

This is a postprint version of the following published document:

Castillo, José Carlos; Castro-González, Álvaro; Fernández-Caballero, Antonio; Latorre, José Miguel; Pastor, José Manuel; Fernández-Sotos, Alicia; Salichs, Miguel A. (2016). Software Architecture for Smart Emotion Recognition and Regulation of the Ageing Adult. *Cognitive Computation*, v.8, 357-367.

DOI: <https://doi.org/10.1007/s12559-016-9383-y>

© Springer Science+Business Media New York 2016

[Click here to view linked References](#)

José Carlos Castillo<sup>1,2</sup>, Álvaro Castro-González<sup>2</sup>, Antonio Fernández-Caballero<sup>1</sup>, José Miguel Latorre<sup>3</sup>,  
José Manuel Pastor<sup>4</sup>, Alicia Fernández-Sotos<sup>5</sup>, and Miguel A. Salichs<sup>2</sup>

## Software Architecture for Smart Emotion Recognition and Regulation of the Ageing Adult

<sup>1</sup> Universidad de Castilla-La Mancha, Instituto de Investigación en Informática de Albacete, 02071-Albacete, Spain

<sup>2</sup> Universidad Carlos III de Madrid, Robotics Lab, 28911-Madrid, Spain

<sup>3</sup> Universidad de Castilla-La Mancha, Instituto de Investigación en Discapacidades Neurológicas, 02071-Albacete, Spain

<sup>4</sup> Universidad de Castilla-La Mancha, Instituto de Tecnologías Audiovisuales, 16071-Cuenca, Spain

<sup>5</sup> Universidad de Castilla-La Mancha, Facultad de Educación de Albacete, 02071-Albacete, Spain and Conservatorio Profesional de Música Maestro Gómez Villa, 30530-Cieza (Murcia), Spain}

\* Corresponding author.

Email address: [jccmontoya@gmail.com](mailto:jccmontoya@gmail.com)

Phone +34 967 599200

Fax +34 967 599224

# Software Architecture for Smart Emotion Recognition and Regulation of the Ageing Adult

José Carlos Castillo<sup>1,2</sup>, Álvaro Castro-González<sup>2</sup>, Antonio Fernández-Caballero<sup>1</sup>, José Miguel Latorre<sup>3</sup>,  
José Manuel Pastor<sup>4</sup>, Alicia Fernández-Sotos<sup>5</sup>, and Miguel A. Salichs<sup>2</sup>

## Abstract

This paper introduces the architecture of an emotion-aware ambient intelligent and gerontechnological project named “Improvement of the Elderly Quality of Life and Care through Smart Emotion Regulation”. The objective of the proposal is to find solutions for improving the quality of life and care of the elderly who can or wants to continue living at home by using emotion regulation techniques. A series of sensors is used for monitoring the elderlies' facial and gestural expression, activity and behaviour, as well as relevant physiological data. This way the older people's emotions are inferred and recognized. Music, colour and light are the stimulating means to regulate their emotions towards a positive and pleasant mood. Then, the paper proposes a gerontechnological software architecture that enables real-time, continuous monitoring of the elderly and provides the best-tailored reactions of the ambience in order to regulate the older person's emotions towards a positive mood. After describing the benefits of the approach for emotion recognition and regulation in the elderly, the eight levels that compose the architecture are described.

## Keywords

Emotion recognition, Emotion regulation, Ambient intelligence, Gerontechnology, Software architecture

## 1. Introduction

Population ageing is a recent concern in developed countries due to the decreasing birth rates and the higher life expectancy. This phenomenon relates to a shift in the structure of the population ages, where basically those groups with greater ages grow whilst younger population is reduced. Although improvements in nutrition and healthcare enables ageing with high quality of life, there are cases in which some health-related issues require close and personalized supervision. Indeed, Ambient Assisted Living (AAL) approaches are gaining importance in recent years because they can provide a personalised support to older people, mainly at home. We believe that perceiving and enhancing the quality of life of

the elderly who lives at home is possible through automatic emotion recognition and regulation. Some insights in these topics are presented in Section 2 from a psychological perspective.

From an individual perspective, quality of life can be considered in terms of wellbeing. It includes emotional (self-esteem, emotional intelligence, mindset), social (friends, family, community) and physical (health, physical safety) aspects in a person's life. Indeed, it has been largely studied that positive emotional states promote healthy perceptions, beliefs, and physical wellbeing [38]. Now, quality of life and care (QL&C) assesses how welfare is affected due to illness or disability. These aspects can only be assessed through proper monitoring of the individuals in the environment. Usually, elder people tend to live at their own home rather than choosing other options. This leads to many challenges in order to adapt some environments to provide safety and home care to the elder, particularly if the elder is in trouble. In recent years, Information Technology and Communications (ICT) has been playing an active role towards mitigating the most common problems found at home (e.g. [9], [12], [4]).

This paper presents the software architecture of a system to guide the ageing adults towards a positive mood. In this field, many terms have been used in the literature with no clear distinction among them: emotion, mood, feelings, or emotional state, among others. For the purpose of this work we will not differentiate between them and use them as synonyms. The study of the subtle differences among these terms is out of the scope of this paper.

Furthermore, the system, described in Section 3, aims at finding solutions to improve the quality of life and care of aging adults who can or want to keep living at home. The approach uses advanced tools and techniques of ICT supplemented with expert knowledge based on experimental techniques from Psychology, Neurobiology and Music about the regulation of emotions.

Computer tools and techniques used in this project are directly linked to the concepts of Ambient Intelligence (AmI). Indeed, we are convinced that AmI is a well-suited area for the design, implementation and deployment of global systems that seek to provide effective solutions to real problems in our society. The general objective of this project results in some specific goals, such as: (1) to analyse the emotional states and the regulation techniques based on the knowledge provided by subject matter experts, (2) to monitor and recognize emotions in Ambient Intelligence environments, (3) to regulate emotions through adapting the environment, (4) to construct an intelligent emotion regulation system, and, (5) to validate the emotion regulation system based on the expert knowledge. The expected project outcome is an AAL system to raise the ageing adult's QL&C.

