An Affect Detection Technique using Mobile Commodity Sensors in the Wild

Aske Mottelson & Kasper Hornbæk
Department of Computer Science, University of Copenhagen
Njalsgade 128, DK-2300 Copenhagen, Denmark
{amot, kash}@di.ku.dk

ABSTRACT
Current techniques to computationally detect human affect often depend on specialized hardware, work only in laboratory settings, or require substantial individual training. We use sensors in commodity smartphones to estimate affect in the wild with no training time based on a link between affect and movement. The first experiment had 55 participants do touch interactions after exposure to positive or neutral emotion-eliciting films; negative affect resulted in faster but less precise interactions, in addition to differences in rotation and acceleration. Using off-the-shelf machine learning algorithms we report 89.1% accuracy in binary affective classification, grouping participants by their self-assessments. A follow up experiment validated findings from the first experiment; the experiment collected naturally occurring affect of 127 participants, who again did touch interactions. Results demonstrate that affect has direct behavioral effect on mobile interaction and that affect detection using common smartphone sensors is feasible.

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

Author Keywords
Affective computing; affect detection; smartphone; touch; crowdsourcing

INTRODUCTION
Affect influences cognitive abilities and motor skill; it also influences human-human and human-computer interaction. As a result thereof, the research field of how computers can assess and respond to human affect has grown. Picard [29] popularized this research field of *Affective Computing*, and since then numerous systems that detect and respond to affect have been proposed (e.g., [6, 12, 23, 26, 34]).

Contemporary techniques for detecting human affect, however, have several limitations. Often these techniques are verified in laboratory experiments with few participants equipped with costly hardware. Another approach has been to study participants in office-settings over long periods of time, resulting in techniques that require long individual training to function.

UbiComp '16, September 12-16, 2016, Heidelberg, Germany
©2016 ACM. ISBN 978-1-4503-4461-6/16/09…$15.00
DOI: http://dx.doi.org/10.1145/2971648.2971654

This paper departs from findings in experimental psychology that provide evidence for a link between affect and movement; Coombes et al. [10] for instance found that exposure to unpleasant images caused greater error and faster performance in a subsequent square-tracing task. We use these findings to present an affect detection technique inspired by emotion psychology theory using commodity sensors, that works in the wild, without per-user training. We therefore address the following limitations in current affect detection techniques:

Specialized Hardware: Several techniques for inferring affect have been proposed, including audio/video approaches, or using specialized hardware, for instance heart rate variability or galvanic skin response (e.g., [27, 34, 36]). These techniques are often either intrusive to user privacy or less suitable for widespread adoption because of the need to acquire and wear custom sensors. We propose an affect detection technique using less invasive measurement methods, namely sensors already present in most commercial smartphones.

Controlled Laboratory Experiments: Previous studies concerning human affect and computer interaction have mostly conducted experiments using artificial tasks in controlled laboratory settings with relatively few participants (e.g., [4, 10, 18, 23]). The external validity of these studies makes it difficult to reason about how effective the proposed techniques are in more real-life settings. We present a crowdsourced method of gathering touch interactions and affective assessments, thus increasing generalizability.

Extensive Individual Training: Previous studies have used commodity hardware sensors to detect affect, such as using keystroke dynamics [14] or smartphone usage [26, 30]. However, these studies conducted extensive experiments lasting weeks to months, resulting in idiosyncratic models that require substantial per-user training to estimate affect effectively. We present an approach that requires 140 seconds of user interaction to assess affect, without any previous training with data from that user.

We report findings from two experiments, where participants recruited through crowdsourcing conducted general touch tasks on their own devices after being emotionally primed using video clips. The results show that the affective impact on touch interaction corroborates psycho-motor theory: Speed and precision of motor control varies with affective states. Using participants' touch data it was possible to model affect using 140 seconds of smartphone sensor data, with 89.1% accuracy for binary (high/low) self-assessed affect, and 1.33 RMSE on a 1-7 positive-negative scale. An additional study using participants' natural occurring affect showed a similar effect, although with less confidence; it was possible to detect binary affect with 69.0% accuracy (1.32 RMSE, 1-7 positive-negative), binary valence with 81.7% (1.61 RMSE, 1-9 SAM), and binary arousal with 67.5% (1.88 RMSE, 1-9 SAM).

BACKGROUND AND RELATED WORK

Research on how emotions influence physical expression has been treated extensively, starting with Darwin's work in the $19th$ century [11]. Darwin proposed that emotions are products of evolution; discrete emotions trigger actions that have been favorable to survival [11, 25]. This widely supported view suggests that emotions are organized around a motivational base such that our state of mind motivates beneficial physical expression. For instance, when a negative or threatening situation occurs, a fast reaction with less emphasis on precision optimizes chances of survival.

A multitude of emotional modalities and their respective physiological responses have been studied. It has been shown that moods influence cognitive performance, general health and well-being, creativity, decision-making processes, and social relationships [5, 20]. Most commonly studied is the relation between emotions and facial expressions [5, 12], but studies have also shown that affect has a significant impact on both motor skills [2, 10] and voice intonation [6, 15], in addition to body movements and body postures [39].

Models of Affect

Popular models of emotions are Plutchik's emotion wheel [31], that offers a hybrid between emotional dimensions and discrete emotions, and Russel's circumplex model of affect [32] which describes linear combinations of two dimensions, valence and arousal, as varying degrees of stimulus (valence) and intensity (arousal). Sometimes these two dimensions are extended by a third dimension (the PAD model [28]), dominance, which describes the degree of control exerted by a stimulus.

The proposed emotional models are rather complex, and their respective self-assessment measures are therefore extensive, making them less suitable for an in-the-wild mobile experiment. Also self-assessment measures such as PANAS [40] or SAM [3] may reveal the purpose of the study to participants filling them out, distorting the elicited affect [41]. In this study we are interested in the direct physiological response to affective stimuli, and we therefore employ the term *affect*, measured on a positive-negative scale, as proposed by Isen et al. [20].

Motion and Emotion

Emotions change our physical behavior; we smile when we are happy and our bodies tremble when angry. Drawing on the Darwinian view that emotions cause biological determined reactions, Ekman [13] proposed his theory of basic emotions. From cross-cultural field studies he found six discrete emotions to cause similar physical response in facial expressions across cultures. Body postures and movements have in a similar way shown to be influenced by emotions [39], which is essential to the core theory of the emergent field of embodied cognition; that cognition as well as affective aspects go beyond the brain and manifest themselves physically in our bodies, such that emotions provide embodied information. This paper draws upon this view: If motor behavior, including physical interaction with mobile devices, encodes affective information, this should be detectable by analyzing the user behavior patterns of interactions with mobile devices.

Previous studies also examined the affective aspects of computer interaction. Cairns et al. [4] studied the influence of emotions on a simple number entry task. The preliminary study showed that participants who were in a more positive emotional state were more accurate at entering numbers on a touch-based number pad. A study investigating the impact of emotions on the performance of computerized motor tasks was carried out by Coombes et al. [10]. The authors had 40 participants perform a computerized square-tracing task after being exposed to affective imagery. The authors concluded that exposure to affective pictures has direct behavioral consequences on speed and precision of performance on motor control.

