How to Use Smartphones for Less Obtrusive Ambulatory Mood Assessment and Mood Recognition

Anja Bachmann  
Pervasive Computing Systems  
KIT, Karlsruhe, Germany  
bachmann@teco.edu

Michael Beigl  
Pervasive Computing Systems  
KIT, Karlsruhe, Germany  
michael@teco.edu

Christoph Klebsattel  
Pervasive Computing Systems  
KIT, Karlsruhe, Germany  
klebsattel@teco.edu

Markus Reichert  
Applied Psychology  
KIT, Karlsruhe, Germany  
markus.reichert@partner.kit.edu

Matthias Budde  
Pervasive Computing Systems  
KIT, Karlsruhe, Germany  
budde@teco.edu

Philip Santangelo  
Applied Psychology  
KIT, Karlsruhe, Germany  
philip.santangelo@kit.edu

Till Riedel  
Pervasive Computing Systems  
KIT, Karlsruhe, Germany  
riedel@teco.edu

Ulrich Ebner-Priemer  
Applied Psychology  
KIT, Karlsruhe, Germany  
ulrich.ebner-priemer@kit.edu

Abstract
We present MoA^2, a context-aware smartphone app for the ambulatory assessment of mood, tiredness and stress level. In principle, it has two features: (1) mood assessment and (2) mood recognition. The mood assessment system combines benefits of state of the art approaches. The mood recognition is concluded by smartphone-based wearable sensing. In a formative study, we evaluated the usability and unobtrusiveness of our mood assessment. A median SUS score of 90 shows a high usability. Subjects reported an easy, fast and intuitive use. The mood recognition was evaluated in terms of classification accuracy. First, we analyzed which features are best for the recognition. Spatio-temporal attributes, i.e. daytime, day of week and location, correlate most with the monitored mood. Based on the identified attributes, we trained personalized classifiers using Naïve Bayes and applied ten-fold-cross validation. The average recognition accuracy was 0.76 which is comparable to related work.

Author Keywords
Mobile ambulatory assessment; mood sensing; mood recognition; smartphone-based wearable sensing

ACM Classification Keywords
H.1.2 [User/Machine Systems]; I.5.m [Pattern Recognition]; J.4 [Social and Behavioral Sciences]
Introduction
Moods are relevant to all humans as they provide the affective background, the emotional color to all that we do [3]. Whereas moods are not necessarily perceived consciously all the time, they still influence our experiences, thoughts, and behavior [18] – in healthy subjects as well as in psychiatric patients. Especially in patients suffering from psychiatric disorders, moods are a matter of concern: emotional dysregulation is a central characteristic of several disorders and thus constitutes a transdiagnostic symptom [19]. In a theory based on clinical observations there is a direct association between mood disturbances and problematic dysfunctional behaviors. For this reason, clinical psychologists and clinical researchers are interested in monitoring moods in order to (1) further foster the understanding of emotional dysregulation, (2) clarifying the affective states which lead up to problematic behaviors, and (3) to intervene in those instances when patients are prone to problematic behaviors.

Up to now, the best applicable way in psychology and psychiatry is to use ambulatory assessment. It refers to the use of computer-assisted methodology for self-reports, behavior records, or physiological measurements, while the participant undergoes normal daily activities [5]. A very desirable option of ambulatory assessment is to take into account environmental factors and context information such as current events and situations or interactions with other people to trigger assessment prompts or to provide additional information to the supervising psychologist. Though, it puts a burden on patients to provide all of these information via questionnaires. Several articles advocate the use of sensor technology for continuous, unobtrusive and objective assessments of variables of interest, three desired beneficial properties in psychological and psychiatric research that are never achieved using self-report measures [13].

We want to take a first step into that direction with MoA²: a smartphone app for Mobile Ambulatory Mood Assessment. The assessment happens in a continuous and unobtrusive way on the subjects’ own smartphones. In addition to mood assessment via a short questionnaire, we also access different sensor sources such as microphone and data sources such as call history and physical activity to gain a broader view of the subjects’ contexts. Such environmental factors and contexts of the subject can, later on, be correlated to the reported moods.

