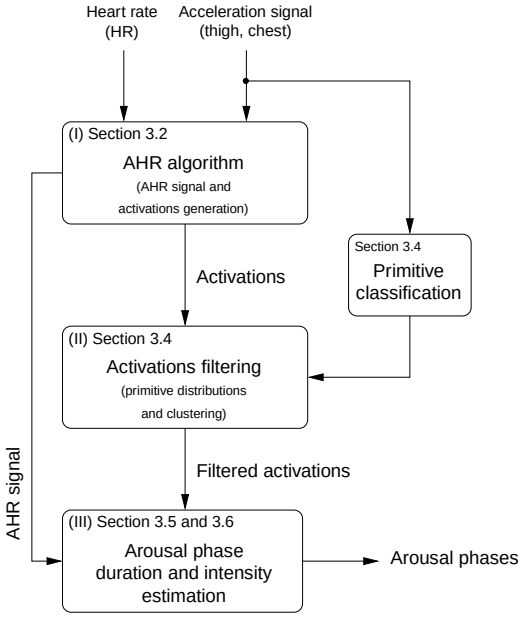

Fig. 2. Schematic of the arousal phase estimation procedure. (I) HR and acceleration data are combined to identify activations (potential arousal onsets) after Myrtek et al. [9]. (II) PA-induced activations are filtered using our primitive modelling. (III) Estimation of duration and intensity to obtain arousal phases.

HR changes [30] and does not provide an estimation of arousal duration and intensity that are essential to quantify accumulated arousal. As AHR uses only PA intensity for its estimation, it cannot cope with particular physiological responses, e.g., due to posture transitions.

Figure 1 illustrates an example of HR and 3D acceleration signals measured at chest and thigh during the transition from sit to stand. HR increased by 15 bpm directly after the transition and remained elevated for about 10 s. The AHR algorithm identified arousal by the HR increase due to the posture change rather than a true arousal condition. In this example, HR changes can be directly attributed to changes in posture.

The remainder of this paper is organised as follows. Section 2 summarises our approach and Section 3 describes our implementation of the arousal phase modelling. Our wearable recording system and data recording procedure is described in Section 4. The results obtained are detailed in Section 5. Finally, we discuss and summarise important findings in Section 6 and Section 7.

2 Approach Summary

We describe a procedure to estimate onset, duration, and intensity of arousal phases from HR and acceleration data. The basis of our approach is the AHR algorithm [9] (as introduced in Section 1.2). We alleviated the PA-related sensitivity of the AHR algorithm by introducing an activity-aware filtering procedure (see Section 3.4). Novel algorithms for duration and intensity estimations were derived (see Section 3.5 and 3.6). We analyse outcomes and limitations of our arousal phase modelling procedure on four participants's daily living data. For this purpose, we gathered 182.5 h of sensor data in natural daily living using a multi-modal wearable monitoring system. We investigated estimated participant-specific arousal with respect to personal daily activities and self-reports, including questionnaires. Using data of four participants allowed us to take interpersonal variation into account, and record and analyse data with reasonable participant burden.

Figure 2 depicts the schematic of our arousal phase estimation procedure comprising of three main blocks: (I) fusion of HR and acceleration signals using the AHR algorithm after Myrtek et al. [9] (see Section 3.2) to estimate arousal in continuous manner (AHR signal) and determine activations (potential arousal onsets), (II) filtering of PA-induced activations (see Section 3.4), and (III) estimation of arousal duration and intensity given the filtered activations and the AHR signal (see Section 3.5 and 3.6). In (I) we aimed at a high sensitivity to capture potential arousal onsets, whereas in (II) we optimised specificity of the arousal onset estimation.

The AHR algorithm is not sufficiently robust to all PA-induced HR increases. We developed an activity-aware activations filtering using clustering. Therefore, we modelled classes of acceleration signal patterns, referred to as *primitives*, and their distributions. We used clustering to automatically identify those activations that exhibited lowest influence of PA, as indicated by changes in primitive distribution patterns. The activations filtering is described in detail in Section 3.4.

To determine arousal duration and intensity, we developed an algorithm that estimates the end point of an arousal phase from the set of activation obtained after filtering and the AHR signal (see Section 3.5 and 3.6).

3 IMPLEMENTATION

3.1 Heart rate and acceleration signal processing

We used a wireless body area network (WBAN) to monitor HR and body acceleration (inputs to algorithm block I in Figure 2, device details in Section 4.1).

A wireless heart rate monitor (HRM) measured the time duration between consecutive heart beats (RR-intervals in ms). Noisy RR-intervals that deviated by more than 20% from their predecessors were excluded [31]. Subsequently, the RR-interval time series was interpolated using a piecewise cubic polynomial. A linear interpolation was used if no samples were received for more than 5 s. Interpolation sampling rate matched the sampling rate of the acceleration samples (in this work 16 Hz, see Section 4.1). From the interpolated RR-interval series, we computed the HR time series in beats·min^{-1} (bpm) by $HR = 6 \cdot 10^4 \cdot RR^{-1}$ (RR in ms). We linearly interpolated the acceleration samples that were not received from the wireless sensors (quantification of missing samples in this work in Section 4.3). To match a common time reference, the restored HR and acceleration data streams were aligned using MATLAB. For the computation of AHR, data streams were segmented into one minute epochs on which the means of HR and PA were computed as detailed below. To limit the influence of activations because of signal restoration, we excluded those AHR activations from the analysis where either a total of ten RR-intervals within the preceding three minutes and the minute of activation or five consecutive RR-intervals within the minute of AHR activation were interpolated (29.5% of potential AHR activations of the data set, see Section 4.3).

3.2 Additional heart rate (AHR) algorithm

The AHR algorithm developed by Myrtek et al. [9] identifies onsets of arousal by the fusion of HR and body acceleration information (see algorithm block I in Figure 2). When the AHR algorithm was developed, a dedicated sensor system (called FMS) was used whose acceleration sensors generated specific units of PA [26]. Here were summarise the AHR algorithm in such a way that it can be more generally applied to other sensing devices, for example, wireless HRM chest belts and 3D accelerometers, as considered in this work.

3.2.1 AHR signal

The AHR signal is a continuous estimation of arousal (resolution 1 min) from HR and PA. Analytically, the AHR signal is the difference of heart rate $HR[i]$ and heart rate level $HR_L[i]$ (three minute HR average) in relation to the required minimum increase of heart rate $HR_+[i]$ for minute i.

$$AHR[i] = (HR[i] - HR_L[i]) \cdot {HR_+}^{-1}[i] \tag{1}$$

$HR[i]$ is computed from the interpolated (sampling frequency f_s) heart rate samples HR_k of RR-interval samples (interbeat interval RR_k in ms) for each minute i.

$$HR[i] = \frac{1}{60 \cdot f_s} \sum_{k=1}^{60 \cdot f_s} HR_k \tag{2}$$
$$\text{with} \quad HR_k = 6 \cdot 10^4 \cdot RR_k^{-1}$$

$HR_L[i]$ is the mean heart rate of the three minutes before minute i.

$$HR_L[i] = \frac{1}{3} \sum_{k=1}^{3} HR[i-k] \tag{3}$$

$HR_+[i]$ is the required minimum increase in heart rate to qualify a minute i as activation. $HR_+[i]$ depends on $PA[i]$, and the parameter K.

$$HR_+[i] = K^{-1} \cdot (90 + PA[i]) \tag{4}$$
$$\text{with} \quad K \in \{1, 2, \cdots, 30\}$$

Combining (1), (3), and (4), yields an equation to compute the AHR signal from $HR[i]$ and $PA[i]$.

$$AHR[i] = \frac{K}{90 + PA[i]} \cdot \left(HR[i] - \frac{1}{3} \sum_{k=1}^{3} HR[i-k] \right) \tag{5}$$

3.2.2 Computation of PA

PA is measured using accelerometers at the chest and the thigh. The analytical expression of PA in the AHR algorithm refers to specific units that were originally generated by the dedicated acceleration sensors of the FMS. To calculate these units, literature [26] defines a logarithmic conversation ($y = a + b \cdot \log(x + c)$, with input x, output y, and constants a, b, and c) of 3D acceleration samples according to (6).

