Utilizing Emoticons on Mobile Devices within ESM studies to Measure Emotions in the Field

Alexander Meschtscherjakov
ICT&S Center, University of Salzburg
Sigmund-Haffner Gasse 18
5020 Salzburg, Austria
++4366280444813
alexander.meschtscherjakov@sbg.ac.at

Astrid Weiss
ICT&S Center, University of Salzburg
Sigmund-Haffner Gasse 18
5020 Salzburg, Austria
++4366280444812
astrid.weiss@sbg.ac.at

Thomas Scherndl
ICT&S Center, University of Salzburg
Sigmund-Haffner Gasse 18
5020 Salzburg, Austria
++4366280444813
thomas.scherndl@sbg.ac.at

ABSTRACT
Assessing emotions in situ while people are using technology is a difficult endeavor. Several assumptions on the concept of emotion exist in the research field of HCI and similarly several methodological approaches how to measure them. In this position paper we propose the usage of emoticons on mobile devices within experience sampling method (ESM) studies to measure emotions in-situ while the mobile device is used. Since ESM studies often require high efforts form the participant in terms of being interrupted several times a day it is especially important for ESM studies to have a means to be able to capture quick emotional user states and responses. We present a set of five emoticons, which cover two dimensions of emotions (strength and arousal) within one scale. To our conviction these emoticons allow an intuitive option for users to state their emotions on a mobile device during an ESM study. Using a case study, we investigated the feasibility of emoticons as answer categories for questions aiming at emotional states and feelings. We found that besides the space-saving aspect of the emoticons, which is an important aspect in conducting mobile studies on small displays, findings were not biased and participants had a positive user experience towards these question types. Furthermore, the usability of the emoticons was evaluated.

Categories and Subject Descriptors
H.5.2 [Information Interfaces and Presentation]: User Interfaces: valuation/methodology

General Terms
Measurement, Design.

Keywords
Emotions, emoticons, experience sampling method, mobile devices, user experience.

However, measuring emotions is difficult. It is even more difficult to measure emotions of users during interactions triggered by these interactions, since these interactions always take place in specific contexts. These specific contexts often have a big influence on the user’s emotions. Emotions do not occur in a vacuum: Emotional responses emerge in a context and are reflected within an ongoing action and interpretation process. When it comes to mobile devices it is even more difficult, since the mobile context is an ever-changing one. To capture user experience and emotions within various contexts we propose to utilize emoticons during experience sampling method (ESM) studies. ESM is an in-situ method to catch emotions, motivations and cognitive processes when they actually occur.

After giving a short overview on related work on the concept “emotion” we will present our approach to utilize emoticons as emotion measure during experience sampling studies. Thereafter we will describe a case study implementing this methodological approach.

2. RELATED WORK
The field of studying emotion has a long tradition and was observed from different perspectives with a focus on specific objectives. Each perspective has its own assumptions about how to define and construct theories of emotion [2]. A wide spectrum of different approaches does coexist, distinguishing between affect, mood, and emotion. There is growing consensus that affect refers to the general valence of an emotional state, emotion refers to specific types or clusters of feelings that occur in response to particular events, and mood refers to relatively enduring and global states of pleasant or unpleasant feelings [1].

Different methods can be utilized to measure the emotional state of a subject [9]. One promising area is to use body responses as a measure of emotions (e.g. skin conductance response (SCR), skin resistance response (SRR), electrocardiogram (ECG), or the size of pupils). Another approach is to monitor the activities of specific regions of the face (e.g. observation of zygomaticus major and corrugator supercilii). So far all of these approaches are only feasible in laboratory studies. Therefore up to now most in-situ studies used retrospective self reporting measures, like questionnaires and interviews to gain information on emotional responses.

In principle two theoretical approaches of measuring emotion can be distinguished: the categorical approach and the dimensional. The categorical approach by Ekman and Friesen [6] proposes that there are some “fundamental” emotions, where the term...
fundamental presents those patterns of responses to the world that evolution has equipped us with, due to their necessity for our survival. All other emotions are somehow derived from this small set of simpler emotions. They postulated the "Big Six" emotions, which are happiness, sadness, fear, surprise, anger, and disgust.

The dimensional model describes emotions in two (or in some cases three) independent dimensions (arousal, pleasure) in a Cartesian Space [10]. Based on these two dimensions, Russell created a 'circumplex of emotions' [11]. In this model, each emotion has a specific location on the circumplex. Emotions therefore are not perceived in categories but in much more complex and fluent manner. They can be measured according to valence and arousal, where valence refers to whether the emotion is positive or negative and arousal refers to the intensity.

To overcome the problem of wording and phrasing to describe emotions, pictorial tools have been developed. They are more likely to capture non-verbal aspects of emotions, are supposed to be less culture dependent, are also applicable to special user groups with restricted possibility to verbalize experiences and emotions such as children, and they allow people to express conflicting emotional responses. Instruments which make use of this approach are the Self-Assessment Manikin (SAM) [2], Emocards [5] or the Product Emotion Measurement Instrument (PrEmo) [4]. The SAM is a non-verbal pictorial assessment technique that separately measures three dimensions of emotions: pleasure, arousal, and dominance. The Emocards consist of 16 cartoon faces with eight distinct emotional expressions (eight male faces and eight female faces). These expressions vary on the basis of the dimensions 'pleasantness' and 'arousal'. A similar instrument is PrEmo, which measures 14 emotions (seven pleasant, and seven unpleasant). Respondents can report their emotions with the use of expressive 14 cartoon animations. In the instrument, each of the 14 measured emotions is portrayed by an animation by means of dynamic facial, and bodily, and vocal expressions.

3. EMOTICONS FOR ESM

Similar to the examples above our approach to measure emotions on mobile devices by means of emoticons is based on the dimensional approach. Contrary to the other approaches one main issue for us was to measure emotions on mobile devices. To use the emoticons during an ESM study we needed an instrument, which

- fits on the relatively small mobile device screen,
- is intuitive, does not need much mental effort for interpretation and
- is capable to be answered with input modalities provided by different mobile devices.

We therefore decided to create a non-verbal self-report measure designed on the basis of commonly known emoticons (see Figure 1). We designed five different emoticons embodying two emotional dimensions: positive and negative emotions (pleasure) and the strength of emotions (arousal). The emoticons were designed following an iterative design procedure.

Contrary to the Emocards and PrEmo, which are arranged in a circle and similar to SAM we decided to arrange our emoticons in a linear order. This order had the advantage of a higher usability on small-scale screens like on mobile devices. It gave us the possibility to suggest arousal not only in the design of the emoticon but also by their arrangement. Similar to Emocards we decided to code emotions with facial expressions, but we did not include male and female faces for the same emotion but used instead emoticons since they are sexual neutral. Contrary to Emocards and PrEmo we did not try to allocate basic emotions to our construct (e.g. ‘annoyed’, ‘euphoric’ for emotions with high levels of arousal; ‘bored’ and ‘content’ for calm emotions; ‘thrilled’ for very pleasant, ‘horrified’ for very unpleasant, and ‘surprised’ for neither pleasant nor unpleasant).

Similar to the SAM we stayed with the categorical approach. Contrary to the SAM, which measures three dimensions separately (pleasure, arousal, and dominance) we only included two dimensions (pleasure and arousal) and merged them onto one scale. This again has the advantage that it needs very little space on the mobile display and is easy to interpret.

As stated above we used these emoticons to retrieve emotional answers to questions during in-situ studies utilizing ESM. The next section describes a case study the above described emoticons were used to capture users’ emotions. Different kinds of questions could be asked to be answered using the emoticons. As an example the user could be asked how he/she feels in a certain moment (e.g. after using a special service). As another example they could be used to ask how satisfied a user is with a certain service. The satisfaction scale can then be visualized by emoticons instead of using answer categories like “very unsatisfied” or “very satisfied”.

4. CASE STUDY

4.1 Study Setting

To investigate the feasibility of the emoticons approach a case study was conducted in the framework of a bigger field study on emotional attachment to mobile devices in May 2007. As mentioned before the field study used the experience sampling method (time triggered) and was organized as follows:

1. Notification: Notification was done through a short message service (SMS). The user got an SMS whenever he/she needed to answer a questionnaire. This was done based on a predefined timesheet.

2. Delivery: The SMS used for notification contained a link to a webpage, which displayed the questions.

3. Capturing: Users’ responses were captured via a webpage submission into a database.

The study lasted for one week (including a weekend). Participants received seven notifications per day, where each sample consisted
of four questions. As time frame for the notification we defined 9:00 until 20:00 during weekdays, and between 10:00 and 21:00 during the weekend. The time-triggered sampling was scheduled for about every second hour. This means, that each participant received a total of 196 questions. Each question sample combined the same question forms, which were the following once: (1) yes/no questions, (2) multiple answer questions, (3) rating questions and (4) emoticons. In the following some example questions are mentioned, which had to be answered by means of emoticons (an example for the visualization on the mobile phone display can be found in Figure 2).

- How do you feel after being available for this person?
- How do you feel about customizing your mobile phone with additional services?
- How do you feel now that it was possible to have had a conversation?

Figure 2: A typical question during the ESM study.

Three different topics were assessed by means of the emoticons approach: usage environment, the mobile device itself and the user. As the ESM study was time triggered the same question had to be answered by the participants several times on different days and times.

Twenty people participated in our study (8 female and 12 male). Their age ranged from 26 to 52 years. Gender was counterbalanced and all participants were experienced in using smart phones with Internet connection.

4.2 Results

For reasons of statistical analysis the emoticons were coded numerically from -2 (very unhappy) to +2 (very happy) in this study. Although a lot of information concerning our main research questions could be derived from the data analyzing the emoticons results in correlation with the ESM logging data influence factors, like day time, used service etc., we report here only results concerning the usability and acceptance of the emoticons. For more detailed research findings concerning usage and emotional attachment toward mobile devices, we recommend to read the full technical report [8].

Our first concern was whether the emoticons could depict the full range of emotions during our field study and help the participants to formulate their current feelings. Indeed, we found that participants really utilized the full range of emoticons indicating that there was no or only minor bias grounded in social acceptability (see Table 1). These results go along with the general finding that participants reported significantly more often positive situations than negative ones, therefore resulting in a generally positive rating.

### Table 1: Distribution of responses (in percent) for different questions

<table>
<thead>
<tr>
<th>Question</th>
<th>-2</th>
<th>-1</th>
<th>0</th>
<th>1</th>
<th>2</th>
</tr>
</thead>
<tbody>
<tr>
<td>How do you feel after being available?</td>
<td>1.4</td>
<td>4.2</td>
<td>42.3</td>
<td>26.8</td>
<td>25.4</td>
</tr>
<tr>
<td>How do you feel after using the device?</td>
<td>2.8</td>
<td>5.6</td>
<td>36.1</td>
<td>36.1</td>
<td>19.4</td>
</tr>
</tbody>
</table>

Additionally to the variability over different situations and participants, individual participants’ ratings also differed over time, situation and context, suggesting that our emoticons were sensitive to changes in participants’ feelings. These results show a certain amount of face validity, as feelings at work during a situation which were perceived as very stressful were rated worse than situations at home which were rated as relaxing.

Figure 3 gives an example how the response behavior of the participants changed on the question: “How do you feel right now while using your mobile phone”.

Figure 3: Change in response behavior

Thirdly, we concentrated on the mean answering times of the questions using emoticons to assess momentary feelings and found no difference towards other question types as other multiple or single choice items (for emoticon question: average answering time was 31.64 sec with a standard deviation of 18.22; for a single choice question with similar length: 34.23 sec with a standard deviation of 12.45). We interpret this finding that participants could answer the questions concerning their feelings as adequately and fast as questions concerning more rationale and cognitive topics.

At last, a final interview with all the study participants revealed that they appreciated the emoticons approach and also approved of its intuitiveness. One participant suggested to add an answer category ‘have no feelings right now’, which indicates that the middle emoticon was correctly interpreted as ‘neutral feeling’, instead of ‘no feeling’. Nevertheless it seems to be quite unlikely that participants had no feelings at all.

To conclude, we suggest that the emoticon approach is suited for assessing emotion in situ, because it is intuitive, easy to understand and handle for the user and space-saving. The obtained results can be coded in different ways to concentrate on the two axis strength of emotion (positive vs. negative) and arousal...
(strong vs. weak), therefore allowing several approaches of research.

5. SUMMARY AND OUTLOOK
In this position paper we propose the usage of emoticons to assess emotions arising from the contextual usage of mobile devices in situ when they occur. Based on two dimensions (pleasure and arousal) five emoticons were used in a linear order to provide mobile phone users with a self-reporting tool to state their current emotion. A first case study could prove that the emoticons did not bias the answer behavior into one direction and are suitable to investigate different research contexts. A final interview with the case study participants revealed that participants experienced the emotions as intuitive and easy to answer on a mobile phone.

