

Informing Intelligent User Interfaces by Inferring Affective States from
Body Postures in Ubiquitous Computing Environments

Peer-reviewed author version

TAN, Chiew Seng Sean; LUYTEN, Kris; SCHOENING, Johannes & CONINX, Karin
(2013) Informing Intelligent User Interfaces by Inferring Affective States from Body
Postures in Ubiquitous Computing Environments. In: Proceedings of the IUI'13
International Conference on Intelligent User Interfaces.

Handle: <http://hdl.handle.net/1942/14533>

Informing Intelligent User Interfaces by Inferring Affective States from Body Postures in Ubiquitous Computing Environments

Chiew Seng Sean Tan Johannes Schöning Kris Luyten Karin Coninx

Hasselt University

tUL – iMinds

Expertise Centre for Digital Media

3590 Diepenbeek, Belgium

firstname.lastname@uhasselt.be

ABSTRACT

Intelligent User Interfaces can benefit from having knowledge on the user's emotion. However, current implementations to detect affective states, are often constraining the user's freedom of movement by instrumenting her with sensors. This prevents affective computing from being deployed in naturalistic and ubiquitous computing contexts. In this paper, we present a novel system called mASQUE, which uses a set of association rules to infer someone's affective state from their body postures. This is done without any user instrumentation and using off-the-shelf and non-expensive commodity hardware: a depth camera tracks the body posture of the users and their postures are also used as an indicator of their openness. By combining the posture information with physiological sensors measurements we were able to mine a set of association rules relating postures to affective states. We demonstrate the possibility of inferring affective states from body postures in ubiquitous computing environments and our study also provides insights how this opens up new possibilities for IUI to access the affective states of users from body postures in a nonintrusive way.

Author Keywords

Affective Computing; Posture Tracking / Detection; Social Behavior; Emotion Recognition; Ubicomp; Intelligent User Interfaces.

ACM Classification Keywords

H5.2 [Information interfaces and presentation]: User Interfaces. – Interaction Styles.

General Terms

Performance; Experimentation; Design; Human Factor.

MOTIVATION

Weiser's vision of ubiquitous computing [35] is not longer a vision as pointed out by Abowd [1] and Rogers [28]. With the increase of computational power that is available in our environments and the quality of sensors, ubiquitous systems that are embedded in our environment can take better decisions on how to support our daily routines and activities. This is exactly the context-awareness that has been most prominently advocated by Dey et al. [12] to frame interaction within the situation of the user. Although the *context of use* in general has been widely explored to improve usability of UbiComp systems, we argue further improvements can be accomplished by taking into account the *context of the user*, more specifically the *affective state* of a user. For interaction in an UbiComp environment, a system, that detects the affective states of the users, should maintain the freedom of movement of the users by not instrumenting them with additional sensors.

Researchers have demonstrated the huge potential of including the affective states of users to inform UIs by approaches such as multimodal fusion [14] and data-centric modeling [21]. However, affective computing has not been deployed in ubiquitous computing environments yet. A leading question for this challenge is: How does one detect and represent the emotions that affective systems should implement in ubiquitous computing environments? This question in turn leads us to determine whether current sensors are able to deliver the advantages of affective computing in our research domain. Recently Aviezer et al. [36] highlighted the role of body postures in expressing and perceiving emotions. They showed in their Science article that body posture is as expressive as facial expression for distinguishing positive and negative emotions.

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.

IUI'13, March 19–22, 2013, Santa Monica, CA, USA.

Copyright © 2013 ACM 978-1-4503-1965-2/13/03...\$15.00.

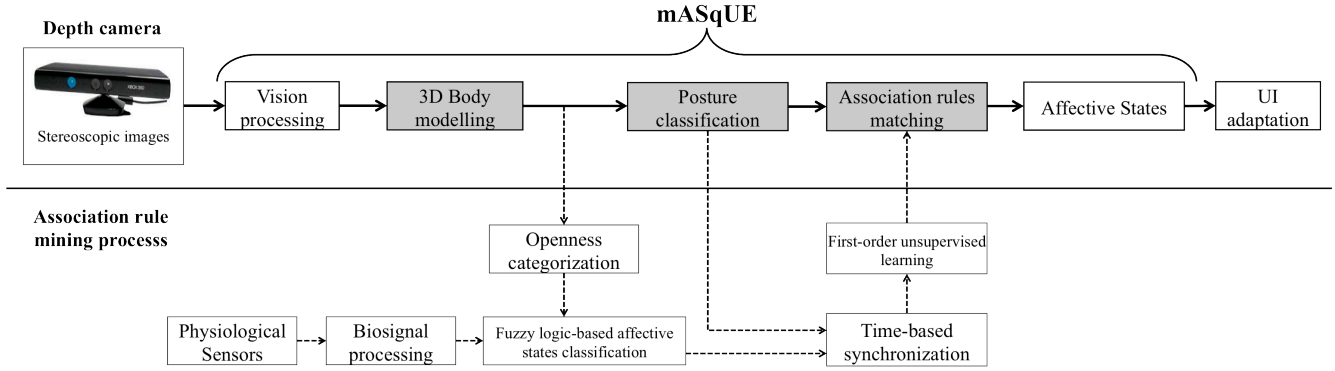


Figure 1: mASQUE affective state detection: By fusing the postural information with physiological sensors measurement (dotted lines) we are able to mine a set of association rules. The three main parts are highlighted with grey filled boxes. This set then can be used to detect affective states with posture detection only (solid lines) and inform UIs.

We identify three practical limitations emerging from current sensor implementations and use them to guide our system design mainly relying on body posture. (1) *Bound to the desktop*: the detection range of the aforementioned input modalities restricts users’ mobility. As an example, systems that use facial expressions or voice intonation analysis compel users to situate themselves directly in front of a camera or microphone, usually mounted on top of a computer screen. For that reason, we decided to use cheap commodity hardware and designed our system to work in ubiquitous computing environments. (2) *Not always “on”*: certain emotional cues of the users are undetected due to the device’s functionality. Similar to the above example, systems that detect and track lip movement or voice intonations depend on the users’ speech formation. Body postures constantly cover many communicative functions. They may replace speech during dialogue, or when speech is not used at all. Similar to body postures, ubiquitous computing systems are also always available to provide their computation functionalities and are constantly engaging with their users. Hence, using ubiquitous computing systems to detect postures ensures the continuous access to user’s emotions cues. (3) *Heavy user instrumentation*: users are also affected by a device’s presence (e.g. wireless sensors) throughout their normal activities. Systems based on physiological information require users to wear specialized instruments such as sensors that can detect galvanic skin response or heart rate, which burden their natural behavioral and mental response. Automatic visual detection of body movements and postures does not distract users, as no wearable instruments are required and also does not require user to constantly focus on a computer screen.