The objective of our approach is to enable a user-friendly and fast mood assessment that yields a basis for a mood recognition system. Therefore, we review established sensor sources and select promising candidates. In a formative user study, we evaluate the usability of our app and gather training data. This data is analysed in terms of correlations between features and mood. Finally, we investigate if our system can serve as a basis for a mood recognition system based. Therefore, we evaluate a basic classifier on the study data in terms of recognition accuracy.

Background
Ambulatory Assessment
Ambulatory assessment is a means to investigate psychological issues, e.g. affective states of patients, in everyday life. To date, clinical interviews and questionnaires are the most commonly used instruments to collect information on patients’ emotional suffering, associated problematic behavior and dysfunctional thoughts [4]. However, there is a large body of research raising doubts about the validity of such retrospective self-reports.
To reduce retrospective biases ambulatory assessment approaches repeatedly monitor subjects and gather reports digitally on a frequent basis. Modern digital devices such as smartphones facilitate the "interactive assessment of subjective, physiological and behavioral data in real-time" [9]. If the subjects use their mobile device then the burden to carry additional devices is avoided and the assessment can be included into everyday life more easily. In addition, mobile devices allow for collecting additional contextual user information. Such information would be beneficial for psychologists since it applies for therapeutic use. Co-appearance of mood and context can be analyzed and triggers for mood changes can be identified more easily. Alternatively, a mood recognition system can be realized based on collected data.

Related Work

In the following, we review state of the art mood and emotion sensing apps and projects in terms of sensor sources and mood assessment.

One mood assessment projects is MoodSense [11], formerly known as MoodScope. This system gathers the number of interactions with the 10 most prominent contacts for each call, SMS and email; the 10 most frequently used (category of) apps and the usage duration; the 10 most visited website domains; the 10 most visited locations and items concerning valence and arousal. This app runs on the subject’s own phones. It only collects data, no mood recognition takes place.

The project StudentLife investigated the impact of means such as mood on educational performance [17]. They collected smartphone data from accelerometer, microphone, light sensor, GPS and Bluetooth and collected self-reports from the students that assessed mood, among others. The subjects were free to use their own smartphones or use a loaned Google Nexus 4.

A huge project in the area of mood sensing is MONARCA [1]. The researches investigated location parameters, wifi, Bluetooth, accelerometer data and voice (smartphone sensors) in addition to data collected from a wrist-worn activity sensor and a physiological sock sensor that gathers GSR and pulse. Mood was not assessed by subjects’ self-reports but by analysing their medical records and by using psychiatric rating scales and is, later on, used as a basis for mood recognition. The project members provided subjects with smartphones for the duration of the study.

One emotion sensing app is Emotion Sense [15]. It gathers the affective state by classifying speech and considers smartphone-based features such as daytime, location, SMS patterns, call patterns, accelerometer data, screen activity, microphone.

Reactions is a twitter-based mobile emotion sensing app [8]. The app tracks usage patterns such as touch gestures and keystroke dynamics and collects sensor data from the accelerometer and gyroscope sensor. Another twitter-related app is the affective twitter client ("AT client") [10]. Ambient light sensor, daytime, weather, location, keystroke dynamics and the "Discomfort Index" (also: "Temperature–Humidity Index") [2] are considered.

Grünerbl et al. focus on recognizing bipolar state changes using smartphone data [7]. They handed out test devices to their subjects and analysed phone call and sound data in addition to GPS and accelerometer as location and activity information, respectively.

1[http://monarca-project.eu/](http://monarca-project.eu/)
Based on Related Work

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Location</td>
<td>Cell ID (CID); Location Area Code (LAC)</td>
</tr>
<tr>
<td>Current app</td>
<td>App running in the foreground; empty if screen</td>
</tr>
<tr>
<td>Microphone</td>
<td>Max. absolute amplitude</td>
</tr>
<tr>
<td>Message history</td>
<td>Unique caller ID, folder (inbox, sent),</td>
</tr>
<tr>
<td>Call history</td>
<td>Unique caller ID, type (incoming, outgoing,</td>
</tr>
<tr>
<td>Ambient light</td>
<td>Light level in Lux</td>
</tr>
</tbody>
</table>