Our next step will be to focus more on the changing of emotional states and thus using these emoticons in a long-term study to investigate if they cannot only be used to measure the current emotional state of the user, but also indicate the development of an emotional attachment. This could be a valuable insight to derive influence factors for emotional attachment in the three different proposed contexts, e.g., phone services that get rated more positively the more often they are used. Another refinement of our approach will be to use context information as a trigger for delivering ESM questionnaires. This will make it possible to research the influence of specific contextual parameters to emotional responses.

6. REFERENCES
Measurement station for recording of different biosignals to detect emotions under mobile conditions

Wigand Poppendieck
Fraunhofer IBMT
Ensheimer Str. 48
D-66386 St. Ingbert
Tel.: +49-6894/980-170
wigand.poppendieck
@ibmt.fraunhofer.de

Roman Ruff
Fraunhofer IBMT
Ensheimer Str. 48
D-66386 St. Ingbert
Tel.: +49-6894/980-176
roman.ruff
@ibmt.fraunhofer.de

Eduardo Fernández
Universidad Miguel Hernández de Elche
Avenida Universidad s/n
E-03202 Elche
Tel.: +34-965222001
e.fernandez@umh.es

Klaus-Peter Hoffmann
Fraunhofer IBMT
Ensheimer Str. 48
D-66386 St. Ingbert
Tel.: +49-6894/980-401
klaus.hoffmann
@ibmt.fraunhofer.de

ABSTRACT
Feelings and emotions influence our daily routine and affect our social behavior. Especially while driving a car we change our emotions very quickly depending on the situation. To understand the different mental situations, the recording of biosignals such as EEG, ECG, blood pressure, eye movements or respiration plays an essential role. Here, a measurement station is presented which can be used to detect a great variety of biosignals from a person sitting in a driving simulator. By developing suitable data processing algorithms, the recorded biosignals can be used to detect and identify the emotional status of a person.

Categories and Subject Descriptors
J.3 [Life and Medical Sciences] – health, medical information systems

General Terms
Algorithms, Measurement, Human Factors

Keywords
Biosignal Recording, Emotions, Driving Simulator

1. INTRODUCTION
Feelings and emotions are permanently influencing our daily lives. In recent years, major research has been undertaken in the field of detection and classification of feelings and emotions for the use in emotional cognitive systems [1]. For instance, Ekman has defined 6 basic emotional states which can be found in every person independent of aspects such as culture or religion [2]. In order to detect the emotional status of a person, various vital parameters can be recorded by the use of appropriate sensors. These parameters include biosignals such as heart rate, galvanic skin response, body temperature, respiration, EEG or blood pressure [3]. By suitable algorithms, the acquired signals can be processed in a multidimensional feature space, in order to reduce the data to the indispensable features which can be used to identify the current emotional state of a person (see Figure 1).

Emotional cognitive systems can be used for a great variety of applications. In this context, we would like to focus on the detection of feelings and emotions in a car. While driving a car, we typically experience a wide spectrum of emotions, ranging from fatigue to anger. The objective is to develop a mobile system which identifies and reacts on the emotional status of a driving person, in order to modulate his emotional response and to increase road safety. For instance, by the use of eye-tracking systems, increased eye blink frequency correlated to fatigue could be detected. In such a situation, an alarm signal could be generated urging the driver to take a break. To detect a state of anger, vital parameters such as heart rate, galvanic skin response, body temperature, respiration of blood pressure could be used. If the system identifies the driver as “angry”, measures can be undertaken to calm him down, for instance by playing appeasing music or applying a certain enjoyable aroma.

Figure 1. Ekman’s basic emotional states [2].

The development of a mobile system is especially challenging due to the high demands caused by the changing environmental conditions, increasingly occurring artifacts and the usually limited resources. In the following, a measurement station is presented which can be used to record a variety of biosignals under mobile conditions, in order to detect and identify the emotional state of a person by appropriate signal processing in a multidimensional feature space.

2. CONCEPT
The concept of the measurement station is presented in Figure 2. The central part of the system is a driving simulator (SimuTech, Bremen), which allows for driving a car with all required actions, such as accelerating, braking, changing gears or steering. The simulator is equipped with three screens to project real driving.
situations, providing a 180° field of view. With this system, various hazardous episodes can be simulated, and a multitude of vital parameters can be simultaneously recorded and processed.

Figure 2. Concept of the measurement station.

The measurement station includes two polygraphy systems (Comlab 44, Schwarzer, München), an electrocardiograph (BT12, Corscience, Erlangen) and an electromyograph (Topas, Schwarzer, München) for multi-channel recording of bioelectrical potentials, such as EEG, EMG, EOG, ECG, evoked potentials and changes in the galvanic skin response. Moreover, signals such as pulse, oxygen saturation, respiration, respiratory effort, body position, body temperature, blood pressure, acceleration etc. can be recorded. Additionally, 16-channel wireless EMG recordings (TeleMyo 2400T G2, Noraxon, Scottsdale, Arizona) can be correlated to three-dimensional movements and movement patterns.

The station also includes a video camera to detect changes in posture and facial expression (SIMI Reality Motion Systems, Unterschleißheim), and an eye-tracking system (iView X, SensoMotoric Instruments, Teltow) with a high-speed camera to record eye movements as well as the size of the pupils and the eyes. Finally, by the use of a thermography system (Varioscan, Infratec, Dresden), information about body temperature and peripheral blood circulation can be gained.

3. REALIZATION

The measurement station was realized as described above. Figure 3 shows a part of the set-up with the driving simulator as the central part.

Figure 3. Realization of the measurement station.

All recorded signals are collected by a central control unit and can be processed with respect to different aspects, such as changes of vital parameters, exceeding of threshold values for individual signals, detection of microsleep, identification of vigilance oscillations and arousal states. Individual signals can also be connected in order to obtain further information. One example is the Pulse Transit Time (PTT), which can be calculated from the ECG and the pulse curve. Moreover, the pulse can be detected by the use of thermography recordings. By applying specially adapted processing algorithms, it is also possible to detect and identify different emotional states. In Figure 4, simultaneous polygraphic recording of different vital parameters is shown as an example of the possibilities of the system.

Figure 4. Simultaneous polygraphic recording of vital parameters.

The recorded parameters form a multidimensional feature space, which can be analyzed by the use of mathematical and statistical methods. By this means, it is possible to isolate the respective indispensable features which are relevant to identify a certain emotional state.

4. DISCUSSION

In this work, the successful design, development and set-up of a measurement station is presented, which can be used for a multitude of applications in the field of Ambient Assisted Living [4]. A special focus was put on mobile application in a car, therefore a driving simulator was used as the central part of the system.

The measurement station can be used to record a great variety of vital parameters and to identify the application-specific indispensable parameters. Thus, by developing suitable algorithms, the set-up provides the detection of emotional states. The use of such a system in a car bears the potential to increase road safety, as states such as fatigue or aggression can be detected and counteracted.

Future development of the presented system includes the optimization of the applied sensor technology, for instance the development of intelligent sensors (see Figure 5) including electronics for signal processing and wireless transmission [5], and capacitive, contactless electrodes [6].
Moreover, in order to use the measurement system for detection of feelings and emotions, the development of suitable algorithms will be a crucial issue in future work. With the set-up, it is possible to create stress situations or certain emotional states (e.g., anger). By evaluating the recorded biosignals, the characteristic features of these states can be identified. Techniques such as interviews and questionnaires can also be used to verify the identification of emotional states. By isolating the indispensable features relevant for emotion detection, it will be possible to reduce the required number of measurement signals. This will facilitate the integration of the system in a real car.

Figure 5. Wireless, self-organizing electrode compound.

5. ACKNOWLEDGMENTS
This work was supported by the European Union (Projects MINET (Measuring the Impossible NETwork, EU-FP6-NEST-043297) and Feel Europe (European Expert Platform for Measuring Human Feelings and Emotions, EU-FP6-NEST-043357)), by the German Federal Ministry of Education and Research (Project Seed (Self-organizing electrode network for electrophysiological diagnostis)), and by the Fraunhofer Society (Project InMuSens (Intelligent multimodal mobile sensor systems for applications in medicine, occupational safety and well-being)).

6. REFERENCES
ABSTRACT
In this paper, I describe open research questions associated with building a persuasive dialog system that can gain rapport with users by recognizing and synthesizing appropriate three dimensional emotions (activation, valence, and power). I also discuss why such an emotionally intelligent system is well suited for mobile devices, specifically those with voice input.

Categories and Subject Descriptors
H.1.2 [User/Machine Systems]: Human factors

General Terms
Human Factors

Keywords
mobile persuasion, persuasive technology, rapport, emotion, emotion dimensions, immediate response patterns, emotion recognition

1. INTRODUCTION
Since the early 90s affective computing has been used in many domains. For example, there have been systems that utilize emotion for customer support[3], providing tutoring [5], and persuasion[2, 7, 6].

While desktop PCs have taken advantage of emotional computing, research in mobile systems that use emotion is still lacking. Speech processing tools, for example speech recognizers, end of utterance detectors, and emotion recognizers, are becoming more accurate and their availability is reaching the mobile community. Embedded processors are becoming more advanced and are able to run applications that were before strictly for use on high end desktop PCs. The next step in mobile HCI is to focus on the application of better serving user needs based on emotion indicators such as prosody in voice.

One area of research that has recently received much attention is mobile persuasion as seen in the rise of venues such as MobilePersuasion since 2007. Fogg [4] mentions that the time and place that information is given is important for persuasion. Mobile devices are intrinsically suited for these types of persuasion because, unlike desktop computers, people have them wherever they go.

The rest of the paper is organized as follows. First I discuss my claims regarding the relationship between persuasion, rapport, and emotion. Next, I describe the development procedures of a persuasive dialog system and how such a system could be implemented for use in a mobile device.

2. CLAIMS

2.1 Emotional sensitivity is needed to gain rapport in mobile applications
When people interact with each other, they share a variety of nonverbal behaviors. Sometimes people that do not react to a speaker’s emotion can be seen as careless and even negative. For this reason, artificial agents must be able to determine user emotional state and react appropriately; some research is working on this. Since mobile devices are small, they are limited by their input methods. Voice is a common choice for these devices because it makes it easy for humans to use, without having to dedicate their hands, eyes, and others. For this reason, emotion sensitivity through voice is key for mobile applications. For example, in customer support applications [3], when a user sounds frustrated, the system will transfer the caller to a human.

Regarding gaining rapport, some may think that rapport is gained automatically after knowing someone for a long time. If this was the case with computers, people that use, for example, answering machines for long periods of time, would feel a sense of rapport with the machine. This is not the case. This means that something is missing; I think it is the ability for the machine to detect emotion and adapt accordingly. In this research I look to gain instant rapport, which means that an artificial agent is able to gain rapport with a user within a short period of time, in my case, within one interaction. Nonverbal behaviors that signal emotions may be key. Spoken dialog systems and artificial agents in general lack emotional intelligence and this makes their artificiality easy to spot.

2.2 Emotional responses are needed to gain rapport in mobile systems
In addition to detecting emotional states of users, it is also important to respond appropriately. If we are to create systems that exhibit more human-like interaction patterns, we must also build systems that respond to users with appropriate actions, dependent on their emotional state. Communication Accommodation Theory [9] says that people change
their nonverbal behaviors in order to reduce social distance. If there are nonverbal behaviors absent, it could be seen as negative. This may be why spoken dialog systems are seen by many, negatively. They seem without emotion, uncaring for the user’s feelings. When dealing especially with persuasion, rapport, gained through emotion, is a key element because it creates a closer relationship with users.

I argue that implementing a system that only adapts information based on user emotion is not suitable. Taking the example of the customer support systems, if the system would simply transfer a frustrated customer to a human person, this may not be enough. It would seem as if the customer was being put aside, and that only now that they are showing frustration, they will be taken seriously. It would be more effective if the spoken dialog system showed emotion throughout and then, when the user seemed frustrated, first apologized in a sympathetic voice, then transferred the user to a human.

2.3 Speech and Dialog is good for persuasion

When people engage in persuasive dialog, their speech is composed of arguments and counter-arguments. This is not possible in one-sided persuasion, such as billboards or even commercials. When persuasion occurs in dialog, people learn about the other person and tailor their arguments. This is probably why we still rely on salesmen, not just advertisements, to sell. We hear arguments tailored to us and colored with emotional prosody.

Although advertisements are good methods of persuasion, dialog may be better, even if the dialog is with a machine. With one-way arguments, people may have preconceived knowledge that may or may not be true. This cannot be dealt with in a one-way argument. Advertisements are effective, but they may mainly serve as a starting point, maybe to catch the interest of the person. Most people find it difficult to, for example, lose weight or commit to an exercise routine, unless they have someone (with whom they probably have rapport) that can encourage them to continuously practice their routine.

