The Human Face of Mobile

Hajar Mousannif, Ismail Khalil

The Human Face of Mobile

Hajar Mousannif[1], Ismail Khalil[2]
[1] Cadi Ayyad University, LISI Laboratory, Faculty of Sciences Semlalia, B.P. 2390, 40000, Marrakesh, Morocco
mousannif@uca.ma
[2] Johannes Kepler University, Institute of Telecooperation, Altenberger Strasse 69, A-4040Linz, Austria
Ismail.khalil@jku.at

Abstract.
As the landscape around Big data continues to exponentially evolve, the « big » facet of Big data is no more number one priority of researchers and IT professionals. The race has recently become more about how to sift through torrents of data to find the hidden diamond and engineer a better, smarter and healthier world. The ease with which our mobile captures daily data about ourselves makes it an exceptionally suitable means for ultimately improving the quality of our lives and gaining valuable insights into our affective, mental and physical state. This paper takes the first exploratory step into this direction by using the mobile to process and analyze the "digital exhaust" it collects to automatically recognize our emotional states and accordingly respond to them in the most effective and "human" way possible. To achieve this we treat all technical, psycho-somatic, and cognitive aspects of emotion observation and prediction, and repackage all these elements into a mobile multimodal emotion recognition system that can be used on any mobile device.[1].

Keywords: Emotion Recognition, Affective Computing, Context, Pattern Recognition, Machine Learning, Reality Mining, Intelligent Systems.

1 Introduction

Who among us never yelled at his/her mobile phone, and cried or laughed franticly at it? Who among us never wished his/her mobile phone was able to react in one way or another to his/her anger, sadness, disgust, fear, surprise or happiness? Sifting through the rubble of the huge amount of data (Big data) our mobiles are collecting about us to automatically recognize and predict our emotional states is definitely one of the very innovative fields of current research with many sorts of interesting implications.

Mobile phones are part of our everyday life. They have the ability to get inside our heads and position our bodies [1]. They can even impact the way we define our identity and therefore represent an exceptionally suitable means for improving the quality of our life. A broad range of functions are already available on mobile phones nowadays, ranging from the basic functions of calling and texting, to more advanced ones such as entertainment and personal assistance. Mobile phones also have many modalities that are equivalent to human senses, such as vision (through cameras), hearing (through microphones), and touch (through haptic sensors). They, however, lack the human ability of recognizing, understanding and expressing emotions in an effort to support humans emotionally whenever and wherever they need it.

A variety of technologies especially in the fields of artificial intelligence and human computer interaction consider and implement emotions as a necessary component of complex intelligent functioning. Many emotion recognition systems which process different communication channels (speech, image, text, etc.,) as well as the outputs from bio-sensors and off-the-shelf sensors of mobile terminals have been developed. Research on integrating many modalities into a single system for emotion recognition through mobile phones is still at a very early stage, not to mention a shortage of literature on real applications that target providing personalized services that fit users' current emotional states.

In this paper, we design a multimodal emotion recognition system that aims at integrating different modalities (audio, video, text and context) into a single system for emotion recognition using mobile devices. We also design a method, we refer to as EgoGenie, which automatically recognizes the emotional states of mobile phone users and reacts to them by serving up personalized and customized content and service offerings based on their emotional states.

While all agree that interpreting human emotional cues and responding accordingly would be highly beneficial, addressing the following problem areas is a challenging task:

[1] The system and method described in this work were registered as Patent Application Ref. 36049 at the Moroccan Office of Industrial and Commercial Property OMPIC (jun. 26th 2013)

1. How can we represent and model emotions?
2. How can we extract the valuable and insightful information that best conveys the emotional states of mobile phone users from the Big data they generate?
3. How can we process users' data despite the limitations and constraints of mobile technology?
4. How to accordingly and "humanly" respond to the identified emotional states of the users?

The rest of the paper is structured as follows. Section 2 presents the emotion-theoretical background and sheds light on the importance of emotions in human's way of reasoning and its decision-making activities. Section 3 reviews some related work in the area of emotion recognition using mobile technology. We classify it into 5 categories: 1°) Emotion recognition in speech, 2°) Facial emotion recognition, 3°) Affective information detection in text the user types, 4°) Emotion recognition from the context around the user, and 5°) Multimodal emotion recognition where many communication channels or modalities are used. In section 4, we describe our proposed multimodal system for emotion recognition. Section 5 introduces our proposed method for emotion recognition and personalized content/service delivery and presents some implementation examples. Conclusions and directions for future work are presented in section 6.

2 Theoretical Background

Over the last years, research on emotions has become a multidisciplinary research field of growing interest. For long, research on understanding and modeling human emotion has been predominantly dealt with in the fields of psychology and linguistics. The topic is now attracting an increasing attention within the engineering community as well. More particularly, emotion is now considered as a necessary component of complex intelligent functioning, especially in the fields of artificial intelligence and human computer interaction. It is well argued that for systems to adapt the human environment and to communicate with humans in a natural way, systems need to understand and also express emotions in a certain degree [2]. Of course, systems may never need to have all of the emotional skills that humans have but they will definitely require some of these skills to appear intelligent when interacting with humans.

Unlike physical phenomena such as time, location, and movement, human emotions are very difficult to measure with simple sensors. Extensive research has been conducted to understand and model emotions. Two famous models have been adopted by the scientific world in order to classify emotions and distinguish between them: the basic emotions model (or a mix of them) and the 2-dimensional Arousal-Valence classification model. While the former tends to classify emotions as a mix of some basic emotions (such as anger, sadness, disgust, fear, surprise and happiness), the latter aims at representing emotions on two axes: valence, ranging from "unpleasant" to "pleasant", and arousal, ranging from "calm" to "excited". We argue that an emotion cannot be properly investigated without the analysis of the following 3 components (Figure 1):

1. The cognitive component: which consists in classifying and understanding the environment, the context, and attended situations that triggered the emotion.
2. The physical component: which includes focusing on the physical response (facial expressions, speech, postures, verbal/non verbal behavior etc.) and physiological response (heartbeats, blood volume pressure, respiration, etc.) that co-occur with the emotion or immediately follow it.
3. The outcome component: This involves analyzing the impact of the emotion, e.g. on behavior, on social communications, on decision making, on performance, on activities, etc.

Additionally, emotion assessment can be exploited in at least two levels: Either as a tool in evaluating attractiveness, appreciation and user experience of a service or product or in bringing the machine closer to the human by making the machine recognize, understand and express emotions in an effort to naturally interact with humans. Many real life applications are making use of emotion assessment in both levels. One could refer to [3] for some examples. Those incorporating mobile technology are the ones that are of much interest to us. Examples include improving mobile education and learning [4], mobile healthcare [5], stress detection [6], predicting and preventing task performance degradation [7] and avatar realism [8] among many others.