$$PA_k = 127.322 + 44.2078 \cdot log(\|\mathbf{a}[k]\| + 0.039) \tag{6}$$
$$\text{with} \quad \|\mathbf{a}[k]\| = \sqrt{{a_x}^2[k] + {a_y}^2[k] + {a_z}^2[k] + {a_{x'}}^2[k]}$$

Hereby, $\|\mathbf{a}[k]\|$ is the L2-Norm of the high-pass filtered (first order Butterworth, cut-off frequency 0.1 Hz) sample k of the 3D chest accelerometer axes (x, y, z), and 1D sagittal acceleration (x') of the thigh accelerometer. By definition, PA is bounded to $|PA| \leq 200$. The logarithmic conversion is computed on each acceleration sample k (sampling rate f_s) with subsequent averaging to obtain $PA[i]$ of each minute i.

$$PA[i] = \frac{1}{60 \cdot f_s} \sum_{k=1}^{60 \cdot f_s} PA_k \tag{7}$$

3.2.3 AHR activations.

Minutes i that fulfil all of the following three conditions are referred to as *activations* (arousal onsets) [26].

$$\text{Condition 1} \quad : \quad AHR[i] \geq 1 \tag{8}$$

$$\text{Condition 2} \quad : \quad PA[i] \leq PA_L[i] + 10 \tag{9}$$

$$\text{with} \quad PA_L[i] = \frac{1}{3}\sum_{k=1}^{3} PA[i-k]$$

$$\text{Condition 3} \quad : \quad HR[i] > HR[i-1] \tag{10}$$

Condition 1 requires the difference between HR and HR level to exceed a threshold with respect to PA. Condition 2 states that the maximum increase in PA during one minute should be bounded with respect to the mean PA level. Condition 3 ensures that only increases in HR are considered activations.

3.3 AHR algorithm sensitivity adjustment

The sensitivity of the AHR algorithms influences the number of identified activations. The parameter K in (5) controls the sensitivity ($K \in \{1, \ldots, 30\}$). Higher sensitivity yields more activations, as a larger values of K put more weight on smaller differences in HR.

Different daily activities can have different proportions, types, and intensities of PA, that can influence HR. Thus, we investigated the sensitivity (selection of K) of the AHR algorithm with respect to daily activities. We introduced the conditional probability $P(A|K,R)$ of an activation A given K and daily activity R to quantitatively describe arousal and the relation to daily activities. R can be any daily activity used in this work. The conditional probability $P(A|K,R)$ is calculated as the ratio between the number of activations $N_R^{(K)}$ and the total duration T_R of a daily activity R (in min) in the data set.

$$P(A|K,R) = T_R^{-1} \cdot N_R^{(K)} \tag{11}$$

3.4 Activity-aware activations filtering

The AHR algorithm can generate PA-induced activations not related to arousal (see Figure 1, Section 1.2). To identify and filter out PA-induced activations, we developed a filtering scheme using classification of primitives (see algorithm block II in Figure 2). To identify the activations that are least likely influenced by PA transitions, clustering partitions differential distributions of classified primitives according to their pattern similarity. We considered activation to be least likely influenced by PA if the activations' primitive distributions change least.

3.4.1 Primitive classification

Based on the influence of body posture and motion on HR, we considered the following six primitives: sit, stand, walk (thigh accelerometer), bend front, lean back, upright (chest accelerometer). One Naïves Bayes classifier for each sensor position at the chest and thigh automatically recognised the corresponding primitives. The two features mean and variance of the horizontal acceleration axis of each sensor position (sliding window size 3 s, step size of 0.5 s) were used as classifier input. To train the classifiers, we collected a primitive data set on individuals performing a sequence of all primitives during a separate recording. Testing the classification performance on the daily activity data would require to annotate the data set with all the primitives that would have implied an unreasonable burden. Consequently, we tested the performance of the primitive classifiers using a 10-fold cross validation on the primitive data set (see Section 5.2).

3.4.2 Computation of differential primitive distributions

We introduced primitive distributions and their temporal changes to include the primitive information in the activations filtering. Through the use of primitive distributions, information of body motion detected at a higher time resolution compared to activations can be incorporated. Using differential distributions allows to identify relative activity transitions and their effects in HR. Primitive distributions also allow for a generic use in the filtering process, as this approach scales with the number of primitives.

If an activation is influenced by transitions in PA, we assumed that this would be related to the minute right before and during the activation. As features for the clustering, we used the primitive distribution differences Δp_n of primitives p_n, with $|p_n|$ being the number of observations assigned the primitive class p_n, N the total number of observations within one minute, and Δ to denote the forward difference operator between the minute before and during activation (12).

$$\Delta p_n = N^{-1} \cdot \Delta |p_n| \tag{12}$$

To also account for changes in PA intensity, we introduced a seventh primitive a_{dyn} that is computed from the mean of the L2-Norm of the 3D acceleration dynamics $\|\mathbf{a}\|$ at both the chest and the thigh (13). The primitive a_{dyn} ranges between [0 1] to match the scaling of the other six primitives.

$$a_{dyn} = 0.5 \cdot \sum_{d \in \{chest, thigh\}} \|\mathbf{a}_d\| \cdot max^{-1}\{\|\mathbf{a}_d\|\} \tag{13}$$

3.4.3 Clustering of differential primitive distributions

The clustering automatically derives a participant-specific decision criteria to identify activations that are least influenced by PA. Therefore, the clustering partitions the differential primitive distribution according to their similarity. A partitioning yields d clusters and the choice of d influences the number of activations obtained after the filtering. The filter automatically retains the cluster c_i (k as index of the activations assigned to c_i)

with the lowest mean sum of the absolute value of the primitive distribution differences $|\Delta p_n^{(k)}|$ (14). All activations that were assigned to the other $d-1$ clusters are discarded.

$$\hat{c}_i = \underset{c_i}{\operatorname{argmin}} \left\{ |c_i|^{-1} \cdot \sum_{k \in c_i} \sum_{n} |\Delta p_n^{(k)}| \right\} \qquad (14)$$

3.5 Arousal phase duration

We developed an algorithm to compute the duration of an arousal phase from the filtered activations and the AHR signal (see algorithm block III in Figure 2). The rationale behind the algorithm is that every arousal phase has a finite duration and that arousal intensity declines after some time. While the activations had been filtered to remove PA causes, causes of the arousal intensity decline can be manifold, e.g., the end of a stressful condition. Therefore, similarly to the formulation of an onset (activation), a formulation of arousal decline was derived. We define a decline in arousal as a decrease in HR below mean HR level: $HR[i] < HR_L[i]$ and, thus, $AHR[i] < 0$, cf. (5). The arousal phase ends when a decline of at least the magnitude of the activation occurs. However, the three minutes windowing effect of the AHR signal needs to be considered, as the arousal decline of a single activation is dispersed across the three minutes following the activation. For example, a single activation with $AHR[i] = 3$ (PA kept constant) can result in an arousal decline of $AHR[i+k] = -1$ for $k = 1, 2, 3$. As a result, we derived the following decision rule to determine the phase duration T_p for an activation at minute i:

$$T_p(i) \leftarrow \underset{k}{\operatorname{argmin}} \{ AHR(i+k) \leq -\frac{1}{3} \cdot AHR(i) \} \qquad (15)$$

As activations can be located within an estimated phase of another activation, called *nested activations*, the phase duration calculation is applied recursively. Nested activations can be regarded as a successive increase in arousal intensity.

Algorithm 1 describes our arousal phase duration estimation in pseudocode notation. As an example, Figure 3 shows two activations at $i = 2$ and $i = 3$ to illustrate the algorithm's functionality. The activation at $i = 2$ is processed first and assigned $T_p = 3$. Thus, the activation at $i = 3$ is a nested activation and processed in the same fashion as the previous activation. The nested activation's phase is assigned $T_p = 2$. To indicate that the same end point A'_N is shared across activations, its AHR signal value $AHR(A'_N)$ is adjusted by the threshold value of the nested activation $threshold(A'_N)$. With the adjusted AHR signal value at $i = 5$ the first activation is processed again and $T_p = 3$ is confirmed.