INTRODUCTION

As outlined above, we present mASQUE (Affective State detection in Ubiquitous Computing Environments), which is aimed to overcome these problems. It is a novel system that addresses the key limitations imposed by typical Ubicomp environments. mASQUE merely examines users’

body postures so that their dominant affective states can be inferred. We present a 3D body-modeling technique tailored for tracking even subtle body postures with an off-the-shelf depth camera. Our system is designed to passively and unobtrusively detect affective states when the whole body is detected and tracked, which usually means the user is standing. Although this can also be considered as a limitation of our system, there are other solutions that work for non-standing positions [18, 23]. mASQUE works on the principle of using a set of association rules that allow inference of someone’s affective state from their body postures (as illustrated in Figure 1). We found several affective states that are expressed through body postures in a naturalistic manner.

We used a mining process involving a series of experiments to obtain the set of association rules. We first looked for clear associations to dominant affective states and postures people take. To find these, we compared posture information with physiological sensor data from the users, which were captured at the same time. In this stage, we used the traditional input data: the user’s arousal and valence, as defined by [29], which have proven to be useful and reliable, together with the openness values. We then transformed this information to affective states and correlated them with the detected postures. Openness is derived from posture information as well. The characterization of openness levels enables us to gain insights of the attitude of the users. Attitudes commonly refer to the general evaluations people hold regarding various objects, issues, and people [25]. With that, we derive a set of association rules to cover ten affective states, consistent with the ones described by Breazeal and Scassellati [6]. The set of ten affective states comprises of five basic affective states: content, excitement, frustration, calm, and boredom, which are also described by Russell’s Affect Grid [29]; and another five affective states (overjoy, accepting, stern, disgust, anger) as described by [6]. Our system, mASQUE, uses the association rules to decide what the most likely dominant affective state is, merely based on body postures. By including “openness” as an additional

categorization of affective states, we are able to use a fine-grained definition of affective states. In contrast to physiological measurements, which are based on distinctive patterns of the autonomic nervous system inside the body, openness is observed through bodily movements, such as postures, to determine for the varying degrees of emotions. A low openness level can indicate an emotion that shows greater intensity of animosity. For example, frustration in a neutral openness level can be intensified to become anger. When the openness level drops, this indicates that the user is being aggressive, introvert or unapproachable.

With these components in place, we have a system that captures the context of the user and provides information on the dominant affective state of the user.

RELATED WORK

Affective systems, equipped with different sensory modalities, are able to detect and recognize emotions of the users. Their applications range from troubleshooting for interactive systems [4] to enhancing game experience with play technology [19]. Different modalities and affective models are being used, such as voice intonation analysis with a corresponding Pleasure-Arousal-Dominance model [14] or physiological instrumentation being used with arousal and valence in addition to a third affective dimension (engagement [23]). For a comprehensive overview on emotion recognition using different modalities and affective models, we refer readers to the survey of Gunes et al. [15].

Interactive systems with affective computing capabilities have been previously studied. Ball and Breese [4] developed an affective system, which utilize a four-dimensional affective model to develop a character-based interface for software troubleshooting. Their affective model includes dominance and friendliness alongside with arousal and valence to model the traits that appear to be most critical to interpersonal relationships. AffectAura, developed by McDuff et al. [23], utilizes a combination of video, speech, physiological, and activity information to provide visualization of the user's affective states in desktop work-related context. They used valence, arousal, and engagement to model user states, where engagement in this case refers to the user's level of work-related commitment. However, both systems are highly dependent on the speech characteristics (and also physiological instrumentation [23]) to determine the classification of emotionally communicative behaviors. This dependency creates practical limitations (mentioned in previous section) for the users in ubiquitous computing environments.

Gilroy et al. [14] used a multimodal fusion approach for video and speech to interpret the dimensional representations of emotion. Their work focuses on the EMG electrode placements to establish the pleasure dimension using Pleasure-Arousal-Dominance model during physiological measurement. Lance and Marsella [20] used

motion capture to collect data of the head, eye, and body movement of actors performing emotionally different gaze shifts. They also used a Pleasure-Arousal-Dominance model. Their work involves gaze behaviors and emotional expressiveness in animation of virtual characters. Wagner et al. also explored fusion methods for multimodal emotion recognition with a special focus on missing data for real-time applications [34].

Bolls et al. [5] used facial electromyography (EMG) and physiological measurements to determine the valence of user and explore the effects of user's arousal on attention and memory. They used arousal, valence and dominance to measure and determine the relationship between valence from radio messages and cognitive functions. These studies, however, mainly focused on improving sensor measurements or performance accuracy and did not strive for incorporating their work in ubiquitous computing environments.

Probably the work most similar to ours is that by Kleinsmith et al. [19], who used posture information to determine a set of affective states in an interactive play context. However, they used potency and avoidance to indicate the level of control to influence emotionally significant people, events, or situations, which is grounded in the work of Mehrabian [24]. Our affective model, which differs from theirs, uses openness to indicate how approachable the users are by the expansiveness or contractiveness of their body postures. Our use of openness is inspired by Ekman and Davidson [11]. We describe openness as one's behavioral closeness to another person in interpersonal space. Furthermore, our approach does not rely on additional instrumentation – body markers in Kleinsmith's case – to acquire visual information. Markers can also be problematic outside of a laboratory context, which is not a problem for mASqUE. Additional benefit of our work is the use of affordable and highly accessible hardware (Microsoft Kinect).