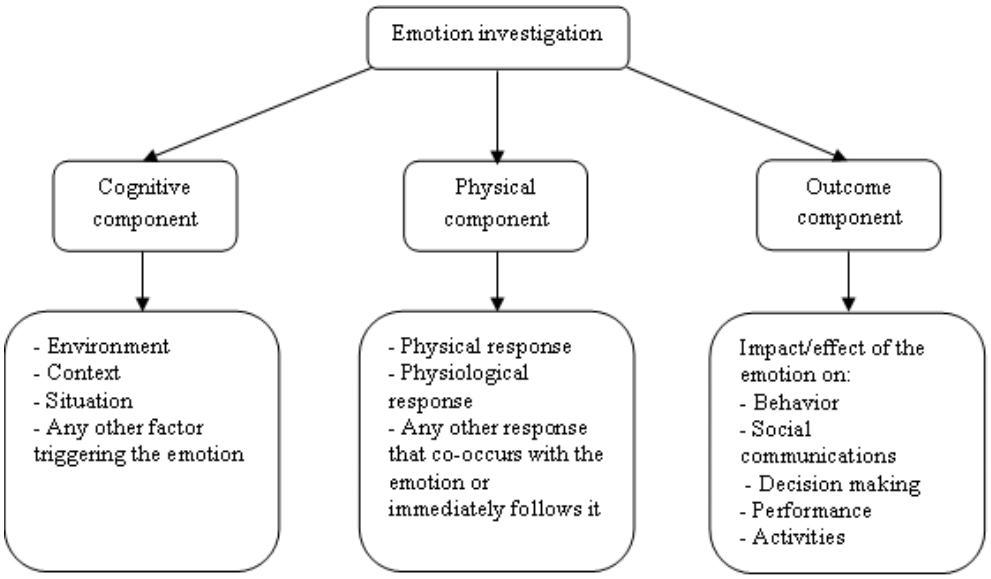

Fig. 1. The three essential components for a better investigation of emotions

3 Emotion Recognition

While general research on Automatic Emotion Recognition (AER) has matured over the last decades especially in the fields of affective computing and Human-Computer Interaction (HCI), research on emotion recognition and processing using mobile technology is still at a very early stage.

Most (if not all) emotion recognition systems are based on either a rule-based system or on a well-trained system that decides which emotion category, dimension or label fits the best. From a scientific perspective, emotion recognition is nothing more than a mapping function from a feature space to emotion descriptors or labels space using solid and analytically-founded machine learning algorithms [9]. Theoretically, any type of information, such as audio, visual, written, mental or even physiological [10], can be used for an accurate selection of features and emotional labels. In practice, integrating all these modalities into a single system for emotion-sensitive analysis is still a very challenging issue. In this section, we start by presenting the most recent state of the art in the field of emotion recognition using mobile technology. We classify it into 5 categories: 1°) Emotion recognition in speech, 2°) Facial emotion recognition, 3°) Affective information detection in text, 4°) Emotion recognition from the context around the user, and 5°) Multimodal emotion recognition. We also describe the challenges related to adopting a multimodal emotion recognition approach and highlight our contribution.

3.1 Emotion Recognition in Speech

In any speech emotion recognition system, three crucial aspects are identified [11]. The first one is the extraction of a reasonably limited, meaningful, and informative set of features for speech representation. Some of these features might be speech pitch [12], spectra [13] and intensity contours. The second is the design of an appropriate classification scheme capable of distinguishing between the different emotional classes it was trained with. This design depends on many aspects such as finding the best machine learning algorithm (neural networks, support vector machines, etc.) to use in constructing the classifier [14], the suitable architecture for the classifier [15], [16], or the proper technique to use when extracting features [17]. The last aspect is the proper preparation of an emotional speech database for evaluating system performance [18]. The number and type of emotions included in the database as well as how well they simulate real-world environment are very important design factors.

Although many combinations of emotional features and classifiers have been presented and evaluated in the literature especially in the context of Human Computer Interaction, little attention has been paid to speech emotion recognition using mobile technology. In [19] for instance, authors propose a speech emotion recognition agent for mobile communication service. They argue that the agent is capable of determining the degree of affection (love, truthfulness, weariness, trick, friendship) of a person, in real-time conversation through a cellular phone, at an accuracy of 72.5 % over five predetermined emotional states (neutral, happiness, sadness, anger,

and annoyance). The system alleviates the noises caused by the mobile network and the environment, and which might cause emotional features distortion, by adopting a Moving Average filter and a feature optimization method to improve the system performance.

Authors in [20] introduce an emotional speech recognition system for the applications on smartphones. The system uses support vector machines (SVM) and a trained hierarchical classifier to identify the major emotions of human speech, including happiness, anger, sadness and normal. Accent and intonation are recognized using a time-frequency parameter obtained by continuous wavelet transforms. Authors argue that their system achieves an average accuracy of 63.5% for the test set and 90.9% for the whole data set. As an application on smartphones, authors suggest to combine the system with social networking websites and the functions of micro blogging services in an effort to increase the interaction and care among people in community groups.

In [21], authors introduce EmotionSense, a mobile sensing platform for conducting social and psychological studies in an unobtrusive way. EmotionSense automatically recognizes speakers and emotions by means of classifiers running locally using Gaussian Mixture methods. According to authors, the platform is capable of collecting individual emotions as well as activities and location measures by processing the outputs from the sensors of off-the-shelf mobile phones. These sensors can be activated or deactivated using declarative rules social scientists express according to the user context. Authors claim that the framework is highly programmable and run-time adaptive.

In [20] and [21] emotion recognition is processed locally within the mobile phone, whereas in [19], the speech signal is transmitted to an emotion recognition server which performs all the computation and processing and reports back a classification result to the mobile agent based on the confidence probability of each emotional state.

3.2 Facial Emotion Recognition

Emotions are highly correlated with facial expressions. When we smile, frown or grimace, thousands of tiny facial muscles are at work making it almost impossible to express emotions without facial expressions.

A typical facial expression detection system would comprise of two essential blocks: the feature extraction block and the feature recognition block. The first block tries to extract some relevant patterns from the images (such as the shape, the texture composition, the movements of the lips, the eyebrows, etc,), and then feed them into either a rule-based or well-trained feature recognition block that is able to recognize the emotions based on the extracted features. A large variety of purely facial emotion recognition systems has been presented in recent years. A detailed overview of these techniques can be found in [22].

Processing images requires a considerable amount of computational resources (CPU) and it is highly time-consuming. If a visual emotion recognition system is to be implemented in mobile devices, it must definitely take into account their limited hardware performance. Very few works in the literature investigate visual emotion recognition using mobile devices. Authors in [23] for instance investigate the deployment of a face detection algorithm to identify facial features on an android mobile platform. The algorithm uses a mixture of statistical approaches to predict different facial regions (eyes, nose and mouth areas) and image processing techniques to accurately locate specific features. Authors claim that less than 3 sec is needed for accurate features detection. No emotion recognition, however, is investigated using the extracted features.