In the computation of arousal duration, we considered phases as entities that could overlap in time (resolution one minute). However, this would result in an ambiguous mapping of minutes under evaluation to arousal phases, as one minute under evaluation could be assigned to multiple arousal phases. To resolve this issue, we identified overlapping phases and assigned each minute under evaluation to at most one phase. If phases overlap during a minute under evaluation,

Algorithm 1: estimateArousalPhaseDuration

Input : Set of activations A, AHR signal AHR
Output: Set of phase end indices K.

while $i \in A,\ A \neq \emptyset,\ i$ *has not been processed* **do**
  `// Set end point condition`
  $threshold(i) \leftarrow -AHR(i)/3$
  repeat
    `// Compute phase duration`
    $T_p \leftarrow \underset{k}{\operatorname{argmin}} \{ AHR(i+T_p) \leq threshold(i) \}$
    `// Process any nested activations`
    $A_N \leftarrow \{ j \mid j \in A : i < j < i + T_p \}$
    $K_N \leftarrow$ estimateArousalPhaseDuration(A_N)
    `// Adjust AHR signal at end points`
    $A'_N \leftarrow \{ j' | j \in K_N : j' \leftarrow j + 1 \}$
    $AHR(A'_N) \leftarrow AHR(A'_N) + threshold(A'_N)$
    `// Compile set of phase end indices`
    $K \leftarrow K \cup \{ i + T_p - 1 \} \cup K_N$
  until $A_N = \emptyset$;
end

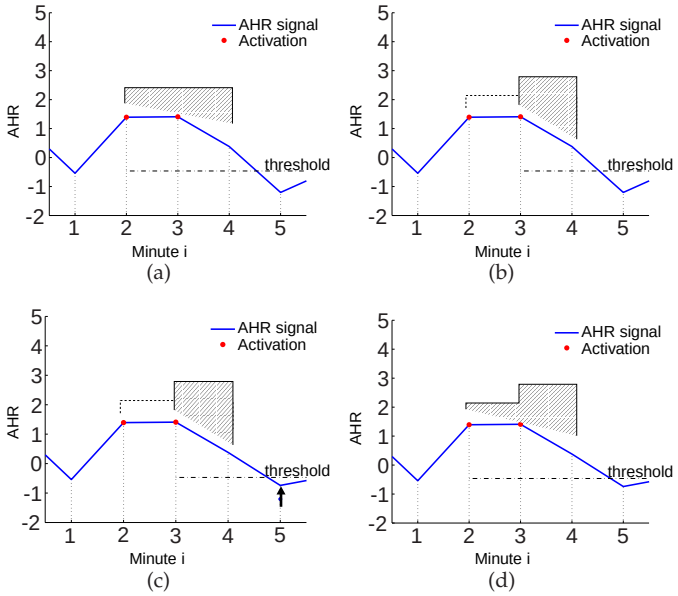

Fig. 3. Example of the arousal phase estimation procedure (see Section 3.5 for details). The hatched area indicates the minutes belonging to a phase: (a) End point of the first activation derived from its threshold, and a nested activation. (b) Recursion step to compute the arousal duration for the nested activation at $i = 3$. (c) Adjustment of the AHR signal to indicate phase end of the nested activation. (d) Recomputation of the phase duration for the first activation.

the phase with the higher intensity is assigned to that minute under evaluation. In the example Figure 3, phase 1 is assigned a duration of 1 min (minute 2) and phase 2 (nested phase with higher intensity than phase 1) is assigned a duration of 2 min (minutes 3 and 4).

From the estimated arousal phases, the probability $P(A_p|R)$ of a person being in an arousal phase A_p given the daily activity R can be computed as the ratio of the sum of durations of the arousal phases $T_p^{(i)}$ relative to the total duration T_R of the corresponding daily activity R.

$$P(A_p|R) = T_R^{-1} \cdot \sum_{i \in R, A_p} T_p^{(i)} \tag{16}$$

3.6 Arousal phase intensity

The intensity I of an arousal phase is the AHR value of its corresponding activation at minute i. According to literature [22], HR can be considered a nearly additive combination of each single stressor's HR contribution. Therefore, for nested phases, the arousal intensity is the sum of the AHR value of its corresponding activation at minute i and the surrounding arousal phase A_p.

$$I = AHR(i) + AHR(A_p) \tag{17}$$

In the example Figure 3, arousal phase 1 has an intensity of 1.4, and arousal phase 2 is a nested phase with an intensity of 2.8 (its activation of 1.4 plus the AHR value of the surrounding arousal phase 1 of 1.4).

3.7 Evaluation strategy

As detailed in Section 1.1.2, ideal references to evaluate arousal in daily living do not exist. Our evaluation approach aims at confirming the method's applicability for naturalistic daily living studies using several complementing strategies. Here, we related estimated arousal phases to annotated daily activities and self-reporting of participants to confirm plausibility of results.

3.7.1 Relation to daily activities.

Daily activities can help in associating situations of arousal experiences and place them into social contexts. Hence, we analysed estimated arousal phases with respect to daily activities. In particular, the following daily activities were considered: *Eating* (any kind of meal intake, except snacking: breakfast, lunch, dinner), *Conversation* (conversation with other people, also talking on the phone or at the office desk), *Housework* (including tidy-up, preparing food, washing and drying dishes), *Hygiene* (any kind of personal hygiene, brushing teeth, using the toilet), *Work* (being at the place of work but not at the office desk), *Office desk* (working at the office desk, excluding conversations), *Office meeting* (attending a business meeting in the office), *Read a book* and *Watch TV* (both leisure activities), *Bicycle*, *Transport* (using public transport such as train, bus, tram), *In transit* (any kind of physical transition between places, except using public transport). *Bicycle*, *In transit*, and *Housework* most likely involve more PA than the other daily activities.

3.7.2 Relation to self-report.

We repeatedly recorded subjectively perceived arousal by a standard questionnaire (multidimensional mood questionnaire (MDBF), short form A, in German) [32]. The questionnaire consists of twelve items each with a 5-point scale (1=not at all, 5=very much) to assess three bipolar dimensions: having good vs. bad mood (GB), feeling awake vs. tired (AT), and feeling calm vs. aroused (CA). Values for each dimension range between 4 (bad mood, tired, aroused) and 20 (good mood, awake, calm). Participants were advised to complete a questionnaire whenever they perceive arousal, and randomly throughout the day. This approach was used to collect self-reports even if participants did not perceive arousal.

To relate questionnaire information and estimated arousal phases, we computed the time between estimated arousal phases and the time the questionnaires were completed (as indicated by the participant). Each estimated arousal phase was assigned to the next questionnaire event following that arousal phase while we assumed that questionnaires completed more than 60 min after an estimated arousal phase were not related to the estimation. In addition, we computed each questionnaire's scores for three dimension GB, AT, and CA.

4 Experimental procedure

To test our arousal phase modelling procedure, we collected a data set during natural daily living using a multi-modal wearable monitoring system.

4.1 Wearable monitoring system

Figure 4 depicts the wearable system and sensor placement. The system comprised a heart rate monitor (HRM) chest belt (Suunto ANT, data transmission frequency: 5 Hz) and two 3D wireless acceleration sensors, called BodyANT [19]. BodyANT was attached with an elastic strap to the right thigh above the knee, and the chest (sampling frequency 16 Hz). The HRM wirelessly transmitted the time duration between consecutive heart beats (RR-interval in ms). The Q-belt integrated computer (QBIC) [33] was used to receive and store the data samples, using the Context Recognition Network Toolbox (CRNT) [34] running on the device.