2.4 Three dimensions of emotion suffice for most dialogs

Using the three dimensional approach provides us with an easy way to characterize prosodic patterns. The three dimensions and their acoustic correlates, described in [8], are defined as follows: activation is activity in voice or sounding ready to take action. Mean pitch, intensity and speaking rate and others are mostly correlated with activation. Valence is either positive or negative, where positive valence is correlated mostly with lower pitch, large pitch range and others. Power is defined as sounding dominant or submissive. The acoustic correlates of power are similar to activation, except that lower pitch represents dominance (whereas higher pitch represents activity). Using the three dimensions, we can build recognizers and synthesizers and not have to account for every aspect of the spoken signal; instead they can focus on these limited correlates in the acoustic signal. This makes it possible to run emotionally sensitive applications on real-time systems and limited resource systems, such as cell phones.

It is known that the three dimensional approach does not account for every emotional category, in other words, there are some emotions that cannot be expressed using only the three dimensions. Also, there seems to be some overlap especially in the activation/power domains. However, it seems that it is more feasible to use the three dimensional approach than categorized emotions because there may be a large amount of emotions that are difficult to annotate as categories. Previous work that uses categories use a small set of broad representations of two to five emotions.

2.5 The most important emotional modeling is local, and can be captured by immediate response rules

My last claim is that instead of having to keep track of emotional state throughout the dialog, adequate information can be gathered from looking only at a small window of the dialog. This would allow for simple emotional modeling systems.

Contrary to choosing an optimal dialog act structure, where it is important to know user beliefs and attitudes learned since the start of conversation, emotion modeling requires much less context and can still be useful as seen in [5, 1, 10].

3. A PERSUASIVE DIALOG SYSTEM

In my work in progress, I use the three dimensional approach to detect emotional exchanges present during persuasive dialog (see [1] for more details). A corpus of 10 conversations between a graduate coordinator and students was analyzed. The corpus consisted of a graduate coordinator persuasively informing students about the graduate school option. The corpus was labeled using the three dimensions of activation, valence, and power, each with values from -100 to +100. Correlation coefficients showed some significant relationships between the emotions in coordinator’s responses and the emotion in the student’s previous utterance. The findings showed that in many cases, if the student sounded positive, the coordinator replied with a positive sounding voice. If the students sounded negative, then the coordinator also sounded negative. The authors also found that there was an inverse relation in the power dimension. If the student sounds dominant, then the coordinator will respond with more submissiveness, but if the student sounded submissive, the coordinator would display more dominance.

Next, more complex immediate response rules were extracted using machine learning methods and slightly better correlation coefficients were achieved. An emotion recognizer was built by extracting acoustic features from the speech signals and using them to predict the labeled emotion levels.

3.1 System realization

Several tools are available to collect speech and recognize both words and emotion in pseudo real-time environments. Here I discuss a dialog system that utilizes a user’s voice to assess emotion and spoken words, and responds with persuasive statements. Different from previous work, I plan to automatically choose an appropriate response and render that response with appropriate emotion through prosody.

Figure 1 shows a dataflow diagram for a persuasive dialog system. First the user speaks into a microphone, which is the receptor in, for example, a cellular phone, and the user’s words and emotional state are recognized by processing the spoken signal. For recognizing words, pocketsphinx is a viable solution because it is optimized for embedded pro-
Emotions are recognized by first computing pitch and energy using the lightweight C program known as dede. The speech is passed into a dialog manager which will decide on the lexical response. The lexical responses will be retrieved from a database. Several embedded databases packages exist that are fit for mobile devices, e.g. SQLite and VistaDB. This step could be distributed to a remote server if the database is too large. The emotion is given to the immediate response pattern module, which is a set of switch statements, in order to determine the appropriate emotional color to add to the lexical content based on the user’s current emotional state. Both the lexical content and the appropriate emotional coloring will be give to a speech synthesizer capable of producing emotional speech. A possible solution is MaryTTS, which can reside on a remote server and provide synthesized emotional speech as compressed audio.

Figure 1: A dataflow diagram for a system that attempts to gain rapport with users

Currently in progress is the evaluation of such a responsive system. Several informal tests have been conducted that suggest that the responsive system is perceived as more adaptive than a baseline system. The persuasive ability of the system will be examined in future work.

4. CONCLUSIONS

I believe that dialog systems should respond to what the user says and also to how the user says it. Emotional state given through prosody is essential to persuasion. Rather than concentrate on methods that use physiological state, face sensors, eye detection, and others, I believe there is valuable data that can be extracted in primary mode of communication in most mobile devices: voice. Mobile devices are at a technological point at which they must utilize emotion as a resource to provide better user experiences.

5. REFERENCES


Emotion Recognition on the Go: Providing Personalized Services Based on Emotional States

Syed Shahbaz Hussain
KTH (Royal Institute of Technology)
Stockholm, Sweden
sshus@kth.se

Christian Peter, Gerald Bieber
Fraunhofer IGD Rostock
Usability and Assistive Technologies
Rostock, Germany
cpeter,gery@igd-r.fraunhofer.de

ABSTRACT
Personalized products and services are the basic need to attract more customers and to gain a competitive advantage as most of the products available in the market place are similar in technical characteristics, quality, and price. Currently, most of the products and services don’t consider the emotional state of the users before offering personalized services which is a big drawback. User emotional state acts as an important context in perception and decision making, and hence should be taken into consideration for offering personalized services. This applies particularly in mobile scenarios, during which people rapidly change places and environments and hence experience rapid changes of emotionally stimulating situations. Adaption to this rapid change of the emotional state promises to be an efficient way to offer personalized emotion-aware services.

This paper reports about an ongoing project investigating the possibility of providing personalized services based on the emotional state of users. The model described in the paper uses emotion-related bio-data and physical activity as input to determine the likely emotional state of the user. Based on the computed emotional state, the model chooses an appropriate personalized service to provide. In order to evaluate the model, an experimental prototype is developed. Successful development and test of the model shows that emotional states can be determined with mobile technology and effectively utilized to offer better personalized products and services. The model can further lead to the possibility of developing emotionally intelligent interactive systems.

Categories and Subject Descriptors
Affective computing, intelligent interactive systems, mobile HCI

General Terms
Design and development of model, Experimental prototype, Human Factors, Personalized product and services.

Keywords

1. Introduction
Research carried out in the field of HCI (Human–computer interaction) shows that humans have an intrinsic affinity to communicate with computers in usual and social way just like they interact with other people in usual and social situations [4, 15]. Recognition of human emotion as part of affective computing has become an increasingly important field in HCI. The ability to recognize, understand and express emotions can play a vital role in human communication and increasingly in HCI [10, 12, 18]. The advantages of emotional computing are manifold. From a scientific perspective, emotions can play a vital role in human decision making, as well as in perception and learning. Furthermore, human emotions have the tendency to influence rational thinking and therefore can become a part of rational agents as proposed by artificial intelligence research [16]. Another possible use of emotions is on creating lively human–computer interfaces like avatars which comprise credible animations of interface agents, cf. [16].

With the advancement of mobile technologies and the need to provide personalized products and services to customers, emotion aware personalized services gain more importance as they are directly based on customer’s current emotional experiences. Customer emotion consideration can enhance the pleasure of buying, owning, and using the products and services offered to the customer. Considering the experiential or emotional quality of products also helps in gaining differential advantage in the marketplace as most of the products nowadays are similar with respect to technical characteristics, quality, and price.

Providing personalized services based on emotions is quite a difficult task faced by the researchers due to the high level of complexity associated with sensing users current emotional state. Human beings in order to assess a person emotion’s uses different information channels like monitoring a person’s recognizable bodily reactions (face, gestures, posture…), individual traits, and causal information context. But, on the other hand, computer systems cannot yet easily evaluate information like individual traits or associate environment information with observed behavior [18].

The paper presents a model for providing personalized services based on emotional state. The model computes the current emotional state of the user through affective computing techniques. The computed emotional state is utilized by the service used in the model to transform itself according to the current emotional state of the user. In order to evaluate the model an experimental prototype system is developed which will provide mobile device services based on the emotional state of the user.

The next chapter will present a short overview of the background and related work to the project. Chapter 3 introduces the model for providing emotion-based personalized services, while chapter 4 shortly describes the experimental system prototype implementing the model.
2. Background & Related Work

The concept of providing personalized product and services emerged from the idea of “Web Personalization” [9]. Web personalization aims to transform the websites to reflect user’s explicit and/or implicit interests and desires which motivates the idea of providing personalized product and services to customers. Personalization observes the user as a physical person, and it develops user profiles which consist of information or personal data about the user. More and more web stores and services use personalization to better serve their customers, Google and Amazon are just two examples.

Researches like Picard [12-14], and Essa and Pentland [7, 8] have been researching on the use of various techniques to deduce a user emotional state for providing personalized services directly based on human emotions. Various researchers have been working with computational perception techniques in order to predict human emotions by matching voice features or facial expressions with some prescribed expressions representing the current emotional state of the human (e.g. smiling, frowning, surprised, etc.) [1, 2, 7, 8, 10, 11]. Though the work is not a 100% correct way to predict human emotions, it clearly suggests the possible improvements and associated perception research by improving the quality of contextual information that can be collected. Picard's work on affective computing [14, 15] proposes a similar purpose of measuring human emotions using bio-electric signals feedback of physiological parameters like: skin conductance, skin temperature, heart beat rate, blood volume pressure (BVP), respiration, electroencephalography (EEG) and muscle tension in correlation to change in emotion. Picard's work is based on theories of emotion and cognition.

3. Model

3.1 Model Overview

The model for providing emotion based personalized services targets mobile devices such as mobile phones. Biosensors are used to measure emotion-related physiological parameters of the user and calculations are performed to deduce the emotional state of the person based on those measurements. The model classifies emotional states based on training data which can contain any emotional categorization like: Ekman’s basic emotions [5], emotional dimensions such as valence (pleasure) and arousal [19], or self-specified ones like: good, bad, ok. The model uses statistical methods to determine the current emotional state of the user. After having calculated the emotional state, the model chooses an appropriate adaptation and passes it to the mobile device service to provide against the measured emotional state of the user. All the collected data is provided with a time stamp and sent to the input preprocessing module.

3.2 Components of the Model

For easy processing of the data, the overall model is built using a modular approach. The figure below represents the different modules involved in the model followed by a brief overview for each module.

3.2.1 Emotional Data Acquisition

This module makes use of any biofeedback device for collecting physiological data, or even vision or sound based input devices like: cameras or microphones for collecting other emotion-related data. Collecting auxiliary data such as activity data of the person or environmental temperature is beneficial to gauge the sensor readings. Also, during the training phase, user input on the currently experienced emotion is acquired to find out the likely emotional state of the user. All the collected data is provided with a time stamp and sent to the input preprocessing module.

3.2.2 Input Preprocessing

During the emotion preprocessing phase, the collected data is synchronized based on their timestamp. User input is associated with emotion-related data like physiological readings, and auxiliary data. After the data has been preprocessed the module passes the preprocessed data to the training module and/or to the emotion recognition engine.

3.2.3 Training Module

The training module is responsible for training the model to recognize emotions based on the emotion-related sensor data (e.g. physiological readings), user input, and auxiliary data received from the input preprocessing module. A training sample consists of different sensor values along with auxiliary and user input data, for a particular emotional state. The collected training data is then processed by statistical algorithms to form emotion classifiers. The emotion classifiers are stored as templates in the emotion database for later use in the emotion recognition process.

3.2.4 Emotion Database

The model stores all the data in the database present on the server. There are separate tables for storing training data, and emotion-to-mobile device service mapping. The table for storing training data contains sensor readings, auxiliary data as well as user input data. The table for emotion classifiers stores necessary information in form of templates for emotion pattern matching. Table for emotion-to-mobile device service mapping stores the mapping of the mobile device service to provide against the measured emotional state of the user.

3.2.5 Emotion Recognition Engine

The emotion recognition engine is the heart of the model it makes use of the statistical approach for analyzing and classifying the
emotion. The recognition engine uses the classifiers generated in the training module to examine the input data. The input data is matched with the emotional templates stored in the database for a particular emotional state. In case of a match the suitable adaption command for that particular emotional state is sent to the mobile.

3.2.6 Classifiers
Multiple classifiers will be created in the training module; each classifier would be classifying the incoming data for particular emotional states. E.g. when going for states labeled “good”, “bad”, and “okay”, three classifiers would be created, one for “good”, one for “bad” and one for “okay”. Possibly a fourth could be used for “no emotion” for cases of no emotion being observable.

3.2.7 Mobile Transformation Module
The mobile transformation module is the output module of the model. Input for this module is the appropriate personalization command received from server as a result of measured emotional state of the user. Based on the received command the mobile device will transform itself accordingly to provide the personalized service based on the received emotional state.

4. Prototype Implementation
4.1 System Overview
The model has been tested by setting up a test prototype. The prototype system consists of three physical components:

i. Emotion sensing device for emotional data acquisition
ii. Mobile phone device collecting the emotional data and affective adaptation according to the received transformation command.
iii. Server which contains database for storing data, input preprocessing module, training module, and emotion recognition module.

The components communicate wirelessly with each other, either through Bluetooth or GSM/UMTS. It is also conceivable to use WLAN to make use of free WLAN Internet access for communicating the data. The figure below shows the overall prototype system and the interaction of components between them.