In [24], authors propose EmoSnaps, a mobile application that captures pictures of one's facial expressions throughout the day and uses them for later recall of momentary emotions. The photo captures are triggered by events where users are paying visual and mental attention to the device (e.g "screen unlock", "phone call answer" and "sms sent"). The photo shots are not used for real-time emotion recognition but for reconstructing one's momentary emotions by inferring them directly from facial expressions. Authors argue that EmoSnaps should better be used for experiences that lie further in the past rather than the recent ones.

A more market-oriented approach is presented in [25] where authors show how mobile devices could help in measuring the degree of people's response to the media. More specifically, authors document a process capable of assessing the emotional impact a given advertisement has on a group of people through Affectiva's facial coding platform Affdex [26]. Affdex provides real-time emotional states recognition by analyzing streamed webcam videos in Affectiva's cloud and tracking smirks, smiles, frowns and furrows to measure the consumer response to brands and media. Emotions are later aggregated across individuals and presented on a dashboard that allows playback of the ad synchronized with the emotion response. Authors also explore how this process can help in providing new insights into ad effectiveness and ad recall.

3.3 Context Emotion Recognition

It is very difficult even for humans to judge a person's emotional state from a short spoken sentence or from a captured image. To interpret an utterance or an observed facial expression, it is important to know the context in which they have been displayed. Therefore, modern Automatic Emotion Recognition systems are influenced by the growing awareness that long-range context modeling plays an important role in expressing and perceiving emotions [27]. Again, there are relatively few focused studies on retrieving emotional cues from the context around the user through mobile devices.

Motivated by the proliferation of social networks such as Facebook and MySpace, authors in [28] propose a mobile context sharing system capable of automatically sharing high-level contexts such as activity and emotion using Bayesian Networks based on collected mobile logs. Low-level contexts from sensors such as GPS coordinates, and call logs are collected and analyzed. Meaningful information is then being extracted through data revision and places annotation. To recognize high-level contexts, Bayesian Networks models are used to calculate the probability of activities (e.g. moving, sleeping, studying, etc.) according to predetermined factors: status factor, a spatial factor, a temporal factor, an environmental factor, and a social factor. Authors later derive specific emotions (e.g. bored, contented, excited, etc) directly from the activity arguing that activity has an influence on user's emotion directly, which, in our sense, might not be accurate since a user's specific activity can be associated with different types of emotion.

Authors in [29] propose a method for generating behaviors of a synthetic character for smartphones. Like in 28, user contexts are also inferred using Bayesian networks to deal with information insufficiency and situations uncertainty. Authors deploy two types of Bayesian networks: one infers valence and arousal states of the user and the other infers the business state of the user by gathering and analyzing information available in smartphones such as contact information, schedules, call logs, and device states. Their work, however, focuses more on how to create the emotions of the synthetic character rather than inferring the emotional state of the user from the context around him/her.

In [30], authors introduce a machine learning approach to recognize the emotional states of a smartphone user in an unobtrusive way using the user-generated data from sensors on the smartphone. The data is first classified into two groups: behavior (e.g. typing speed, frequency of pressing a specific key, etc) and context of the user (location, time zone, weather condition, etc). Specific emotions are then identified using a Bayesian Network classifier and used for enhancing affective experience of Twitter users. Their experimental results, which were obtained using a developed Twitter client installed in the mobile device, showed an average classification accuracy of 67.52% for 7 different emotions. During the data collection process, participants are required to self-report their current emotion, whenever they feel it at the certain moment in their everyday life, via the Twitter client using some short text messages. This might not be practical in our sense since an emotion recognition system should be able to identify emotions in a more autonomous and stand-alone way.

3.4 Emotion Recognition from Text

It is generally rather difficult to extract the emotional value of a pure written text (whether it is containing emoticons or not). One reason could be the contextual ambiguity of sentences and their lack of subtleness. In fact, it is easier to express emotions through facial expressions rather than putting them into words. Authors in [31] showed how a banal conversation through instant messaging may turn into a fight by giving the example of someone who is getting increasingly annoyed in a conversation until he/she finally 'shouts' at the conversation partner that he/she had enough of it. The emotion detection system would suddenly change the perceived emotion from "neutral" to "outraged", making it difficult for the conversation partner to understand this sudden outburst of anger.

Emotion recognition from text involves concepts from the domains of both Natural Language Processing and Machine Learning. Text-based emotion recognition techniques can be classified into Keyword Spotting Techniques, Lexical Affinity Methods, Learning-based Methods and Hybrid Methods [32]. These techniques as well as their limitations are surveyed in [33].

We found no literature record of a mobile phone system or architecture that is able to recognize emotional states from pure textual data. Text input is usually combined in a multimodal setup with other communication channels such as audio, video or context.

3.5 Multimodal Emotion Recognition

A multimodal emotion recognition system is a system that responds to inputs in more than one modality or communication channel (e.g., speech, face, writing, linguistic content, context, etc.). Many earlier works have

already proved that combining different modalities for emotion recognition provides complementary information that tends to improve recognition accuracy and performance. A comprehensive survey about these works in a HCI context can be found in [34].

Mobile phones have many modalities that are equivalent to human senses. They can see us through their cameras, hear us through their microphones, and touch us through their haptic sensors. Moreover, they are becoming more and more powerful since all sort of data are stored and accessed through them. The most surprising issue regarding multimodal emotion recognition is that despite the huge advances in processing modalities (audio, video, text, context, etc.) separately and despite the fact that great improvement of recognition performance could be achieved if these channels are used in a combined multimodal setup, there were only a few research efforts which tried to implement a multimodal emotion analyzer. Further, there is no record of a research effort that aims at integrating all modalities into a single system for emotion recognition through mobile devices.

In [35], authors propose a model that uses emotion-related data, obtained through biosensors, along with physical activity data, obtained through motion sensors integrated in the mobile phone, as input to determine the likely emotional state of the user. Authors claim that their model chooses an appropriate personalized service to provide based on the computed emotional state but, in the experimental prototype, only the mobile device ring tone is modified to match the user's current emotional state. Also, authors only vaguely describe how emotions are identified. The model has also the limitation that users should always wear a device in form of a glove for emotion-related data to be gathered, which can be quite intrusive.

A similar approach is presented in [36] where authors developed an emotion recognition engine for mobile phone. The engine uses a sensor enabled watch for the gathering of bio signals and environmental information. Compensation methods were applied for the sensor enabled watch to increase emotion decision accuracy. Based on one of the two identified emotional states: pleasant and unpleasant, differentiated multimedia contents, through IPTV, mobile social network service, and blog service, are provided to the user. However, practical limitations like in [35] are to be mentioned since wearing a biosensor watch all the time may cause inconvenience to users.