4.2 Participants and recording procedure

One female and three male individuals, aged between 24 and 34 (no known medication or history of cardiac or psychological diseases), participated in our recordings. We trained the participants to put the system on and off, and to operate the system by themselves. The participants could suspend the recording when it was

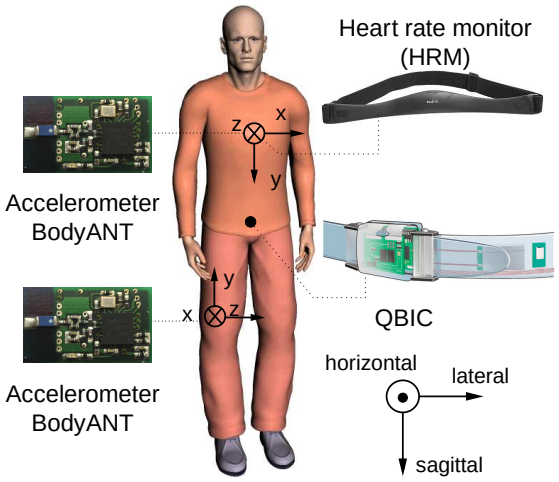

Fig. 4. Wearable monitoring system and sensor placement: Heart rate monitor (HRM) and two 3D wireless acceleration sensors (BodyANT), attached to the right thigh and the chest. Arrows indicate the coordinate system of the BodyANT sensors. The Q-belt integrated computer (QBIC) received and stored the data samples.

not favoured or hindering, e.g., during the night while asleep, while having a shower, or during sports. Every participant maintained a log of daily activities (temporal resolution: 2 min) during the data recording. In addition, the participants were requested to fill out questionnaires using paper and pen (see Section 3.7) during the day and add additional comments on salient situations. For data analysis, we added the daily activity annotation and questionnaires (time of completion and scores) to the data set.

4.3 Data set

Using the wearable sensor system, we recorded data from four participants during their normal working days from the morning after getting dressed until the evening before undressing and going to bed (Participant 1: 31.9 h, Participant 2: 44.7 h, Participant 3: 64.0 h, Participant 4: 41.9 h, total: 182.5 h). Figure 5 illustrates the relative duration of the annotated daily activities in the dataset for all four participants.

During daily living, the participants spent most of their time in the activities *Office desk* followed by *Conversation*. Not all participants engaged in the daily activities *Office meeting*, *Read a book*, *Watch TV*, and *Bicycle*. The average data reception rates (sensor-specific ratio of received data samples to total number of expected samples during the recording time) were 95.9% for the HRM, 98.8 % for the chest sensor, and 97.3 % for the thigh sensor. In addition to the daily activity diaries, all participants returned a total of 77 questionnaires (Participant 1: 12, Participant 2: 23, Participant 3: 19, Participant 4: 23).

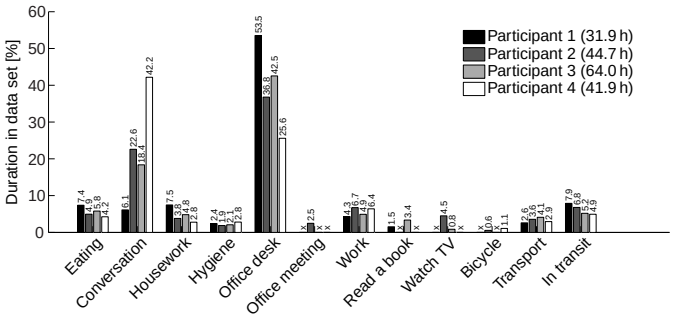

Fig. 5. Relative duration of daily activities in the data set. Daily activities that were not recorded for a participant are marked with 'x'.

5 Results

Participants followed their personal daily activities and encountered individual arousal situations. We conducted a participant-specific analysis to account for inter-individual variation in behaviour and situations to confirm applicability of our algorithms for naturalistic daily living studies.

5.1 AHR algorithm sensitivity adjustment

We investigated the participant-specific sensitivity of the AHR algorithm with respect to daily activities (cf. Section 3.3). Figure 6 shows the probability of an activation $P(A|K,R)$ with respect to daily activity R and sensitivity parameter K.

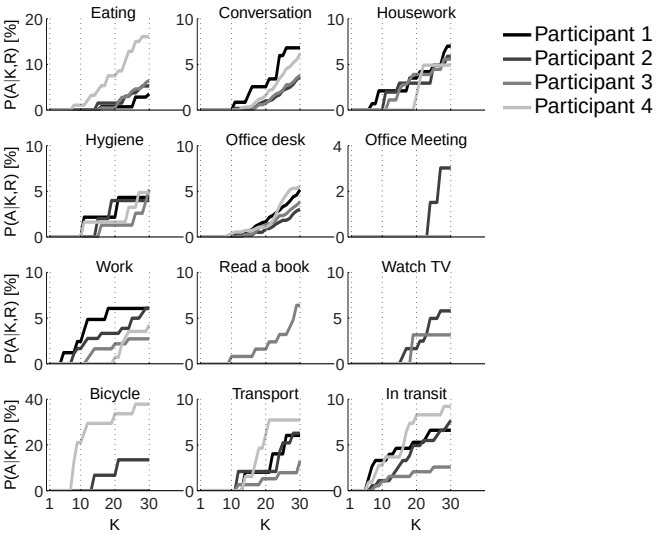

Fig. 6. Participant-specific probability $P(A|K,R)$ of an activation A with respect to daily activity R and sensitivity K.

As expected, $P(A|K,R)$ varied across daily activities and participants, suggesting that participant-specific arousal characteristics exist during individual daily activities. The maximum probability $P(A|K,R)$ was obtained for the highest sensitivity ($K = 30$), mostly

between 3 and 10%. The total maximum probability $P(A|K,R)$ of 39% was observed for *Bicycle*. Since for this sensitivity analysis we considered the AHR output without the activity-aware activations filtering, we assume that PA-induced arousal may have resulted in *Bicycle* receiving the total maximum probability $P(A|K,R)$. From the sensitivity analysis, we could not establish a rationale on how to select K specifically for each daily activity. Thus, we used the highest sensitivity (K=30) for each participant in the subsequent analyses.

5.2 Activity-aware activations filtering

We applied our activity-aware filtering procedure introduced in Section 3.4 to remove PA-influence in the AHR-based activation estimation.

5.2.1 Primitive classification

For the encoding of primitives, we trained Naïve Bayes classifiers, using a separate primitive data set of Participants 3 and 4 (see Section 3.4.1). A 10-fold cross-validation achieved a subject-independent accuracy of 99.9 % for the three activity primitives of the chest (bend front, upright, lean back), and 99.5 % for the thigh (sit, stand, and walk). The trained classifiers were subsequently applied to encode primitives in the data set.

5.2.2 Computation of differential primitive distributions

Using the encoded primitives, we generated differential primitive distributions (see Section 3.4.2). Figure 7 illustrates an example of one participant's differential primitive distribution with respect to daily activities. Positive values (bright) indicate an increase, negative values (dark) a decrease in the relative primitive distribution differences. Primitive distribution changes mainly related to the transition between bend front and upright, sit and stand, and an increase in body motion dynamics (Var.L2). In this example in Figure 7, these changes occurred during almost all daily activities

5.2.3 Clustering of differential primitive distributions

Differential primitive distributions were clustered by pattern similarity. In Figure 7, e.g., Cluster 9 and 10 reflect the transition from bend front to upright with no changes for the other primitives. Cluster 20 aggregated the most number of activations and exhibited the lowest mean primitive distribution differences. The filtering procedure automatically retained the activations of Cluster 20 for subsequent arousal phase estimation and discarded the activations assigned to the other 19 clusters.