In Addition

<table>
<thead>
<tr>
<th>Sensor</th>
<th>Features</th>
</tr>
</thead>
<tbody>
<tr>
<td>Connectivity type</td>
<td>Wifi, mobile or none</td>
</tr>
<tr>
<td>Calendar entries</td>
<td>ID of current entry; calendar name</td>
</tr>
<tr>
<td>Activity type</td>
<td>The user’s physical activity</td>
</tr>
</tbody>
</table>

There is also one project that does not provide one single app, but a framework to build experience sampling apps to assess user information: the movisensXS\(^3\) system. In our context, such an app might sample the affective state of a user. So far, those apps can log, but not recognize mood. A movisensXS app is able to access the smartphone’s GPS signal and receives additional measurements such as accelerometer or barometer data from an external sensor. Such sensor data can be used to define location or activity-based event-triggers.

\(\text{MoA}^2\) combines benefits of these approaches. We select sensor sources that were prominent in different related works to cover a variety of contexts of the subject. Our app is designed to run, as some but few presented apps, on the subjects’ own smartphone and exclude additional devices as they could infer a burden. To our best knowledge, none of the related approaches has an event-based triggering that depends on smartphone data other than location or activity. We apply not only time-based, but also event-based triggering of self-report prompts. This context-awareness of our app is a crucial difference to related work and a first step towards a better understanding of subject’s daily living.

\(\text{MOA}^2\): Analysis and Design

\(\text{MoA}^2\) is an Android app for mood assessment that is designed primarily for experts in health care scenarios, e.g. for monitoring patients. It is intended to be run on the patients own smartphone to reduce the burden of carrying additional devices and getting to know a new device as suggested by Wang et al. [17].

Data Assessment for Mood Recognition: Analysis & Design

Mobile devices offer a wide range of sensors to choose from. Based on related work we decided to use the sensor sources and features depicted in Table 1.

We also introduce new sensor sources. The connectivity type can reveal where a user is, e.g. at home or work with wifi connection or jogging in the woods without any connection. Calendar entries can reveal how many appointments a person has and if they are of private or business nature. The current physical activity\(^4\) of the user can contain relevant information as well.

We sample all information at 1Hz. If there is no data available, e.g. because there is no telephone call at the moment, the entry remains empty. We use cell ID instead of GPS as location information for enhanced privacy protection. Related work suggest to hash personal data [17] to ensure data protection. In \(\text{MoA}^2\), both unique caller ID and location are hashed using the SHA-1 algorithm\(^5\). So far, we are only interested in finding patterns and hence do not need to know the exact person or location. In addition, data is only stored locally.

Mood Assessment: Analysis and Design

In addition to the before-mentioned data sources, the subjects have to to provide their current affective state via self-reports. Schimmack et al. argue that a 3-dimensional description of affect is necessary [16]. Based on former studies conducted by co-authors, we decided to assess the subjects mood, tiredness and stressfulness.

\(\text{4Based on the Google Activity Recognition API (https://developer.android.com/reference/com/google/android/gms/location/DetectedActivity.html); one out of “on foot”, “in vehicle”, “on bycicle”, “running”, “still”, “tilting”, “walking” and “unknown”.

\(\text{5Based on the NIST SHA-1 algorithm (http://csrc.nist.gov/publications/fips/fips180-4/fips-180-4.pdf)\).}

Table 1: Selected sensor sources and derived features.

\(^3\)http://www.movisens.com/en/products/movisensxs/
In our app, the subject has to select one out of five values (cf. MoodSense) for each of the following dimensions:

- **Mood**: How do you feel? (horrible, bad, neutral, good, great)
- **Tiredness**: Do you feel tired? (tired, exhausted, neutral, active, awake)
- **Stress Level**: Do you feel stressed? (stressed, strained, neutral, calm, relaxed)

For each dimension, we offer two positive and two negative options and included a neutral option [8, 10]. We decided to use Likert scales as they are a common method in psychology. Figure 1 shows MoA’s user interface.

It is important to find a suitable prompting type and sampling frequency. Fischer and To give a good overview of different prompting types: random, time-triggered and event-triggered [6]. We intended to use the mood within a specified time interval and hence chose to have **time-triggered** prompts. A notification which reminds the subject to provide their information appears at every full hour between 9 a.m. and 8 p.m. (i.e. 12 prompts).