Authors in [37] propose an emotion recognition framework for smartphone applications. Using an emotional preference learning algorithm, the difference between two quantified (prior and posterior) behaviors is learnt for each entity in a mobile device (e.g. downloaded applications, media contents and contacts of people). Their conducted experiment showed that touching behavior (e.g. tapping, dragging, flicking, etc.) can also give clues about smartphone users' emotional states. No explicit response to predicted emotions is investigated by the authors.

3.6 Mobile Multimodal Emotion Recognition

With respect to all related efforts presented above, most authors focus on emotion recognition via mobile phones using only one communication channel (speech, image, context, etc.). The very few of them who tried to integrate many modalities into a single system desperately struggled with both hardware performance and energy limitations of mobile devices. As we mentioned earlier, processing many modalities locally within the mobile phones requires a considerable amount of computational resources and it is highly time and energy consuming. Moreover, most of them fail to provide immediate responses or reactions to recognized emotions in an effort to support mobile phones users emotionally.

The present work comes to overcome such limitations. The key contributions of this work can be summarized as follows:

- We design a multimodal emotion recognition system that aims at integrating different modalities into a single system for emotion recognition using mobile devices. All emotion recognition processing and their related computing complexity issues are dealt with in the cloud [38], while mobile devices simply transmit different inputs' data to the cloud which reports back the emotion recognition decision to the mobile device.
- We design a method, we refer to as EgoGenie, which automatically recognizes the emotional states of the user through his/her mobile phone and reacts to them accordingly by presenting personalized content and service offerings to support him/her emotionally and effectively, whenever and wherever the user needs it.

It is important to point out that both the proposed system and method could be implemented on any interactive device, PC or tablet, but since mobile phones are part of our everyday life and are almost all the times with us, the emotional impact of the proposed system and method on users is higher when deployed through mobile devices. Next sections provide extensive details about both the proposed system and method.

4 Mobile Multimodal Emotion Recognition System

This section explores the design of the mobile multimodal emotion recognition system and describes its major building blocks. The system is depicted in Figure 2. It includes client devices, a data collection engine, an emotion recognition engine, a content delivery engine, a network, an online service provider, an advertiser, and an input/output unit. The method which will be described in the next section, and which is illustrated in Figure 4, orchestrates the interaction between all the components in the system.

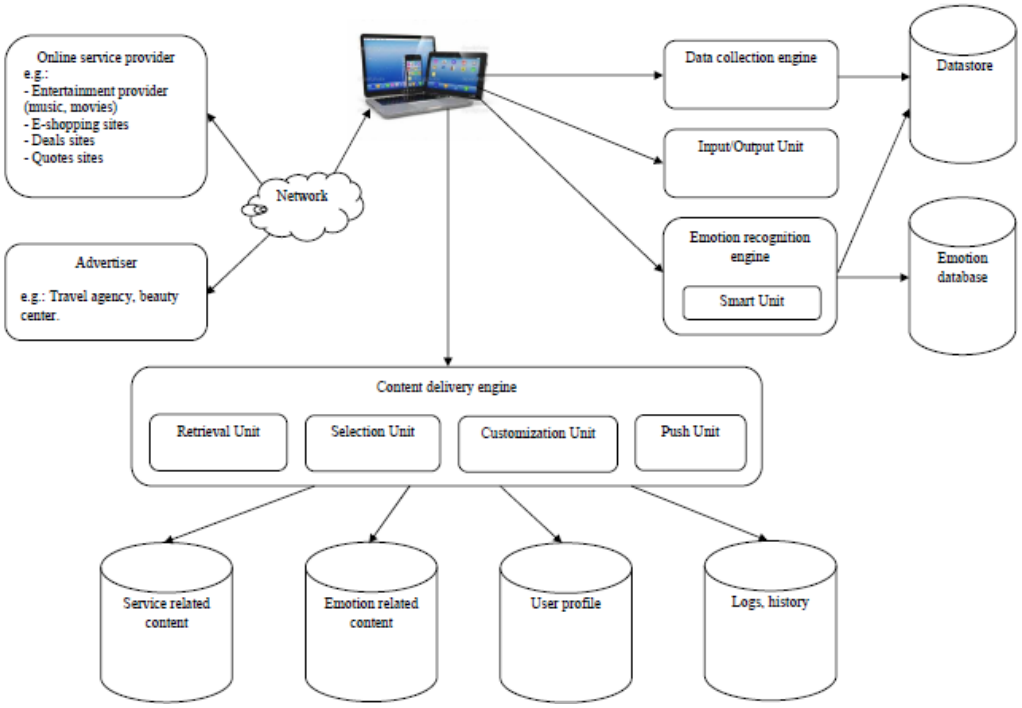

Fig. 2. Mobile Multimodal Emotion Recognition System Overview

The system comprises a data collection engine to collect expressive, behavioral and contextual data associated with an emotional state, analyze the collected data, and extract patterns associated with (or triggering) specific emotional states, an emotion recognition engine to map the patterns to different emotion categories and automatically recognize the emotional states, a content delivery engine to deliver to the user personalized content based on his/her emotional state, and an Input/output unit to interact with the user and the network. Some functions of the mentioned engines may be performed locally on the mobile phone or deported to the cloud for more convenience.

4.1 Data Collection Engine

The data collection engine may collect and analyze all types of data described in FIG. 3 including visual data, vocal data, written data, and contextual data. It also extracts relevant expressive, behavioral and contextual patterns and features to be used by the emotion recognition engine. Contextual data may include both low-level and high-level context. Low-level context is retrieved from both internal and external sensors. Internal sensors may provide information about location (through GPS), motion (through accelerometer), touch behavior, e.g. device shake frequency, typing speed, etc. (through haptic sensors). Additional information may be provided by phone logs (e.g. call logs, SMS logs, tweets, etc.). External sensors may include bio sensors that collect physiological signals (such as skin conductivity, blood volume pressure, respiration, EMG, etc.). High-level context such as user activity is also collected. User agenda is used to determine user activity (e.g. in a meeting, exercising, etc.).

The data collection engine uses collected data to extract relevant expressive, behavioral and contextual patterns and features to be used by the emotion recognition engine.