4.2 Data Acquisition
For gathering emotion-related data we use the Fraunhofer EREC System [6, 12]. EREC is a wearable device in form of a glove and allows online and offline access to physiology data. The version used in this project sends out physiological readings (skin temperature, skin resistance, and pulse), plus hand movement data via Bluetooth to any receiving device, which in our case is a mobile phone device. Figure 3 below shows the EREC system used in the project.

4.3 Input Preprocessing
The preprocessing of the input data received takes place at the server. The server receives all the data from the mobile device via GPRS/WLAN. During the emotion preprocessing, all the emotion related data is synchronized with user input data based on their timestamp. The input preprocessing module passes preprocessed data to training module and or to the emotion recognition engine for determining the current emotional state of the user.

4.4 Training Module
The training module is responsible for training the model so that it performs emotion recognition afterwards. The training modules receives training samples from input preprocessing module which consists of different sensor values for physiological data attributes along with movement, annotation and activity data for a particular emotional state. The collected training data is then processed by statistical algorithms to calculate different statistical values like: mean (m) and standard deviations (s) of all the sensors values for the current emotional state entered by user. The results are stored as emotion classifiers inform of templates for later use in the emotion recognition process.

Figure 2. Prototype system overview

Figure 3. The ERECBT system used.
4.5 Emotion Database
The model stores all of the data in the emotion database present on the server. We use a MySQL database in the project. The table for training data stores all the training data along with user current emotional state. The table for emotion classifiers stores e.g. mean (m) and standard deviations (s) of all the sensors for an emotional state along with the emotion id of a template to represent the particular emotional state. This table is used for emotion pattern matching. There is also a third table for emotion-to-mobile device service mapping which contains emotion id and the corresponding adaption to provide. The recognized emotion is matched with the emotion id present in this table to pick the suitable adaption against that particular emotional state.

4.6 Emotion Recognition Engine
The emotion recognition engine runs on the server and it classifies the emotion based on the received data. Emotion recognition engine classifies a particular emotion by matching the data with the emotion template stored in the emotion database created during the training phase. In case a match is found an appropriate personalization command is picked from the database against that particular emotional state and sent to the mobile device.

4.7 Mobile Transformation Module
The mobile transformation module is the output module of the model. The mobile device receives the appropriate personalization command from the server via UMTS and provides personalized service based on it. In our model depending on the current emotional state, the mobile device sets the appropriate ring tone for the user. The mobile device ring tone volume is set high for happy state, low for sad state and normal for normal emotional state of the user.

5. Conclusion
The core idea behind this project work is to present a model for providing personalized services and products based on the user current emotional state. The model measures the user current emotional state by collecting and analyzing emotion related data from the user. Based on the input sensor data the model determines the current emotional state of the user. Depending on that the model decides what type of personalized service should be offered to user. The service is offered in form of mobile device ring tone service and is directly based on user current emotional state.

Considering the customer emotion in providing personalized product and services on one hand can enhance the pleasure of buying, owning, and using the products and services offered to the customer while on the other hand considering the experiential or emotional quality of products can help in gaining differential advantage in the marketplace as most of the products now days are similar with respect to technical characteristics, quality, and price.

6. References
ABSTRACT
Mobile emotion measurement (MEM) through physiological signals is a promising tool for both experiments and application. We provide 1) an overview of unobtrusive physiological sensors and 2) a review of studies that have tried to infer emotions from physiological signals. This review shows that there is a lack of general standards, low accuracy, and a doubtful validity of the results. To overcome these problems, we provide three guidelines for future research on MEM: validation, triangulation, and a physiology-driven approach.

CATEGORIES AND SUBJECT DESCRIPTORS
H.1.2 [Models and Principles]: User/Machine Systems—Human factors; J.4 [Social and Behavioral Sciences]: Psychology; J.7 [Computers in Other Systems]: Consumer products; I.2.m [Miscellaneous]: []

GENERAL TERMS
Experimentation, Human Factors, Measurement, Performance, Reliability, Standardization

KEYWORDS
Emotion, Physiology, Wearable, Affective computing, Physiological computing

1. INTRODUCTION
Would it not be great if a computer could warn us when we are under too much stress, if a tutoring system could monitor a student’s attention, or if music could automatically be selected based on how we feel? These are typical examples of the next step in machine intelligence, which require an unobtrusive method for measuring one’s mental state [19]. One of the promising ways of measuring these mental states is through physiological signals. The physiological counter-parts of psychological phenomena have been researched for over a century. This has resulted in an enormous body of literature that describes physiological responses to all kinds of psychological states: mental workload, attention, pain, emotions, and dreams, to name but a few [4]. However, findings of studies manipulating psychological states and measuring physiological signals are typically inconsistent [10]. Nonetheless, the amount of research employing physiological signals with machine learning tools to predict mental states has exploded in the last decade; see also Table 2. Together with technological advances in physiological sensors, this leads the way to true mobile emotion measurement (MEM).

MEM would be of great benefit for mobile HCI. In the first place, it would enable real-time unobtrusive objective emotion measurements in experimental settings. This has advantages over subjective methods like questionnaires, which can only be done post-hoc and are very obtrusive. Additionally, it is doubtful whether subjective emotion reports always reflect the actual emotion. In contrast to physiological measurements, they are not free from social masking.

Currently, several wearable devices are being developed that can conduct physiological measurements in a unobtrusive, real-time fashion. This enables physiology-based MEM, as opposed to facial affect recognition that cannot be done through wearable devices. Moreover, they can be used anytime, where emotional speech processing only works in situations where one is speaking [19]. Hence, for studies using a mobile setting, emotions are best captured by physiological signals. Furthermore, MEM has not only methodological advantages, it also provides numerous opportunities for mobile applications. Possible applications include an affective mp3 player, continuous emotion communication, atmosphere creation, or even emotional jewelry. In turn, this might prove essential in realizing true ambient intelligence and forms next step in wearable and mobile computing [19].

In this paper, we will first provide an overview of mobile devices that measure physiology. In addition, we give an overview of state-of-the-art emotion prediction from physiological signals. We show that this prediction has not reached a satisfying level for MEM. Furthermore, we identify several methodological shortcomings of the studies done so far. To overcome these problems, we provide three guidelines that will help to further the development of successful MEM.
Table 1: General concerns with mobile emotion measurement (MEM).

1) Affective signals are typically derived through non-invasive methods to determine changes in physiology and, as such, are indirect measures. Hence, a delay between the actual change in emotional state and the recorded change in signal has to be taken into account, especially with mobile measurements.

2) Mobile measurements make physiological sensors sensitive to movement artifacts and differences in bodily position.

3) Most sensors are obtrusive, preventing their integration in real world applications.

4) Affective signals are influenced by (the interaction among) a variety of factors [4]. Some of these sources are located internally (e.g., a thought) and some are among the broad range of possible external factors (e.g., a signal outside). This makes affective signals inherently noisy, which is prominent in mobile measurements in real world environments.

5) Physiological changes can evolve in a matter of milliseconds, seconds, minutes or even longer. Some changes hold for only a brief moment, while others can even be permanent. Although seldom reported, the expected time windows of change are of interest [19]. In particular since changes can add to each other, even when having a different origin, as is often the case with mobile measurements.

6) Affective signals have large individual differences. This calls for methods and models tailored to the individual. It has been shown that personal approaches increase the performance of affect recognition; e.g., [3].

2. THE STATE-OF-THE-ART
A broad range of affective signals are used in affective sciences. Over the last decade, several unobtrusive devices for the processing of such signals have been developed. For instance, sensors for heart rate measurements have been integrated into a chair [1]. Furthermore, Healey and Picard [6] integrated heart rate and skin conductance measurements into a car. Ark et al. [2] have developed skin conductance sensors in a mouse, and Paulos et al. [13] combined wearable wireless heart rate sensors in a wrist band. For a broad range of techniques, applications, and discussions concerning unobtrusive emotion measurement, we also refer to [20]. This shows the wide variety of possibilities for integrating physiological sensors in all kinds of everyday objects in an unobtrusive manner. After the signals have been captured, they have to be processed in real-time. When processing such signals, some general issues have to be taken into consideration, as are denoted in Table 1.

Typically, studies attempting to predict mental states using physiological signals conduct an experiment in which participants are brought into distinct mental states. A wide range of physiological signals are monitored from which numerous features are extracted. After feature extraction, machine learning techniques are employed to see if correct mental state predictions can be made based on the extracted features. See Table 2 for a concise review of such studies.

As illustrated by Table 2, a variety of physiological signals and machine learning techniques have been explored. Nonetheless, both the recognition performance and the number of emotions that the classifiers were able to discriminate are disappointing. Moreover, comparing the different studies is problematic because of the different settings the research projects in an unobtrusive manner. After the signals have been integrated into a chair [1]. Furthermore, Healey and Picard [6] integrated heart rate and skin conductance measurements into a car. Ark et al. [2] have developed skin conductance sensors in a mouse, and Paulos et al. [13] combined wearable wireless heart rate sensors in a wrist band.

For a broad range of techniques, applications, and discussions concerning unobtrusive emotion measurement, we also refer to [20]. This shows the wide variety of possibilities for integrating physiological sensors in all kinds of everyday objects in an unobtrusive manner. After the signals have been captured, they have to be processed in real-time. When processing such signals, some general issues have to be taken into consideration, as are denoted in Table 1.

Typically, studies attempting to predict mental states using physiological signals conduct an experiment in which participants are brought into distinct mental states. A wide range of physiological signals are monitored from which numerous features are extracted. After feature extraction, machine learning techniques are employed to see if correct mental state predictions can be made based on the extracted features. See Table 2 for a concise review of such studies.

As illustrated by Table 2, a variety of physiological signals and machine learning techniques have been explored. Nonetheless, both the recognition performance and the number of emotions that the classifiers were able to discriminate are disappointing. Moreover, comparing the different studies is problematic because of the different settings the research was applied in, ranging from controlled lab studies to real world testing, the type of emotion triggers used, the number of target states to be discriminated, and the signals and features employed. Moreover, the general concerns are often disregarded, as denoted in Table 1. To conclude, there is a lack of standards, low prediction accuracy, and inconsistent results. For MEM to come to fruition, it is eminent to start dealing with these issues. This illustrates the need for a set of guidelines for MEM, as is done in the next section.

3. GUIDELINES
We identify three guidelines for MEM: 1) validity; 2) triangulation of measurements, and 3) a physiology-driven approach.

3.1 Validity
In the pursuit to trigger emotions in a more or less controlled manner, a range of methods have been applied: actors, images (IAPS), sounds (e.g., music), (fragments of) movies, speech, commercials, games, agents / serious gaming / virtual reality, real world experiences, and reliving of emotions. However, how to know which of these methods actually triggered participants’ true emotions? This is a typical concern of validity, which is a crucial issue for MEM. Validity can be best obtained through four approaches: content, criteria-related, construct, and ecological validation, as we will discuss in this section.

Content validity refers to a) The agreement of experts on the domain of interest; e.g., limited to a specific application or group of patients; b) The degree to which a feature (or its parameters) of a given signal represents a construct; and c) The degree to which a set of features (or their parameters) of a given set of signals adequately represents all facets of the domain. For instance, employing only skin conductance level (SCL) will lead to a weak content validity when trying to measure emotion, as SCL is known to relate to the arousal component of an emotion, but not to the valence component. However, when trying to measure only emotional arousal, measuring only SCL may form strong content validity.

Criteria-related validity handles the quality of the translation from the preferred measurement to an alternative, rather than to what extent the measurement represents a construct. Emotions are preferably measured at the moment they occur, as is feasible with MEM. However, how to know which of these methods actually triggered participants’ true emotions? This is a typical concern of validity, which is a crucial issue for MEM. Validity can be best obtained through four approaches: content, criteria-related, construct, and ecological validation, as we will discuss in this section.

Content validity refers to a) The agreement of experts on the domain of interest; e.g., limited to a specific application or group of patients; b) The degree to which a feature (or its parameters) of a given signal represents a construct; and c) The degree to which a set of features (or their parameters) of a given set of signals adequately represents all facets of the domain. For instance, employing only skin conductance level (SCL) will lead to a weak content validity when trying to measure emotion, as SCL is known to relate to the arousal component of an emotion, but not to the valence component. However, when trying to measure only emotional arousal, measuring only SCL may form strong content validity.

Criteria-related validity handles the quality of the translation from the preferred measurement to an alternative, rather than to what extent the measurement represents a construct. Emotions are preferably measured at the moment they occur, as is feasible with MEM. However, measurements before (predictive) or after (postdictive) the particular event are sometimes more feasible; e.g., through subjective questionnaires. The quality of these translations are referred to as predictive or postdictive validity. With emotion measurement, this is especially relevant for obtaining a reliable ground truth. The closer the ground truth measure to the actual emotion, the more reliable it becomes. A third form of criteria-related validity is concurrent validity: a metric for the reliability of measurements applied in relation to
the preferred standard. For instance, the more emotions are discriminated, the higher the concurrent validity.