These studies together show the correlation between movements and emotions, and provide evidence for the link between computer interaction and affect.

Affect Detection

Among real-time affective predictors, audio and vision-based techniques are by far the most common and robust (see Zeng et al. [42] for a review). Emotional states can also be inferred using physiological sensors; equipment measuring for instance heart rate variability or galvanic skin response have been used to infer stress levels [27, 36] and emotional states [34].

These sensors may be disturbing to their users (such as common galvanic skin response sensors), and the equipment is generally absent from home and office settings. Therefore, several studies have employed non-specialized equipment available at most home or office settings to detect affective states; such as using keystroke dynamics [14], touch-based gameplay strokes [18], computer mouse tasks [36] or smartphone usage [26, 30].

Gao et al. [18] studied mobile touch activity as an indicator of emotional states by extracting finger-stroke features from 15 participants during a *Fruit Ninja* game. Self-assessed emotional states coupled with touch strokes led to an 89.7% accuracy in binary arousal classification, with almost similar rates for valence. The small sample size and the specific task studied limit the generalizability of the results, and thus further work is needed to shed light on the implication of affect on general touch interaction.

Another example of utilizing commodity computer equipment for affective computing was conducted by Sun et al. [36] who inferred stress measurements through common computer mouse operations. In a study with 49 participants, physiological measurements and stress self-reports were measured, and the data collected were used to train a stress detection system with a stress rate detection accuracy of 70%.

LiKamWa et al. [26] leveraged smartphone usage to estimate participants' affective states. Thirty-two participants partook in a field study, where self-assessed mood was linked to phone activity. The authors failed to create a generic robust affective model, but reported 93% accuracy in affective classification using a personalized model with two months of training data.

By analyzing the rhythms of 12 participants' typing patterns on a standard keyboard in a field study, Epp et al. [14] reported the correlation between emotions and keyboard typings. The authors reported 77-88% binary classification accuracy for 15 emotional states.

Limitations of Earlier Work & Our Approach

Although the emotional influence on human motor aspects has been studied extensively, few studies concern the influence of affect on human-computer interaction. Promising results in affective detection depend on either specialized hardware or personalized models that require prolonged per-user training to function. Also, a prevalent shortcoming of the previous work on affective modeling stems from the use of relatively few participants in controlled experimental settings.

The intention of this paper is to study the affective impact on HCI on more immediate use, in more ecologically valid settings, using common sensing hardware, with more participants than related research. To do so, we report findings from two crowdsourced user studies providing evidence for the feasibility of in-the-wild affect detection from mobile interaction, in addition to an analysis of interaction patterns and their relation to affect. The overall reasoning behind the experimental approach used is to increase the external validity, and thereby the generalizability of the findings in comparison to previous affect and HCI related papers. To do so, we designed a set of general purpose mobile touch tasks covering the bulk of mobile interaction strategies employed in most touch based graphical user interfaces. To increase quantity and representativeness of the participants, we did online recruiting; participants installed our experimental software on their own devices, and followed on-screen instructions.

EXPERIMENT I: EMOTION ELICITATION

Based on psycho-motor theory we envisioned that the physical properties of mobile interactions would vary with affect. To manipulate emotion as an independent variable to study the contribution to mobile interaction made by affect, we employed emotion elicitation (see Coan and Allen [9]). This way we also enforce a bigger variance in affect among participants. The purpose of the experiment was therefore to gather information about participants' mobile interactions and couple them to their elicited affective states. The collected data was then used to train a classifier.

To gather data from mobile interactions in the wild, we developed a mobile application that participants installed on their own devices. The application collected demographic information, and elicited either neutral or positive affect using video. Subsequent to elicitation, participants conducted three touch tasks. We refrained from eliciting negative affect, due to ethical concerns.

Participants

We crowdsourced 276 participants who partook in the experiment for US $1. Half of the participants were from the USA, the rest were from other native English speaking countries or Western Europe. Ages ranged from 18-70 (M=30.5), with 54% males, and 92% right-handed.

Apparatus

The application was implemented as an Android application, targeting Android ≥ 4.0. The app sent relevant user metrics over HTTP every 30 seconds to a server application created using the Python-based web application framework `webapp2` deployed at Google App Engine. The application forced a full-screen landscape orientation.

Procedure

Participants installed our experimental application on their own smartphones, and followed the same experimental procedure (see Figure 1). Half of the participants were placed in the positive group, and the other half in the neutral. The order of the touch tasks was randomized.

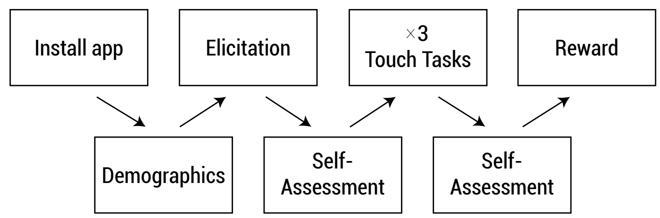


Figure 1: Overview of study procedure.

Design

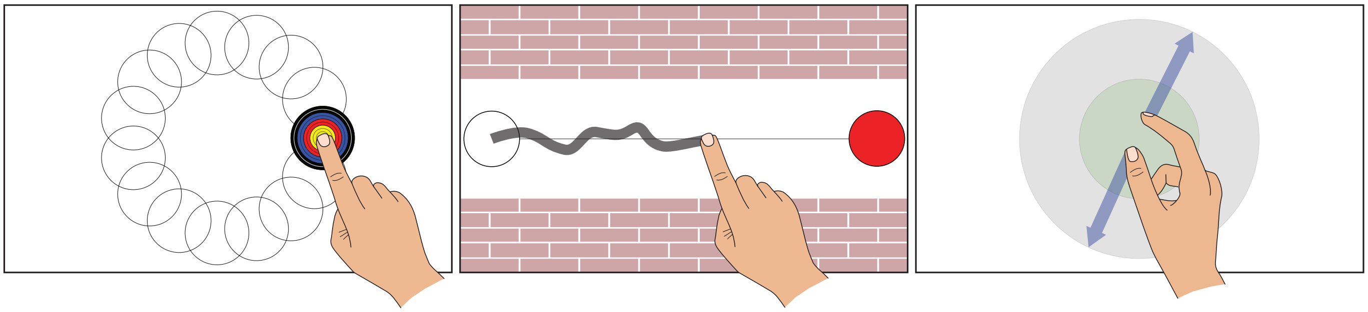
The experiment used a between-subject design where participants, after being elicited with either positive or neutral affect, conducted three touch tasks. We collected affective self-assessments following both elicitation and touch tasks. We employed the self-assessment protocol proposed by Isen et al. [20]: five affective differentials, measured on 7-point likert scales, with four filler items (refreshed vs. tired, calm vs. anxious, alert vs. unaware, and amused vs. sober) and one deliberate item (positive vs. negative).

Emotion Elicitation

Emotion elicitation techniques or mood induction procedures (MIPs), are methods that allow for scientific investigations of emotions through experimentally controlling emotions [9]. A comparative study of MIPs by Westermann et al. [41] found that showing movie clips had a larger effect size compared to other procedures: *Film/Story + Instruction is significantly more effective than all other MIPs* [41]. It is also fairly simple to include video content into a mobile application, making movie clips a suitable MIP for this experiment.