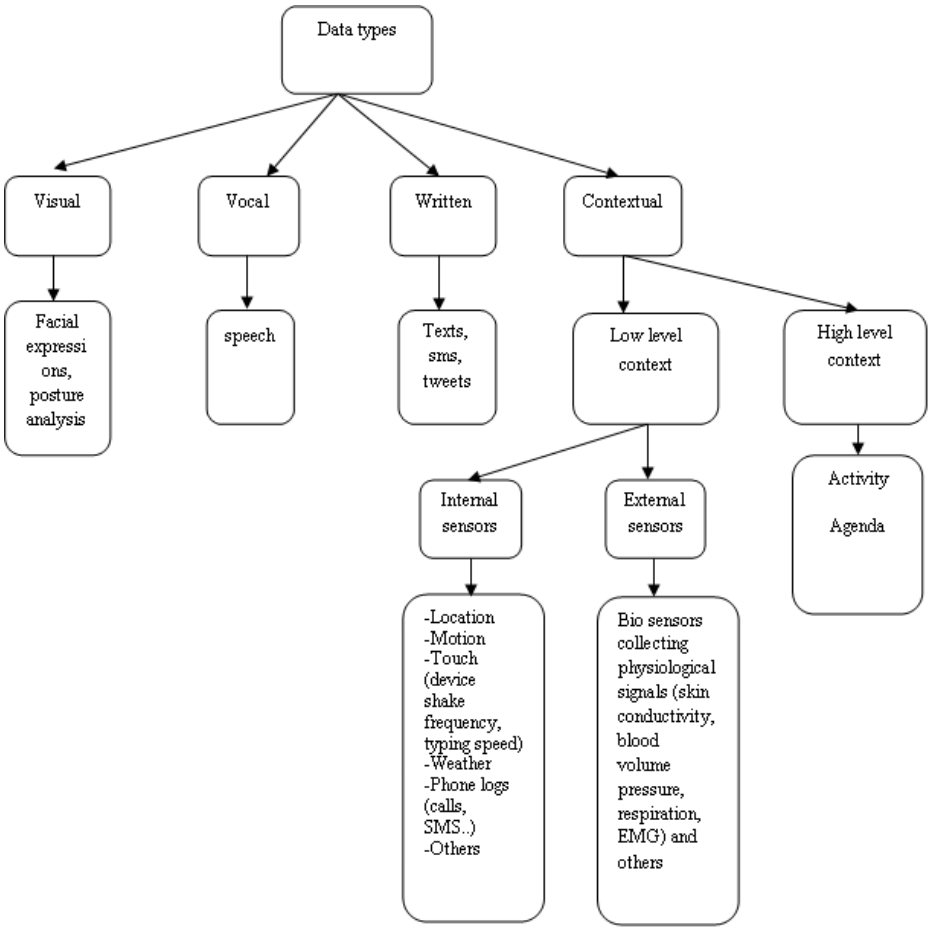

Fig. 3. Data types collected by the Data collection engine

4.2 Emotion Recognition Engine

The emotion recognition engine maps the patterns extracted by the data collection engine to different emotional state categories. The emotion recognition engine also includes a smart unit that may automatically recognize emotional states. The accuracy of automatic emotion recognition depends on the confidence on the patterns the emotion recognition engine is trained with. A detailed description on how the emotion recognition engine works will be described in section 5.

4.3 Content Delivery Engine

The content delivery engine reacts to the recognized emotional states by pushing to the user personalized content to support him/her emotionally. The content may include text, pictures, animations, videos or any combination of them. The content may also include service offerings such as vacation and SPA deals, shopping suggestions, beauty and fitness sessions, etc. Many contents may be served as a response to one particular emotional state. In neutral emotional states, specific content including amusing, entertaining, inspirational, ego-boosting, and motivating is served.

The content delivery engine comprises a content retrieval unit to retrieve different contents, a content selection unit to pick up the most appropriate content based on both user emotional state and user profile and preferences, a content customization unit to make the content look more personal and increase its emotional impact on the user, and a push unit to send the content to the user. The content delivery engine also interacts with a user profile database to store user-specific information such as name, gender, date of birth, preferences and hobbies, a logs database to store the emotional states of the user along with the pushed content, an emotion related content database and a service related content database to store contents or service offerings that may impact the user emotionally. Detailed description of these elements is provided bellow.

The content retrieval unit may retrieve content either from the emotion related content database or the service related content database. The emotion related content database is populated with content that may support the

user emotionally, boost his ego, make him feel better about himself and cheer him up. An example of this content may be a quote, a poem, or a joke, retrieved from the online service provider. The service related content database contains service offerings which may impact the user emotionally or improve his/her emotional state. Example of these services include dating websites (suggested by the system to people with broken hearts), travel offerings and SPA/massage sessions offerings (suggested to stressed people), beauty centers (suggested to women feeling ugly or in a bad mood), psychologists (suggested to very depressed people), and hospitals (in case the monitoring of the emotional states of some patients is critical like in cardio-vascular illnesses where patients should not get too angry to avoid heart attacks for example). These services are offered by the advertiser.

The content selection unit allows users to adjust both the nature and the frequency of the content to be served. It may also use the context around the user, retrieved through the method described in Figure 4, the user profile and preferences, and the user activity to refine content selection. For instance, a new hairstyle at a beauty center cannot be suggested to an old man.

The content customization unit uses user-specific information and preferences to customize the content in order to increase its emotional impact on the user. User-specific information may be retrieved from user profile database.

The push unit pushes content to users according to their emotional states. It also selects the most appropriate time to push the content based on the activity of the user, which is retrieved from the collected contextual data.

5 EgoGenie

In this section, we present a method we refer to as EgoGenie, and which can be built on top of the mobile multimodal emotion recognition system described in the previous section. EgoGenie automatically recognizes the emotional states of the mobile phone users with a certain confidence and reacts accordingly to them by pushing to the user personalized contents and service offerings. We start by presenting EgoGenie and then we provide some examples of its implementation.

5.1 Presentation

EgoGenie is a method to automatically recognize the current emotional states of the mobile phone user by analyzing different modalities such as: facial expressions (e.g. smiles, frowns or grimaces, etc.), speech (e.g. pitch, intensity contours, etc.), text the user types (e.g. instant messaging, tweets, etc.) and context around the user (e.g. location, weather condition, etc.). Once the method determines how the user is feeling, it will subsequently serve up personalized content and service offerings on his mobile phone to promote his well-being and boost his ego. Our idea is triggered by finding solutions to problems, such as daily life stresses, lack of productivity at work, task performance degradation, lack of self-confidence and self-esteem, and even more. The nature of the content pushed to the mobile phone varies according to the emotional state of the mobile phone user.

5.2 EgoGenie Logic

EgoGenie performs the following basic actions:

- Collecting contextual expressive and behavioral data and extracting patterns associated with an emotional state;
- Mapping expressive, behavioral and contextual patterns to an emotion category;
- Automatically deciding the emotional state of the user;
- Selecting and customizing the content or service offering, as a response to the recognized emotional state;
- Delivering the content or service offering.

The operating logic of EgoGenie is illustrated in Figure 4. Referring to Figure 4, upon method initialization, the emotional state is set to neutral. Neutral is the default emotional state in case no other emotional state is recognized.