Determining the number of clusters for filtering. Selecting a too low number of clusters d can result in underfitting, where PA-induced activations are not sufficiently separated. On the other hand, a too high number of clusters can result in overfitting, where each activation is assigned to one cluster. Theoretically, the

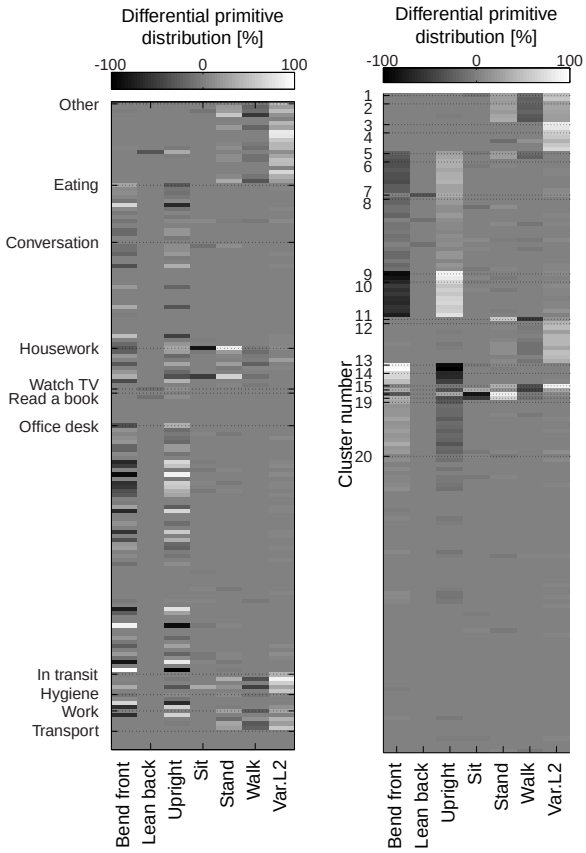

Fig. 7. Illustration of one participant's differential primitive distribution. Left: Differential primitive distribution sorted by daily activities. Right: Corresponding clustering of the differential primitive distribution using 20 clusters. Activations assigned to Cluster 20 were retained for the phase duration and intensity estimation, as these activations showed the least influences of PA changes.

number of combinations of differential primitive distribution patterns $\hat{C}^{(q)}$ can be estimated using (18), with q denoting the number of primitive domains, and N_i the number of primitives in the respective primitive domain i.

$$\hat{C}^{(q)} = \prod_{i=1}^{q} \left(1 + \sum_{k=2}^{N_i} \binom{N_i}{k} \right), \quad N_i > 1 \tag{18}$$

In (18), the product term considers the combination of primitive domains, the sum term considers the number of primitive combinations within each domain, while the case of no PA-change () is accounted for by adding one. The estimation in (18) is an underestimate of possible differential primitive distribution patterns, because the value for each primitive can vary continuously between -100 and 100% resulting in many more combinations. According to (18), there are 50 possible combinations for the configuration of primitives used in this work (two primitive domains chest and thigh with three classes each, and the primitive L2.Var: $\hat{C}^{(3)} = (1+3+1) \cdot (1+3+1) \cdot (1+1) = 50$).

However, not all possible combinations actually occurred in the data, as shown by the clustering in Figure 7, e.g., transitions involving lean back and walking. To determine the number of clusters d for filtering, we varied d between 2 to 30 and performed the filtering.

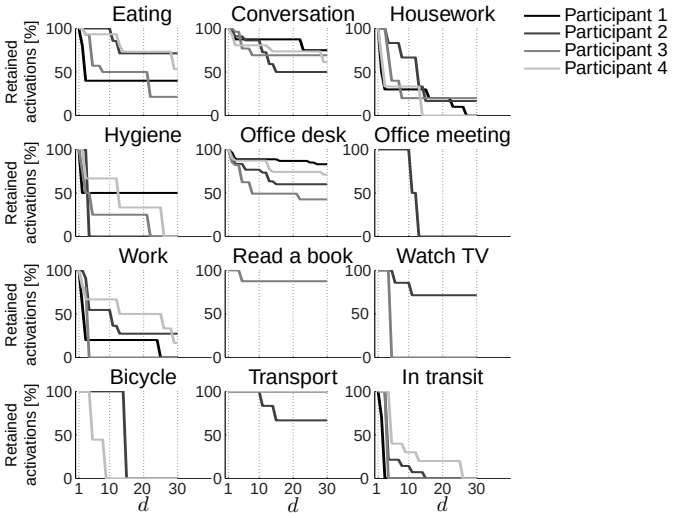

Fig. 8. Participant-specific relative number of retained activations after filtering with respect to daily activities and number of clusters d used for filtering.

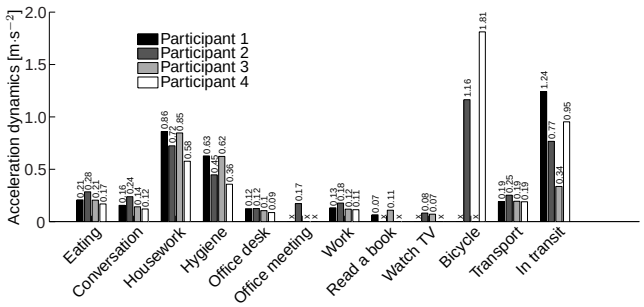

Fig. 9. Median acceleration dynamics of the chest sensor position during daily activities. Daily activities that were not recorded for a participant are marked with 'x'.

Figure 8 illustrates the relative number of retained activations after filtering with respect to daily activities and number of clusters d. The relative number of retained activations after filtering decreased with an increase in the number of clusters d. Daily activities *Housework*, *In transit*, and *Bicycle*, showed a drop in filtered activations by more than 70% for less than 20 clusters. We attribute this drop in filtered activations to an effective filtering of PA-induced activations, as these daily activities showed elevated levels of PA in comparison to the other daily activities. Figure 9 shows the measured median acceleration dynamics of the chest sensor position with respect to daily activities. Other daily activities, e.g., *Conversation* and *Office desk* showed a smaller decrease in filtered activations even when the number of clusters used for filtering was increased. Between 16 to 20 clusters, we observed no further decrease in the relative number of

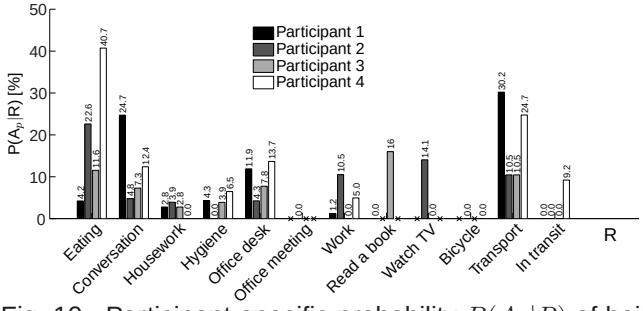

Fig. 10. Participant-specific probability $P(A_p|R)$ of being in an arousal phase A_p during daily activity R. Daily activities that were not recorded for a participant are marked with 'x'.

retained activations after filtering across all participants. We attributed the increase in the number of clusters for filtering to an increased sub-division of clusters. As changes in differential primitive distributions occurred at different magnitudes, we expected that this sub-division of clusters would be related to PA-induced activations, gradually grouping transition of similar distribution.

In subsequent analyses, we set $d = 20$ and the cluster-based filtering retained on average 51% (Participant 1: 58%, 2: 39%, 3: 45%, 4: 61%) of the original activations that were retrieved by the AHR algorithm.

5.3 Arousal duration and intensity estimation

We applied our arousal phase duration and intensity modelling procedure introduced in Section 3.5 and 3.6 to the retained activations after filtering and computed the participant-specific probability $P(A_p|R)$ according to (16). Without a firm ground truth and with expected high inter-individual variations in experienced arousal situations, we investigated estimated arousal phases with respect to participant-specific daily activities and self-report.

5.3.1 Relation to daily activities

Figure 10 shows the probability $P(A_p|R)$ of being in an arousal phase A_p during daily activity R. $P(A_p|R)$ differed across daily activities and participants. The amount of total PA did not explain the difference among participants, as PA was comparable among the participants during these daily activities (see Figure 9). For Participant 1 and 4, the probability $P(A_p|R)$ was higher during *Conversation*, *Office desk*, and *Transport* than during other daily activities. Among all daily activities and participants, *Eating* showed the highest probability $P(A_p|R) = 40.7\%$ for Participant 4. Participant 4 also had the shortest duration of *Eating* in the data set (see Figure 5). From this participant's activity diary, we observed that mealtimes varied and repeatedly coincided with *Transport* or *Office desk*, that likewise showed elevated $P(A_p|R)$. Hence, $P(A_p|R)$ might be elevated because the participant was engaged in multiple tasks

at the same time. For most of the participants, *Eating* was a sociable activity and arousal through interpersonal communication could have increased $P(A_p|R)$ during *Eating*. $P(A_p|R)$ was zero for daily activities involving high amounts of body motion, e.g., *Bicycle* and *In transit* (except Participant 4). We attributed this finding to an effective filtering of PA-induces activations.