Choice of Movies

A study by Schaefer et al. [33] reported mean affective assessments followed by watching a large variety of movie clips. We conducted a small-scale between-subjects movie-survey (N=43) surveying four movie clips eliciting the highest mean positive affect from [33]. The results indicated that a scene



(a) (b) (c)

Figure 2: Three touch tasks: (a) Tapping, (b) Steering, and (c) Scaling.

from *There's Something About Mary* elicited the most positive affect (M=6.13, SD=1.13, on a 1-7 likert scale, 1=negative, 7=positive), and that the neutral movie clips (two clips from *Three Colors: Blue* and one from *The Lover*) scored significantly lower (M=4.33, SD=1.14), Cohen's d = 1.40; consequently, this movie configuration was chosen. The duration of the positive and neutral clips were 01:48 and 01:52, respectively. The neutral clips were shown with a 2 seconds black still in between. The order of the neutral movies was randomized.

Tasks

We wanted to design tasks that characterize the bulk of common operations on touch devices. Studies on human psychomotor modeling using general computer mouse tasks include point-and-click, drag-and-drop, and steering through straight, narrowing, and spiral tunnels [1, 17, 35, 36]. Touch differs from mouse interaction as drag-and-drop is almost identical to steering and as touch interaction rarely requires complex steering. In addition, touch interfaces commonly employ multi-finger interactions. We therefore end up with three tasks: tapping [35], steering [1, 36], and scaling [37].

1. Tapping

This task (see Figure 2a) presented a series of circles one at a time, located at different locations in a circular formation. The participant had to tap the circular targets as fast and accurately as possible. The task is a common Fitts's Law exercise described by ISO 9241-400 [21], used in numerous previous studies. The specific setup, such as order of sequence and number of targets, adhered to MacKenzie [35] who reported best practices for this task. This touch interactivity corresponds to regular taps, frequent when users dial numbers using a number pad or enter text using a soft keyboard.

2. Steering

In this task participants were asked to draw a line through a straight tunnel from left to right (see Figure 2b), as fast and accurately as possible. The trajectory of the participant corresponds to a drag-and-drop activity, as [36]. Steering behavior is used in mobile contexts for instance when reordering home screen applications, panning in maps or scrolling web pages.

3. Scaling

In this task (see Figure 2c) participants were asked to scale a circle as accurately and fast as possible. Scaling was done by expanding the distance from the origin of the two fingers' initial positions, similar to [37]. Two-finger behavior is common when browsing the internet (zooming), or navigating maps (rotating and zooming).

Task Repetitions

To ensure the same task difficulty for all participants regardless of phone size, we used Fitts's Law [17] to calculate appropriate target sizes using constant ID's, hardcoded in the application. We used Fitts's Law settings as described by MacKenzie [35]. To use as much of the limited screen size as possible, and thus also ensuring largest possible target widths, we maximized the distance, D, according to the phones' screen sizes. This means that we could calculate appropriate target width sizes, by solving the Shannon formulation for W:

$$ID = log_2(D/(W+1)) \Rightarrow W = D(2^{ID} - 1)$$

To keep the experiment as short as possible, while ensuring validity we chose to present participants for the same condition 15-16 times per task, as shown in Table 1.

Task	*IDs*	*Targets*	*Rep./ID*	*Actions*
Tapping	8 (2 - 4.1)	15	1	120
Steering	8 (2 - 4.1)	2	8	128
Scaling	8 (2 - 3.05)	2	8	128

Table 1: Summary of experimental settings used to ensure the same level of difficulty across devices.

Experimental Conditions

The independent variable was binary affect (positive, neutral), ensured through emotion elicitation.

While some previous studies proposed behavioral predictors trained on any available mobile data; such as time of day, network-strength, battery-level and so on, we collected measurements related to the physical properties of mobile device interactions, devised from the theoretical link between movement and affect. We measured among other things speed and precision of participants' touch tasks, see Table 2 for a complete list of dependent variables.

Sensor	*Measurement*	*Unit*
Touchscreen	Finger position	(x,y)
Touchscreen	Touch area	mm^2
Touchscreen	Precision	$]0,1] \in \mathbb{R}$
Pressure sensor	Pressure applied	$]0,1] \in \mathbb{R}$
Timer	Duration of action	ms
Accelerometer	Acceleration	m/s^2
Gyroscope	Change of orientation	rad/s
Screen	Width $\times$ height	pixels
Phone	Brand, model	name
Questionnaire	Age	years
Questionnaire	Handedness	left/right
Questionnaire	Gender	male/female
Self-assessment	5 differentials	1-7 ($\times$2)

Table 2: Data from mobile sensors collected during the experiment, as well as user-reported measures.

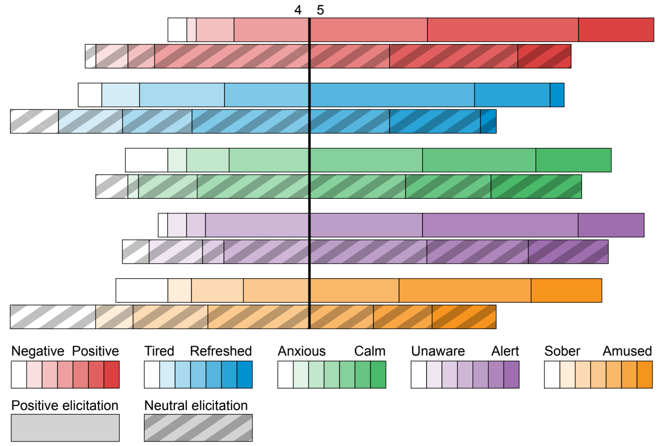


Figure 3: Self-assessments immediately after elicitation: The positive group reported higher assessments for all five differentials (1-7). The width of the bars represent the percentage of participants who reported the same value. The graph is fixed at the value 4 (mid value of 1-7). The neutral group (striped) reported lower likert values for all differentials, and in general had a larger percentage of self-reports at the value 4.

Hypotheses

Coombes et al. [10] reported an effect of affective stimuli on both speed and precision in a square-tracing task. Their findings suggest that negative valence causes greater haste and/or reduced precision in motor tasks, which corresponds to Fitts's Law [17]; that speed and precision are inversely proportional. Taken above in to consideration in regards to the outcome of present experimentation, we hypothesize the following:

H1: Exposure to positive affective stimuli will decrease participants' speed when performing touch tasks compared to exposure to neutral affective stimuli.

H2: Exposure to positive affective stimuli will increase participants' precision when performing touch tasks compared to exposure to neutral affective stimuli.

H3: Tasks completed immediately after exposure to affective stimuli cause bigger variance across the experimental groups, than tasks conducted later in the experiment.

Due to the lack of previous studies in this field, it is difficult to hypothesize about whether, and to what extent, mobile sensor data such as acceleration and rotation correlate with affect.

Analysis

Anomalies

We removed participants with zero variance in their self-assessments (i.e., all differentials were answered the same). We also removed participants who due to technical issues spent unreasonable long time watching the movies. After removing disqualified participants, 194 participants' data remained.