1 In this article, we introduce a gerontechnological software architecture able to deal with the problem  
2 of detecting and regulating emotions through ambient intelligent is described in Section 4. The proposal is  
3 prepared to accept any sensor and actuator for continuously monitoring of the elder, providing the best-  
4 tailored reactions to regulate their emotional state towards a positive mood. Section 5 presents the  
5 limitations of the system and some general issues that must be considered. Finally, the conclusions are  
6 summarized in Section 6.  
7  
8  
9  
10

## 11 **2. Fundamentals of Emotion Recognition and Regulation**

12 For emotion interpretation and regulation we heavily rely on Ambient Intelligence (AmI) [37] which  
13 proposes the creation of intelligent environments to suit the needs, tastes and interests of people that live  
14 inside them. It is worth highlighting that AmI integrates three major research areas: sensing of the  
15 environment and/or people, application of computing technologies, and acting on the real and/or virtual  
16 environment. Moreover, the relationship between emotions and AmI is called “emotion-aware AmI”  
17 (AmE) [50], which could be defined as emotion conscious AmI. AmE exploits concepts from Psychology  
18 and Social Sciences to adequately analyse the state of the individual and enrich the contextual  
19 information. AmE achieves this objective by extending the AmI devices through a collection of improved  
20 sensors able to recognize human emotions from facial expressions [46] and human behaviours, such as  
21 hand gestures, body movements and speech [42], [48]. The exploitation of these new emotion context-  
22 aware devices allows systems to deliver a highly personalized and dedicated collection of services to  
23 support users and improve their personal care.  
24  
25  
26  
27  
28  
29  
30  
31  
32  
33  
34  
35  
36  
37  
38  
39

40 Automatic monitoring of emotional states is a valuable tool in the areas of Health Sciences and  
41 Rehabilitation, Clinical Psychology, Psychiatry and Gerontology. Current research in wireless area  
42 networks (WSNs) [5] and body area networks [20] enables the inclusion of advanced monitoring devices.  
43 Thus, energy efficient small size devices are being dedicated to body feature measurements. It is  
44 important to consider a multisensory approach, which combines multiple sources of information  
45 presented by various sensors to generate a more accurate and robust interpretation of the environment  
46 [35]. The availability of new types of sensors for monitoring tasks poses new challenges in multisensory  
47 data fusion. Now it is possible to construct models of the environment and to diagnose situations through  
48 the analysis of a sequence of sensory information.  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

1 Ultimately, we are talking about semantic interpretation from multi-sensed and video-controlled  
2 environments [16], [11]. Such interpretation is defined as the need to recognize situations, activities and  
3 interactions between different actors involved [13]. In these environments in which information comes  
4 from different sensors or a combination of them, the problem is to link the physical signals received with  
5 the interpretation of their meaning.  
6  
7

8  
9 In our case, the goal is to develop a software architecture to recognize the activity and mood of older  
10 people living alone in their homes. Indeed, we believe that the ability to monitor changes in the emotional  
11 state of people in their own context allows implementing regulatory strategies for reducing negative  
12 affect. Despite emotion interpretation in humans has traditionally been an area of interest in disciplines  
13 such as Psychology and Sociology, there is a lack of real applications that relate emotion with human  
14 behaviour. This fact is mainly due to the large number of theoretical models proposed and the complexity  
15 of human emotions [27]. Often, emotions appear as an affective phenomenon, i.e., as a state of mind, an  
16 interpersonal attitude or a personality trait. Moreover, other authors identify emotions by intensity and  
17 duration and in combination with other emotions to result into more complex ones [41].  
18  
19

20 Existing emotion recognition technologies are divided into four major categories depending on what  
21 kind of data is analysed: physiological signals, facial expressions, behaviour or voice [24], [21].  
22 Physiological emotion recognition shows acceptable performance but has some critical weaknesses that  
23 prevent its widespread use; they are obtrusive to users and need special equipment or devices [36]. The  
24 capture of physiological data is increasingly done through body sensor networks. These allow continuous  
25 measurement in the daily life of a person of physiological parameters such as heart rate, muscle tension,  
26 skin conductance, breathing rate, and so on. When combined with contextual information extracted from  
27 the environment through WSNs, such as the ambient temperature or humidity, these parameters can be  
28 used to infer emotions [31]. Facial analysis includes a number of processing steps that attempt to detect or  
29 track the face, to locate characteristic facial regions such as eyes, mouth and nose on it, to extract and  
30 follow the movement of facial features. This is one of the least intrusive processes of automatic emotion  
31 recognition. In recent years, various methods have been developed for extracting and analysing the facial  
32 features from a description of facial expressions. The Facial Action Coding System [7] encodes all  
33 possible facial expressions as unitary actions (UAs) that may occur individually or in combination. In  
34 fact, facial expressions associated with emotions are generally described as a set of UAs [45]. Let us  
35 highlight the FaceSense research project [10] developed by the Affective Computing group of  
36  
37  
38  
39  
40  
41  
42  
43  
44  
45  
46  
47  
48  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

Massachusetts Institute of Technology. Speech signal conveys a large amount of information being also a non-intrusive technique for emotion recognition. Two main problems are addressed: (1) finding the set of features in the speech signal that are most significant in conveying emotions and (2) finding the best classification algorithm that can indicate emotional expression, based on the extracted features [6].

Emotion regulation refers to a set of processes that either stop the emotion from emerging or prevent it from being expressed once it is triggered [18]. Bottom-up emotion generation refers to the elicitation of emotion by the presentation of a stimulus that is thought to have simple physical properties that are inherently emotional [30]. Antecedent emotion regulation strategies apply while the emotion is still unfolding and has not reached its peak. In our case, we are interested in response-focused emotion regulation, which tries to aim at altering and controlling the experiential, behavioural and physiological response to the fully established emotion. There is a great deal of work based on the use of music [1], colour [8] and lighting [34] in emotion regulation.

Recently, different innovative strategies have been developed to improve mood through external stimulation. In the field of gerontology, some techniques have been used to aid reminiscence and life review that claim to help the elderly to remember autobiographical events in an organized manner by using auditory stimuli such as music and visual stimuli such as photographs or videos. Specifically, “Life Review Based on Specific Positive Events” (ReVISEP) [42] is a technique aimed at the recovery of especially positive memories occurred throughout life. The main effect of this technique is a better mood and increased life satisfaction, both in older people with and without major depression.

For a long period in the history of psychological research, emotion and cognition were studied independently as if they were irrelevant to one another. However, after the appearance of a pioneering work [40], renewed interest in the study of the relationship between cognition and emotion has led to the development of a wide range of techniques that allow temporary induction of different mood, both positive and negative. The first modern technique is the Velten mood induction procedure [47]. Other developed techniques are induction through music [32], film sequences [19] and autobiographical memory [3]. It is considered more useful to combine two or more induction techniques simultaneously, since multiple inductions contribute additively to mood [2]. In the same way that emotional states are induced for research, it is also possible to induce emotional states in order to regulate emotions.