Consistently, we estimated the highest arousal intensity of each participant in the office environment with an intensity of $I > 4$ (HR increases by at least 12 bpm during activation) and phase durations T_p of 3 to 8 min. Regarding other daily activities of all participants, arousal phase duration T_p of 2 to 5 min were more likely than longer durations and arousal phase intensity I remained below 3 (HR increases by at least 9 bpm during activation) for phase durations of more than 10 min.

The differences in estimated arousal duration and intensity suggest that different arousal stimulating situations exist in daily activities, confirming a participant- and daily activity-specific analysis. Nonetheless, we could attribute plausible reasons from daily activity information of individual behaviour to the estimated arousal phases, e.g., the multiple engagements of Participant 4 during *Eating*.

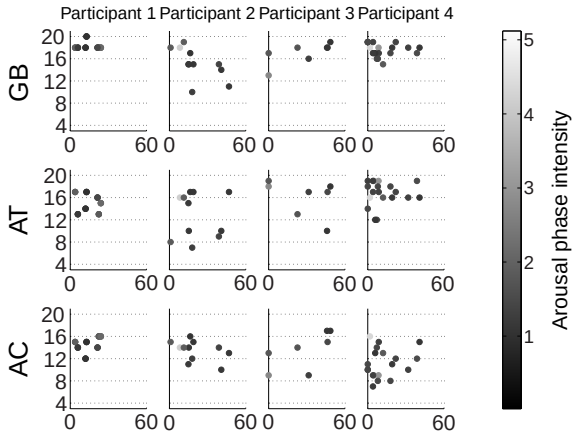

Fig. 11. Relation of questionnaire and arousal phase timing. x-axis: time difference between questionnaire event and the nearest estimated arousal phase. y-axis: questionnaire scores of the three dimensions good vs. bad mood (GB), alert vs. tired (AT), and aroused vs. calm (AC). Scores range between 4 (bad mood, tired, aroused) and 20 (good mood, awake, calm). Brightness reflects the estimated intensity of the arousal phases showing that few but high intensity phases were followed by a questionnaire within a few minutes.

5.3.2 Relation to self-report

We introduced a method to relate self-report to estimated arousal phases. Figure 11 illustrates the relation between estimated arousal phase timing, including its intensity, and questionnaire scores of the three dimensions good vs. bad mood (GB), alert vs. tired (AT), and aroused vs. calm (AC). Arousal phase timing relates to the time difference between the time a questionnaire was completed and the nearest estimated arousal phase preceding that questionnaire.

The majority of questionnaires (92.2%) were completed with a time lag of at least 1 minute. Remarkably, some questionnaires (7.8%) were completed at the time when an arousal phase was estimated. From additional comments on salient situations in the activity diaries, we could associate two estimated arousal phases to distinct events: Participant 3 had perceived arousal because of anger and Participant 4 because of happiness, both during *Office desk*. As an example, the situation of Participant 3 was known to be Office desk based on activity annotations. Increased HR while PA was limited to sitting indicated an arousal phase (estimated duration 2 min). During the arousal phase, Participant 3 completed a self-report and commented on a salient situation, clearly indicating a negatively valenced arousal situation. For other estimated arousal phases, there was no additional information available.

Perhaps, the dispersion between estimated arousal phases and questionnaires could be related to the fact that participants were adviced to also complete questionnaires randomly without any perceived arousal during the recordings could have contributed to this result. In addition, limited congruence between subjectively perceived arousal, as assessed with questionnaires, and objective measurement, in particular for natural environments, was reported earlier [22]. Our results further support this finding.

6 Discussion

The identification of arousal phases in natural daily living is a challenging problem. Besides on-body data acquisition and interpretation, arousal phase modelling is needed. This work approaches these challenges and introduces an activity-aware arousal phase modelling.

6.1 Continuous wearable sensing of HR and PA

Our multi-modal wearable system approach comprising a belt computer, miniature accelerometers, and a chest-worn HRM allowed for day-long operation without interfering with daily activities. The goal pursued in this work was to maintain a wearable monitoring system conforming to the AHR sensing modalities HR and PA. Today, HR and PA data could be obtained from commercial sensors and mobile phones. To facilitate deployment of sensor ensembles in natural settings in the future, clothing-integrated solutions could be used, e.g., [35].

In contrast to the accelerometer-based modelling of PA in the AHR algorithm, respiratory parameters measured with straps at chest and abdomen were reported to provide more accurate estimates of PA [36]. However, such respiration measurements would require additional sensors and personal calibration using a spirometer.

Complementary to HR, which is used in our approach for arousal estimation, HR variability (HRV) is often

cited to reflect psychological influences on cardiac activity. Recently, HRV has been proposed as a trigger to request self-reports on mobile phones during stressful situations [37]. However, HRV as an indicator of arousal is known to be a matter of conjecture especially in non-clinical, natural settings [22]. Moreover, HR responses to stressors, when measured in natural settings, were reported to be higher than in the laboratory setting and thus improve the estimation [38]. To maintain data alignment, we interpolated missing RR-interval samples, in total 4.1%. As a result, 30% of activations were probably affected by data interpolation. We excluded these activations from the analysis to only consider the most reliable activations.

6.2 Arousal phase estimation

Our arousal phase estimation procedure is based on the AHR algorithm. Although AHR had been used in psychological studies, a systematic evaluation and optimisation of its functional elements and parameters had not been performed. To the best of our knowledge, no arousal estimation algorithm comparable to AHR exists, which could be applied in natural daily living.

6.2.1 AHR algorithm sensitivity adjustment

AHR sensitivity influenced the activations estimation. Variations in the conditional probability $P(A|K,R)$ (see Figure 6) appeared larger between participants than between daily activities which confirmed our participant-specific analysis approach. As there was no rationale on how to select the AHR sensitivity for each individual, we used the highest sensitivity setting ($K = 30$) for all participants. As a consequence of the lacking personal calibration, differences in participants's physical fitness level could result in different HR responses for the same amount of PA and cause PA-induced activations. Nonetheless, physical fitness level could also influence the HR response to stressors [39].

6.2.2 Activity-aware activations filtering

Our results showed that AHR lacks robustness regarding PA-induced HR changes, specifically for body transitions. Activity primitives and a cluster-based filtering complemented AHR and yielded plausible arousal estimations. Our filtering approach is robust against individual differences in physical fitness. The primitives are, compared to daily activities, simple body motion patterns that require only limited amount of training data for statistical classification. Our empirically selected cluster count was smaller than the number of possible primitive combinations, as not all possible combinations occurred, e.g., lean back and walk are mutually exclusive. We surmise that not all primitive distribution patterns may elicit equal PA-induced arousal. Thus, our approach could be expanded by weighting patterns of primitive distribution differences according to the severity of PA-induced activations. As we currently lack evidence to establish a reasonable weighting scheme, we automatically selected one cluster, which aggregated activations that were the least likely related to PA-induced arousal. All other activations were discarded. In that, our current implementation may represent a rigorous filter setting to exclude PA-induced activations.

6.2.3 Arousal duration and intensity estimation.

To estimate an arousal phase's end point, a segmentation procedure was developed that evaluated the AHR signal with respect to an activation's AHR value. If PA set in only after an activation, the duration of an estimated phase might be potentially overestimated, as PA masks psychological arousal by delaying the return of HR below mean HR level. Therefore, the duration estimation could be optimised by weighting phases according to their amount of PA after activation.

Arousal phase intensity corresponds to the AHR value at the minute of activation. For nested activations, we considered the affect to be additive, as reported by literature [22]. However, we cannot completely rule out that non-linear effects might be in play, as participant-specific daily activities could have exhibited other stressors compared to those investigated in the literature.