A construct validation process aims to develop a nomological network, or possibly an ontology or semantic network, build around the construct of interest. Such a network requires theoretically grounded, observable, operational definitions of all constructs and the relations between them. Such a network aims to provide a verifiable theoretical framework. The lack of such a network is one of the most pregnant problems physiological emotion measurement is cop- ing with. A frequently occurring mistake is that emotions are denoted, where moods (i.e., longer object-unrelated affective states with very different physiology) are meant. This is very relevant, as it is known that moods are accompanied by very different physiological patterns than emotions are. Moreover, different signals relate to different emotional properties. For instance, arousal is strongly related to skin conductance and valence is thought to be reflected by heart rate variability.

Ecological validity refers to the influence of the context on measurements. As emotions are easily contaminated by contextual factors, using a similar context as the intended application for initial learning is of vital importance. Hence, emotion measurements done in controlled laboratory settings, are poorly generalizable to real-world applications. Hence, the need for MEM to make longitudinal real-world studies possible is pressing.

### Table 2: A summary of 11 studies that have tried to infer a mental state from physiological signals. They all employed a similar approach: first, a certain mental state (e.g., stress, certain emotions, mental workload) is induced in participants, while a number of physiological signals are measured. Subsequently, a variety of features is extracted and pattern recognition and machine learning techniques are employed to enable the automatic classification of the emotional states.

<table>
<thead>
<tr>
<th>Source</th>
<th>Signals</th>
<th>Features</th>
<th>Selection/Reduction</th>
<th>Classifiers</th>
<th>Target</th>
<th>Result</th>
</tr>
</thead>
<tbody>
<tr>
<td>[14]</td>
<td>C,E,R,M</td>
<td>40</td>
<td>SFS, Fisher</td>
<td>LDA</td>
<td>8 emotions</td>
<td>81 %</td>
</tr>
<tr>
<td>[12]</td>
<td>C,E,S</td>
<td>3</td>
<td></td>
<td>kNN, LDA</td>
<td>6 emotions</td>
<td>69 %</td>
</tr>
<tr>
<td>[16]</td>
<td>C,E,B</td>
<td>18</td>
<td></td>
<td>SVM</td>
<td>6 emotions</td>
<td>42 %</td>
</tr>
<tr>
<td>[9]</td>
<td>C,E,S</td>
<td>10</td>
<td></td>
<td>SVM</td>
<td>3 emotions</td>
<td>78 %</td>
</tr>
<tr>
<td>[11]</td>
<td>C,E,S</td>
<td>12</td>
<td></td>
<td>kNN, LDA, ANN</td>
<td>6 emotions</td>
<td>84 %</td>
</tr>
<tr>
<td>[5]</td>
<td>G</td>
<td>3</td>
<td>PCA</td>
<td>ANN</td>
<td>4 emotions</td>
<td>90 %</td>
</tr>
<tr>
<td>[22]</td>
<td>C,G,S,P</td>
<td>11</td>
<td></td>
<td>SVM</td>
<td>2 stress levels</td>
<td>90 %</td>
</tr>
<tr>
<td>[21]</td>
<td>C,E</td>
<td>20</td>
<td>ANOVA</td>
<td>SVM, ANN</td>
<td>2 fun levels</td>
<td>70 %</td>
</tr>
<tr>
<td>[8]</td>
<td>C,E,M,R</td>
<td>15</td>
<td></td>
<td>SVM, ANFIS</td>
<td>4 affect states</td>
<td>79 %</td>
</tr>
<tr>
<td>[18]</td>
<td>E,M</td>
<td>10</td>
<td>ANOVA, PCA</td>
<td>kNN, SVM, ANN</td>
<td>4 emotions</td>
<td>61 %</td>
</tr>
</tbody>
</table>

Notes: C: Cardiovascular activity; E: Electrodental Activity; R: Respiration; M: Electromyogram; S: Skin temperature; P: Pupil Diameter; ANN: Artificial Neural Network; RT: Regression Tree; BN: Bayesian Network; SVM: Support Vector Machine; LDA: Linear Discriminant Analysis; kNN: k Nearest Neighbors; ANFIS: Adaptive neuro-fuzzy inference system; PCA: Principal Component Analysis; SFS: Sequential Forward Selection.

3.2 Triangulation

We propose to adopt the principle of triangulation on MEM, as applied in social sciences and human-computer interaction. This may deal with the noisy physiological signals inherent to MEM. For example, movement artifacts and corruption due to other signals can be major problems.

Heath [7] defines triangulation as “the strategy of using multiple operationalizations of constructs to help separate the construct under consideration from other irrelevancies in the operationalization”. Using this strategy provides several advantages: 1) Distinct signals can be used to validate each other; 2) Extrapolations can be made based on multiple data sets, providing more certainty. In turn, corrections can be made to errors in a result set that clearly defy from other results; and 3) More solid ground is obtained for the interpretation of signals, as multiple perspectives are used.

Triangulation was, for example, successfully employed by [3], who showed that combining physiological signals and facial expressions leads to better predictions than using one of them. Also, [19] showed that the combination of a physiological parameter (i.e., heart rate variability) and speech parameters can provide more robust emotion recognition than either of them separately. Hence, we advise to record multiple affective signals, as is facilitated through MEM. Moreover, qualitative and subjective measures should accompany the signals; e.g., questionnaires, video recordings, interviews, and Likert scales. Systematic, well-controlled research exploring the plethora of possible affective signals should increase the grip on the meaning of these signals.

3.3 A physiology-driven approach

A third guideline stems from the idea that physiological emotion measurement can never be entirely based on psychological changes. As discussed, there are many factors outside one’s affective state that contaminate affective signals. Besides validation and triangulation, a physiology-driven perspective could be taken to deal with this [17].

Instead of expressing goals of MEM directly in terms of affective states, they can often be stated in terms of the affective signals themselves. For instance, instead of inferring an air-traffic controller’s stress level, thresholding skin conductance level might be sufficient. Note that there always remains an
interpretation in affective states. Then, the use of syntactic or structural pattern recognition should be explored. Its hierarchical approach to simplifying complex patterns in affective signals could be valuable for MEM.

4. CONCLUSION
This paper described MEM through physiological signals and explained the lack of its success. Next, three guidelines were introduced from which MEM is expected to benefit significantly: validation, the principle of triangulation, and a physiological-driven approach.

With the guidelines provided and the future’s progress ahead, we envision embedding of MEM in various professional and consumer settings, as a key factor of our daily life; cf. [19]. MEM fits the ambient intelligence vision perfectly. Although combining wearable and intelligent devices into smart environments is a great challenge, we strongly believe this holds a great promise for future technology and lifestyle. Would it not be an appealing idea to live in an empathic surrounding that adapts to your mood and emotions, which can even calm you or can help you concentrate when required?

Acknowledgments: The authors thank Joyce H.D.M. Westerink and Marjolein D. van der Zwaag, both Philips Research, The Netherlands.

5. REFERENCES
Possibilities of Psychophysiological Methods for Measuring Emotional Aspects in Mobile Contexts

Eva Ganglbauer*, Johann Schrammel*, Arjan Geven*
*CURE - Center for Usability Research and Engineering
Hauffgasse 3-5, 1110 Vienna, Austria

Manfred Tscheligi‡
‡HCI-Unit, ICT&S Center,
University of Salzburg
Sigmund-Haffner-Gasse 18, 5020 Salzburg, Austria

{ganglbauer, schrammel, geven, tscheligi}@cure.at

ABSTRACT
Traditional methods to access the emotional experience of a user such as subjective reports have certain disadvantages. Participants have to be asked for their feelings and emotions which interrupts the process of experience and flow. Psychophysiological methods offer data throughout the process of emotional experience, which unfolds new possibilities for user experience (UX) evaluation. In this paper we provide a short overview of applied psychophysiological methods for human computer interaction (HCI) and the findings of our examinations for mobile motion contexts. Our outcomes will be discussed considering the possibilities, challenges and feasibility of these methods in the area of interaction with systems and emotions. Based on our experience we think that psychophysiological measurements provide important possibilities for applications in the field and can help to deepen and expand the insights gathered by traditional methods.

Keywords
User Experience, Psychophysiology, Emotion, Human Computer Interaction, Evaluation, Movement;

INTRODUCTION
Over the past years there has been increasing interest in emotional aspects of mobile applications. Pleasurable mobile products have to offer more than functionality and usability, they address other aspects such as aesthetics, beauty, or playability, which enrich our experience with interactive systems. In recent years the field of HCI and industry acknowledged the design and architecture knowledge that stimulating experience of a product plays an essential role for mobile UX. However trying to “measure the experience” is a nontrivial task by definition and also regarding a practical approach. UX is embedded in a situational, temporal, individual and product context and difficult to comprehend [12]. Experience with technology as characterized by McCarthy and Wright consists of a sensual, an emotional, a compositional and a spatio-temporal thread [15]. Hence, accessing the emotional state of users is crucial for developing satisfying mobile products that are rich in experience, although emotion is only one out of more aspects.

Methods to Assess Emotions
Traditional methods include questionnaires, interviews, narrative techniques and contextual inquiry. In general, these methods are based on self-reporting and record a user’s perception, such as sensual evaluation [10]. The added value of these methods is the possibility of an insight into the person’s feelings and preferences. Other possibilities are observational techniques and video analysis, where the evaluation of the material is a very long and laborious process. Moreover, the intersubjectivity and validity of this approach is hard to guarantee. Beside the described techniques, there are approaches to automate expression recognition of speech, faces or gestures [7]. Combinations of these automatic expression recognition methods result in a broader spectrum of emotions that can be detected [5, 8]. The problem that remains unsolved is how these methods support accessing emotions in a mobile context.

Psychophysiological Methods in HCI
Psychophysiological recordings have been shown to be valuable approaches for measuring valence (a quality for positive and negative emotions) and arousal throughout the process of an experience. Asking a participant about its emotional state is crucial, but interrupting the flow of interaction and experience, which is not necessary when psychophysiological methods are employed. Additionally, psychophysiological measurements can help uncovering social masking, which means people often say positive things about a product because they hesitate to be negative. Furthermore, it’s possible to analyse data for special situations during the evaluation, e.g. the moment when a participant won a game. Psychophysiological methods support the analysis of certain crucial situations of an experience that are essential for emotional experience, but also provide summative analyses over time [11]. Subjective
reports are prone to the fact that emotions are not always easy to put into concrete words and small but important details are sometimes forgotten.

Psychophysiological methods have some disadvantages as well. They are still costly and complex to apply and people are fitted up with cables and electrodes and therefore restricted in moving and acting freely and without restrictions. Making it more difficult, the setting may cause arousal or different emotions in people and therefore alter the results. Much time is needed to postprocess and interpret all the data that has been acquired during evaluation. Nevertheless, the possibility of obtaining data throughout the whole test is very advantageous. To give more insight into the world of psychophysiology, the following section shall give a short overview of methods we applied for emotional evaluation:

Electromyography (EMG) measures muscle activity by detecting surface voltages that occur when a muscle is contracted. To find out about positive emotions the activation of the *zygomaticus major* muscle, which is activated while laughing, is recorded. Simultaneously, negative emotions are measured by the *corrugator supercilii* muscle, which is activated while frowning. EMG was used in a lot of studies to access the valence of emotions [3, 11, 14, 15]. On the one hand, EMG is more accurate than facial expression recognition with video analysis, because low evocative emotions are difficult to recognize visually. On the other hand, sensors with cables are attached in the face, which is obtrusive for participants.

Electrodermal activity (EDA) measures the activity of the eccrine sweat glands and is said to be a linear correlate to arousal [4]. Although room temperature, humidity, participants activities and the correct attachment of the electrodes has to be carefully considered, tonic EDA is a well researched and valid method to record arousal and was used for measuring emotions for interaction with systems [19, 15].

Respiration can be used as measurement for negative valence and arousal [7]. More important, changes in respiration rate affect other psychophysiological metrics such as EDA or cardiovascular functions.

The cardiovascular system offers several measuring options to determine valence or arousal: Blood Volume Pressure (BVP) indicates a correlation between greater dilation in the blood vessels with less arousal [19]. The heart rate (HR) is correlated with arousal as well and variability of the heart rate (HRV) is used as a metric for assessing the positive or negative valence of an experience [1]. HRV is also used as a measure for mental workload. Nevertheless, measuring HR can raise privacy and intimacy issues, as traditional electrocardiography requires the attachment of electrodes in the chest area.

It’s important to mention that these methods should not be applied unimodal. A multimodal approach is more accurate and results in a broader spectrum of aspects of emotions, but has the disadvantage that multiple channels have to be combined, analysed and finally interpreted. Every method has its strengths and weaknesses, also strongly depending on the evaluation context.