The most studied field is the induction of positive emotional states through music [22], which has been tested with some success in various mental health conditions like depression [25]. Emotional

1 responses to music are born of different musical features, including mode, harmonic complexity, tempo  
2 and intensity [26]. Although there are not many works with elderlies, we consider feasible to translate the  
3 principles of emotion regulation through music to ageing adults. One of the most used sets of stimuli in  
4 experimental research on emotions is the International Affective Picture System (IAPS) [23]. The IAPS is  
5 a collection of more than 1000 colour photographs that represent objects, people, landscapes and  
6 everyday situations. It allows accurate selection of the stimuli according to their position in the affective  
7 space defined by the dimensions of valence, arousal and dominance.  
8  
9  
10  
11  
12  
13  
14  
15

### 16 **3. A New AmE Approach for Emotion Recognition and Regulation**

17 In this paper, we propose a new system to improve the elder's QL&C living at their home. The  
18 system is based on emotion recognition and regulation considering the expertise of a multidisciplinary  
19 group of subject matter experts. The general overview of the proposed system is shown in Figure 1.  
20  
21  
22  
23  
24  
25

26 Figure 1. General layout of the Smart Emotion Recognition and Regulation project.  
27  
28  
29

30 As depicted in Figure 1, the four main modules of the smart ambient are "Emotion Recognition",  
31 "Emotion Regulation", "Ambient Adaptation" and "Expertise Provision". Each one of these modules is in  
32 charge of performing different tasks.  
33  
34  
35  
36  
37

#### 38 **3.1. Emotion Recognition**

39 This module encompasses two main tasks: "Emotion Monitoring & Recognition" and "Data Fusion"  
40 (see Figure 2). The former is devoted to information acquisition and processing from the different sensors  
41 placed in the environment to recognize elder's emotions: camera(s) to acquire the elder's face and  
42 gestures, camera(s) to track the elder's behaviour, microphone(s) for emotional speech recognition, and  
43 body sensors to acquire physiological signals. Reduced sensor awareness is crucial not only to increase  
44 elder acceptance but also to get unbiased emotion recognition. For this reason, our goal is to miniaturize  
45 the needed body sensors to turn them as less cumbersome as possible from the elder's perspective. With  
46 the information acquired from the sensors, face and gesture detection as well as speech recognition are  
47 crucial steps towards obtaining an emotion interpretation in terms of positive/negative mood. The activity  
48 recognition evaluates the user's behaviours as normal or abnormal; and the affect recognition is obtained  
49  
50  
51  
52  
53  
54  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

from the analysis of the physiological data. Some technical details on the ongoing work have been recently presented [15], [28] [29].

The previous task generates a linguistic label related to an emotion for each sensory source, namely: ambient information, face, gesture recognition, speech recognition, activity detection and physiological data. This single information sources might detect different emotions (even in contradiction with others). Thus, a higher level task, “Data Fusion”, is crucial to achieve a sensible output. Indeed, a big deal of work is performed in data fusion, as emotion detection and regulation is only viable if all the data recorded from all the sensors provides enough consensual evidence on the emotion detected. More information on the data fusion architecture is available in a recent article [43].

Figure 2. Main tasks of the system.

### 3.2. Emotion Regulation

The next module is “Emotion Regulation” (middle block of Figure 2). Its input is the detected emotion as obtained from “Data Fusion”, and its objective is to provide the best-suited conditions of music performance, as well as colour and light to attain the desired emotion in the elderly. Following the idea presented by Bachovik et al. [2], we propose to simultaneously use multiple induction techniques to improve the regulation of the elder’s mood.

The task present in this module, “Emotion Interpretation and Regulation”, is in charge of analysing the emotion recognized by the previous module. Experts play a fundamental role in this task, as it is necessary to have a precise insight of how stimuli affect emotions in order to decide the emotion-tailored music, colour and lighting conditions for each situation.

Here, the system automatically decides the proper reaction in response to the detected emotion. The knowledge that is used to make these kinds of decisions is provided by subject matter experts. Decision making can vary from one user to other and also in accordance with the evolution of the elder. Then, it is important to provide an easy way to update the knowledge where the decisions are based on. Some initial results of the proposal have been published so far in relation to music [14] and colour and light [44], [33].

### 3.3. Ambient Adaptation

“Ambient Adaptation” uses the emotion-tailored music, colour and lighting conditions as an input to build up an AmI system that integrates the whole process (see bottom part of Figure 2). “Ambient Intelligence” creates an intelligent system capable of adapting the ambience towards regulating emotions. Apart from interacting with the actuators in the environment, the “Ambient Intelligence” task is able to transmit alarms to physicians, care providers and relatives in case a potentially dangerous situation is detected.

The implementation of this task involves controlling several domotic technologies (e.g. colour led, shutters, or piped music). This task includes the development of software to control the different elements altering the ambient, also referred as actuators. Consequently, the “Ambient Adaptation” module works as a bridge to translate the commands generated from the “Emotion Regulation” to different technologies involved.

### 3.4. Expertise Provision

“Expertise Provision” is a module that incorporates all the roles that the subject matter experts play in the system (see Figure 3). Subject matter experts are present in all modules and they play a transversal role providing the expert knowledge that is needed for the success of the system. Experts are involved in all the main tasks that are present in several parts of the system and they differ depending on the module where the expert takes action.

1. “Emotion Recognition” module. Psychologists have a vast experience in assessment of emotional states (e.g. emotion or mood) based on physiological signals and the observation of the individual’s behaviour [20], [41], [46]. Experts have to transfer this knowledge to the engineers who translate it into algorithms. Therefore, the algorithms to evaluate the sensorial data for emotion recognition rely on the expert knowledge.
2. “Emotion Regulation” module. Physicians and psychologists decide how to act in the ambient to induce particular emotional reactions in the ageing adult. They provide the guidelines to adapt the environment based on the recognized emotional states.
3. “Ambient Adaptation”. After ambient adaptation takes effect, the experts evaluate the immediate (short-term) elder’s reaction. For example, assessing the elders’ behaviour while a relaxing music is played.

4. “Emotion Recognition and Regulation Assessment”. The results of the system have to be evaluated in the long term. Experts will analyse historical data collected along the continuous operation of the system. They will ultimately decide the performance of the system based on the results of the analysis.

We are happy to count on a multidisciplinary team composed of computer scientists, electrical engineers and, as subject matter experts, psychologists, neurobiologists, physicians and musicians.