Emotion Elicitation

The effect of the elicitation was less significant compared to our movie survey, but the positive group did report higher positive-negative self-assessments (M=5.10, SD = 1.48), than the neutral group (M=4.77, SD=1.46); Cohen's d = 0.22. Figure 3 depicts the variance in participants' self-assessments for every differential.

Participants' self-assessments showed to differ immediately after elicitation, but ended up almost identical after the experiment (after approximately 10 minutes, see Figure 4), showing a relatively fast decrease of the emotional effect after elicitation. The relative modest, although consistent, increase in self-assessments compared to the positive group are attributable to the absence of a negative elicited group.

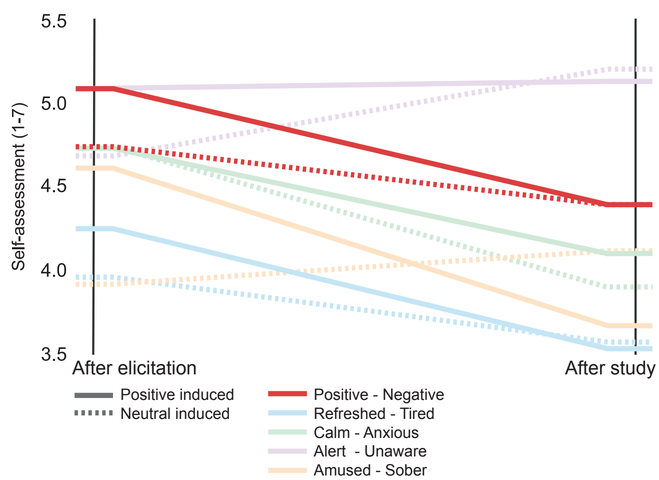


Figure 4: Temporal development of self-assessments: Difference between affective differentials (1=negative, 7=positive) immediately after elicitation, and after the task completion (with emphasis on the positive-negative differential). The trend is that the groups on average self-report differently immediately after elicitation, but have similar assessments towards the end of the study, showing the temporal equalization to the affective base level.

Durations

The neutral group on average completed all three tasks slower than the positive group. There was a close to significant difference in the task durations for the positive group (M=420s,

Feature	*Description*	*Interpretation*
speed	Distance traversed divided by duration, in px/ms	Higher values equal faster finger interactions
speedID	Speed divided by index of difficulty	Speed normalized by task difficulty
precision	Precision of activity: tap task uses distance to center, steering task average distance to center line, two fingers task distance to target scaling	1.0 equals perfect hit, 0.0 exactly on target edge, and $\leq$-1.0 completely off target
precisionID	Precision divided by index of difficulty	Precision normalized by task difficulty
accelerationY	Horizontal acceleration force, in m/s^2	Movements to left or right
accelerationX	Vertical acceleration force, in m/s^2	Up- and downward movements
Δacceleration	$\Delta\sqrt{x^2+y^2+z^2}$	Difference in aggregated acceleration, high values suggest constant shaking
rotationα	Rotation around z-axis, in rad/s	Rotations around axis pointing towards the participant
rotationβ	Rotation around x-axis, in rad/s	Rotations around the short edge of the phone
rotationγ	Rotation around y-axis, in rad/s	Rotations around the long edge of the phone
Δrotation	$\Delta\sqrt{\alpha^2+\beta^2+\gamma^2}$	Difference in aggregated rotation, high values suggest constant rotation
pressure	Applied finger pressure, from 0-1	High values indicate harder pressure
pressureDecline	Difference in pressure between beginning and end of interaction	Higher values indicate bigger pressure differences
devAngle	Difference in angle between fingers and centroid in beginning and end of interaction	Higher values indicate bigger differences in angles
centerAngle	Angle between horizontal line intersecting the centroid and line intersecting centroid and tap, $\Rightarrow \cos^{-1}(x/r)$	0-180 indicates activity on top half of target, and 180-360 on lower half.
approachDirection	Position of tap corrected for approach direction	Same as centerAngle but corrected for direction from last interaction
tapMovement	Movement of finger during tap	Usually very low, indicates slippage in pixels of finger during tapping
fingerDistance	Distance between two fingers	Distance between the two fingers in pixels

Table 3: The best features selected using recursive feature elimination. Features may be represented by their maximum, minimum, average, median and standard deviations measured throughout the experiment. The above 18 feature types represent a total of 46 features, out of the initial 352.

SD=119s) and neutral group (M=472s, SD=255s); $t(190) = -1.859, p = 0.065$.

Data Reduction

The data anomaly removal resulted in 194 participants out of initial 251. On this data set we were able to achieve a binary classification accuracy of 67%. However, data analysis and visual inspection led us to believe that differences in hardware among participants distorted the analysis because of non-comparable scales and granularity of sensors. Another difficulty we encountered was normalizing touch interactions between phones with difference in pixel depth and aspect ratios. We therefore limited the data analysis to only include common phone models with similar sized screens, thus removing all tablets and any phone used by less than 3% of the participants. The resulting data contained 55 participants using seven phone manufacturers: Samsung, Asus, Google, LG, Motorola, and Sony. Whereas this dramatically reduced the sample, it allowed us to compare measurements obtained from phone sensors.

Feature Selection

Participants accounted for roughly 6MB of raw data each, primarily because of the very comprehensive capturing of motion, touch, and timing data. To facilitate the use of machine learning on this data, every participant's data needed to be represented by a number of features. The strategy was to include features found in related work, features derived from emotion theory, in addition to conceivable features computable using the gathered data.

Some features are applicable for all three tasks (such as applied finger pressure or finger size), while some only apply to a specific task (such as distance between fingers in the scaling task). Each feature is represented by several sub-features: the minimum, maximum, average, and median value, with some variation based on the applicability of the specific feature. Distances were normalized over screen sizes. We extracted a total of 352 features, predominantly computed using motion sensor measurements.

We used the checklist for optimizing variable and feature selection provided by Guyon and Elisseeff [19], which amongst other things suggests normalization (using l_2-norm), variable ranking, and outlier detection.

Feature Relevance

There are several reasons to estimate the individual relevance of the set of features: To facilitate visualization and understanding of the data, reduce storage requirements, reduce training time, and improve prediction performance [19].

We used recursive feature elimination with five-fold cross-validation on a linear SVM to do automatic tuning of the number of selected features. This yielded 46 as the optimal number of features. Table 3 lists the features with the highest discriminative powers.

Classifier Selection and Optimization

We tried a variety of classification methods; both linear, non-linear, and ensemble methods to classify affect. Specifically we compared k-Nearest Neighbor (k-NN), Support Vector Machines (SVM) with Radial Basis Function (RBF) and Linear kernels, Decision Tree (DT), Random Forest (RF), AdaBoost, Naive Bayes (NB), and Linear- and Quadratic Discriminant Analysis (LDA, QDA). The RBF kernel SVM showed to be the most promising predictor of the inspected algorithms. To find optimal parameter values, we used grid search on some bandwidth parameters calculated using the Jaakkola's heuristic [22]. The decision boundary created using the before-mentioned SVM can be seen in Figure 5.