The data collection engine starts by collecting contextual data. The main objective from collecting contextual data is to classify and understand the environmental, contextual, and attended situations that contributed or triggered the emotional state. A similar context is more likely to trigger the same emotional reaction from the

user, in the future. Understanding the context allows the emotion recognition engine to automatically recognize the emotional state of the user without any self-reporting from the user himself.

After contextual data is collected, it needs to be analyzed. The data collection engine extracts contextual patterns. These patterns may include any pattern susceptible to be associated with (or trigger) an emotional state. For example, the approaching deadline for an exam or a project (provided through Agenda) is a pattern that is likely to generate a "stressed" emotional state. Period approaching for a female (also provided through Agenda) is a pattern that is likely to generate a "Fragile" emotional state. Being in a hospital or a cemetery (retrieved through location sensors of the client device) is a pattern that is likely to be associated with a "sad" emotional state. A user shouting during or after receiving a phone call (retrieved through the microphone of the client device) is a pattern that may indicate that the user is in "upset" or "angry" emotional state. Many patterns could be fed into the emotion recognition engine. The accuracy of emotion recognition depends on the degree of confidence in the patterns. This is obtained through the training of the emotion recognition engine during the self-reporting phase.

The client device allows the user, through the Input/output unit, to self-report his/her emotions very much like as if the user is talking or writing to a close friend and seeking comfort or advice. The progress of the algorithm will depend on whether the user self-reports his/her emotional state or not.

If the user self-reports the emotional state, the emotion recognition engine determines the emotion category by analyzing the written/vocal input stream. Concurrently, the data collection engine collects and extracts expressive and behavioral patterns associated with the emotion by analyzing data captured from the different sensors of the client device (e.g. camera, microphone, etc.). This consists in determining how the user expresses a specific emotion. Expressive patterns may include speech patterns (e.g. speech pitch, spectra, intensity contours, etc.), facial patterns (such as the shape, the texture composition, the movements of the lips, the eyebrows, etc.), or any combination of them (e.g. laughing, crying, shouting, etc.). Behavioral patterns may include body posture and attitude (e.g. holding head, hiding eyes, etc.). These patterns are used to train the smart unit to automatically detect the user's emotional state even if the user does not explicitly self-report it.

The data collection engine checks whether a new pattern is identified, by verifying whether the pattern has or has not been previously stored as associated with an emotion category in the emotion database. If the pattern is new, the emotion recognition engine maps it to the emotion category which was identified earlier and increases the confidence in that pattern. If the pattern is not new, the emotion recognition engine will simply increase confidence in the pattern. Let us assume, for example, that a person reported his sad emotional state to the client device, the emotion recognition engine analyzes what the user said and set the present emotional state of the user to "sad". The data collection engine collects expressive and behavioral patterns associated with the "sad" emotional state of the user, and concludes that the pattern:"In tears" is set. The emotion recognition engine checks whether that pattern is already associated with the "sad" emotional state in the emotion database. If not, it will map it along with the other patterns and increase confidence in that pattern. If yes, it will simply increase confidence in the pattern.

The content delivery engine retrieves and select appropriate content/service offering associated with the identified emotional state and customize it to increase its emotional impact on the user. Finally, the content delivery engine pushes the content or service offering to the user.

The extracted contextual patterns are periodically checked by the smart unit to identify any match with an already mapped emotional state. Since some contextual patterns can be associated with more than one emotional state, a high blood volume pressure for example can be associated with both "Angry" and "Happy" states, expressive and behavioral patterns are recollected by the data collection engine to better identify the emotion. Hence, the smart unit identifies, with a certain confidence, the emotional state of the user based on contextual, expressive and behavioral patterns it was trained with during the self-reporting phase. The confidence in emotional state recognition is correlated with the confidence in the patterns.

In case no match between contextual patterns and emotion category is identified, the emotional state is maintained to "Neutral" and the content delivery engine serves up to the user specific content associated with "Neutral" states, including entertaining, amusing, inspirational, ego-boosting, and motivating.