Figure 3. Expertise provision.

#### 4. Software Architecture for Emotion Recognition and Regulation

Nowadays, there is a need for architectural solutions to develop complex software systems. In this sense, a set of software blocks is essential so that the programmers can use, extend or modify them to fit specific applications. This mechanism allows the decomposition of an application into a set of independent modules describing a set of interfaces in order to ease communications. Such flexible approaches enable the quick creation and deployment of applications. Under this scheme, components are defined as reusable and independent blocks that are combined with other components to build a specific system. From a developer perspective, building an application consists of assembling some existing components and a few of their own.

Among the number of advantages, the biggest one is that a modular architecture reduces the time in developing any software as it provides a set of blocks that users can directly use or improve. This means that users are abstracted from the lower level implementation and focus the effort on programming their required module. The main disadvantage of such approaches is the development complexity. For small projects, it is faster to code directly rather than to implement a whole architectural infrastructure. In contrast, for big projects, like the one proposed here, these solutions are an appropriate option.

Real-time emotion recognition and regulation involves a wide diversity of technologies. This results in a complex system that requires a powerful, flexible software/hardware framework. Designing a new solution presents crucial challenges:

- *Deciding the purpose of the system.* This simple decision involves several issues related to the scope of the problem to be solved. For instance, a solution for web development has nothing to do with the

one suggested in this paper, so the requirements of each one must be carefully analysed before taking further steps.

- *Assessing the levels.* After having confirmed the purpose, it is necessary to establish a division to allow a structured development. Finding an optimal decomposition involves both a deep analysis of the problem as well as subject matter experts.
- *Determining the grain size.* This problem is related to the levels one. At an early design stage the functionalities included in each level have to be clearly defined at design time.
- *Avoiding being restrictive.* Finally, the system must be adaptable to a growing problem, but not too wide to lose its purpose. This is an important trade-off that must also be considered at the design stage.

After considering these aspects, several works have presented solutions for emotion recognition. A system called Smart Sensor Integration considerably boosts the development of multimodal on-line emotion recognition systems. Furthermore, Gehrig and Ekenel present a common architecture for real-time action unit detection and emotion recognition [17]. For these tasks, they employ a local appearance-based face representation approach using discrete cosine transform. Santos et al. come up with a proposal that performed real-time stress detection based only in two physiological signals: heart rate and galvanic skin response [39].

In relation to emotion regulation, the literature does not offer many software architectures. Emotion regulation is mostly inspired by works from the area of Psychology. Psychologists have explored traditional frameworks (defined as a set of concepts, practices, and criteria to deal with a particular kind of problem) for diagnosis and treatment of deceases. An example is the work presented by Werner and Gross, which provides a framework for researchers and clinicians interested in understanding the role of emotion regulation processes in Psychopathology [49]. We will base on these works to provide a proper reaction of the ambience. Lastly, another paper provides a framework for researchers and clinicians interested in understanding the role of emotion regulation processes in Psychopathology [46].

#### 4.3 Levels of the Software Architecture

In this section we deal with the organization of the software modules involved in the proposed smart ambient system. We suggest a hierarchy of software levels intended to deal with the problems that commonly arise in real environments (hardware abstraction, noise in the data, or real-time processing).

Next, the software architecture describing the functions performed at each level is presented. Each level is defined by the information that it consumes (receives), how this incoming data is transformed, and the resulting data it provides (sends) to the next level. The data flows are detailed for each level in the next subsections. Figure 4 shows the most outstanding elements to be considered in the proposed architecture for emotion recognition and regulation.

Figure 4. Software Architecture for the Emotion Recognition and Regulation.

#### 4.3.1. Level 0: Hardware - Sensors and Actuators

The proposed software architecture is composed of software elements located at different levels. However, when dealing with data coming from the real world, also referred to as the environment, the hardware elements need to be placed at the lowest level. In our case, this is Level 0. This layer represents the hardware components in charge of perceiving the environment and performing actions. No software module is considered here, although software from other levels directly interacts with it.

Sensors are devices that capture data from the environment and convert the information into electronic format. They can be roughly classified according to the information they provide, among others, visual data (RGB cameras, IR cameras, or depth sensors), audio (microphones), biometric data (blood pressure, perspiration, breathing rate, and so on), ambient conditions (temperature, humidity, light). An example of a possible sensor deployment is shown in Figure 5.

Actuators are devices able to alter the environment after receiving instructions and transforming them into actions. Some actuators considered in this project are: projectors, light bulbs, and sound systems.

Figure 5. Conceptual description of the sensor deployment: Left, ambient sensors, the areas shaded in red are considered as private (e.g. bathroom or bedroom). No sensors are placed there; Right, body sensors, a wristband is a low intrusive option for sensor placement

#### 4.3.2. Level 1: Acquisition

Software at this level handles the communication with the sensors. For each sensor, there is a module in this level that understands the data received. Since this is the lowest processing level, it does not consume data from other levels. It reads the low-level, raw data from the sensors and provides it to the next levels.

#### 4.3.3. Level 2: Feature Extraction

This is one of the key levels. Here the raw data provided by the sensors is interpreted by different algorithms. This means that the input data is transformed into more informative, numerical features that summarize the raw data. Each feature represents an important characteristic of the data that is of special interest.

From a computer vision perspective, this level is in charge of detecting, recognizing and tracking the ageing adult in his/her home. In this sense, features may be understood as the track and shape of the elderly. Regarding speech recognition, some features coming from the elder's utterances can be extracted, such as the sentiment (positive, negative, neutral), the gender of the speaker, or the topics involved within the speech. So, this level consumes raw data and provides features to the upper level that will have to interpret them.

#### 4.3.4. Level 3: Multimodal Fusion

A system operating in the real world should not rely completely on single or separate information sources due to problems such as noise, false positives or occlusions, a fusion step is crucial in order to reduce the uncertainty associated to the extracted features. Indeed, feature extraction is performed over data coming from very different sources. In our scenario, we have information coming from cameras, wearable sensors sending biometric data, microphones, and others. Different features are extracted for each kind of data. All these heterogeneous features are combined in order to provide higher-level abstraction of information. This is known as multimodal fusion.