6.3 Evaluation strategy

6.3.1 Data recording in natural environments.

Data collection in natural daily living is not a trivial task. We focussed on arousal phase modelling and exemplarily analysed applicability of algorithms for naturalistic daily living studies. As there are no references on how much data in daily life studies are most suitable, we collected and analysed 182.5 h of sensor data of four participants' natural daily living (on average 45.6 h of daily living data per participant). Using four participants allowed us to analyse interpersonal variations, limit the participants's burden to record data and annotate activities. Nonetheless, our developed arousal phase estimation model could be applied to further investigate groups of participants that share particular characteristics and, thus, allow generalising results using a larger pool of participants in the future.

Comparable investigations using multi-modal sensor systems for psychological monitoring, not bound to dedicated environments, such as the home or clinical settings are rare. For example, Healey et al. [24] collected data of 19 participants with an average of 23.7 h of data per participant, however data loss due to sensor failure was $\sim$50%. The data analysis was focusing on the temporal relation of self-reports and measured patterns. Plarre et al. [25] collected 24.8 h of data per participant of which only 33.6% was valid data and could be used to apply a two-class partitioning (stress or no-stress). The discrimination was not evaluated, e.g, using self-reports. Prill et al. [30] recorded 20.1 h on average per participant during free living of 40 participants to investigate blood pressure changes during AHR activations.

6.3.2 Relation to daily activities and self-reports.

Since an ideal reference for assessing arousal in daily living does not exist, we explored alternative techniques. We combined evaluation strategies and related estimated arousal phases to annotated daily activities, the MDBF questionnaire, and notes on salient situations. In addition, we used the AHR algorithm as basis of our work, which had been evaluated in Psychophysiology studies.

We analysed estimated arousal phases with respect to annotated daily activities that can be associated to specific social engagements and PA levels, both being known sources of arousal responses. Across daily activities and participants, duration and intensity of estimated arousal phases differed. In particular, we observed an increased probability of arousal during conversations in contrast to other daily activities. This observation conforms to literature reports on cardiac responses, especially during emotional conversations [40]. Changes in respiration pattern and, hence, speaking can modulate HR (respiratory sinus arrhythmia, RSA), and could have influenced activation estimations as well. However, Spearman correlation of $r = -0.4$ between total arousal duration during conversation and total time spent in conversations did not confirm a systematic influence of speaking on estimated arousal duration.

Questionnaires and additional notes on salient situations from participants showed some coincidences with estimated arousal phases. These coincidences could be related to the rare arousal situations that are sufficiently salient to be perceived during daily living [22]. It was suggested that while self-reports can be used in daily living, responses depend on situational circumstances and individual cognitive schemes. Our multi-modal approach fused sensory signals of HR and PA in conjunction with self-reports. Further research is required to investigate the relationship between arousal phases estimated from sensor data and its manifestation in self-report during daily living. In addition, multidimensional self-reports used in this work would allow investigating the influence of other affective dimensions on arousal, e.g., mood. Vocal input could further facilitate self-reporting in natural environments.

7 Conclusions

In this work, we explored options for arousal modelling and estimation in daily living. The arousal estimation was based on the AHR algorithm, which was complemented by activity-aware activations filtering using activity transition information. The activity transition information was obtained by encoding generic activities and body part postures in primitives followed by a cluster-based distribution modelling. With this approach we could identify and filter out PA-induced activations. Moreover, we extracted arousal phase information (duration and intensity) that has not been considered in the original AHR. We tested our procedure on natural daily living data collected using a multi-modal wearable system. Estimated arousal phases from sensor data were investigated regarding their plausibility to participant-specific daily activities and self-reports, including the MDBF questionnaire and notes on salient situations. Our filtering procedure identified PA-induced activations during daily activities, including *In transit* and *Bicycle*. Arousal was most likely during conversations or using public transport, and least likely during daily activities involving elevated levels of PA. Arousal phases with highest intensities were found during office work for all participants.

Arousal duration and intensity were weakly correlated, with a tendency towards shorter durations of 2-5 min. Participant-specific self-report coincided in some cases with estimated arousal and allowed us to relate estimated arousal to a specific contextual origin. Differences between sensor-based measurements and perception of arousal in daily living were largely confirmed. We concluded that these evaluation strategies provided complementing indications on the plausibility of our arousal phase estimation and account for the fact that no absolute reference information exists.

We consider our approach capable of providing reasonable quantification of arousal in daily living and in the presence of PA. While our approach was used to model arousal phases from HR and body acceleration, further research is required to attribute arousal phase patterns to specific arousal triggers. The analysis of arousal phase patterns, e.g., by machine learning methods, could help discerning positively and negatively valenced emotional states based on context in the future [41]. Additionally, multi-modal sensing could exploit other physiological sensors and reuse our activity-aware arousal filtering approach investigated for HR in this work, e.g., EDA with PA-filtering.

Acknowledgements

The authors thank the volunteers for their participation in the data recording.

References

[1] J. Russell, "Core affect and the psychological construction of emotion." *Psychological review*, vol. 110, no. 1, pp. 145–172, 2003.
[2] A. C. Guyton and J. E. Hall, *Textbook of medical physiology*, 11th ed. Elsevier, 2006, ch. 60, pp. 748–760.
[3] R. Picard, *Affective Wearables*, ser. Affective Computing. Cambridge, MA: The MIT press, 1997, ch. 8, pp. 227–252.
[4] D. Almeida, "Resilience and vulnerability to daily stressors assessed via diary methods," *Current Directions in Psychological Science*, vol. 14, no. 2, p. 64, 2005.
[5] P. Kohn, K. Lafreniere, and M. Gurevich, "Hassles, health, and personality," *Journal of Personality and Social Psychology*, vol. 61, no. 3, pp. 478–482, 1991.
[6] M. Kusserow, O. Amft, and G. Tröster, "Analysis of heart stress response for a public talk assistant system," in *European Conference on Ambient Intelligence*. Springer, 2008, pp. 326–342.
[7] R. S. Lazarus, *Stress and emotion: a new synthesis*, R. S. Lazarus, Ed. Springer Publishing Company, 2006.
[8] M. A. Nicolaou, H. Gunes, and M. Pantic, "Continuous Prediction of Spontaneous Affect from Multiple Cues and Modalities in Valence-Arousal Space," *IEEE Trans. on Affective Computing*, vol. 2, no. 2, pp. 92–105, 2011.