All these described methods seem to have great potential to be used in UX research, but the question remains whether they can be used in a mobile context. Mobility demands people to be in motion, and this has not only effects on the applied methods. Common sense and results from UX research suggests that the interaction with products, tools, and artefacts can be enriched by allowing people to move naturally and unrestrictedly [17, 2]. If people express themselves with their whole body, they immerse into another world more naturally and easily. There is evidence that there is a strong relation between movement and emotions [13]. With the following report we try to shed some light on the world of psychophysiology for mobile UX research.

**AN EXPERIENCE REPORT**

Implementing psychophysiological methods has to be done very carefully and with great care due to the many variables that can alter the results. Temperature, humidity, attachment of electrodes, individual differences, differences concerning gender (women even differ depending on the menstrual cycle), age, time of the day, consumed stimulants such as coffee or energy drinks, medicaments, drugs, etc can cause different reactions in sensors and in people. Therefore the former consumption of stimulants from test participants has to be clarified. Care has to be taken for sensor attachment as well: the skin should be shaved if very hairy and not have stains of makeup and skin creme.

In order to access the practicability of such approaches we tried several methods, including EMG, EDA and Respiration, using the ProComp Infiniti System from Thought Technology. Generally, it’s a simple task to attach the electrodes and recording the signals compared to the difficulty of psychophysiological signal processing and interpretation of the signals. The signals have to be interpreted but also why participants reacted the way they did in certain situations.

Due to the motion aspect in mobile situations, we conducted research on movement and psychophysiological measuring methods. Our findings and examination of methods suggest that facial EMG is a viable and reliable method to measure positive or negative emotional states, even when participants are moving. Analysis of the signals clarifies explicitly when certain muscles are activated or not (figure 1 and figure 2). Fortunately talking doesn’t activate the muscle responsible for laughing in general. Although it can happen that people activate the *zygomaticus* during mimic expressions.

EDA was proven as a viable method to access arousal in several studies. In a movement context, the number of peaks is increasing the more a participant is in motion (see
electrodermal activity in figure 1 and figure 2). Therefore it is very hard to tell the difference between greater arousal because of the movement or enhanced emotions. Our current approach is to record a “movement baseline” to calculate the difference, but further research has to be conducted to shed some light on this matter.

Figure 1: Zygomaticus, Currogator and Electrodermal Activity without movement

Figure 2: Zygomaticus, Currogator and Electrodermal Activity with participant moving

From our point of view, psychophysiological methods should be supplemented by subjective reports. We used EmoCards, which represent eight faces distributed over the valence arousal space [6], asked participants about discrete emotions such as happiness and anger, and conducted qualitative interviews.

Our next steps will be to employ EMG, EDA and Respiration really in the field. The ProComp Infinity System can store the data also on Flash Memory.

CONCLUSIONS

Based on our first hand experience we think that psychophysiological methods are very well applicable in gaming and entertainment industry, as during gameplay or watching a movie intense emotion and their dramaturgy are a crucial element. This fact makes it easier to analyse and interpret the emotional scope. Psychophysiological measurements of the user’s emotions and experiences can help to finetune gameplay and plot composition in the process of developing mobile entertainment systems. However, psychophysiological measuring techniques are not appropriate for all contexts and the decision of which methods to choose has to be adapted in accordance to the system or product that has to be evaluated. Most important of all is to consider the setting, the system and the context before choosing which psychophysiological methods to apply. Not every method is suitable for evaluation of a certain type of interactive system, as invasiveness, privacy and other aspects of psychophysiological methods should be taken under consideration.

A lot of studies aim at assigning discrete emotions (anger, happiness, fear, pleasure, etc.) from psychophysiological recordings. From our point of view emotions are so complex themselves and vary individually and culturally, so that it’s vague to create discrete models. Moreover, discrete emotions are not clear to differentiate on a psychophysiological level, and an emotional experience consists of more than one emotion. For those reasons we think that psychophysiological methods are inappropriate for measuring discrete emotions.

Sooner or later psychophysiology will hopefully work non-invasively (advanced video analysis, speech recognition and other refined future technologies) and hence offer great opportunities for mobile emotional experience evaluation. Though, also other areas of interest such as affective computing would benefit from such technical solutions. During an interaction with an affective system the emotional state of users could be determined and the system will react to it appropriately. This could help to enhance technology acceptance for human robot interaction, increase the UX or even a learning process with a system. An example for non-invasive measuring is Anttonen’s EMFi chair (a chair equipped with electromagntical film), where attachment of an electrode is reduced to the ear to record the heart rate [1]. HR would also be potentially interesting for mobile UX, as there are special systems for conducting the heart rate of patients in the field of cardiology. These systems could be used for measuring HR(V) of participants.

Regarding mobile contexts, it’s crucial to provide methods that don’t restrict users in moving, feeling and interacting freely.

Psychophysiological methods are not ideally applicable for such contexts because movement of participants alter the signals such as increased heart rate and electrodermal activity. Furthermore, almost all electrodes and sensors are sensitive to movement. Despite of these drawbacks, our findings suggest that facial EMG is a viable and reliable method to measure positive or negative emotional states, even when a participant is in movement. We are working on the implementation of other methods such as EDA as well,
if possible to get viable results even for moving participants.

Psychophysiology in the field of mobile UX is in its infancy and further research is necessary to meet future challenges. We think that EMG is ready to be implemented in certain areas such as the mobile entertainment and gaming industry. At the moment we consider psychophysiological methods as a valuable complement to qualitative and quantitative subjective reports and observational analyses. It has to be taken into account, that psychophysiological methods are restricted in their scope. They enable to measure certain aspects, but not the holistic emotional experience in all its complexity. Further research has to be done to improve these methods and develop non-invasive technologies.

REFERENCES


MOLMOD: Analysis of Feelings based on Vital Information for Mood Acquisition

Jumpei Yamamoto†
Graduate School of Media and Governance, Keio Univ.
jum-p@ht.sfc.keio.ac.jp

Mizuki Kawazoe
Graduate School of Media and Governance, Keio Univ.

Kazunori Takashio
Faculty of Environment and Information Studies, Keio Univ.
Graduate School of Media and Governance, Keio Univ.

Jin Nakazawa
Faculty of Environment and Information Studies, Keio Univ.

Hideyuki Tokuda
Faculty of Environment and Information Studies, Keio Univ.
Graduate School of Media and Governance, Keio Univ.

ABSTRACT
We often sense the mood around us, for example, in a cozy restaurant, in a pleasant park, in a tense classroom, and so on. Such mood can affect our behavior, thus is important to build context-aware ubiquitous computing applications. However, there is no common model to recognize mood in the real world. In this paper, we propose the metrics to represent mood, and a system, called MOLMOD, for mood acquisition. We define mood as "the mass of feelings that we have in a particular place". MOLMOD distinguishes the feelings based on human vital information. The paper shows our study of the feeling-based mood calculation, and the initial prototype system that uses wearable vital sensors to acquire feelings. It then shows, based on experiments, that our feeling-based approach can actually represent the mood in several different places.

Categories and Subject Descriptors
H.1 [Models and Principles]: User/Machine Systems—Human information processing; J.4 [Social and Behavioral Sciences]: Psychology

Keywords
mood of a certain location, psychology of emotion, human vital information, ubiquitous computing, context aware.

1. INTRODUCTION
We often sense mood in places in our daily lives, and it can be used as one of the factors to extract context of users in the places. For example, if a computer system can detect the mood of coffee shops (e.g., relaxed, cozy, noisy, etc) around a user in a city, the system might recommend one based on the user’s desired mood. If the system can detect the mood of a meeting (e.g., tense, active, etc), the system could record the mood as a piece of life log information. If it can sense the mood of a home (e.g., fun, pleasure, etc), the system would control home information appliances, such as a music player, automatically to augment the mood. Such a mood detection system, thus, is needed to create human-centered context-aware applications.

There have been several works to this extent. Our previous work is on a mood modeling and visualization system using vital information[1]. It classifies user’s feelings in four different colors on a map using vital data. "SHOJI" is another system that visualizes mood[2]. It expresses the mood by analyzing user’s emotion, and visualizes the emotion and environmental data using lights of a terminal. Another study tries to acquire the mood using voice information in a meeting[3]. ComSlipper measures emotion, presence, and users’ thought, and sends messages to other persons[4]. However, these researches acquire mood in limited situations. That is to say, we have not established a system that can acquire mood in daily situations.

In this study, we focus on availability of moods and try to construct a new technique for mood acquisition. We define that mood is described by proportions of feelings that people have in a certain situation. In this paper, we propose a system called "MOLMOD" that models mood of a place with feelings of humans gathered by vital sensors attached on humans. We then evaluate correctness of the definition, and suitability of the technique to acquire feelings. In the rest of the paper, we define mood in Section 2, and propose the system in Section 3. We describe the evaluation in Section 4, and conclude the paper in Section 5.

2. DEFINITION OF MOOD
This section shows the abstraction of mood, and basic algorithm to detect it.

2.1 Abstraction of Mood
There are a number of factors that influence the mood. For example, environmental conditions, such as the brightness, the temperature, and the noisiness, may affect the mood
2.2 Mood Description by Feelings

There are many theories [5][7][8][9] about classification of feelings. Among them, we use Russell’s circumplex model for our feeling model. Russell explains that each of us monitors the current psychological state of other people in our face-to-face encounters [9]. We inevitably judge how happy or unhappy, and how sleepy and agitated others appear. The model shows how a feeling can be classified by these two dimensions. Based on the Russell’s idea, we extract the following eight feeling as the major dimensions to calculate moods (feeling model: Fig3): surprise, excitement, happiness, calmness, sleepiness, depression, sadness, and stressfulness.

Based on the feeling model, we define mood as proportions of the eight feelings of all people in a place. The calculation is conducted as follows. First, each dimension of the feeling model has a discrete scale. Second, a person’s feeling is detected along the pleasure and arousal dimensions. We use skin temperature and heart rate sensors for the detections respectively. Section 3 details how we do this. Let us suppose, for example, a case where a user’s happiness (pleasure dimension) is 6 and surprise (arousal dimension) is 4. Then, as shown in Fig.4, the distance between the origin and the point, and those from the nearest two axes are calculated. We then acquire the feeling proportion; in this case, the system recognizes user’s feeling as 5.34 excitement and 1.88 happiness. Finally, mood is calculated as a proportion of the feeling of all people in the place. After calculating these feelings proportions, it sum up the data to acquire the mood (Tab.1). In this way, the mood is described by proportions of 8 feelings.

3. MOLMOD SYSTEM

The major issue in this study is detecting feelings of people. This section describes a system called "MOLMOD: MOod Labeling and MODelling based on vital information". It uses wearable vital sensors, and translates sensor readings into the value in the feeling model.

3.1 Feeling Distinction using Vital Information

MOLMOD system uses vital information to judge arousal and pleasure levels. These years, research on emotions has revealed the relationship between changes in emotion and vital information. Studies have shown that skin temperature rises when we have a positive feeling, and declines when we have a negative feeling [6]. A pulse increases on tense/ excited feeling, and decreases on relaxed/sad feeling [7][8]. We thus use skin temperature and pulse to detect pleasure and arousal levels respectively. The system maps skin temperature and pulse, respectively, on the horizontal and vertical axes of the feeling model as shown in Fig.5. The origin of the graph represents a person’s normal values of skin temperature and pulse. It calculates the normal value of pulse from the average of a day, and that

<table>
<thead>
<tr>
<th>Stress: 5%</th>
<th>Calmness: 25%</th>
</tr>
</thead>
<tbody>
<tr>
<td>Excitement: 10%</td>
<td>Depression: 2%</td>
</tr>
<tr>
<td>Happiness: 28%</td>
<td>Sadness: 5%</td>
</tr>
<tr>
<td>Surprise: 5%</td>
<td>Sleepiness: 20%</td>
</tr>
</tbody>
</table>

Table 1: Mood Data
of skin temperature from the average of temperature difference between skin and the air. The actual value of arousal and pleasure levels at a certain moment are achieved from the difference between the recorded values and the normal ones. The difference values are discretized with the unit of two beats for the pulse, and 0.18 Celsius degrees for the skin temperature.

3.2 Wearable Vital Sensors
To reduce user’s stress, the system distinguishes feelings using wearable vital sensors. Recent years, many researchers have tried to analyze user’s feeling using vital information\cite{6}\cite{7}\cite{8}. Some of them used the change of vital information using large vital sensors, and other used cameras to analyze facial expression. However, these technique limits user’s action in that the users are forced to carry large sensors, or they need to be in the camera’s fixed angle. Therefore, we use wearable vital sensors because the system should be able to recognize feelings in our daily lives.

We chose to use an RF-ECG made by Medical Electronic Science Institute Co., Ltd.\cite{Fig.7} for pulse detection. RF-ECG can measure electrocardiogram, skin temperature, and accelerate. It sends sensor data to a PC where a software calculates the average value of a minute and send the data to our system. We also use SunSPOT made by Sun Microsystems Inc.\cite{Fig.6} with an additional air temperature sensor, which is used to calculate the temperature difference between skin and the air. SunSPOT inputs the sensor readings and transmit them every 10 seconds to the PC where our system is running. The system calculates the average of temperature difference every minute.