In our particular problem, the recognition of the elder's emotion in the scenario is crucial. For this task, our architecture uses features related to the users' speech, their blood pressure, and their face. The fusion of all these features results on richer information at the time of assessing the elder's emotion. This level consumes the features extracted in Level 2, and provides a combination of features in a time

1 window. It is important to consider the time constraints in order to determine how often the fusion process  
2 is performed. For example, if it is detected that the elder is talking about a sad event, how long do we  
3 have to consider this feature? Therefore, the frequency of the fusion process will determine the lifespan of  
4 the features. The fusion process gives temporal coherence to the features.  
5  
6  
7  
8  
9

#### 10 **4.3.5. Level 4: Event Detection**

11 This level receives a set of features and processes them. The result is a high-level description of the  
12 situation (e.g. the elder is happy). In our problem, the emotion analysis logic is at this level. Its inputs are  
13 the array of features sent by the multimodal fusion. The output is the events related to the emotion of the  
14 elder. This means that the emotion recognition algorithm corresponds here. Moreover, our architecture is  
15 not limited to detect the ageing adult emotion. The algorithms intended to detect other events have to be  
16 implemented at this level and the inputs would be the same.  
17  
18  
19  
20  
21  
22  
23  
24  
25

#### 26 **4.3.6. Level 5: Decision Making**

27 Here is where the *intelligence* of the system lies: considering the events published by the Level 4 as  
28 well as other external factors such as the ambient information (e.g. temperature, humidity) provided by  
29 the Level 0, the system reacts and decides the action to execute. For example, considering that our system  
30 reacts to the elder's mood, once this is detected, the system will vary the illumination, move the shutters,  
31 or change the music. This level consumes events and provides actions to be executed by the different  
32 actuators in the scenario.  
33  
34  
35  
36  
37  
38  
39  
40  
41

#### 42 **4.3.7. Level 6: Multimodal Fission**

43 This level defines the particular result of an action. A multimodal fission component accepts actions  
44 and splits them into several commands. Each command is directed to a different actuator in the  
45 environment. For instance, if the action is to create a relaxing environment, the multimodal fission will  
46 launch commands addressed to dim the lights, project warm colours, and play pleasant music.  
47  
48  
49  
50  
51

52 The fission process is tuned according to the user's likes. Hence, some information associated to the  
53 elderlies is welcome. Knowing some of their preferences, some environmental conditions, such as the  
54 background music and the projections, are selected accordingly. This results in a faster achievement of  
55  
56  
57  
58  
59  
60  
61  
62  
63  
64  
65

the desired effect. This level consumes actions from the decision making level, and provides commands to the proper actuators.

#### 4.3.8. Level 7: Actuation

This level is the counterpart of Level 1. Here the modules communicating with the actuators send the proper commands. There is one module in charge of managing each actuator. The modules at this level consume commands published by the Multimodal Fission level. There is no information sent to other levels but they write low-level instructions (or primitives) to hardware components. As an example, a text-to-speech module receives the command *say hello* that is transformed into low level instructions that synthesize “hello” through speakers.

### 5. Discussion

The project presented in this paper aims to improve the elders’ QL&C by leading them towards positive moods. Considering the current state of the art, it is not possible to build a reliable system to quantify a person’s QL&C. However, it is not possible to automatically assess the benefit that such an AmE system will cause in the elder’s daily life. Consequently, we need subject matter experts to assess the performance of our system. These experts are in charge of providing the initial expert knowledge to the system. Moreover, they have to evaluate the elders’ mood in order to adjust the system when the results are not satisfactory. So, the success of the project is highly dependent of the involvement and participation of subject matter experts (psychologists, or physicians).

The emotion recognition and regulation system requires a fine tuning of the parameters and configurations. That is, for example, the proper coloured lights or sounds that result in a positive mood can vary from one elder to other. Therefore the system has to be adjusted for each senior according to his/her preferences. This implies a previous analysis of the participants to obtain as many details as possible. This results in the elaboration of individual profiles with personal and private data. These profiles have to be carefully handled according to the current regulation and ethical rules. Therefore, private data are not disseminated and participants are thoroughly informed about how their data will be treated and used.

In line with the ethical aspects, monitoring a person in his/her home must be considered a critical issue. The use of heterogeneous sensors to monitor an elder generates large volumes of sensitive data (e.g.

images or audio recordings). These sensitive data are considered as private data and are treated similarly to the personal profiles mentioned above. Moreover, some areas, such as the bathroom or bedroom are considered as especially private and, therefore, the system would not be deployed there. The location where the system is installed is conceived and agreed with the participants, also taking into account that sensors are placed in strategic positions seeking good coverage of the monitored area. The devices to be installed will be selected considering their impact in the environment. Consequently, we aim at installing non-intrusive devices with a friendly appearance.

It is a complex matter to ensure high performance when detecting the emotional state of elder people from a system endowed with multiple information sources that acquire ambient data. Moreover, as the system is designed to be as less intrusive as possible, it is expected that single sensors will get high quality information just during short periods of time. If we consider face detection, the algorithms will perform better when the user faces a camera at a determined range. For this reason, we propose the user of multisensory fusion and long-term analysis. Therefore, a person mood is not considered as an instantaneous phenomenon but the result of the analysis of the sensor information along a period of time combined with expert knowledge. We believe that this is the main strength of the system, constant monitoring and analysis to allow practitioners to assess the evolution of elder's mood through long periods of time.

Besides, we believe that the success of the project is closely related to the attitude of the ageing adults towards the system. If elders feel uncomfortable at their own home, it will be almost impossible to induce a positive mood. The use of intrusive sensors and the deployment of noticeable sensors in the house do not help. As a result, the use of reduced body sensors and their proper placement in the house are major issues that should not be underestimated. Actually, we are studying the possibility of using body sensors that look like jewels or to integrate other sensors into the furniture or structure of the house.

## **6. Conclusion**

This paper has described a project named "Improvement of the Elderly Quality of Life and Care through Smart Emotion Regulation". The objective of the project is to find solutions for improving the quality of life and care of the elderly who can or wants to continue living at home by using emotion regulation techniques. For this sake, we have introduced the four main modules that make up the smart ambient system. These are "Emotion Recognition", "Emotion Regulation", "Ambient Adaptation" and

“Expertise Provision”. Heterogeneous vision and body sensors are used for monitoring the elderly and detecting his/her emotions, whilst music, colour and light are the simulating means to regulate the emotions towards a positive and pleasant mood.

This paper has also justified the need of a specific software architecture for emotion recognition and regulation tasks. Besides, we have introduced a new gerontechnological approach for monitoring the elderly at home. In first place, the goal is to detect the elder's emotions by analysing their physiological signals, facial expression, behaviour and voice. Then, the system provides the best-tailored reactions of the ambience to regulate the elder's emotions towards a positive emotion. The current state-of-the-art in emotion regulation through music, colour and light is used with the final goal of enhancing the quality of life of elder people living alone at their homes. After describing the benefits of our automated solutions for emotion recognition and regulation in the elderly, we have detailed the eight levels of the proposed architecture for implementing the smart ambient system.