In addition, we wanted to assess mood after certain events occurred and therefore also chose to have **event-triggered** prompts. The app sends prompts via notifications whenever the subject (a) sends or receives a message or (b) finishes a call, (c) a calendar entry starts or ends or (d) the device loses the connection to a wifi or mobile network or re-establishes the connection. To avoid too many notifications, we added restrictions for (d). They can only trigger a notification once per five minutes and only if the screen is unlocked. If the subject does not react to any MoA notification within five minutes, the notification will be deleted and re-sent to trigger another signal (vibration, beep or LED illumination).

**Formative User Study**

We conducted a formative study to assess the usability of our system and to receive qualitative feedback.

**Study Design**

We recruited nine subjects aged between 21 and 27, four of them female. One third of them already had experience with a mood assessment system. Three of them possessed non-Android phones and used a Google Nexus 4 phone that we provided. This allows us to assess how well non-Android users get used to our app. Two of them used their own SIM card with the loaned device and installed the apps they needed so that they were able to use the device like their own. One subject rejected to use an own SIM card and preferred to carry two devices.

The subjects participated voluntarily and were not paid. They received a summary of their mood and an overview of correlated features as a report.

The study took place from Thursday noon to Monday noon, i.e. including two weekdays and two weekend days in total.

At the first meeting, we explained the background of the study (5 min) and the data that is collected (3 min) and we installed the app on the subject’s mobile device (5 min). During the study, the subjects were supposed to use their smartphones normally but in addition run our app whenever a notification was sent or whenever they wanted to share their mood voluntarily.

At the end of the study, we handed out a questionnaire.

First, we collected demographic data. Second, we applied the **standardized System Usability Scale (SUS)** [2] to capture the usability. Third, we asked the participants to state how comfortable the usage of MoA was in terms of...
We counted the answers of the comfortableness questions. The results are visualized as bar graphs with confidence intervals in Figure 3. The high values indicate that MoA$^2$ was perceived convenient, easy and fast to be answered.

**Figure 2:** Feedback on the system’s usability obtained using the SUS.

(1) duration of mood assessment, (2) complexity of mood assessment and (3) overall convenience by rating each of from 1 (uncomfortable) to 5 (comfortable). Last, we assessed the user experience qualitatively by interview questions about the benefits and drawbacks of our application and general remarks.

We noticed that there was barely any call or SMS data. Hence, we asked the subjects to name telephoning and messaging services they use alternatively.

In the app itself, we logged the task completion time, i.e. how long it took the subjects to complete a report.

**Study Results**

We calculated the SUS for each subject and visualized the results in form of a boxplot (see Figure 2). Our approach gained a median score of 90 and can hence be considered excellent [1].

There is only one subject with a SUS score below 80. This is the subject who used a loaned smartphone, rejected to use their own SIM card and hence carried two devices. We assume that the usability of MoA$^2$ was lower for this subjects as there was only a minor benefit (expected incentive) and an additional burden. This confirms our assumption that letting the users use their own devices enhances compliance and usability.

In addition, we asked the subjects for general feedback.

No one had the impression that there were to many prompts in total. Even though 3 mentioned that at some point in time there were two prompts nearby which confused them. This behavior of the app was caused by a combination of time-triggered prompt (full hour) and event-triggered prompt (SMS). The results indicate that the sampling strategy is good, but we have to implement restrictions, e.g. a cooldown time, for every event-trigger and not only for changes of the connectivity status.

8 out of 9 (89%) were positive about how simple, fast and intuitive the mood sampling was. Only the subject who used a loaned phone without own SIM did not perceive the app as useful. She also stated that her smartphone behavior changed drastically which was caused by carrying and using an additional device. We conclude that our app is very user-friendly and convenient and should be run on a subject’s own smartphone.
Regarding alternative telephoning and messaging services, the most common apps were Skype (3 mentions out of 5 answers) for telephoning and Whatsapp (7 mentions out of 9 answers) and Facebook Messenger (3 mentions out of 9 answers) for messaging.

The investigations regarding the task completion times show that on average it took the subjects 8.78s (±1.30s) to open the app, select values for mood, tiredness and stress level and submit the information. This time is fairly low and is in line with the subject’s feedback regarding the fast and easy self-reporting.