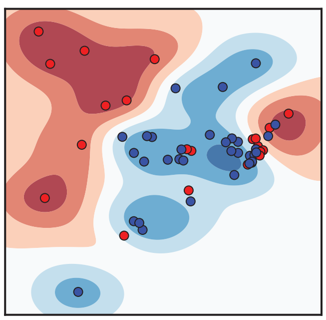


Figure 5: Data reduced to two dimensions using PCA with the decision boundary, created using an RBF-kernel SVM. Red and blue dots represent neutral and positive elicited participants, respectively.

Results

Classification

Table 4 shows the classification accuracies. We conducted three classifications: (1) a binary classification using the elicited affect, (2) a regression using the 7-point self-assessments protocol, and (3) a binary classification grouping participants by their self-assessments. Using the assessed affect (above/below median of assessment) results in a better accuracy than using the experimental groups. That is not very surprisingly, since participants' assessments should be closer to the actual affective state, than due to the elicitation. Overall the results are rather promising, showing that mobile device interactions are fairly efficient indicators of affect.

Affective Impact on Touch Performance

Analysis showed that 11 features had a significant difference among elicited groups, see Table 5. A t-test showed that all tasks provided measures that are influenced by affect. These results are consistent with the theoretically predicted directions of the parameters, specifically that speed decreases for positive affect, and that precision increases (see Table 5).

Variable	*Classes*	*Chance*	*Result*
Elicited affect (P/N)	2	60.0 %	87.3 %
Self-assessment (1-7)	7		1.33 RMSE
Grouped by median	2	56.4 %	89.1 %

Table 4: Classification accuracies for participants with similar hardware, $N = 55$. Results are obtained using an SVM with an rbf-kernel (regression or binary classification, measured using RMSE or Leave-one-out).

Feature	*task*	*t*	*df*	*p*
speedID_min	3	3.556	39.748	.001
deltaAccelerationAgg_max	3	2.854	45.128	.006
rotationB_max	3	-2.669	38.076	.011
speed	3	2.618	45.192	.012
speedID_avg	3	2.518	43.919	.016
accelerationY_avg	3	2.313	34.883	.027
speedID_med	3	2.250	45.703	.029
accelerationY_med	3	2.233	39.937	.031
speedID_std_dev	1	2.080	21.431	.050
precision_avg	1	-2.077	21.497	.050
precision_min	2	-2.010	51.332	.050

Table 5: 2-tailed t-test results for features with significance at .05, with equal variances not assumed.

Speed

The neutral elicited participants on average had faster interactions speeds for several tasks and measures (see Table 5). On the contrary, the neutral group on average finished all tasks slower than the positive induced participants due to lower precision (and therefore an increased amount of repeated interactions), most significant for the scaling task.

Acceleration

Neutral elicited, compared to positive elicited participants, performed tasks with higher aggregated acceleration $t(45) = 2.854$, $p = .006$. This was particular dominating on the y-axis. This implies that participants exposed to positive stimuli performed the tasks with smaller horizontal movements; that is positive affect led to steadier control of the device.

Rotation

Positive elicited participants account for higher max values of β-rotations (roll), $t(38) = -2.669$, $p = .011$. This means that the highest sudden rotational value around the x-axis for participants, were higher on average for participants exposed to positive stimuli. Rotations around the other axes followed same direction, but were not found significant.

Precision

Participants in the positive group performed tasks with higher precision than the neutral, most predominant for the tapping task, $t(21) = -2.077$, $p = .05$.

Two fingers

Positive elicited participants performed scaling tasks with bigger distance between their fingers. The angle between the two fingers and the center of the scaling target was also higher, although not statistically significant, $t(38) = 1.918$, $p = .063$.

Taps

The amount of pixels the finger drifted while performing taps showed not to have a statistical significant difference, $t(28) = 1.057$, $p = .3$. The position of the tap showed not to significantly differ, nor if it was normalized for approach direction.

Pressure

The data from force sensors showed no significant correlation to affect. The difference in decline of force throughout experimentation was not significant either.

The above descriptions presented several significant differences of features caused by affect, with speed, acceleration and precision as the most dominating affective indicators. The analysis showed that speed and precision of mobile interaction follow psycho-motor theory; that positive stimuli cause slower and more accurate motor behavior. Additionally, the results show that positive stimuli led to bigger movement (although less change of orientation) with the devices.

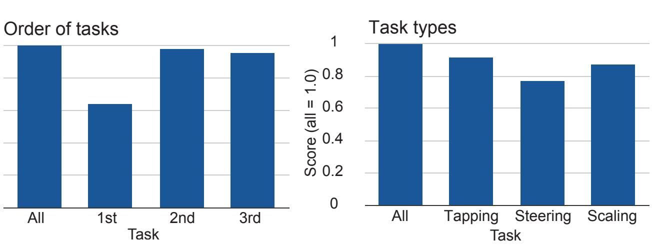
Insights

To get insights into the contribution to classification accuracy of each task type and order of tasks, we compared classification accuracies from predictors trained with different subsets of the data.

As evident in Figure 6a, all tasks individually provided classification accuracies over chance level, with tapping as the most robust task of estimating affect, and steering the worst. The combination of data from all tasks provided higher accuracy than any of the tasks alone.

Since the order of the tasks was randomized, each task was conducted at different times, relative to the other tasks. Analysis showed that the first task regardless of type of task, contrary to the hypothesis **H3**, scored relatively lower than the second and third (see Figure 6b). An explanation to this could be that the first task imposed a bigger challenge since participants were not completely aware of the application's mechanics, and therefore caused noisier data.

The accumulated classification accuracy did not increase when increasing the number of tasks from two to three, suggesting limiting the number of tasks for pragmatic affect detection.



(a) (b)

Figure 6: The relative contribution to classification accuracy of (a) task order, and (b) type of task (1.0=standardized classification accuracy of a predictor trained with data from all tasks combined).

A reduction in affect over time was evident, probably both due to the natural temporal reduction of the elicited affect and because the tasks themselves influenced participants' affective states (towards neutral). It is therefore natural to question which time window of data provides the optimal affective model, at least in terms of classification accuracy. Comparing classification accuracies of 80 different amounts of aggregated participant data, showed the highest classification accuracy at 89.1% for 140 seconds of data (see Figure 7).

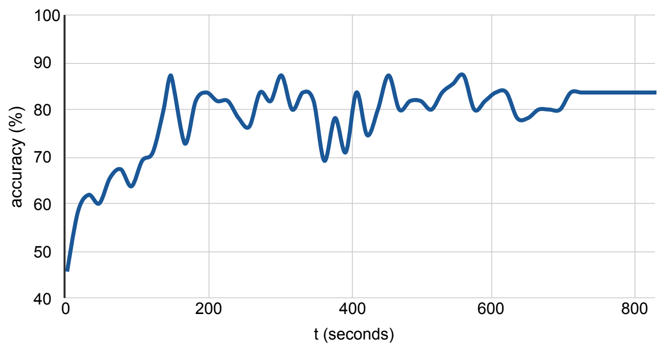


Figure 7: Classification accuracy as a function of the temporal amount of data the predictor is trained with. Classification accuracy peaks at 140 seconds with 89.1%.

EXPERIMENT II: NATURALLY OCCURRING AFFECT

The purpose of the second experiment was to understand the influence of naturally occurring affect on touch interaction in the light of the insights from the first experiment, with more participants, in order to validate the features used.