SYSTEM OVERVIEW OF MASQUE

Our system, mASqUE, consists of three main parts as can be seen in Figure 1 (grey filled boxes). The first part of mASqUE is the **3-dimensional body model** (for kinematic analysis of the whole body using vision processing techniques) that also discriminate the openness level of the users. The second part is the **posture classifier** (for body posture classifications using 3-dimensional body modeling) that maps the kinematical and semantical features to an extended range of postures commonly observed in naturalistic, conversational settings. The third part is an **association rule matcher**, which is used to deduce the affective states from the detected postures. The association rules are derived from a mining process that we will describe later in this section.

mASqUE's approach for converting (sets of) postures into affective state is based on the stages in affective signal

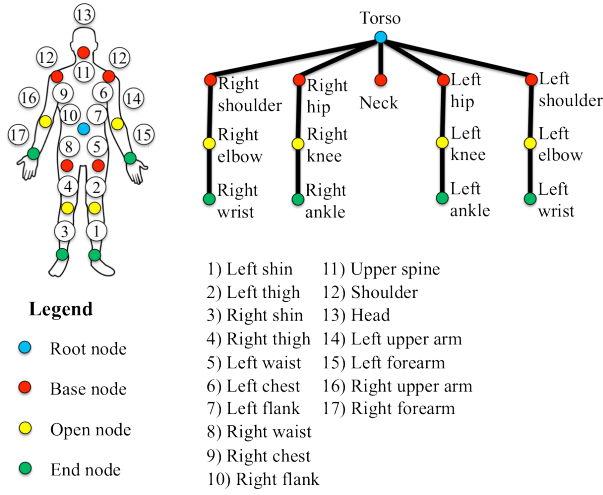


Figure 2. Link model with joints represented by hierarchical nodes. The 14 joints (shown as colored nodes) are connected to the 17 numerated body parts.

processing defined by Picard [26]. First, the low level information (postures) is captured by a sensing device (a MS Kinect). Next, we detect patterns in these postures that indicate a specific prominent affective state. The affective states we are able to classify are deduced from Russell’s Affect Grid [29]. Finally, we convert these states into high-level representations of goals and preferences for emotion synthesis. This final stage gives us symbolic data that can be used by affective systems to reason about their situations. In the following we discuss the three main parts of the mASqUE system in detail.

Vision-based approach for 3D Body Modeling

We delineate a set of postures typically identified in psychology literature [30]. The postures can be described with a combination of primitives in the body lean, arms, head, and legs [3]. The eleven postures we use are Hand-on-chin, Hand-near-hip, Hand-behind-neck, Arm Crossed, Limbs Restrained, Limbs Relaxed, Legs Crossed, Foot Forward (Left), Foot Forward (Right), Lean Forward, and Lean Sideward. We developed a 3D body model for identifying body kinematics that can be used to correlate the abovementioned postures with different types of limbs.

The skeletonization method to detect the body structure from the human silhouette [13] is the basis for our approach. In contrast to Fujiyoshi and Lipton [13] we used a 17-links model compared to their simpler “star” skeleton to represent the human body. However, for capturing more subtle posture cues (e.g. to enable posture tracking during human conversations) we extract additional features related to the quality of body movement. Based on this information, we introduce a model-based kinematic framework that is able to characterize body language in a conversational setting. Our 3D body model is represented

with a link model that is made from a set of limbs that are segments being connected together by means of joints. Our 17-links model comprises of fourteen body joints and seventeen limbs (as can be seen in Figure 2). We measure rotations and transitions of fourteen monitored body joints using the kinematic framework. Details on posture classification are provided in the next subsection, which is based on the measurements of body articulations (i.e. joint angle) and quality of body movement (i.e. Region of Interest movement).

Classification based on Postures

As an input for the posture classifier we use depth images from the Microsoft Kinect depth camera system. Basic skeletonization and tracking is provided by the OpenNI SDK from PrimeSense. The OpenNI SDK also allows us to detect significant body parts for our 3D body modeling including the head, tips of the feet, and tips of the hands, hips, knees, elbows, and the torso. We build a tracking algorithm on top of OpenNI that is tailored for detecting postures. The algorithm includes a combination of temporal scaling and spatial transformation parameters for tracking the significant body parts.

To increase the precision of our 3D body model, when subtle changes in postures occur, we implemented the 2D image tracking algorithms of Comaniciu [9] (i.e. kernel-based object tracking to monitor movement of the joints that are being tracked at per-pixel level) and color histogram model matching to measure the background noise caused by the transitioning of body postures. With our 3D body model, we can extract the angles of fourteen monitored body joints and track the following kinematic and semantical features: the body articulation measurement (i.e. flexion and pivot angles of each monitored body joint) and quality of body movement (i.e. quantification assessment of the tracked Region of Interest (ROI) that is superimposed on each joint).

Flexion refers to the rotation in the direction of the limb and pivot refers to the rotation laterally across the limb. The quantification assessment for quality of body movement uses a kernel-based object tracking to determine the number of Meanshift iterations for the ROI movement from the previous position at per-pixel level. Color histogram model matching is also included in the quantification assessment to measure the Mahalanobis distance, which is used to detect outliers (background noise) in data association of the color histogram patterns along image sequences. All of these features are combined to create a 44-element feature vector that is updated every 200 milliseconds. A feed-forward neural network with a backpropagation learning algorithm executes the posture recognition algorithms (one for each defined posture).