We investigated the number of time and event-triggers for self-reports and voluntary self-reports. The share of reasons for self-reports vary from subject to subject and is summarized in Table 2. Most reports were time-triggered, but there is a non-negligible number of event-triggers and voluntary reports. We conclude that event-triggers are of importance.

### Mood Recognition: Correlation Analysis

We analysed potential correlations between mood and all other features of the data collected during the user study.

First, we calculated correlation matrices for the three affective dimensions mood, tiredness and stress level.

In a next step, we preprocessed the data by replacing the timestamp by the weekday (day of the week, e.g. Monday, Tuesday), the daytime (e.g. noon, evening, night) and the daytype (type of day, weekday vs weekend) and by factorizing the lux values of the light level into categories (e.g. pitch black, dark indoors, dim outdoors, direct sunlight).

We calculated correlation matrices for each subject and the three affective dimensions and using the Pearson correlation coefficient. Table 3 gives an overview of the correlations between mood and the other dimensions for each subjects and aggregated over all participants.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Mood x Tiredness</th>
<th>Mood x Stress Level</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>−0.21</td>
<td>−0.49</td>
</tr>
<tr>
<td>2</td>
<td>−0.38</td>
<td>−0.51</td>
</tr>
<tr>
<td>3</td>
<td>−0.49</td>
<td>−0.69</td>
</tr>
<tr>
<td>4</td>
<td>−0.12</td>
<td>−0.15</td>
</tr>
<tr>
<td>5</td>
<td>−0.47</td>
<td>−0.37</td>
</tr>
<tr>
<td>6</td>
<td>−0.35</td>
<td>−0.39</td>
</tr>
<tr>
<td>7</td>
<td>−0.30</td>
<td>−0.50</td>
</tr>
<tr>
<td>8</td>
<td>−0.31</td>
<td>−0.42</td>
</tr>
<tr>
<td>9</td>
<td>−0.26</td>
<td>−0.34</td>
</tr>
<tr>
<td>Avg.</td>
<td>−0.32</td>
<td>−0.43</td>
</tr>
</tbody>
</table>

Table 3: Overview of Pearson correlation factors for mood x tiredness and mood x stress level.

Afterwards, we investigated correlations between mood and each feature for every subject by:

1. Building pruned J48 Decision Trees (DT)
2. Calculating the Information Gain (IG)
3. Generating Association Rules (AR) using the Apriori algorithm
4. Building Random Forests (RF)
5. Applying Recursive Feature Selection (RFS)

### Results

We calculated correlation matrices for each subject and the three affective dimensions and using the Pearson correlation coefficient. Table 3 gives an overview of the correlations between mood and the other dimensions for each subjects and aggregated over all participants.

The correlation values are mostly low and hence do not have a high expressiveness. Though, as expected, all values are negative, indicating that high tiredness or a high stress level relates to a low mood.
In addition, we investigated correlations between mood and the smartphone-based features. The results of our analyses are summarized in Table 4. In the table we provided the number of subjects for which features:

- **DT**: yielded an accuracy higher than or equal to the accuracy of guessing the classes
- **IG**: gained an information gain higher than 0.2
- **AR**: were antecedent of a rule with mood as consequent and had a confidence higher than 0.7
- **RF**: resulted in a mean decrease on gini index of more than 3.0
- **RFS**: were selected by the algorithm based on their high accuracy

Features that were relevant for at least 7 subjects in at least one method are printed in bold to emphasize their importance.

All five investigated algorithms showed that the spatial feature location and the temporal features daytime and type of day are highly correlated with mood. This confirms the results of Lee et al. who identified location as an important feature [10]. The correlation to temporal features is also natural and was expected as each subject has an individual biorhythm. The benefit of this information is that we can access it automatically without asking the user about it.

The results of RF and RFS indicate that the activity type indeed has a correlation for some subjects as well. Unfortunately, the Google Activity Recognition API often did not provide any data so that the entries remained empty. We suggest to build an own accelerometer or GPS-based activity recognition system that provides recognitions with higher frequency.

Other features showed correlations for some subjects only.

**Mood Recognition**

We investigated whether an accurate, MoA²-based mood recognition system can be built that is competitive with state to the art approaches.