Design

The first experiment showed that an optimal classification accuracy was found using 140 seconds of participant data, and that the tapping task accounted for the highest discriminative power among the tasks; consequently this experiment employed only the Fitts's Law tapping task. To study the detection of participants' natural occurring affective states, we ran this experiment without emotion elicitation.

We extended the affective assessments to also include valence and arousal assessment using the pictorial 9-point Self-Assessment Manikin (SAM) [3] to allow higher dimensional affective assessments. Also, this assessment protocol would not sensitize participants to the study purpose because of the absence of elicitation in this study design. Experiment II allowed only a specific list of comparable phone models, ensured through strict settings at the Android app market. In summary Experiment II constituted the following:

- Within subjects design
- 1 task: Fitts's Law tapping task
- 140 seconds intended duration
- No induction; naturally occurring affect
- Removed demographics questions
- Five affective differentials and additionally SAM (9-point)
- Only comparable devices

Participants

127 participants participated for US $0.5, of which 29 also participated in the first experiment 75 days before. Seven were discarded because of zero variance in their self-assessments (5.5%), leaving N = 120.

Procedure

Because this experiment did not involve emotion elicitation and only contained one touch task, the procedure was simpler than the first experiment (see Figure 8).

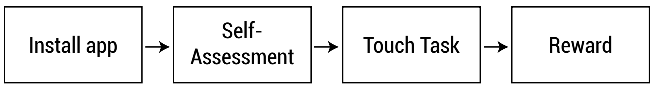


Figure 8: Overview of study procedure of Experiment II.

Results

Again an RBF-kernel SVM showed the most promising accuracies, with results over chance level for all protocols. Table 6 shows the accuracies of classifying binary affect ($\geq$ median) for the three protocols. This way, both groups are of approximately equal size in all classifications. Additionally results from predicting affects on the full likert scales, using a RBF-kernel Support Vector Regression (SVR) is shown. The best accuracy was found for valence detection, with 81.7%, or 1.61 RMSE for regression (valence of 1-9).

Variable	*Classes*	*Chance*	*Result*
Positive-negative (1-7)	7		1.32 RMSE
Valence (1-9)	9		1.61 RMSE
Arousal (1-9)	9		1.88 RMSE
Binary affect	2	54.2 %	69.0 %
Binary valence	2	51.7 %	81.7 %
Binary arousal	2	50.8 %	67.5 %

Table 6: Classification accuracies from Experiment II.

Kory and D'Mello [5] noted that affect detection using natural affect usually results in less accurate models than those constructed from elicited affect; results from this study reflect this finding. It is encouraging that it was possible to reach classification accuracies above chance level for all protocols, when training the model on data only from self-assessments of non-elicited participants.

This follow-up study showed that a rather short generic mobile touch task can estimate human affect with above chance-level accuracy, although less robustly than having a set of general tasks. It also showed that natural occurring affect is detectable using measurements from mobile interaction, although with less accuracy than elicited. We conclude that it is possible to design a generic affect detection task that detects affect above chance level, and that a standard Fitts's law task with data collection using mobile sensors is likely a good candidate for such a task.

DISCUSSION

Experiment I showed that the way humans interact with mobile devices does encode affective information. The affect is detectable by statistically analyzing features representing physical behavior, inspired by the literature on psycho-motor theory [2, 10]; speed, acceleration, and precision of touch operations showed to be indicators of affective states. Using this information we developed an affect detection technique that showed an 89.1% accuracy in binary classification of affect. Experiment II showed that the insights from Experiment I could be used to develop a more generic affect detection mechanism with only one touch task. Without emotional elicitation we were able to link touch activity to participants' self-assessed affective states with accuracies well above chance level for all assessed protocols.

Together the findings show that movement during mobile interaction corroborates psycho-motor theory, and that in-the-wild and affordable affect detection can be implemented using already common mobile sensors. The technique described in this paper differs from related research by training a model on data gathered in the wild using commodity hardware, which in turn provides a more generic affect detection protocol that does not require per-user training.

Hypotheses

Based on findings in experimental psychology (e.g., [2, 10]) we hypothesized that (1) positive affect would decrease speed of touch tasks, and (2) positive affect would increase precision of touch tasks, and (3) tasks completed immediately after elicitation would better indicate affect than tasks completed later.

H1: Positive Affect Decreases Speed

Results corroborated this hypothesis: The speed of finger movement was significantly slower for participants who were exposed to positive eliciting stimuli. Normalizing the speeds with the index of difficulty of the tasks resulted in better indicators of affect. We also found that the overall task completion times for the positive elicited group were shorter than the neutral group; due to a higher tendency among positive elicited participants in completing tasks in the first attempt.

H2: Positive Affect Increases Precision

Results corroborated this hypothesis: Precision of tasks (distance to ideal touch activities) was found significantly higher for participants who were exposed to positive eliciting stimuli. This follows Darwinian emotion theory: That accuracy of movement decreases with more negative affect to rapidly prepare the organism to respond to threatening behavior [25].

H3: Encoded Affect is Stronger Immediately after Stimuli

This hypothesis was not confirmed: Features computed from the first task conducted immediately after exposure to emotion eliciting stimuli provided worse affective detection accuracies than data from the two subsequent tasks. We believe this is likely caused by the effect of learning and the difficulty in conducting tasks that are non-familiar.

Towards a Generic and Unobtrusive Technique

For generic affect detection we pursued a technique that ideally works independent of people, situations and devices. The participants we recruited showed a non-significant difference to the population of the world: comparing age, gender, and handedness with *The World Factbook* [8], we find $\chi^2 = 1.36$, $p < .51$, in line with crowdsourcing literature [16]. The proposed technique was verified using data gathered from uncontrolled settings, such that the whereabouts of participants were unknown, suggesting an at least wider situational applicability than the laboratory. By normalizing interactions for screen size, and index of difficulty we envisioned that cross-device comparison of interactions would be possible. Nevertheless we struggled to model affect robustly device-independently. Between devices comparability of mobile interactions is highly dependent on the mobile devices' physical properties and form factors. And since different sensing hardware offer different frequency and granularity of measurements, a device-independent affect model based on touch interactions is cumbersome to achieve. The results from the studies presented in this paper suggest that obtaining high accuracies in affect detection from mobile interaction requires predictors trained individually by phone model.

We employed established HCI practices [1, 17, 35, 38] in designing the mobile tasks that worked as affective indicators. We envision that the the proposed affect detection technique could be implemented in virtually any mobile interface that offers different IDs: The validity of Fitts's Law has been verified at numerous occasions and interfaces both in experimentally controlled settings and in the wild [7]. As the proposed technique uses interactions originating from artificial touch tasks, it is uncertain to what extent the experimental tasks conducted by the participants in our study represent the actual bulk of touch interactions performed on touch devices. Therefore, it would be interesting to implement the affect detection technique proposed in this paper in the background of common applications such as text messaging or other popular applications such as social media or news applications, making the detection completely subtle. This way learning and boredom of conducting the tasks would not influence results either, since the affect detection would run on top of existing interfaces users are already engaging with. Work is still needed on the proposed affect prediction technique in order to seamlessly and robustly provide applications with information about the user's current affective state, needed by most real-life applicative scenarios. Our technique showed to peak in accuracy at 140 seconds, although providing above chance-level accuracies from 15 seconds of touch interaction and onwards (60% accuracy at 15 seconds, see Figure 7). While faster than existing models that employ related subtle affective predictors (e.g., [14, 26]), the duration of the calibration is relatively long for many real life scenarios. Ideally the technique would deliver instant affective predictions – further work is needed to achieve this accurately.