## Acknowledgements

This work was partially supported by Spanish Ministerio de Economía y Competitividad / FEDER under TIN2013-47074-C2-1-R grant. José Carlos Castillo was partially supported by a grant from Iceland, Liechtenstein and Norway through the EEA Financial Mechanism, operated by Universidad Complutense de Madrid.

## Conflict of Interest

The authors declare that they have no conflict of interest.

## References

1. Bachorik JP, Bangert M, Loui P, Larke K, Berger J, Roew R, et al. Emotion in motion: investigating the time-course of emotional judgments of musical stimuli. *Music Perception* 2009;26:355–364.
2. Bower GH. Mood and memory. *American Psychologist* 1981;36:129–148.
3. Brewer D, Doughtie EB. Induction of mood and mood shift. *Journal of Clinical Psychology* 1980;36:215–226.
4. Castillo JC, Fernández-Caballero A, Castro-González Á, Salichs MA, López MT. A framework for recognizing and regulating emotions in the elderly. *Lecture Notes in Computer Science* 2014;8868:320-327.
5. Chen M, Gonzalez S, Vasilakos A, Cao H, Leung VCM. Body area networks: a survey. *Mobile Networks and Applications* 2011;16:171–193.

6. Devillers L, Lamel L. Emotion detection in a task-oriented dialogs. In: Proc. IEEE International Conference on Multimedia & Expo, Baltimore, MD, USA, July 2003; vol. 3, pp. 549–552.
7. Ekman P, Friesen WV, Hager J C. Facial Action Coding System [E-book]. Salt Lake City, UT: Research Nexus , 2002.
8. Elliot AJ, Maier MA. Color and psychological functioning. *Current Directions in Psychological Science* 2007;16:250–254.
9. Esposito A, Fortunati L, Lugano G. Modeling emotion, behavior and context in socially believable robots and ICT interfaces. *Cognitive Computation* 2014;6:623-627.
10. FaceSense. <http://affect.media.mit.edu/projects.php?id=2523>, 2013 (last access July 29, 2015).
11. Fernández-Caballero A, López MT, Saiz-Valverde S. Dynamic stereoscopic selective visual attention (DSSVA): integrating motion and shape with depth in video segmentation. *Expert Systems with Applications* 2008;34:1394-1402.
12. Fernández-Caballero A, Latorre JM, Pastor JM, Fernández-Sotos A. Improvement of the elderly quality of life and care through smart emotion regulation. *Lecture Notes in Computer Science* 2014;8868:348-355.
13. Fernández-Caballero A, Castillo JC, López MT, Serrano-Cuerda J, Sokolova MV. INT3-Horus framework for multispectrum activity interpretation in intelligent environments. *Expert Systems with Applications* 2013;40:6715–6727.
14. Fernández-Sotos A, Fernández-Caballero A, Latorre JM. Elicitation of emotions through music: the influence of note value. *Artificial Computation in Biology and Medicine* 2015:488-497.
15. García-Rodríguez C, Martínez-Tomás R, Cuadra-Troncoso JM, Rincón M, Fernández-Caballero A. A simulation tool for monitoring elderly who suffer from disorientation in a smart home. *Expert Systems* 2015;32:676-687.
16. Gascueña JM, Fernández-Caballero A. On the use of agent technology in intelligent, multi-sensory and distributed surveillance. *The Knowledge Engineering Review* 2011;26:191–208.
17. Gehrig T, Ekenel HK. A common framework for real-time emotion recognition and facial action unit detection. In: Proc. IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops, Colorado Springs, USA, June 2011; pp. 1–6.
18. Gross JJ, Barrett LF. Emotion generation and emotion regulation: one or two depends on your point of view. *Emotion Review* 2011;3:8–16.
19. Gross JJ, Levenson RW. Emotion elicitation using films. *Cognition & Emotion* 1995;9:87–108.
20. Jovanov E, Milenkovic A. Body area networks for ubiquitous healthcare applications: opportunities and challenges. *Journal of Medical Systems* 2011;35:1245–1254.
21. Kim JB, ParkJS, Oh YW. Speaker-characterized emotion recognition using online and iterative speaker adaptation. *Cognitive Computation* 2012;4:398-408.

22. Kim J, André E. Emotion recognition based on physiological changes in music listening. *IEEE Transactions on Pattern Analysis and Machine Intelligence* 2008;30:2067–2083.
23. Lang PJ, Bradley MM, Cuthbert BN. International affective picture system (IAPS): affective ratings of pictures and instruction manual. Technical Report A–8, University of Florida, USA, 2009.
24. Lee H, Choi YS, Lee S, Park I.P. Towards unobtrusive emotion recognition for affective social communication. In: *Proc. IEEE Consumer Communications and Networking Conference*, Las Vegas, NV, USA, January 2012; pp. 260–264.
25. Livingstone SR, Palmer C, Schubert E. Emotion response to musical repetition. *Emotion* 2012;12:552–567.
26. Livingstone RS, Thompson WF. The emergence of music from the theory of mind. *Musicae Scientiae* 2009;17:83–115.
27. López JM, Gil R, García R, Cearreta I, Garay N. Towards an ontology for describing emotions. *Lecture Notes in Computer Science* 2008;5288:96–104.
28. Lozano-Monazor E, López MT, Fernández-Caballero A, Vigo-Bustos F. Facial expression recognition from webcam based on active shape models and support vector machines. *Ambient Assisted Living and Daily Activities* 2014:147–154.
29. Martínez-Rodrigo A, Zangróniz R, Pastor JM, Latorre JM, Fernández-Caballero A. Emotion detection in ageing adults from physiological sensors. *Ambient Intelligence-Software and Applications* 2015:253–261.
30. McRae K, Misra S, Prasad AK, Pereira SC, Gross JJ. Bottom-up and top-down emotion generation: implications for emotion regulation. *Social Cognitive & Affective Neuroscience* 2012;7:253–262.
31. Medjahed H, Istrate D, Boudy J, Baldinger JL. A pervasive multi-sensor data fusion for smart home healthcare monitoring. In: *Proc. IEEE International Conference in Fuzzy Systems*, Taipei, Taiwan, June 2011; pp. 1466–1473.
32. Niedenthal PM, Halberstadt JB, Setterlund MC. Being happy and seing happy: Emotional state mediates visual word recognition. *Cognition & Emotion* 1997;11:403–432.
33. Ortiz-García-Cervigón V, Sokolova MV, García-Muñoz RM, Fernández-Caballero A. Led strips for color- and illumination-based emotion regulation at home. *Ambient Assisted Living. Development and Testing of ICT-based Solutions in Real Life Situations* 2015 (in press).
34. Pail G, Huf W, Pjrek E, Winkler D, Willeit M, Praschak-Rieder N, et al. Bright-light therapy in the treatment of mood disorders. *Neuropsychobiology* 2011;64:152–162.
35. Pavón J, Gómez-Sanz JJ, Fernández-Caballero A, Valencia-Jiménez JJ. Development of intelligent multi-sensor surveillance systems with agents. *Robotics and Autonomous Systems* 2008;55:892–903.
36. Picard RW, Vyzas E, Healey J. Toward machine emotional intelligence: analysis of affective physiological state. *IEEE Transactions on Patterns Analysis and Machine Intelligence* 2001;23:1175–1191.
37. Sadri F. Ambient intelligence: A survey. *ACM Computing Surveys (CSUR)* 2011;43(4): pp.36:1-36:66.