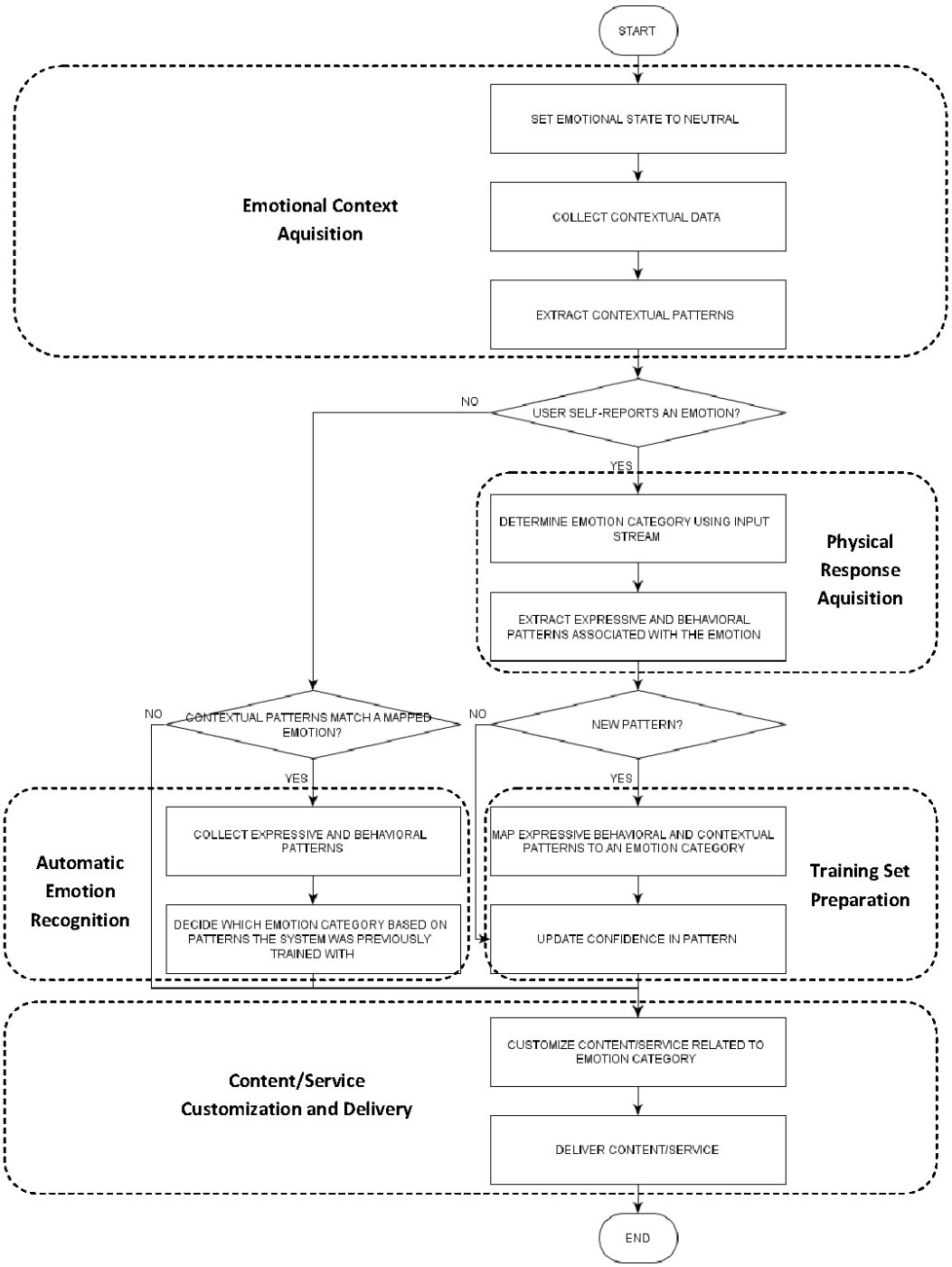

Fig. 4. EgoGenie operating logic

5.3 Implementation Examples

This section provides two implementation examples of EgoGenie: an example of content display based on emotion self-reporting and an example of content display based on automatic emotion recognition.

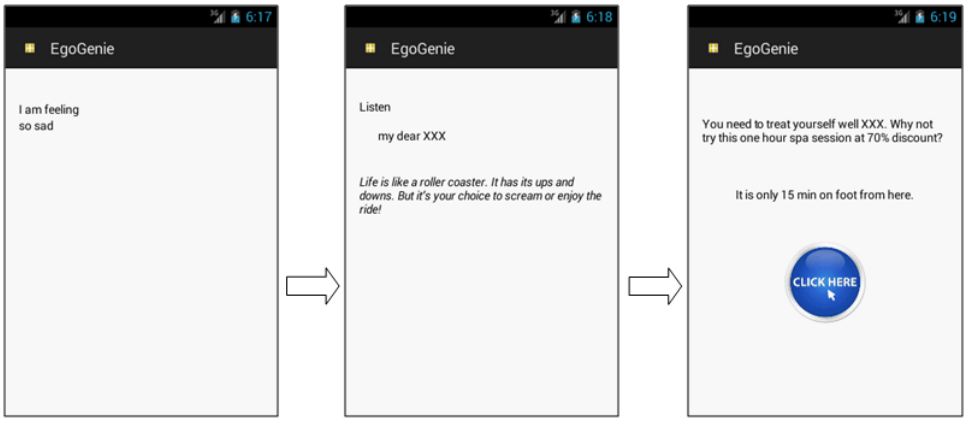

Fig. 5. Example of content display based on emotion self-reporting

As illustrated in (a) of Figure 5, the user of the client device self-reports his emotional state through a textual input. The client device analyzes the textual input and assigns a "sad" emotional state to the user based on the word "sad" used by the user in the text. In (b) of Figure 5, the client device reacts to the identified emotion by interacting in a friendly manner with the user through the sentence "my dear XXX" where XXX is the first name of the user obtained from the user profile database. The emotion-related content which corresponds in this example to a quote, retrieved from the online service provider through the emotion-related content database, is pushed to the user to support him/her emotionally. The pushed content, as illustrated in (c) of Figure 5, can also be a service-related content (SPA session in this example) retrieved from either the online service provider or the advertiser through the service-related content database. Location information may be used to better select the services to be offered to the user. A user is more likely to go for a "SPA session at 70% discount" at a place which is 15 min on foot from where he is, than to go for a "massage session" at a place which is 30 min driving.

Figure 6 illustrates an example of content display based on automatic emotion recognition. In this example, the client device automatically recognizes the emotional state of the user through the method described in Figure 4. The client device sympathizes with the user by pushing textual content conveying its ability to feel the user. The client device recognizes that the emotional state of the user is "stressed", and again serves up appropriate content to ease his stress. In this example, the pushed content is a video showing how abdominal breathing is performed.

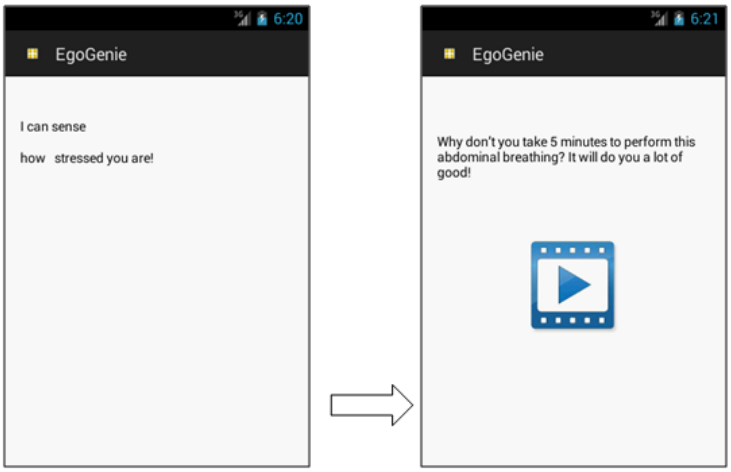

Fig. 6. Example of content display based on automatic emotion recognition

6 Conclusions and Future Work

In this paper, we showed how EgoGenie added a layer of emotional intelligence to mobile phones and made them go beyond the basic functions of calling and texting to fulfill more human and complex roles such those of recognizing, understanding and expressing emotions to support humans emotionally whenever and wherever they need it. We believe that the more mobile phones learn about us, the more they will become emotionally smart and the more they become emotionally smart, the more we will connect to them and maybe fall in love with them! We will work on making EgoGenie smarter while preserving the privacy and the autonomy of the psychic life of the user. We will also work on improving the interaction between the user and the mobile phone so it appears more natural and spontaneous.