Future Work

Experiment I showed that accelerations on the y-axis (horizontal in landscape) and rotations around the x-axis (vertical in landscape) were significantly different across emotion elicitation. The reason why activity at these axes stood out, and conversely – why the others did not – is still a question. There is much room to deeper understand the mechanics of moving/shaking/rotating the devices and its attribute to affect. Because the affective computing literature previously has not concerned movements in these dimensions, the relation remains uncertain.

Results from Experiment II showed that valence classification resulted in a higher accuracy compared to arousal classification (81.7% vs 67.5%), contrary to related work (e.g., [18, 24]), probably due to the previous studies' domains being clearer indicators of arousal: Gao et al. [18] inferred arousal and valence from touch strokes in a gaming context, and Kleinsmith et al. [24] recognized affect from body movements and postures. Both studies reported classification accuracies contrary to the results presented in this paper – arousal classifications rates were consistently higher than valence. Intuitively one would also think that the level of arousal (intensity of emotion) would be stronger encoded in human movement or device interaction than the level of valence (pleasure of emotion), and thus easier detectable. Since previous work, contrary to this, confirmed this intuition, future investigations in the detectability of the emotional dimensions are needed: Is this difference due to the boredom of the tasks studied or the approach to the detection technique?

Out of an ethical concern we decided not to elicit negative affect in the wild, and Experiment I therefore grouped participants by either positive or neutral elicitation. We believe that having the full spectrum of affective elicitation would enforce a bigger discrepancy between mobile interaction behaviors, thus likely providing better classification results.

CONCLUSION

This paper presented an affect detection technique departing from psycho-motor theory, that equip smartphones with the ability to assess human affect, thereby allowing the devices to employ more human-human like interaction styles. The presented technique addresses limitations in contemporary affect detection techniques: (1) by using only commodity mobile sensors, the proposed technique avoids requiring specialized hardware; (2) by doing experimentation in the wild instead of in a laboratory with more participants, the external validity increases; (3) by employing a more generic affect detection technique, we achieve a model that does not require per-user training. In this paper we reported findings from two crowdsourced experiments that studied the implications of human affect on general purpose touch-based mobile interaction. The results of the two empirical studies presented in this paper, reflect findings in experimental psychology, and together confirm that affect has direct behavioral consequences for interactions with mobile devices. We show that it is possible to detect mobile users' affective states using off-the-shelf machine learning techniques. Results show encouraging affect detection accuracies, revealing at most 89.1% accuracy for binary affect classification.

ACKNOWLEDGEMENTS

This work was supported by the European Research Council, grant no 648785.

REFERENCES

1. Johnny Accot and Shumin Zhai. 1997. Beyond Fitts' Law: Models for Trajectory-based HCI Tasks. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems (CHI '97)*. ACM, New York, NY, USA, 295–302. DOI: http://dx.doi.org/10.1145/258549.258760

2. Benoît Bolmont. 2005. *Causes, Role, and Influence of Mood States*. Nova Biomedical Books, Chapter III: Role and Influence of Moods Including Anxiety on Motor Control, 57–75.

3. M. M. Bradley and P. J. Lang. 1994. Measuring emotion: the Self-Assessment Manikin and the Semantic Differential. *Journal of Behavior Therapy & Experimental Psychiatry* 25 (1994), 49–59. DOI: http://dx.doi.org/10.1016/0005-7916(94)90063-9

4. Paul Cairns, Pratyush Pandab, and Christopher Power. 2014. The Influence of Emotion on Number Entry Errors. In *Proceedings of the 32Nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 2293–2296. DOI: http://dx.doi.org/10.1145/2556288.2557065

5. Rafel A. Calvo, Sidney K. D'Mello, Jonathan Gratch, and Arvid Kappas. 2015. *The Oxford Handbook of Affective Computing* (1st ed.). Oxford University Press.

6. Keng-hao Chang, Drew Fisher, John Canny, and Björn Hartmann. 2011. How's My Mood and Stress?: An Efficient Speech Analysis Library for Unobtrusive Monitoring on Mobile Phones. In *Proceedings of the 6th International Conference on Body Area Networks (BodyNets '11)*. ICST (Institute for Computer Sciences, Social-Informatics and Telecommunications Engineering), ICST, Brussels, Belgium, Belgium, 71–77. DOI: http://dx.doi.org/10.4108/icst.bodynets.2011.247079

7. Olivier Chapuis, Renaud Blanch, and Michel Beaudouin-Lafon. 2007. *Fitts' Law in the Wild: A Field Study of Aimed Movements*. Technical Report. CNRS-Université Paris Sud. https://hal.archives-ouvertes.fr/hal-00612026 LRI Technical Repport Number 1480, Univ. Paris-Sud, 11 pages.

8. CIA. 2016. The World Factbook. (2016). https://www.cia.gov/library/publications/the-world-factbook/geos/xx.html.

9. J.A. Coan and J.J.B. Allen. 2007. *Handbook of Emotion Elicitation and Assessment*. Oxford University Press, USA.

10. Stephen A. Coombes, Christopher M. Janelle, and Aaron R. Duley. 2005. Emotion and Motor Control: Movement Attributes Following Affective Picture Processing. *Journal of Motor Behavior* 37, 6 (2005), 425–436. DOI: http://dx.doi.org/10.3200/JMBR.37.6.425-436

11. Charles Darwin. 1998. *The Expression of the Emotions in Man and Animals* (3rd ed.). New York: Oxford University Press. (Original work published 1872).

12. L.C. De Silva and Suen Chun Hui. 2003. Real-time facial feature extraction and emotion recognition. In *Information, Communications and Signal Processing, 2003 and Fourth Pacific Rim Conference on Multimedia. Proceedings of the 2003 Joint Conference of the Fourth International Conference on*, Vol. 3. 1310–1314 vol.3. DOI: http://dx.doi.org/10.1109/ICICS.2003.1292676

13. Paul Ekman. 1992. An argument for basic emotions. *Cognition and Emotion* (1992), 169–200.

14. Clayton Epp, Michael Lippold, and Regan L. Mandryk. 2011. Identifying Emotional States Using Keystroke Dynamics. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. ACM, New York, NY, USA, 715–724. DOI: http://dx.doi.org/10.1145/1978942.1979046

15. Raul Fernandez. 2004. *A Computational Model for the Automatic Recognition of Affect in Speech*. Ph.D. Dissertation. Massachusetts Institute of Technology.

16. Rebecca A. Ferrer, Emily G. Grenen, and Jennifer M. Taber. 2015. Effectiveness of Internet-Based Affect Induction Procedures: A Systematic Review and Meta-Analysis. *Emotion* May 2015 (2015). DOI: http://dx.doi.org/10.1037/emo0000035

17. Paul M. Fitts. 1954. The information capacity of the human motor system in controlling the amplitude of movement. *Journal of Experimental Psychology* 47, 6 (1954), 381–391.