- 1 38. Salovey P, Rothman AJ, Detweiler JB, Steward WT. Emotional states and physical health. *American*  
2 *Psychologist* 2000;55:110–121.
- 3 39. de Santos A, Sánchez C, Guerra J, Bailador G. A stress-detection system based on physiological signals and  
4 fuzzy logic. *IEEE Transactions on Industrial Electronics* 2011;58:4857–4865.
- 5 40. Schachter S, Singer JE. Cognitive, social and physiological determinants of emotional state. *Psychological*  
6 *Review* 1962;69:379–399.
- 7 41. Scherer K. What are emotions? And how can they be measured? *Social Science Information* 2005;44:695–729.
- 8 42. Silva PRD, Osano M, Marasinghe A, Madurapperuma AP. Towards recognizing emotion with affective  
9 dimensions through body gestures. In: *Proc. 7th IEEE International Conference on Automatic Face and Gesture*  
10 *Recognition*, Southampton, United Kingdom, April 2006; pp. 269–274.
- 11 43. Sokolova MV, Fernández-Caballero A, López MT, Martínez-Rodrigo A, Zangróniz R, Pastor JM. A distributed  
12 architecture for multimodal emotion identification. *Trends in Practical Applications of Agents, Multi-Agent*  
13 *Systems and Sustainability* 2015:125–132.
- 14 44. Sokolova MV, Fernández-Caballero A, Ros L, Fernández-Aguilar L, Latorre JM. Experimentation on emotion  
15 regulation with single-colored images. *Ambient Assisted Living. Development and Testing of ICT-based*  
16 *Solutions in Real Life Situations* 2015 (in press).
- 17 45. Soleymani M, Lichtenauer J, Pun T, Pantic M. A multi-modal affective database for affect recognition and  
18 implicit tagging. *IEEE Transactions on Affective Computing* 2012;3:42–55.
- 19 46. Susskind JM, Littlewort G, Bartlett MS, Movellan J, Anderson AK. Human and computer recognition of facial  
20 expressions of emotion. *Neuropsychologia* 2007;45:152–162.
- 21 47. Velten E. A laboratory task for induction of mood states. *Behaviour Research and Therapy* 1968;6:473–482.
- 22 48. Vogt T, André E, Wagner J. Automatic recognition of emotions from speech: a review of the literature and  
23 recommendations for practical realisation. *Lecture Notes in Computer Science* 2008;4868:75–91.
- 24 49. Wagner J, André E, Jung F. Smart sensor integration: a framework for multimodal emotion recognition in real-  
25 time. In: *Proc. 3rd International Conference on Affective Computing and Intelligent Interaction and*  
26 *Workshops*, Amsterdam, The Netherlands, September 2009; pp. 1–8.
- 27 50. Zhou J, Yu C, Riekkä J, Kärkkäinen E. AmE framework: a model for emotion aware ambient intelligence. In  
28 *Proceedings of the second International Conference on Affective Computing and Intelligent Interaction*  
29 *(ACII2007): Doctoral Consortium*, September 2007.

Figure 1. General layout of the Smart Emotion Recognition and Regulation project.

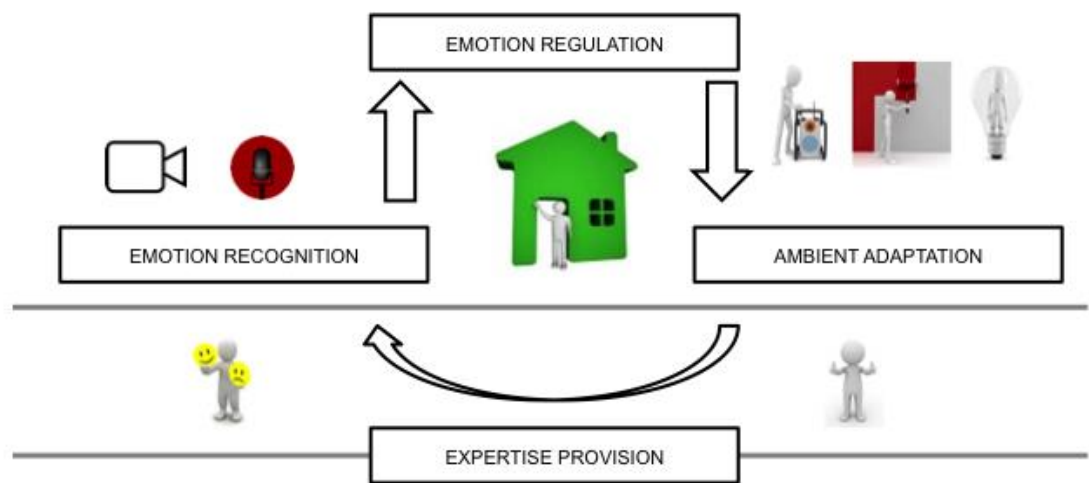


Figure 2. Main tasks of the system.