Association rule matching to infer affective state

Statistical analyses are used by the association rule matcher to make inferences about the affective state in real-time. We

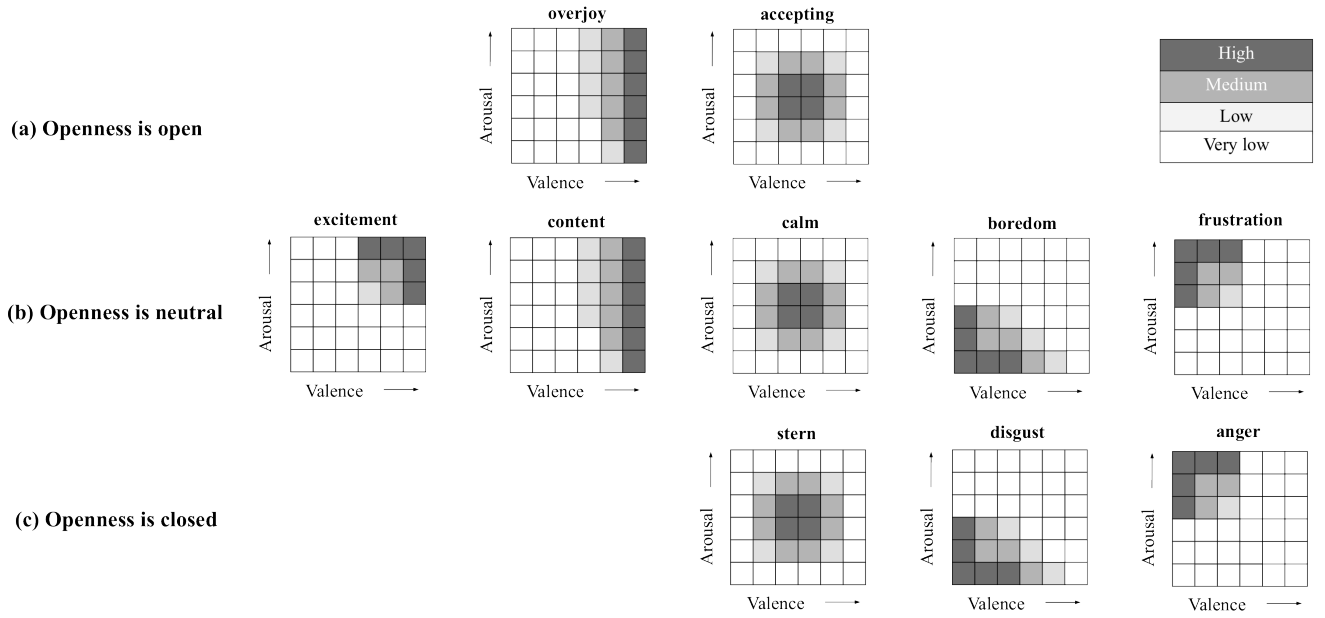


Figure 3: Representation of levels of emotions in arousal-valence space. (middle row) The results with openness being neutral are same as the results by Mandryk [19]. (top and bottom rows) Using the same set of representation of levels of emotions, we defined additional emotions with different openness levels (i.e. open and closed).

accumulate the posture information every 200 milliseconds over a moving window of 2 seconds and rank the postures according to their posterior probabilities. The three postures with highest posterior probability are then selected to determine a corresponding match with the association rules.

PROCESS FOR MINING ASSOCIATION RULES

The association rules are derived by a mining process that took the results we obtained from a study on postures as input. This study was conducted primarily to observe the postures during different emotionally-led conversations.

We developed an affective state classifier, which uses fuzzy logic to combine posture information and physiological measurements to identify an extended range of affective states.

Classification for an extended range of Affective States

Our affective state classifier is based on the combination of two classification approaches: vision-based openness categorization and physiological-based arousal-valence classification. The affective state classification uses the continuous values of arousal, valence and openness to mathematically model a set of emotions (see Figure 1).

Openness categorization based on Postures

We use the posture information from our vision-based posture recognition approach to provide the openness categorization in the detected postures. The transformations from posture to openness value, which provide a symbolic representation of an individual's behavior, is derived from Burgoon [7] and Clore [8]. They have evaluated different body behaviors to identify those that can distinguish a

positive attitude from a negative one. We then relate these body behaviors (such as expansive arm and body movement, postural relaxation, direct body orientation and forward lean of the upper body) to our posture descriptions. This openness categorization can also potentially be implemented in game systems to provide appropriate empathic feedback for the players to improve their game experience. For example, the game feedback can adapt to the openness level (such as players are showing negative signs of approachable attitude) and it provides the relevant support to the players with encouraging comments after a game when they are slouching their shoulders instead of making harsh cynical comments.

To confirm the level of openness conveyed by each of the postures suitable for naturalistic and ubiquitous computing contexts, we asked human observers to appraise their emotional content. We created hypothetical scenarios using guidelines from Scherer et al. [31] to capture non-acted postures from 10 pre-test participants. We then manually selected a total of 57 variations of the user postures and annotated them as test patterns. A total of 60 observers have been recruited from Amazons Mechanical Turk services to rate each test pattern on its level of openness from 1 (not approachable at all) to 7 (very approachable). The faces in the test patterns were pixelated to prevent observers from interpreting emotion from facial expression. The between-groups multivariate ANOVA results show that the postures can be grouped into three distinct levels. Postures with high levels of openness ($M = 4.43$, $SD = 1.54$) were indeed rated significantly higher than postures with mid level of openness ($M = 3.85$, $SD = 1.57$), $F(59) = 0.476$, $p < .001$; $r = 0.002$. Similarly, postures with low level of openness (M

= 3.36, SD = 1.49) were rated significantly lower than postures with mid level of openness, $F(59) = 0.467$, $p < .001$; $r = -0.113$. We then related the three levels of openness (high, mid, and low) to the respective labels: open, neutral, and closed.

Fuzzy logic implementation for Affective State Classifier

The affective state classifier uses a fuzzy logic implementation for determining the affective patterns that can be used to differentiate a set of emotions. Fuzzy logic provides the advantages of modeling uncertainty and also resembling the way humans think, reason and perceive about its situational information [2]. A typical fuzzy logic system uses a collection of fuzzy membership functions and rules to handle inexact, uncertain and vague concepts in an appropriate manner to reason about data.

Our classifier combines both physiological-based arousal-valence classification and vision-based openness categorization. This approach is similar to Mandryk's [22] which also utilized two fuzzy logic models for determining user emotions: the first model transforms physiological signals as a direct indication of user experience into arousal and valence, and the second model transforms the changes in arousal and valence according to the openness (category) level into a set of ten affective states.