[9] M. Myrtek, G. Brügner, A. Fichtler, K. König, W. Müller, F. Foerster, and V. Höppner, "Detection of emotionally induced ECG changes and their behavioural correlates: a new method for ambulatory monitoring," *European Heart Journal*, vol. 9, no. Supplement N, pp. 55–60, 1988.
[10] C. Kirschbaum, K. Pirke, and D. Hellhammer, "The "Trier Social Stress Test" - a tool for investigating psychobiological stress responses in a laboratory setting," *Neuropsychobiology*, vol. 28, no. 1–2, pp. 76–81, 1993.
[11] D. Wu, C. G. Courtney, B. J. Lance, S. S. Narayanan, M. E. Dawson, K. S. Oie, and T. D. Parsons, "Optimal Arousal Identification and Classification for Affective Computing Using Physiological Signals: Virtual Reality Stroop Task," *IEEE Trans. on Affective Computing*, vol. 1, no. 2, pp. 109–118, 2010.
[12] R. A. Calvo and S. D'Mello, "Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications," *IEEE Trans. on Affective Computing*, vol. 1, no. 1, pp. 18–37, 2010.
[13] H. Gunes, B. Schuller, M. Pantic, and R. Cowie, "Emotion representation, analysis and synthesis in continuous space: A survey," in *International Conference on Automatic Face & Gesture Recognition and Workshops*. IEEE, 2011, pp. 827–834.
[14] J. Fahrenberg, M. Myrtek, K. Pawlik, and M. Perrez, "Ambulatory Assessment - Monitoring Behavior in Daily Life Settings," *Europ. Journal of Psychological Assessment*, vol. 23, no. 4, pp. 206–213, 2007.
[15] F. H. Wilhelm and P. Grossman, "Emotions beyond the laboratory: Theoretical fundaments, study design, and analytic strategies for advanced ambulatory assessment," *Biological Psychology*, vol. 84, no. 3, pp. 552–569, 2010.
[16] E. André, "Experimental methodology in emotion-oriented computing," *IEEE Pervasive Computing*, vol. 10, no. 3, pp. 54–56, July–September 2011.
[17] R. Fletcher, K. Dobson, M. Goodwin, H. Eydgahi, O. Wilder-Smith, D. Fernholz, Y. Kuboyama, E. Hedman, M. Poh, and R. Picard, "icalm: Wearable sensor and network architecture for wirelessly communicating and logging autonomic activity," *IEEE Transactions on Information Technology in Biomedicine*, vol. 14, no. 2, pp. 215–223, 2010.
[18] M. Kusserow, O. Amft, H. Gubelmann, and G. Tröster, "Arousal pattern analysis of an Olympic champion in ski jumping," *Sports Technology*, vol. 3, no. 3, pp. 192–203, 2010.
[19] M. Kusserow, O. Amft, and G. Tröster, "BodyANT: Miniature wireless sensors for naturalistic monitoring of daily activity," in *Proc. of the 4th Int. Conference on Body Area Networks (BodyNets)*. ICST, 2009, pp. 1–8.
[20] H. Gunes and M. Pantic, "Automatic, dimensional and continuous emotion recognition," *International Journal of Synthetic Emotions (IJSE)*, vol. 1, no. 1, pp. 68–99, 2010.
[21] S. Koelstra, A. Yazdani, M. Soleymani, C. Mühl, J.-S. Lee, A. Nijholt, T. Pun, T. Ebrahimi, and I. Patras, "Single trial classification of eeg and peripheral physiological signals for recognition of emotions induced by music videos," in *Brain Informatics*, ser. Lecture Notes in Computer Science, Y. Yao, R. Sun, T. Poggio, J. Liu, N. Zhong, and J. Huang, Eds., vol. 6334. Springer Berlin / Heidelberg, 2010, pp. 89–100.
[22] M. Myrtek, *Heart and emotion: Ambulatory monitoring studies in everyday life*, M. Myrtek, Ed. Hogrefe & Huber, 2004.
[23] N. Bolger, A. Davis, and E. Rafaeli, "Diary methods: Capturing life as it is lived." *Annual Review of Psychology*, vol. 54, no. 1, pp. 579–616, 2003.
[24] J. Healey, L. Nachman, S. Subramanian, J. Shahabdeen, and M. Morris, "Out of the Lab and into the Fray: Towards Modeling Emotion in Everyday Life," in *Proc. of the 8th Int. Conference on Pervasive Computing*, ser. Lecture Notes in Computer Science, vol. 6030. Springer, 2010, pp. 156–173.
[25] K. Plarre, A. Raij, S. Hossain, A. Ali, M. Nakajima, M. al'Absi, E. Ertin, T. Kamarck, S. Kumar, M. Scott, D. Siewiorek, A. Smailagic, and L. Wittmers, "Continuous inference of psychological stress from sensory measurements collected in the natural environment," in *Proc. of the 10th Int. Conference on Information Processing in Sensor Networks (IPSN)*. IEEE, 2011, pp. 97–108.
[26] M. Myrtek, F. Foerster, and G. Brügner, *Freiburger Monitoring System*, ser. Psychophysiologie in Labor und Feld, M. Myrtek, Ed. Peter Lang, 2001, vol. 9, in German.
[27] P. Grossman, F. Wilhelm, and M. Spoerle, "Respiratory sinus arrhythmia, cardiac vagal control, and daily activity," *American Journal of Physiology-Heart and Circulatory Physiology*, vol. 287, no. 2, pp. H728–H734, 2004.
[28] J. Turner, D. Carroll, J. Hanson, and J. Sims, "A comparison of additional heart rates during active psychological challenge calculated from upper body and lower body dynamic exercise," *Psychophysiology*, vol. 25, no. 2, pp. 209–216, 1988.
[29] A. S. Blix, S. B. Stromme, and H. Ursin, "Additional heart rate–an indicator of psychological activation." *Aerospace Medicine*, vol. 45, no. 11, pp. 1219–1221, 1974.
[30] T. Prill and J. Fahrenberg, "New methods in ambulatory blood pressure monitoring: Interactive monitoring and detection of posture and movement patterns," *Behavior Research Methods*, vol. 39, no. 3, pp. 390–398, 2007.
[31] R. Kleiger, J. Miller, J. Bigger Jr, and A. Moss, "Decreased heart rate variability and its association with increased mortality after acute myocardial infarction," *The American Journal of Cardiology*, vol. 59, no. 4, pp. 256–262, 1987.
[32] R. Steyer, P. Schwenkmezger, P. Notz, and M. Eid, *Der Mehrdimensionale Befindlichkeitsfragebogen (MDBF) [Multidimensional Mood Questionnaire]*, Hogrefe, Germany, 1997.
[33] O. Amft, M. Lauffer, S. Ossevoort, F. Macaluso, P. Lukowicz, and G. Tröster, "Design of the QBIC Wearable Computing Platform," in *Proc. of the 15th Int. Conference on Application-Specific Systems, Architectures and Processors (ASAP)*, 2004, pp. 398–410.
[34] D. Bannach, O. Amft, and P. Lukowicz, "Rapid prototyping of activity recognition applications." *IEEE Pervasive Computing*, vol. 7, no. 2, pp. 22–31, 2008.
[35] C. Zysset, K. Cherenack, T. Kinkeldei, and G. Tröster, "Weaving integrated circuits into textiles," in *Proc. of the 14th Int. Symposium on Wearable Computers (ISWC)*, 2010, pp. 1–8.
[36] F. Wilhelm and W. Roth, "Using minute ventilation for ambulatory estimation of additional heart rate," *Biological Psychology*, vol. 49, no. 1-2, pp. 137–150, 1998.
[37] M. Morris and F. Guilak, "Mobile Heart Health: Project Highlight," *IEEE Pervasive Computing*, vol. 8, no. 2, pp. 57–61, 2009.
[38] Y. Zanstra and D. Johnston, "Cardiovascular reactivity in real life settings: Measurement, mechanisms and meaning," *Biological Psychology*, vol. 86, no. 2, pp. 98–105, 2011.
[39] L. J. van Doornen and E. J. de Geus, "Aerobic fitness and the cardiovascular response to stress," *Psychophysiology*, vol. 26, no. 1, pp. 17–28, 1989.
[40] W. Linden, "A microanalysis of autonomic activity during human speech," *Psychosomatic Medicine*, vol. 49, no. 6, pp. 562–578, 1987.
[41] M. Kusserow, O. Amft, and G. Tröster, "Monitoring stress-arousal in the wild - measuring the intangible," *IEEE Pervasive Computing Magazine*, 2012, in press.

Martin Kusserow is a PhD candidate of the Wearable Computing Lab at ETH Zurich, Switzerland. His research interests include wearable computing, physiological monitoring, sports, and real-world sensing. Kusserow has an MSc in information technology and electrical engineering from ETH Zurich. He's a member of the IEEE. Contact him at kusserow@ife.ee.ethz.ch.

Oliver Amft is an assistant professor at TU Eindhoven, The Netherlands, and a senior research advisor at ETH Zurich's Wearable Computing Lab. His research interests include ubiquitous sensing, embedded systems, and pattern recognition for activity, behaviour, and context-awareness. Amft has an MSc from Chemnitz University of Technology, and a PhD from ETH Zurich. He's a member of the IEEE. Contact him at amft@ieee.org.

Gerhard Tröster is a professor and head of the Wearable Computing Lab at ETH Zurich, Switzerland. His research interests include wearable computing for healthcare, and production, smart textiles, sensor networks, and electronic packaging. Tröster has PhD in electrical engineering from the Technical University Darmstadt. He's a senior member of the IEEE. Contact him at troester@ife.ee.ethz.ch.