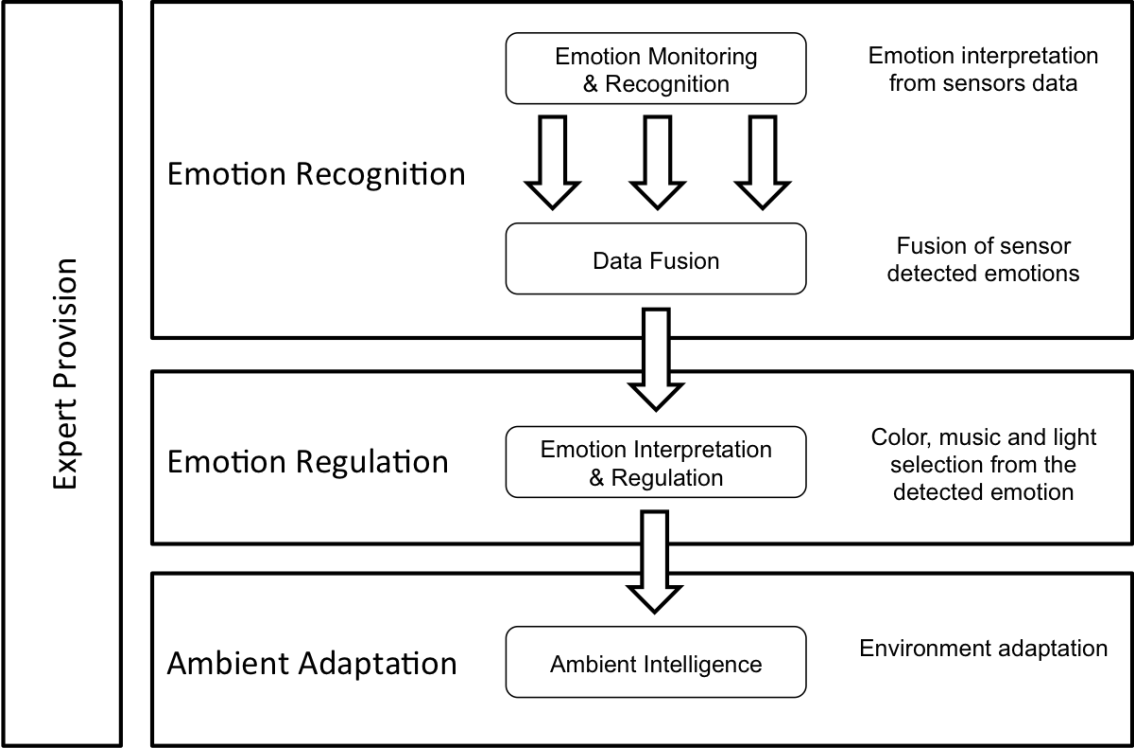


Figure 3. Expertise provision.

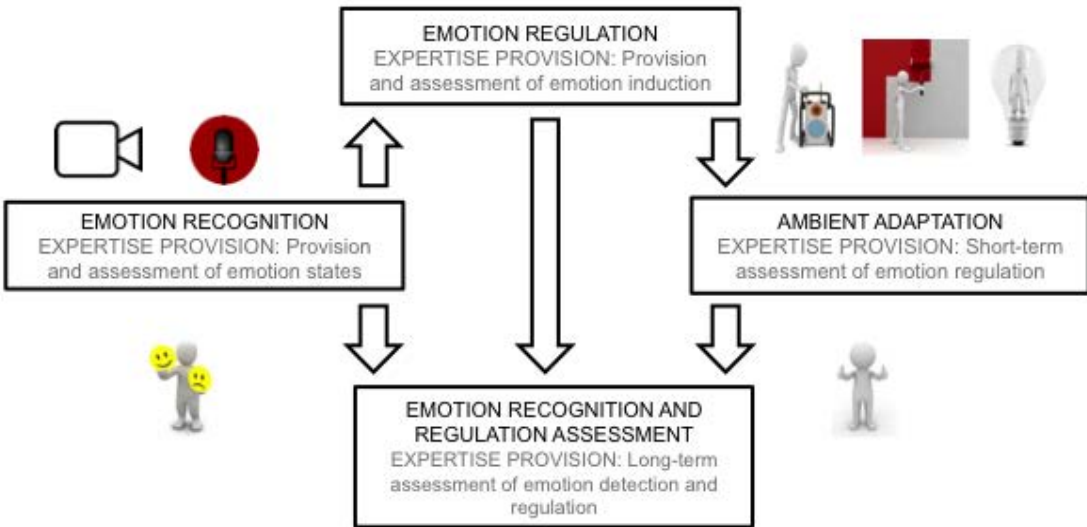


Figure 4. Software Architecture for Emotion Recognition and Regulation

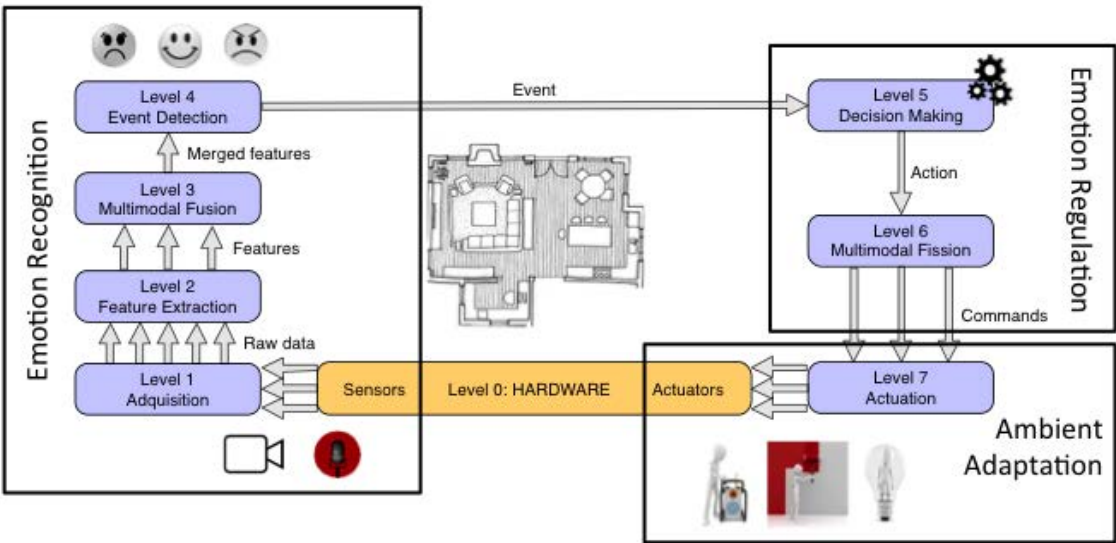


Figure 5. Conceptual description of the sensor deployment: Left, ambient sensors, the red shaded areas are considered as private (e.g. bathroom or bedroom); hence, no sensors are placed there. Right, body sensors, a wristband is a low intrusive option for sensor placement