4. EVALUATION
We evaluate the definition of mood and the technique of feelings acquisition.

4.1 Relationship between Feeling and Mood
In this paper, we abstracted a mood as mass of feelings of people in a place. We first evaluate correctness of this abstraction.

4.1.1 Experiment Methodology
We have done the experiment in the following four different places in Japan: at the Miura beach, in a windsurfing school, in a pastry shop, and at an author’s home. During this experiment, we ask each examinee about his/her current feeling, and map it on the feeling model. We then calculate the proportions of eight feelings from the gathered feelings data. The proportions mean the mood in the corresponding place. Examinees are finally asked to rate the accuracy of the calculated mood on a scale of one to ten. We use average ratings as the metric of correctness. Higher value thus means our abstraction of mood correctly represents the real world.

4.1.2 Results and Discussion
Figure 8 shows the mood observed at Miura beach. The beach is closed (swimming is prohibited due to high waves) at the moment of examination. We observed much happiness, calmness, and excitement, and a little depression and sadness. Similarly, Figure 9, 10, and 11 shows the mood observed in other places. Table 2 shows the result of evaluation done by the examinee in terms of the correctness of mood shown above. The number of examinees are shown in arcs, while the former number is those attended to feeling observation, and the latter is those attended to mood evaluation. The method worked well in the Miura beach where many people exist. Therefore we can conclude that the proposed abstraction of mood well represents the real world in most cases.

4.2 Feeling Acquisition
In this section, we evaluate our system in terms of feeling acquisition using vital sensors. We examine the use of two
Table 2: Accuracy of Moods

<table>
<thead>
<tr>
<th>Situation</th>
<th>Average of Results</th>
</tr>
</thead>
<tbody>
<tr>
<td>Miura Beach</td>
<td>8.50(19, 8)</td>
</tr>
<tr>
<td>Windsurfing School</td>
<td>9.67(2, 3)</td>
</tr>
<tr>
<td>Pastry Shop</td>
<td>7.83(5, 6)</td>
</tr>
<tr>
<td>Home</td>
<td>4.00(2, 3)</td>
</tr>
</tbody>
</table>

Table 3: Mood Color

- Stress: Black
- Calm: Green
- Excitement: Pink
- Depression: Gray
- Happy: Yellow
- Sad: Blue
- Surprise: Red
- Sleepy: Water Blue

Figure 12: Daily Mood by Daily Ave.
Figure 13: Daily Mood by Music Ave.

4.2.1 Experiment Methodology
This experiment involves only one examinee, and collect his skin temperature and pulse at his laboratory office for two days. The system is examined with the following two types of calibrated normal values: long-term calibration and short-term calibration. Long-term (Daily Ave.) is the average of vital sensor readings in a day, and short term (Music Ave.) is the average of those while listening to the music for 10 minutes. The former contains values that are resulted by multiple different practical activities of the examinee, while the latter contains those resulted by asking him to be calm during the calibration. Throughout the test, the system acquired 5 moods in different situations; daily mood, before the submission of bachelor’s thesis, playing the game, sleeping, and watching the video. Readers should be noted that experiments with more examinees is in progress.

4.2.2 Result and Examination
Figure 12 shows the spread of daily moods using the long-term (daily) calibration, while Figure 13 shows that using the short-term calibration. Table 3 shows color mappings. First, one can see that the amount of observed moods differ according to the normal value. The average skin temperature differs between Daily Ave. and Music Ave., therefore these methods make difference to the pleasure dimension. Also, we can’t define validity of calculation for normal vital information through the test. For example, the system with the method of Daily Ave. recognizes stress at the situation of the submission. However, Music Ave. have high validity in the other situation. In addition, feeling distinction lacks in validity, because sad feelings were recognized when the examinee felt happy watching the video. We discuss their causes as follows.

- Acquisition of user’s normal data
  User’s normal skin temperature value differs between 1st and 2nd method, so normal data influences the results of acquisition mood. However, validity of 2 different normal values and discuss the effectiveness of the use of vital sensors.

5. SUMMARY
We proposed the MOLMOD system as a first step for mood acquisition. This system distinguishes user’s feeling, and acquires the mood in proportion of eight feelings. We evaluated the accuracy of the definition that mood is the mass of feelings. In addition, we discussed the issues of feeling distinction. We are currently evaluating further on feeling acquisition using vital sensors with more examinees. We also consider the use of other information to acquire feelings such as heart rate variability, acceleration, and facial expression.

Acknowledgement
This research is partially supported by the Ubiquitous Service Platform Research of Ministry of Internal Affairs and Communications, Japan.

6. REFERENCES
ABSTRACT

A multimethod approach for capturing a broad perspective of user experience (UX) of mobile media content and technology is presented in the current paper. In this research design, a variety of standardized research instruments are implemented and complemented by structured instant recall (SIR) interviews and/or unstructured interviews for targeting in-depth qualities of users’ subjective mobile media experiences as well as usability issues. Methods selected and explored for measuring emotional experiences are described in further detail. The ambition is to create a toolkit of pragmatic evaluation methods, simultaneously targeting both usability and users’ actions, reactions, attitudes and affect of different mobile media content and technology.

Categories and Subject Descriptors

H.5.2 [User interfaces] Evaluation / methodology
H.1.2 [User/Machine Systems]: Software psychology

General Terms

Human Factors, Measurement

Keywords

User experience, usability, evaluation methods

1. INTRODUCTION

The joy-factor is a key to success for almost any media solution today. Not to say that usability is not important. However, it is merely not enough. Media users of the 21st century are critical in their judgments of systems and services they choose to use. They take functionality for granted, but also entertainment; if something does not entertain them, they simply go somewhere else. Hence, a media encounter needs to give rise to a good user experience (UX). However, there is still no official definition of UX [10]. In fact, there are several theories and models describing and defining the concept of UX. Battarbee [1] divides these into person-centered frameworks that focus on human needs and/or the relationship people have with products, product-centered frameworks that focus on quality of the design, i.e. product attributes, and interaction-centered frameworks that focus on the actions of the user. Despite the lack of a unified model or definition of the concept, there seems to be a general agreement that UX is something more than mere usability [12]. It extends beyond technical and usability-oriented aspects (i.e. product-centered), into human emotions and needs (i.e. person-centered). The concept of UX emphasizes the totality of emotion, motivation, and action in a given physical and social context [14].

Users’ experiences of media products and media content are affected by both product-centered aspects, such as functionality and aesthetics, as well as person-centered aspects, such as personal motivation and expectations [11].

Hence, learning about users’ media experiences is a powerful tool, and an essential determinant for improving product design [4], including products designed for mobile use. We have recently witnessed an increase in the demand from the industry for testing the UX and usability of mobile media products (i.e. mobile media content and technology) and we are currently involved in a development project where usability and UX of a mobile application for personal health care is tested. Accordingly, our current commercial and scientific ambition is to see how different research methods and instruments can be synchronized and fine-tuned for measuring users’ attitudes, affect, actions, and reactions from mobile media encounters. Our approach is to both illuminate different aspects of UX in mobile media related contexts, but also to see how different methods and research instruments can be utilized and triangulated. We want to know what lies behind our choices of using or not using mobile media products, to gain a broader and deeper understanding of the essence of UX of mobile media interaction, and also which is the best way of measuring mobile UX. The development work includes exploring and developing valid methods for conducting commercial and scientific research of UX of mobile media in laboratory studies as well as in-situ.

The current paper presents a multimethod approach, which is explored in our attempt to develop standardized research procedures and methods for testing both product experiences and emotional experiences of mobile media content and technology. This approach has successfully been utilized in several commercial studies investigating the usability and/or UX of both print and web based media content (unpublished). Selected research methods targeting emotional experiences are also described in further detail.

2. AMULTIMETHOD RESEARCH DESIGN

As our perspective of UX is holistic, we have developed a standardized procedure for investigating both usability and user experience with a wide range of validated methods and instruments. Earlier studies conducted at our audience research lab (iDTV Lab) have confirmed that a combination of methods guarantees a broad and deep understanding of how humans experience media interactions [9]. By combining different methods, it is possible to study a broader range of research
questions and to produce a more complete picture of UX. It is also possible to provide stronger evidence for a conclusion through convergence and corroborate findings, to increase the generalizability of the findings, and to complement any weaknesses of a single method [2, 5].

Our UX research design includes methods targeting both quantitative and qualitative data about attitudes, affect, actions, and reactions of media users. The selected methods can further be categorized into subjective and objective measurement techniques. The subjective measures include standardized questionnaires based on Likert-scales, questionnaires with open-ended questions, unstructured interviews, and various forms of structured and semi-structured interviews that are based on instant recall techniques. The objective measures include psychophysiological data collection such as monitoring heart rate, and skin conductance. Objective measures also include behavioral data recordings (e.g., eye movements, screen recordings, and recording the person within the physical environment while interacting with a product/media solution). Figure one presents an outline of chosen methods and how these are distinguished regarding what they measure and what kinds of data they generate. The three arrows in the figure point to the targets of SIR (Semi-structured Instant Recall) interviews: attitudes and affect (mostly questionnaire items), reactions (psychophysiological measuring), and actions (behavioral observations). We emphasize the development and fine-tuning of selected research methods, especially the triangulation of methods for answering specific research questions in accordance with demands.

The chosen methods for measuring emotional experience are employed and combined according to the following procedure: Participants’ behaviors and psychophysiological reactions are first monitored during task performance after which the questionnaires are filled in. The interviews are then carried out. Here, the SIR (Semi-structured Instant Recall) interviews, in which each participant is subjected to instant recall stimulated by their own replies regarding their emotions as well as their bodily reactions and behavioral actions, complement the standardized research instruments by targeting more in-depth qualities of users’ emotional experiences. In an Attitude & Affect-Targeted Interview, for instance, an in-house developed computer-based research tool, IRIS (Instant Reduction of Items into Scales), sums up the scaled items of the selected questionnaires, and calculates the means of scales. These means signal the direction of the areas that need to be improved in a product. It further provides us with a useful instrument to structure interviews by, in which a large number of questionnaire variables are down-sized into an easily manageable number. By using the data gathered with the questionnaires, we can target interviews in order to gain a deeper understanding of users’ experiences. An Action & Reaction-Targeted Interview, on the other hand, is based on psychophysiological data (heart rate and skin conductance) and behavioral data (activities). Hence, the targets are actions, as well as reactions. Video-recordings of the user’s actions and reactions during the test situation are used as a stimulus for this part of the SIR interview. Here the in-house developed research tool eValu8 is used, which allows a simultaneous interpretation of video-recordings/screen recordings and psychophysiological reactions. These interviews are again video-recorded in order to get as much information as possible for the analyses of stimuli of reactions and actions. The instant recall interviews and triangulation of data are ways of validating findings from a wide selection of objective and subjective measurements.

The technical research equipment (including psychophysiological measurement device, eye-tracking camera and other) and the methods selected for studying usability and UX at our audience
research lab are portable and can be employed in field settings. However, the mobility of mobile media poses new challenges regarding methods and procedures for collecting data on user experience and usability. Consequently, one of the ambitions of our research and method development is to investigate the applicability of the subjective methods described in this paper in natural mobile user contexts.

3. CONCLUSION

In the present paper, we described a combination of objective and subjective instruments for measuring the emotional experience of mobile media interaction, along with a standardized research procedure. However, as we see emotional experience as one of several aspects of UX, a multimethod approach for investigating both usability and UX of mobile media by employing a wide battery of research instruments that cover attitudes, affect/emotions, psychophysiological reactions, and users’ actions/behaviors was also presented. The main goal of our method development is to guide product development, so that the end-product answers to the needs and demands of targeted users, as in the development project of a mobile application for personal health care mentioned above. By exploring the implementation of the variety methods described earlier in a mobile media context (both in experimental and natural settings), we expect to gain a broader and deeper understanding of the users’ experiences and usage of mobile media content and technology. The multimethod approach enables validation, triangulation and refinement of selected methods, which might result in a new methodology for measuring mobile media experiences. The employment of multiple methods is further a way of verifying research results, which is essential in the collaboration with development projects of mobile applications where audience testing is carried out at several phases of the design process. However, the applicability of existing UX methods needs to be carefully explored in the field of mobile media interaction [3], as the existence of different frameworks of UX and the complexity of the concept along with the diverse and dynamic nature of mobile media consequently have an impact on how different aspects of UX of mobile media can be measured.