To capture the physiological information, we used Thought Technology's ProComp Infiniti hardware and Biograph software. Arousal and valence values were then derived from these four physiological signals: Galvanic Skin Response (GSR), Heart Rate (HR), Electromyography of the corrugator supercilii muscle for frowning (EMG_{frown}), and Electromyography of the zygomaticus major muscle for smiling (EMG_{smile}). Similar to [22], we also faced the issue that there are no established guidelines for transforming arousal and valence according to the openness categorization into a set of emotions. We defined the membership functions and the rules to translate arousal, valence, and openness using the basis of the circumplex model of emotion [29], related works from Mandryk [22] and Breazeal [6]. For the transformation involving arousal-valence space at different levels of openness, fifteen trapezoidal membership functions that are evenly distributed are used for the inputs of arousal, valence and openness. Arousal and valence are each represented in six levels: veryLow, low, midLow, midHigh, high, and veryHigh. Openness is represented in three levels: closed, neutral, and open. For each of the three openness levels (see section above), the physiological information is mapped on the two-dimensional arousal-valence space, forming a set of 2D points that can be used as input to a fuzzy logic model.

There are 156 fuzzy logic rules to transform the input values from arousal, valence and openness into the ten affective states. The ten affective states comprise of five new affective states in addition to the set of basic affective states. The set of basic affective states can be derived from

Posture combination	Affective States
Foot Forward (Right) + Lean Sideward Lean Forward + Lean Sideward + Foot Forward (Right)	Overjoy
Hand-near-hip + Lean Sideward	Excitement
Foot Forward (Right) + Hand-on-chin Foot Forward (Right) + Hand-behind-neck + Hand-on-chin	Content
Foot Forward (Right) + Hand-near-hip + Legs Crossed	Accepting
Limbs Constrained + Legs Crossed Arms Crossed + Lean Sideward + Limbs Relaxed Lean Sideward + Legs Crossed + Hand-on-chin	Calm
Foot Forward (Left) + Limbs Constrained + Lean Sideward Arms Crossed + Hand-on-chin Arms Crossed + Hand-near-neck + Lean Sideward	Stern
Arms Crossed + Limbs Relaxed + Lean Sideward Arms Crossed + Limbs Constrained Limbs Constrained + Foot Forward (Right) Hand-on-chin + Hand-behind-neck	Boredom
Limbs Constrained + Hand-near-neck Hand-near-hip + Legs Crossed Legs Apart + Limbs Constrained + Legs Crossed	Frustration
Arms Crossed + Hand-on-chin Hand-near-neck + Limbs Constrained	Disgust
Hand-on-chin + Arms Crossed Limbs Constrained + Hand-on-chin + Foot Forward (Right) Arms Crossed + Hand-near-hip	Anger

Table 1. This table shows the association rules representative of correlation between postures and affective states. We assumed that the interaction partner is located at the direction where the foot forward is pointing to the right side i.e. Foot Forward (right).

the transformation of arousal and valence, when openness is neutral, based on a set of 71 fuzzy logic rules. We assumed the default for the openness level is neutral for these 71 fuzzy logic rules as they can be described and depicted on an arousal-valence plane. The fuzzy logic rules were generated to map the levels of arousal and valence to the five basic affective states: content, excitement, frustration, calm, and boredom. The details for this representation of the levels of emotion are shown in Figure 3 (middle row, b), which is based on our data and replicate the results of Mandryk. Note that we replace the emotional category of Challenge from [22] with Calm, as Challenge is derived from Csikszentmihalyi's [10] and Mandryk's [22] application domain of gaming whereas Calm is more suited for general contexts.

The details for the representation of levels of the new set of emotion are shown in Figure 3 (top and bottom rows, a and c). The five additional affective states are: overjoy, accepting, stern, disgust, and anger. As such, we reused the 71 fuzzy logic rules, which are described earlier for the set of five basic affective states. The additional 85 rules are derived from the first set of 71 fuzzy logic rules by

		Inferred from the mASQUE System									
		Overjoy	Excitement	Content	Accepting	Calm	Stern	Boredom	Frustration	Disgust	Anger
Fuzzy logic-based affective state classifier	Overjoy	276	0	0	5	0	0	0	0	0	0
	Excitement	9	445	0	0	14	2	0	0	0	0
	Content	0	0	37	0	0	0	0	0	0	0
	Accepting	0	0	0	393	0	1	0	0	0	0
	Calm	0	0	0	0	304	0	0	0	0	0
	Stern	0	1	1	0	5	488	0	0	0	1
	Boredom	0	0	0	0	0	0	220	0	0	0
	Frustration	0	0	0	1	4	0	0	507	1	1
	Disgust	0	0	0	0	0	1	0	0	66	0
	Anger	0	0	0	0	0	3	0	0	0	188

Table 2. The confusion matrix shows the performance of prediction of the affective states we derived from participants using postures only. The numbers indicate the frequency of the different detected affective states.

factoring two other openness levels (i.e. open and closed). For example, stern (when openness is closed) has the same representation of levels of emotion as calm (when openness is neutral).

Association rules mining in a naturalistic and ubiquitous computing context

The procedure of this study consisted of pairs of users engaging in a series of unrehearsed, minimally structured conversations that invoked emotions, both negative and positive. Eighteen participants were recruited within our research facility and subdivided in nine pairs. All participants participated out of free will, were in good health and were informed on the goals and procedure of the tests, as well as on the data being captured. The participants included 8 postgraduate students and 10 researchers, all of who had worked together in projects. Participants were paired in a fashion that maximized the frequency of daily interaction within each pair. In each pair, one participant was monitored (physiological measurement and posture tracking). Nine participants were monitored (six male and three female, ages ranged from 25 to 39). Before each trial, we explained to the participants that there would be two conversational sessions with a brief intermission. Both participants were encouraged to maneuver freely and express themselves physically with no restrictions on bodily movements or postures. Each conversational session took 15 minutes. The participants were debriefed at the end of the experiment and informed on the details of the experiment. We collected over 9 hours of video recordings for synchronizing the results from both the physiological measuring equipment and stereoscopic camera system. This

resulted in over 3600 samples where postures were identified with the corresponding physiological measurements captured during that posture.

RESULTS AND ANALYSIS

From the association rules mining process, we extracted a set of association rules using rule-induction. The association rules mining process is illustrated by the dotted lines in Figure 1. The body postures were first synthesized with expressions of emotions in the time-based synchronization. Of the 3606 samples collected from our experiment, only 2974 instances were able to be validated for the completeness of data. The resulting affective states were then used to facilitate a response directly from the posture input. Next, we implemented a first-order unsupervised learning technique, which employs the Tertius and Predictive Apriori algorithms in the open source WEKA machine learning software package [16]. The result of the association rules is as shown in Table 1.