References

1. Agger, B.: Everyday Life in Our Wired World, in The Virtual Self: A Contemporary Sociology, Blackwell Publishing Ltd, Oxford, UK. (2008), doi: 10.1002/9780470773376.ch1
2. Salmeron, J. L.: Fuzzy cognitive maps for artificial emotions forecasting, Applied Soft Computing 12 (2012) 3704–3710
3. Calvo, R.A., D'Mello, S.: Affect Detection: An Interdisciplinary Review of Models, Methods, and Their Applications. Affective Computing, IEEE Transactions on , vol.1, no.1, pp.18,37, Jan. 2010
4. Alepis, E., Virvou, M., Kabassi, K.: Mobile education: Towards affective bi-modal interaction for adaptivity. Third International Conference on Digital Information Management, ICDIM 2008, pp.51-56
5. Klasnja, P., Pratt, W.: Healthcare in the Pocket: Mapping the Space of Mobile-Phone Health Interventions, J Biomed Inform. 2012, 45(1): 184–198.
6. Carneiro, D., Castillo, J. C., Novais, P., Fernández-Caballero, A., Neves, J.: Multimodal behavioral analysis for non-invasive stress detection, Expert Systems with Applications 39 (2012) 13376–13389
7. Cai, H., Lin, Y.: Modeling of operators' emotion and task performance in a virtual driving environment, Int. J.Human-Computer Studies 69 (2011) 571–586
8. Kang, S.-H., Watt, J. H.: The impact of avatar realism and anonymity on effective communication via mobile devices, Computers in Human Behavior 29 (2013) 1169–1181
9. Busso, C., Bulut, M., and Narayanan, S.: In Social emotions in nature and artifact: emotions in human and human-computer interaction, S. Marsella J. Gratch, Ed. Oxford University Press, New York, NY, USA, 2012 (press)
10. Quintana, D. S., Guastella, A. J., Outhred, T., Hickie, I. B., Kemp, A. H.: Heart rate variability is associated with emotion recognition: Direct evidence for a relationship between the autonomic nervous system and social cognition, International Journal of Psychophysiology 86 (2012) 168–172
11. El Ayadi, M., Kamel, M. S. and Karray, F.: Survey on speech emotion recognition: Features, classification schemes, and databases. Pattern Recogn. 44, 3 (March 2011), 572-587. DOI=10.1016/j.patcog.2010.09.020
12. Yang, B., Lugger, M.: Emotion recognition from speech signals using new harmony features, Signal Processing 90 (2010) 1415–1423
13. Busso, C., Lee, S., Narayanan, S.: Analysis of emotionally salient aspects of fundamental frequency for emotion detection. IEEE Trans. Audio Speech Lang. Proc. 17, (2009) 582–596
14. Oudeyer Pierre-Yves: The production and recognition of emotions in speech: features and algorithms, Int. J. Human-Computer Studies 59 (2003) 157–183
15. Albornoz, E. M., Milone, D. H., Rufiner, H. L.: Spoken emotion recognition using hierarchical classifier, Computer Speech and Language 25 (2011) 556–570.
16. Lee, C-C., Mower, E., Busso, C., Lee, S., Narayanan, S.: Emotion recognition using a hierarchical binary decision tree approach, Speech Communication 53 (2011) 1162–1171
17. Koolagudi, S. G. et al. : Real Life Emotion Classification using Spectral Features and Gaussian Mixture Models, Procedia Engineering 38 (2012) 3892 – 3899
18. Ververidis, D. and Kotropoulos, C.: A Review of Emotional Speech Databases, http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.98.9202
19. Yoon, W-J., Cho, Y-H., and Park, K-S.: A Study of Speech Emotion Recognition and Its Application to Mobile Services J. Indulska et al. (Eds.): UIC 2007, LNCS 4611, pp. 758–766, 2007
20. Tarng, W., Chen, Y.-Y., Li, C-L., Hsie, K-R., and Chen, M.: Applications of Support Vector Machines on SmartPhone Systems for Emotional Speech Recognition, World Academy of Science, Engineering and Technology 48 2010.
21. Rachuri, K. K., Musolesi, M., Mascolo, C., Rentfrow, P. J., Longworth, C., Aucinas, A.: EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research, UbiComp '10, Sep 26-Sep 29, 2010, Copenhagen, Denmark
22. Tian, Y., Kanade, T., Cohn, J.F.: Handbook of Face Recognition, Ch. Facial Expression Analysis, Springer, London, 2011, pp. 487–519
23. Mawafo, J. C. T., Clarke, W.A. and Robinson, P.E.: Identification of Facial Features on Android Platforms, Industrial Technology (ICIT), 2013, pp 1872-1876

24. Niforatos, E., Karapanos, E., EmoSnaps: A Mobile Application for Emotion Recall from Facial Expressions, CHI'13, April 27 – May 2, 2013, Paris, France.
25. Swinton, R. and El Kaliouby, R.: measuring emotions through a mobile device across borders, ages, genders and more, ESOMAR 2012.
26. Affdex, http://www.affectiva.com/affdex/#pane_overview (accessed on May 9, 2013)
27. Barrett, L.F., Kensinger, E.A.: Context is routinely encoded during emotion perception, Psychol. Sci. 21 (2010) 595–599
28. Oh, K., Park, H-S., and Cho, S-B.: A Mobile Context Sharing System using Activity and Emotion Recognition with Bayesian Networks, 2010 Symposia and Workshops on Ubiquitous, Autonomic and Trusted Computing, pp 244-249
29. Yoon, J.-W., Cho, S-B.: An intelligent synthetic character for smartphone with Bayesian networks and behavior selection networks, Expert Systems with Applications 39 (2012) 11284–11292
30. Lee, H., Choi, Y. S., Lee, S. and Park, I. P.: Towards Unobtrusive Emotion Recognition for Affective Social Communication, The 9th Annual IEEE Consumer Communications and Networking Conference- Special Session Affective Computing for Future Consumer Electronics, pp 260-264
31. Tetteroo, D.: Communicating emotions in instant messaging, an overview, the 9th Twente Student Conference on IT, Enschede, June 23 th, 2008
32. Fragopanagos, N., Taylor, J.G.: Emotion recognition in human–computer interaction, Department of Mathematics, King‟s College, Strand, London WC2 R2LS, UK Neural Networks 18 (2005) 389–405 march 2005
33. Kao, E.C.-C., Liu, C-C, Yang, T-H., Hsieh, C-T., Soo, V-W.: Towards Text-based Emotion Detection A Survey and Possible Improvements. International Conference on Information Management and Engineering, ICIME '09., vol., no., pp.70,74, 3-5 April 2009 doi: 10.1109/ICIME.2009.113
34. Sebe, N., Cohen, I., Gevers, T., Huang, T.S.: Multimodal approaches for emotion recognition: a survey. Proc. SPIE 5670, 56–67 (2005)
35. Hussain, S. S., Peter, C., Bieber, G.: Emotion Recognition on the Go: Providing Personalized Services Based on Emotional States, in proc. Of the 2009 Workshop: Measuring Mobile Emotions: Measuring the Impossible?, Bonn, Germany, 15th September 2009
36. Lee, S., Hong, C-s., Lee, Y. K., Shin, H.-s.: Experimental Emotion Recognition System and Services for Mobile Network Environments, in proc. of IEEE SENSORS 2010 Conference, pp-136-139
37. Kim H.-J. and Choi, Y. S.: Exploring Emotional Preference for Smartphone Applications, The 9th Annual IEEE Consumer Communications and Networking Conference - Special Session Affective Computing for Future Consumer Electronics, 2012, pp 245-249
38. Mousannif, H., Khalil, I. and Kotsis, G: The cloud is not "there", we are the cloud!, International Journal of Web and Grid Services, vol. 9, issue 1, pp 1-17, 2013