4. REFERENCES


Combining Worthless Sensor Data
Cornelia Setz, Johannes Schumm, Claudia Lorenz, Bert Arnrich, Gerhard Tröster
ETH Zurich, Wearable Computing Laboratory
Gloriasstrasse 35, 8092 Zurich
setz@ife.ee.ethz.ch

ABSTRACT
Previous work on emotion recognition from physiology has rarely addressed the problem of missing data. However, data loss due to artifacts is a frequent phenomenon in practical mobile applications. Discarding the whole data instance if only a part is corrupted results in a substantial loss of data. To address this problem, we investigated two methods for handling missing data: imputation and reduced-feature models using ensemble classifier systems. The five emotions amusement, anger, contentment, neutral and sadness were elicited in 20 subjects by film clips while six physiological signals (ECG, EMG, EOG, EDA, respiration and finger temperature) were recorded. Results show that classifier fusion increases the recognition accuracy in comparison to single classifiers using imputation by up to 16.3%. We were able to analyze 100% of the data even though only 47% of the data was artifact free. Since more artifacts are expected in a mobile environment than in the laboratory, the proposed methods are especially beneficial for mobile settings.

Categories and Subject Descriptors
G.4 [MATLAB]

General Terms
Algorithms, Measurement, Experimentation.

Keywords
Ensemble classifier systems, classifier fusion, emotion recognition, missing data, artifacts, physiology

1. INTRODUCTION AND MOTIVATION
Applications for emotion recognition are predominantly found in the field of Human-Computer Interaction (HCI). By including emotions, HCI shall become more natural, i.e. more similar to human-human interactions where information is not only transmitted by the semantic content of words but also by emotional signaling in prosody, facial expression and gesture. Several research groups have therefore employed pattern recognition methods in order to automatically distinguish between different emotional states of the user. Modalities which have been used to detect emotions include facial expression (e.g. [1]) speech (e.g. [8]) and physiological signals ([5], [6], [7], [9]).

Our work originates from the European research project SEAT which aims at increasing comfort for airplane passengers by developing a smart seat which is able to detect the emotional state of the passenger. Since the subjective appraisal of comfort is very important, emotion recognition can help in determining appropriate adjustments of the seat, the environment, the entertainment system or the attendance of the stewardess for each passenger individually. We therefore integrated physiological sensors into an airplane seat in order to infer the emotional state of the passenger.

When equipping the airplane seat with sensors, we aimed at minimizing discomfort induced by the sensor attachment. These unobtrusive sensors are generally more prone to artifacts and therefore lead to a lower signal quality. This trade-off is not restricted to our airplane seat but is relevant for any system to be used in real life. When moving from the static laboratory setting to mobile real-life, a lot of “worthless sensor data” is expected. Missing values therefore represent a serious problem for practical applications. However, this problem has so far gained little attention in emotion recognition from physiology. It is current practice to discard the corresponding episodes when encountering artifacts in the data. As an example, 20% of the data in [7] was partly corrupt and had to be discarded. In our research, we are investigating methods for handling missing values to reduce the amount of data which needs to be discarded during run-time. By one of our experiments we show how all the data in a data set can be used even though only 47% of the episodes are free of artifacts.

In the following section we present common kinds of artifacts which can be detected by plausibility analyses. In section 3 we describe methods for handling this artifact affected data. An emotion recognition experiment is described in Section 4 and the results are presented and discussed in Section 5.

2. ARTIFACTS
Artifacts are a common problem encountered when investigating physiological signals. There are different kinds of artifacts. Physiological signals may be corrupted by power line interference, motion artifacts or electrode contact noise. These artifacts can be encountered in laboratory experiments, but they will become far more problematic in mobile settings.

An example of a corrupted EDA signal is shown in Figure 1. Finger movement renders the signal unusable.

Figure 1. Artifacts in EDA signal due to finger movement.
Since we aim at recognizing emotions in everyday life, the employed sensors need to be unobtrusive. However, there is usually a trade-off between unobtrusiveness and signal quality of a sensor: the less obtrusive a sensor, the lower the expected signal quality. The RWTH Aachen has developed a novel comfortable ECG system which measures the ECG capacitively without skin contact, even through clothing [14]. Such a system was integrated in our airplane seat’s backrest. Figure 2 shows a comparison between the ECG signal of the comfortable system and the normal ECG signal (measured simultaneously with wet electrodes on the chest). The comparison illustrates the trade-off between comfort and signal quality: the signal of the contactless (unobtrusive) system cannot be used when the subject moves while the normal ECG is less affected.

Figure 2 also shows how plausibility analyses can determine whether the calculated R-peaks are trustworthy (marked by a green circle) or not (marked in red). By comparing each calculated RR-interval with the RR-intervals in its vicinity, implausible intervals are detected and corrupted sections are identified in a reliable way. This method is based on G.D. Clifford’s work [2]; refer to [3] for Matlab Code available under the GNU public license.

Figure 2. Movement artifacts in normal and contactless ECG.

3. METHODS
This section describes methods for handling missing values in emotion classification. We consider these methods especially promising for mobile applications.

3.1 Handling missing values
In [12], several courses of actions for handling missing data are summarized: Discard instances (in a mobile setting, this means to discard a large amount of data), acquire missing values (infeasible in automatic emotion recognition), imputation or employing reduced-feature models. The latter two are investigated in the following. Imputation uses an estimation of the missing feature or of its distribution to generate predictions from a given classifier model. A simple possibility used in practice comprises the substitution of missing values by the corresponding mean value of the uncorrupted samples.

Imputation is needed if the applied classifier employs a feature whose value is missing. The reduced-feature models technique represents an alternative approach: Instead of imputation, a new model (i.e. a new classifier) is trained which employs only the available features. A simple way to create a reduced-feature model is to use ensemble classifier systems, as described in the following subsection.

3.1.1 Ensemble classifier systems
Ensemble classifier systems consist of several classifiers whose decisions are fused to arrive at a final decision [10]. The idea behind is to consult “several experts”. It can be compared to the natural behavior of humans to seek a second (or third) opinion before making an important decision. The classifier ensemble system used in this work consists of one classifier per signal modality; each signal modality thus represents an “expert”. In case of encountering a missing feature in a certain signal modality, no classifier is trained for that modality and the final decision is only based on the modalities which are artifact free.

Numerous ways to combine the decisions of several classifiers exist. We have chosen and evaluated the following two simple approaches:

1. Majority voting: The class that receives the highest number of votes is chosen. If several classes receive an equal number of votes, the corresponding classifiers are identified as candidate classifiers and the classifier with the highest confidence among them provides the deciding vote.

2. Confidence voting: The class of the classifier with the highest confidence is chosen.

For the two described classifier fusion schemes, a suitable underlying classifier is needed. Linear and Quadratic Discriminant Analysis (LDA and QDA) with diagonal covariance matrix estimation were chosen since they need only few parameters to be estimated. The estimated posterior probability for the selected class was taken as confidence measure for the classifier fusion.

4. EXPERIMENTS
This section describes one of our emotion recognition experiments to show the performance of the discussed methods for handling missing data. The full experiments and results have been accepted for publication in [13].

The emotions to be recognized were chosen according to the well-known 2-dimensional emotion model of arousal and valence often used in emotion recognition studies [4]. One emotion in each quadrant plus neutral were selected as shown in Figure 3: amusement (high arousal, pos. valence), anger (high arousal, neg. valence), contentment (low arousal, pos. valence), sadness (low arousal, neg. valence) and neutral (medium arousal, zero valence).

Film clips were chosen for emotion elicitation because they are capable of eliciting strong emotional responses under standardized conditions and generate a rather long stimulation (1-10 minutes) [11].
For each of the five emotions, one film clip was selected as shown in Table 1. The used film clips are either suggested in [11] (marked with * in Table 1), proposed in another paper (John Q from [6]) or self chosen.

Table 1. Film sets used in the study: Films marked with * originate from [11]. John Q was suggested in [6].

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Film Clip</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sadness</td>
<td>John Q</td>
</tr>
<tr>
<td>Amusement</td>
<td>When Harry met Sally*</td>
</tr>
<tr>
<td>Neutral</td>
<td>Fireplace</td>
</tr>
<tr>
<td>Anger</td>
<td>Cry Freedom*</td>
</tr>
<tr>
<td>Contentment</td>
<td>Winged Migration</td>
</tr>
</tbody>
</table>

Twenty (12 male, 8 female) participants were recruited for the study. The five film clips were presented to the subjects by video glasses (see Figure 4) which were not taken off during the whole experiment. After each film clip, a three minute recovery period followed. The screen went black and the subjects were requested to clean their mind and calm down.

4.1 Signals and Features

The following physiological signals have been recorded: The Electrocardiogram (ECG), the vertical component of the Electrooculogram (EOG) to detect the blinking rate, the Electromyogram (EMG) of the Musculus Zygomaticus (muscle between mouth and eye, contracted when smiling), the tonic and phasic part of the Electrodermal activity (EDA), the respiration and the finger temperature.

Respiration and EDA were recorded with custom electronics integrated into the airplane seat. The respiration was measured by a stretch sensor attached to the seat belt (see Figure 5). The electronic circuits for the EDA directly filter and split the signal into a tonic and a phasic part.

The ECG, the vertical EOG, the EMG of the Musculus zygomaticus and the finger temperature were measured by the commercial Mobi device of TMS International [15].

A total of 53 features was calculated from the six signals. The features were computed for the periods of the film clips (emotion phases) and for the preceding recovery periods. The features calculated during the recovery periods served as baseline values for calculating relative features by division or subtraction.

The recorded data set contained a rather large set of corrupted segments, due to technical problems and artifacts. The artifacts included: (1) EDA signal reached saturation of the amplifier during the course of the experiment; (2) R-peaks in ECG signal were erroneously detected because of a high T-wave or due to motion artifacts; (3) blinking was not visible in EOG signal due to dry skin. The problematic data sets were identified and declared as invalid. This resulted in 96% correct data for EOG, 78% for EDA, 64% for ECG and 100% for the remaining modalities. The percentage of data containing valid features of all modalities amounted to 47%.

5. Results and Discussion

In this section we present how imputation techniques and classifier fusion methods handle the case of missing feature values. The classifiers presented in the following were evaluated in a leave-one-person-out cross-validation and all the six signal modalities were used for this analysis.
Table 2. Comparison of ensemble classifier systems and single classifiers (with imputation) for LDA and QDA. The results are presented for 5 and for 4 classes (neutral excluded).

<table>
<thead>
<tr>
<th>Fusion method</th>
<th>Classifier</th>
<th>Imputation</th>
<th>Acc. 5 classes</th>
<th>Acc. 4 classes</th>
</tr>
</thead>
<tbody>
<tr>
<td>None</td>
<td>LDA</td>
<td>Yes</td>
<td>45.0%</td>
<td>52.5%</td>
</tr>
<tr>
<td>Maj. Voting</td>
<td>LDA</td>
<td>No</td>
<td>41.0%</td>
<td>55.0%</td>
</tr>
<tr>
<td>Confidence</td>
<td>LDA</td>
<td>No</td>
<td>47%</td>
<td>58.8%</td>
</tr>
<tr>
<td>None</td>
<td>QDA</td>
<td>Yes</td>
<td>35%</td>
<td>40%</td>
</tr>
<tr>
<td>Maj. Voting</td>
<td>QDA</td>
<td>No</td>
<td>49%</td>
<td>56.3%</td>
</tr>
<tr>
<td>Confidence</td>
<td>QDA</td>
<td>No</td>
<td>48.0%</td>
<td>56.3%</td>
</tr>
</tbody>
</table>

The results for all 5 classes and for 4 classes (all except neutral) are presented in Table 2. Clearly, classifier fusion always yields a considerable benefit in comparison to single classifiers using all the features (with imputation). In the best case (4 classes, QDA) the increase in accuracy amounts to 16.3%.

Another interesting observation is that confidence voting always performs better than majority voting for LDA, while the tendency is reversed for QDA. The best classifier fusion technique thus seems to depend on the underlying classifiers.

A problem of our majority voting scheme was identified during this analysis: Since the majority voting is initially performed without considering the confidence values, it can happen that several “weak” classifiers agree on a wrong decision and thereby dominate another classifier which might exhibit high confidence value. A possibility to circumvent this problem would be to use the confidence values as weights during the majority voting.

5.1 Conclusion

In practical applications, data loss due to artifacts occurs frequently. Many of these artifacts can be detected automatically by plausibility analysis (e.g. unrealistic RR-intervals). Often the artifacts do not occur in all physiological signals simultaneously and discarding all the data instances containing invalid features results in a substantial amount of usable data. In our experiment, more than half of the data would have been lost if no strategy to handle missing values had been employed. With the proposed methods we were able to analyze 100% of the data of the laboratory experiment. In a mobile environment, we anticipate even more artifacts than in the laboratory. We therefore expect our methods to be beneficial especially for mobile settings.

Two methods for handling missing features in combination with two classifier fusion approaches have been investigated. Classifier fusion has been shown to significantly increase the recognition accuracies. A maximum increase in accuracy of 16.3% was observed when comparing an ensemble classifier system to a single classifier using imputation. Whether majority or confidence voting performs better depends on the underlying classifier.

6. ACKNOWLEDGMENTS

This project is sponsored by the European Commission DG H.3 Research, Aeronautics Unit under the 6th Framework Programme under contract Number: AST5-CT-2006-030958.

7. REFERENCES