A standard 10-fold cross-validation approach was used for testing. We used the J48 implementation of the C4.5 decision tree [27] to construct a pruned decision tree based on the data from our experiment. Our pruned tree consists of 50 nodes and 45 leaves. The decision tree outputs a set of temporal decision rules, which we used to supplement our classification of affective states. We then generated a hierarchy of general to specific categories that is useful for inferring the affective states from postures. Table 2 presents the confusion matrix of the decision tree, which shows the results of the fuzzy logic-based affective state classification and the results inferred from the mASQUE system. The

match instance rate is 98.32%. As can be seen in Table 2, mASQUE returns similar results as using the fuzzy logic-based affective state classifier. It is interesting to see, that e.g. for “excitement” similar affective states were inferred such as “calm” or “overjoy”. Note that there is no generally accepted baseline for comparing system that detects affective states. We showed that our system can reproduce the same results as using physiological measurements for each openness levels.

MASQUE EVALUATION

We conducted a user evaluation to determine the performance of mASQUE, involving 10 participants (3 female, 7 male); aged are between 23 and 31. The participants are a mix of 5 undergraduate students, 2 graduate students, and 3 research staff members. The participants were asked to evaluate mASQUE for 30 minutes by acting out the postures in a simulated conversation with a research assistant. The purpose was for the participants to provide their postures as input to the mASQUE system and observe the inferred affective states. After the evaluation we conducted a semi-structured interview.

The quantitative responses of the users about their affective states from the semi-structured interviews were averaged. The means are presented in Figure 4 (bottom). We performed comparisons using ANOVA on the affective states inferred from mASQUE. Comparing the frequency of occurrence with the Likert scale ratings of the users (comparing top with bottom results in Figure 4), we found the results to be consistent for seven affective states (overjoy, excitement, content, boredom, frustration, disgust, anger); however, we found out that accepting, calm, and stern are being significantly different ($p < .05$) from the seven affective states mentioned above, $F(1,20) = 6.163$, $p = .006$.

The consolidated qualitative responses from the interviews show that the appropriate mapping of postures to the intended affective states can be clearly inferred based on the association rules (which in our case map the affective states from a limited combination of postures as shown in Table 1). Our result is very much in line with the notion that emotion spaces are multi-dimensional and can reliably distinguish basic emotions. Although the emotions that we have represented in Figure 3 occupy distinct points in the arousal-valence emotion space, this is still not an exhaustive description. For a more definitive model for mapping postures to affective states, we posit that a comprehensive study on body postures in different social domains will be needed to obtain a set of more generalized, reliable association rules. Regarding the performance of mASQUE, three participants were skeptical and considered the selection of the ten emotions in Table 1 as still too narrow. The other seven participants expressed general satisfaction with the degree of correctness for characterizing the multi-facet emotions in their overall user

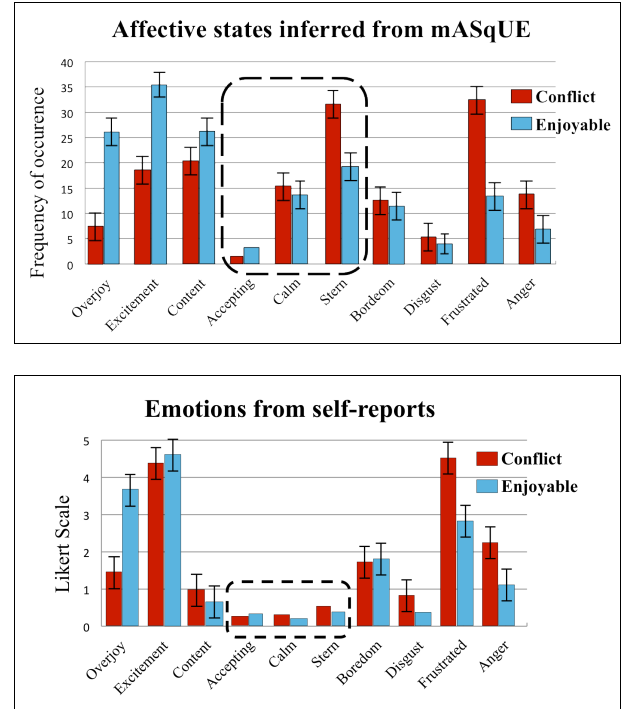


Figure 4. Mean values of affective states inferred from mASQUE (top) and emotions from self-reporting (bottom). The error bar show one standard deviation across the 9 monitored participants. (Bottom) Error bar for Accepting, Calm and Stern are not displayed. Their error rates are 0.03, 0.03, 0.01 and 0.15, 0.03, 0.01 respectively for conflict- and enjoyable oriented conversations. Notice that the mean values are only significantly different for Accepting, Calm, and Stern comparing the mASQUE system to the self-reports as indicated with the dotted box.

experience. One of these seven participants made the following comments “Personally, I do not have a (mental) set of postures for specific emotion. Knowing the affective states that I am having, will allow me to discriminate my self-expression.” Another participant stated: “The extensive range of emotions allows me to try out different possibilities of body movement and understanding their impact on my perceived emotion”.

DISCUSSION

An analysis of the user evaluation showed that mASQUE is suitable for deployment in ubiquitous computing environments as its rich, extensive range of emotion representations (i.e. affective states) is able to inform intelligent user interfaces about the user’s emotion. This is especially important for evaluating user experience in ubiquitous computing environments because the spontaneous affective response of the user can be determined during the process of interaction in real-time, not the outcome of verbal conversation. We have shown that our system can reproduce the same results as using physiological measurements for each openness level.

Besides that, the following two subsections also discuss the potential impacts of affective states detection in ubiquitous computing environments within the area of IUI.