To get a first glance at the expressiveness of the gathered data we used a Naïve Bayes algorithm and ten-fold cross validation (see Table 5).

We reviewed state of the art recognition accuracies to gain a baseline for a mood recognition system. Unfortunately, there are only very few approaches that evaluated the recognition accuracy and published the results. MoodSense is one of them. Their system achieves an initial accuracy of 0.61 using a generalized classifier and up to 0.91 for personalized classifiers for a training period of three weeks, both gathered using four-fold cross validation [11]. A stress recognition approach by Muaremi et al. achieves 0.59 using smartphone data and 0.61 using

<table>
<thead>
<tr>
<th>Features</th>
<th>Analysis Method</th>
</tr>
</thead>
<tbody>
<tr>
<td>Activity Type</td>
<td>DT IG AR RF RFS</td>
</tr>
<tr>
<td>Calender Entries</td>
<td>1 0 3 6 4</td>
</tr>
<tr>
<td>Call History</td>
<td>3 2 1 4 2</td>
</tr>
<tr>
<td>Connectivity Type</td>
<td>2 0 3 0 2</td>
</tr>
<tr>
<td>CurrentApp</td>
<td>2 0 0 2 4</td>
</tr>
<tr>
<td>Daytime</td>
<td>3 4 5 9 7</td>
</tr>
<tr>
<td>Daytype</td>
<td>2 0 3 0 3</td>
</tr>
<tr>
<td>Light Level</td>
<td>4 1 1 3 3</td>
</tr>
<tr>
<td>Location</td>
<td>3 8 1 8 2</td>
</tr>
<tr>
<td>Message History</td>
<td>2 0 0 0 0</td>
</tr>
<tr>
<td>Noise Amplitude</td>
<td>2 0 0 0 0</td>
</tr>
<tr>
<td>Weekday</td>
<td>2 6 4 8 7</td>
</tr>
</tbody>
</table>

Table 4: Number of subjects for which the given feature showed correlations with mood using the given algorithm.
smartphone data and heart rate variability [14]. Referring to their recognition results, we infer a baseline of 0.61. The mean accuracy of our approach is 0.76 (±0.16). Hence, our approach exceeds the baseline and can keep up with state of the art approaches.

**Conclusion**

We have introduced our Mobile Ambulatory Mood Assessment app MoA². The app combines benefits of state of the art approaches in terms of sensor sources and mood models and is designed primarily for experts in health care scenarios.

In a formative user study, we evaluated the usability of our app and gathered training data. Study results showed a very high usability and minor obtrusiveness. Compared to a former assessment method, subjects perceived MoA² as more convenient, much less complex and much faster to be answered. 89% of the subjects liked how simple, fast and intuitive the mood sampling was. A SUS median score of 90 confirms high usability.

Most self-reports during the study were time-triggered (0.48%), though event-triggers (0.3%) and voluntary self-reports (0.22%) also have a high share. Event-triggers shall not be neglected and should be investigated further.

The data gathered in the user study was analysed in terms of correlations between features and mood using common machine learning algorithms. The results reveal that spatio-temporal features correlated with the stated mood the most. Striking features are daytime, day of week and location. This implies that subjects are influenced by their daily schedule and biorhythm.

Finally, we build a preliminary mood recognition system based on the study data. We trained personalized classifiers using Naïve Bayes and applied ten-fold-cross validation. The average recognition accuracy was 0.76 which is comparable to related work. We suggest to test other algorithms and to automatically select the best features and classifiers, e.g. using cross validation on known data, before recognizing the current mood.

Based on the subjects answers, we will consider using alternative data sources such as WhatsApp, Facebook and Skype in future experiments. We intend to improve the event-triggered sampling by refining the rules and by limiting the number of samples per quarter of an hour. In addition, we plan a user study with a higher number of subjects and with subjects with severe psychological disorders such as BPD or depression. We will further investigate correlations and patterns among mood and context information to build a system to recognize the subject’s mood automatically.

**REFERENCES**


11. Robert LiKamWa, Yunxin Liu, Nicholas D Lane, and Lin Zhong. 2011. Can your smartphone infer your mood?. In *PhoneSense workshop*.