18. Yuan Gao, Nadia Bianchi-Berthouze, and Hongying Meng. 2012. What Does Touch Tell Us About Emotions in Touchscreen-Based Gameplay? *ACM Transactions on Computer-Human Interaction (TOCHI)* 19, 4, Article 31 (Dec. 2012), 30 pages. DOI: http://dx.doi.org/10.1145/2395131.2395138

19. Isabelle Guyon and André Elisseeff. 2003. An Introduction to Variable and Feature Selection. *J. Mach. Learn. Res.* 3 (March 2003), 1157–1182. http://dl.acm.org/citation.cfm?id=944919.944968

20. A. M. Isen, K. A. Daubman, and G. P. Nowicki. 1987. Positive affect facilitates creative problem solving. *Journal of Personality and Social Psychology* 52, 6 (June 1987), 1122–1131. DOI: http://dx.doi.org/10.1037/0022-3514.52.6.1122

21. ISO. 2007. 9241-400:2007, Ergonomics of human–system interaction – Part 400: Principles and requirements for physical input devices. (2007). http://www.iso.org/iso/home/store/catalogue_ics/catalogue_detail_ics.htm?csnumber=38896.

22. Tommi Jaakkola, Mark Diekhans, and David Haussler. 1999. Using the Fisher kernel method to detect remote protein homologies. In *In Proceedings of the Seventh International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, 149–158.

23. IftikharAhmed Khan, Willem-Paul Brinkman, and Robert Hierons. 2013. Towards estimating computer users' mood from interaction behaviour with keyboard and mouse. *Frontiers of Computer Science* 7, 6 (2013), 943–954. DOI: http://dx.doi.org/10.1007/s11704-013-2331-z

24. A. Kleinsmith, N. Bianchi-Berthouze, and A. Steed. 2011. Automatic Recognition of Non-Acted Affective Postures. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)* 41, 4 (Aug 2011), 1027–1038. DOI: http://dx.doi.org/10.1109/TSMCB.2010.2103557

25. Michael Lewis, Jeannette M. Haviland-Jones, and Lisa Feldman Barrett. 2008. *Handbook of Emotions* (3rd. ed.). Guilford Press, New York.

26. Robert LiKamWa, Yunxin Liu, Nicholas D. Lane, and Lin Zhong. 2013. MoodScope: Building a Mood Sensor from Smartphone Usage Patterns. In *Proceeding of the 11th Annual International Conference on Mobile Systems, Applications, and Services (MobiSys '13)*. ACM, New York, NY, USA, 389–402. DOI: http://dx.doi.org/10.1145/2462456.2464449

27. Hong Lu, Denise Frauendorfer, Mashfiqui Rabbi, Marianne Schmid Mast, Gokul T. Chittaranjan, Andrew T. Campbell, Daniel Gatica-Perez, and Tanzeem Choudhury. 2012. StressSense: Detecting Stress in Unconstrained Acoustic Environments Using Smartphones. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing (UbiComp '12)*. ACM, New York, NY, USA, 351–360. DOI: http://dx.doi.org/10.1145/2370216.2370270

28. A. Mehrabian. 1980. *Basic Dimensions for a General Psychological Theory: Implications for Personality, Social, Environmental, and Developmental Studies*. Oelgeschlager, Gunn & Hain.

29. Rosalind W. Picard. 1997. *Affective Computing*. MIT Press, Cambridge, MA, USA.

30. Martin Pielot, Tilman Dingler, Jose San Pedro, and Nuria Oliver. 2015. When Attention is Not Scarce - Detecting Boredom from Mobile Phone Usage. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '15)*. ACM, New York, NY, USA, 825–836. DOI: http://dx.doi.org/10.1145/2750858.2804252

31. R. Plutchik. 2001. The Nature of Emotions. *American Scientist* 89 (July 2001), 344. DOI: http://dx.doi.org/10.1511/2001.4.344

32. J.A. Russell. 1980. A circumplex model of affect. *Journal of personality and social psychology* 39, 6 (1980), 1161–1178.

33. Alexandre Schaefer, Frédéric Nils, Xavier Sanchez, and Pierre Philippot. 2010. Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers. *Cognition and Emotion* 24, 7 (2010), 1153–1172. DOI: http://dx.doi.org/10.1080/02699930903274322

34. Jocelyn Scheirer, Raul Fernandez, Jonathan Klein, and Rosalind W Picard. 2002. Frustrating the user on purpose: a step toward building an affective computer. *Interacting with Computers* 14, 2 (2002), 93–118. DOI: http://dx.doi.org/10.1016/S0953-5438(01)00059-5

35. R. William Soukoreff and I. Scott MacKenzie. 2004. Towards a Standard for Pointing Device Evaluation, Perspectives on 27 Years of Fitts' Law Research in HCI. *Int. J. Hum.-Comput. Stud.* 61, 6 (Dec. 2004), 751–789. DOI: http://dx.doi.org/10.1016/j.ijhcs.2004.09.001

36. David Sun, Pablo Paredes, and John Canny. 2014. MouStress: Detecting Stress from Mouse Motion. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '14)*. ACM, New York, NY, USA, 61–70. DOI: http://dx.doi.org/10.1145/2556288.2557243

37. Jessica J. Tran, Shari Trewin, Calvin Swart, Bonnie E. John, and John C. Thomas. 2013a. Exploring Pinch and Spread Gestures on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 151–160. DOI: http://dx.doi.org/10.1145/2493190.2493221

38. Jessica J. Tran, Shari Trewin, Calvin Swart, Bonnie E. John, and John C. Thomas. 2013b. Exploring Pinch and Spread Gestures on Mobile Devices. In *Proceedings of the 15th International Conference on Human-computer Interaction with Mobile Devices and Services (MobileHCI '13)*. ACM, New York, NY, USA, 151–160. DOI: http://dx.doi.org/10.1145/2493190.2493221

39. Harald G. Wallbott. 1998. Bodily expression of emotion. *European Journal of Social Psychology* 28, 6 (1998), 879–896. DOI: http://dx.doi.org/10.1002/(SICI)1099-0992(1998110)28:6<879::AID-EJSP901>3.0.CO;2-W

40. D. Watson, L. A. Clark, and A. Tellegen. 1988. Development and validation of brief measures of positive and negative affect: the PANAS scales. *Journal of Personality and Social Psychology* 54 (1988), 1063–1070.

41. Rainer Westermann, Kordelia Spies, Gunter Stahl, and Friedrich W. Hesse. 1996. Relative Effectiveness and Validity of Mood Induction Procedures: A Meta-Analysis. *European Journal of Social Psychology* 26 (1996), 557–580. DOI: http://dx.doi.org/10.1002/(SICI)1099-0992(199607)26:4<557::AID-EJSP769>3.0.CO;2-4

42. Z. Zeng, M. Pantic, G.I. Roisman, and T.S. Huang. 2008. A Survey of Affect Recognition Methods: Audio, Visual, and Spontaneous Expressions. *IEEE Transaction on Pattern Analysis and Machine Intelligence* 31, 1 (2008), 39–58. DOI: http://dx.doi.org/10.1109/TPAMI.2008.52