Benefits of extending the emotion space

By adding the openness categorization to complement the arousal-valence emotion space, we are able to extend the arousal-valence emotion space from five basic affective states to ten affective states. We notice in our study that the complexity in making choices increases when allowing participants to choose from the ten emotional categories for qualitative evaluation in their self-report. We can refer to [33] for the explanation that people tend to simplify their decision-making processes using some form of heuristics, which in turn introduces judgmental errors and makes the assessment of own behaviors erroneous. This underlies an important role that body posture has for informing the IUI with five more affective states using the openness categorization.

Our system takes a global approach for emotion scaling, so a user's affective state is given as a percentage of the maximum value (i.e. Excitement). As seen in Figure 4, frustration and anger are significantly lower as measured by the frequency count for affective states compared to overjoy and excitement, which are only moderately lower. Although self-reporting from post evaluation interview had deviated from our mASqUE system, we can use the subjective self-reports to provide an explanation for our observation. Subjects taking part in the experiments were feeling excited in general with our expectation of certain exotic and novel experiences during the experiment. Although a user may be angry, and may rate anger as fairly high on a 5-point scale, this rating will be relatively low compared to the feeling of overjoy throughout the experiment.

Enriching the affective evaluation

In addition, mASqUE is able to characterize valuable information about affective states that cannot be extracted from the subjective self-reporting approach. In other words, mASqUE is able to detect affective states that are negligibly rated by the self-report. As shown in Figure 4, mASqUE shows a significantly higher distribution of accepting, calm and stern, whereas in self-reporting these emotions are negligibly rated.

From a biological perspective, these three emotions (accepting, calm and stern), which we describe with calm descriptions in the emotion representation (Figure 3), result in a delay of emotional response as they take a longer time to process cognitively. These emotions could be overtaken by other emotions with shorter response time (e.g. anger, frustration, excitement, and overjoy), which are linked to the biological emotion stimuli. However, it is important to detect these calm-related emotions as they have a positive and beneficial effect on a variety of decision-making processes [17].

Future research is necessary to be carried out in order to define how powerful the influence of a posture in the recognition of an emotion actually is in ubiquitous computing environments. An application of mASqUE, is Bro-Cam [32]. With Bro-Cam we explored the possibility to provide empathic feedback for videogame players. The body posture of a player at the end of a game round was mapped (this mapping was simply done using a MT survey) to humorous statement to improve the players game experience. Results from a user study [32] also validate the positive effects that this empathic feedback (with a very simple mapping) had on the player's game experience. To underline the potential of mASqUE we describe two scenarios where our system can exploit its full potential in the area of IUI in the following section.

INTERACTION SCENARIOS

With this work we provide additional information to the users that need to communicate with each other. Especially when users collaborate over a distance, the information we get from each other's body language can get lost. For tasks that require a tight collaboration between users, being aware of each other's affective state can make a huge difference. Even when co-located, users might not be familiar with each other and thus have difficulties in recognizing each other's affective states. In this section we illustrate how our approach can improve collaborations by means of two example scenarios that apply mASqUE for improving collaboration. The first example covers the support for information-sensitive conversations, while the second provides an example for intense remote collaboration.

1) During information-sensitive conversations, such as during doctor-patient conversations, ensuring sufficient empathy makes a huge difference. Trained professionals can do this during co-located conversations, but when doing them remotely using e.g. video conferencing, this is often problematic. mASqUE provides clues on the affective states and openness of conversation partners thus are able to guide a doctor when and how to reveal sensitive information and follow-up on the effects of her message. mASqUE tells us, besides the basic affective states, when someone is open for new information and ideas. During a video conferencing session, mASqUE presents to the doctor its readings using a straight-forward interface: it tells the doctor what the most dominant emotion is being deduced and what the level of openness is presented as a number ranging from 1 (not approachable) to 7 (very approachable). The doctor can subsequently steer a conversation in such a way the patient will be most comfortable in receiving new information.

2) Similar to the previous example, when people collaborate during intense tasks, mASqUE helps to make this collaboration more efficient and reliable. When a lab technician has to perform a critical task for example in a nuclear lab and is being guided by an expert remotely (since experts are sparse in this area), providing the most complete

view of the lab technician will make guidance fit better to the personal needs and individual state of mind. In this case mASqUE will notify the expert when the lab technician will be insecure or anxious when performing a certain activity. This will ensure that the expert pays extra attention at those stages since these require more help. In the opposite case, mASqUE can reveal when the lab technician becomes too relaxed and can intervene (anticipate) on activities that require the full attention of the lab technician.

These two short examples described above present possible application domains with the field of IUI for which we envision mASqUE to play a crucial role. The subtle information emitted by our body language can be used by mASqUE to determine affective states in various Ubicomp environments.

CONCLUSION

With mASqUE we demonstrated a new way to infer affective states from body postures in a ubiquitous computing environment. Using an unobtrusive technique, such as posture tracking with a depth camera, mASqUE opens up a huge design space for further IUI applications. By freeing the users from the desktop, having the system “always on” and with no user instrumentation we overcome the main limitations as described in the introduction. As also described later, one can imagine various scenarios, such as remote lab training, human-to-human conversations, point-of-sales interactions, or games, where this system is applicable. By adding vision-based posture recognition, we enable a new affective categorization (i.e. openness) to determine the degree of user’s behavioral closeness to another person or object in the interpersonal space. Our fuzzy logic-based affective state classifier for combining both physiological-based arousal-valence and vision-based openness values is able to scale beyond the limited range of basic emotions. The analysis also shows mASqUE is able to detect affective states (accepting, calm, and stern) – with positive and beneficial effect on decision-making processes – that are negligibly rated in the subjective self-reports. mASqUE opens up new possibilities for informing IUI about the affective states of users from their body postures in a nonintrusive way.

ACKNOWLEDGMENTS

We would like to thank Mieke Haesen for her valuable feedback and comments on this paper.

REFERENCES

1. Abowd, G.D. What next, ubicomp? celebrating an intellectual disappearing act. In *Proc. Ubicomp 2012*, ACM Press (2012), 31-40.
2. Abraham, A. Rule Based Expert Systems. In *Handbook for Measurement Systems Design*, Sydenham, P. and Thorn, R. (Eds.), John Wiley and Sons, London, UK, 2005, 920-931.
3. Argyle, M. *Bodily Communication*, Taylor and Francis, Abington, UK, 1988.
4. Ball, G. and Breese, J. Relating Personality and Behavior: Posture and Gestures, *Affective Interactions*, Paiva A.M (Ed.), LNAI 1814 (2000), 196-204.
5. Bolls, P.D., Lang, A. and Potter, R.F. The Effects of Message Valence and Listener Arousal on Attention, Memory, and Facial Muscular Responses to Radio Advertisements. *Communication Research* 28, 5 (2001), 627-651.
6. Breazeal, C. and Scassellati, B. A Context-Dependent Attention System for a Social Robot. In *Proc. IJCAI 1999*, Morgan Kaufmann Publishers (1999), 1146-1153.
7. Burgoon, J.K., Buller, D.B., Hale, J.L. and de Turck, M.A. Relational messages associated with nonverbal behaviors. *Human Communication Research* 10, 3 (1984), 351-378.
8. Clore, G.L., Wiggins, N.H. and Itkin, S. Gain and loss in attraction: Attributions from nonverbal behavior. *Journal of Personality and Social Psychology* 31, 4 (1975), 706-713.
9. Comaniciu, D., Ramesh, V. and Meer, P. Kernel-Based Object Tracking. *IEEE Trans. Pattern Anal. Mach. Intell.* 25, 5 (2003), 564-575.
10. Csikszentmihalyi, M. *Flow: The Psychology of Optimal Experience*, Harper Perennial, New York, USA, 1991.
11. Davidson, J. and Ekman, P. *The Nature of Emotion: Fundamental Questions*, Oxford University Press, New York, USA, 1994.
12. Dey, A.K. Understanding and Using Context. *Personal Ubiquitous Comput.* 5, 1 (2001), 4-7.
13. Fujiyoshi, H. and Lipton, A.J. Real-time Human Motion Analysis by Image Skeletonization, In *Proc. WACV'98*, IEEE Computer Society (1998), 15-21.
14. Gilroy, S.W., Cavazza, M.O. and Vervondel, V. Evaluating multimodal affective fusion using physiological signals. In *Proc. IUI 2011*, ACM Press (2011), 53-62.
15. Gunes, H., Schuller, B., Pantic, M. and Cowie, R., Emotion Representation, Analysis and Synthesis in Continuous Space: A Survey. In *Proc. FG'11, Workshop on EmoSPACE 2011*, 827-834.
16. Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P. and Witten, I.H. The WEKA data mining software: an update. *SIGKDD Explor. Newsl.* 11, 1 (2009), 10-18.
17. Isen, A.M. Positive Affect and Decision Making. Lewis, M. and Haviland-Jones, J.M. (Eds.) *Handbook of Emotions*, New York, USA, 2000.
18. Kapoor, A., Burleson, W. and Picard, R.W. Automatic prediction of frustration. *Int. J. Hum.-Comput. Stud.* 65, 8 (2008), 724-736.

19. Kleinsmith, A. and Bianchi-Berthouze, N. Recognizing Affective Dimensions from Body Posture. In *Proc. ACII 2007*, Springer-Verlag (2007), 48-58.
20. Lance, B. and Marsella, S.C. Emotionally Expressive Head and Body Movement During Gaze Shifts. In *Proc. IVA 2007*, Springer-Verlag (2007), 72-85.
21. Legaspi, R., Fukui, K., Moriyama, K., Kurihara, S., Numao, M. and Suarez, M. Addressing the problems of data-centric physiology-affect relations modeling. In *Proc. IUI 2010*, ACM Press (2010), 21-30.
22. Mandryk, R.L. and Atkins, S. A fuzzy physiological approach for continuously modeling emotion during interaction with play technologies. *Int. J. Hum.-Comput. Stud.* 65, 4 (2007), 329-347.
23. McDuff, D., Karlson, A., Kapoor, A., Roseway, A. and Czerwinski, M. AffectAura: an intelligent system for emotional memory. In *Proc. CHI 2012*, ACM Press (2012), 849-858.
24. Mehrabian, A. Pleasure-arousal-dominance: A general framework for describing and measuring individual differences in Temperament. *Current Psychology* 14, 4 (1996), 261-292.
25. Petty, R.E. and Cacioppo, J.T. *Attitudes and persuasion: Classic and contemporary approaches*, Westview Press, Colorado, USA, 1996.
26. Picard, R.W. *Affective Computing*. The MIT Press, Cambridge, Massachusetts, 1997, Ch. 5,6,and 7.
27. Quinlan, J.R. *C4.5: programs for machine learning*. Morgan Kaufmann Publishers Inc., California, USA, 1993.
28. Rogers, Y. The Changing Face of Human-Computer Interaction in the Age of Ubiquitous Computing. *Usability for e-Inclusion*, LNCS 5889 (2009), 1-19.
29. Russell, J.A., Weiss, A. and Mendelsohn, G.A. Affect Grid: A single-item scale of pleasure and arousal. *Journal of Personality and Social Psychology* 57, 3 (1989), 493-502.
30. Scheflen, A.E. and Scheflen, A. *Body Language and the Social Order: Communication As Behavioral Control*. Prentice-Hall, New Jersey, USA, 1972.
31. Scherer, K.R., Wallbott, H.G. and Summerfield, A.B. *Experiencing Emotion: A Cross-Cultural Study*, Cambridge University Press, New York, USA, 1986.
32. Tan, S., Schöning, J., Schneider Barnes, J., Luyten K. and Coninx, K. *Bro-cam: Improving game experience with empathic feedback using posture tracking*. In *Proc. of the 8th International Conference on Persuasive Technology*, (2013)
33. Timmermans, D. The impact of task complexity on information use in multi-attribute decision making. *Journ. Behavioral Decision Making* 6, 2 (1993), 95-111.
34. Wagner, J., Andre, E., Lingenfelser, F. and Kim, J. Exploring Fusion Methods for Multimodal Emotion Recognition with Missing Data. *IEEE Trans. Affect. Comput.* 2, 4 (2011), 206-218.
35. Weiser, M. The computer for the 21st century. *Scientific American*, 1991, 94-104.
36. Aviezer, H., Trope, Y., and Todorov, A. Body cues, not facial expressions, discriminate between intense positive and negative emotions. *Science* 338, 6111 (2012), 1225-1229.