Chapter 10
Emotion Detection and Regulation from Personal Assistant Robot in Smart Environment

José Carlos Castillo, Álvaro Castro-González, Fernándo Alonso-Martín, Antonio Fernández-Caballero and Miguel Ángel Salichs

Abstract This paper introduces a proposal for integrating personal assistant robots with social capacities in smart environments. The personal robot will be a fundamental element for the detection and healthy regulation of the affect of the environment's inhabitants. A full description of the main features of the proposed personal assistant robot are introduced. Also, the multi-modal emotion detection and emotion regulation modules are fully described. Machine learning techniques are employed for emotion recognition from voice and images and both outputs are merged to achieve the detected emotion.

10.1 Introduction

Smart environments evolve from ubiquitous computing following the idea of "a physical world that is richly and invisibly interwoven with sensors, actuators, displays, and computational elements, embedded seamlessly in the everyday objects of our lives, and connected through a continuous network" [1]. Smart environments are composed of several heterogeneous sensors placed throughout the environment, thus providing great amounts of data. Scalable and flexible platforms integrate such devices and provide applications with the necessary interfaces to interoperate with the information coming from the available resources. The main goal of a smart environment is using this information to achieve a more comfortable life for its inhabitants.

Some recent previous works of our research teams hold the objective to achieve emotion detection and regulation in smart environments through the incorporation of

J.C. Castillo (✉) · Á. Castro-González · F. Alonso-Martín · M.Á. Salichs
Department of Systems Engineering and Automatic, University Carlos III of Madrid, Getafe, Spain
e-mail: jocastil@ing.uc3m.es

A. Fernández-Caballero
Departamento de Sistemas Informáticos, Universidad de Castilla-La Mancha, Ciudad Real, Spain
e-mail: Antonio.Fdez@uclm.es

A. Costa et al. (eds.), *Personal Assistants: Emerging Computational Technologies*, Intelligent Systems Reference Library 132, DOI 10.1007/978-3-319-62530-0_10

several sensing and actuation technologies. The ultimate aim is to maintain a healthy affective state of the subject. In our opinion, the inclusion of a robotic platform in such smart environments opens new possibilities for perceiving the inhabitant's emotional state and properly acting on his/her mood. In particular, social robots and humans tend to establish affective bonds that could be exploited in a smart environment. Therefore, these robotic platforms are not only a mere set of mobile sensors and actuators as robot companions but also a way to offer easy and amiable interaction between humans and the smart environment.

This paper introduces a proposal to incorporate a personal assistant robot with social skills into a smart environment for the sake of complementing intelligent affective detection and regulation strategies. According to the European Commission, "personal assistant robots would be able to learn new skills and tasks in an active open-ended way and to grow in constant interaction and co-operation with humans" [2]. Here, it is our intention to establish the foundations for designing social personal assistant robots as an outstanding part to enhance the capabilities of smart environments.

The rest of the proposal is described next. Section 10.2 introduces the description of the proposed personal assistant robot. This section shows the main features that the social robot has to build-in, that is, it has to be mobile, it has to wear *ears* and *eyes*, and it has to incorporate expressive capabilities. Afterwards, in order to achieve affective detection, Sect. 10.3 describes the main features of the multi-modal emotion detection module included in the social robot. Emotion detection is done in this proposal through applying machine learning techniques to both voice and video streams, analysing the sensory data separately and then merging both outcomes. Next, the proposed affective regulation module is described in Sect. 10.4. Lastly, our conclusions are summarised in Sect. 10.5.

10.2 The Personal Assistant Robot

As described before, in previous works we have been designing smart environments endowed with perception and action capabilities to detect and regulate the user's emotional state [3, 6]. However, this approach presents several drawbacks and limitations. Firstly, the environment has to be altered resulting in high economic cost of installation and maintenance. Besides, the installed devices/sensors are usually perceived by the user as moderately to very intrusive; and therefore the smart system could be rejected. Finally, the location of the sensors is static. Thus, the monitoring elements are located far from the user (walls or ceiling) most of the time, lowering the quality of the data captured.

The use of a social robot mitigates these problems. Social robots are intended to interact with people following established behavioural norms [7]. These robots coexist in daily environments (e.g. homes, schools, hospitals or museums) helping humans to perform particular tasks, assisting patients with their therapies, or just accompanying people. This kind of activities requires the robot appearance to be appealing and friendly, unlike traditional industrial robots for instance. Consequently social robots need to be carefully designed. Traditionally, these robots present an external look similar to animals (dog robot [8], cat robot [9], a seal robot [10]), cartoons (big eyes, or round shapes), or even a mix of both, such us the little, furry DragonBot [11].

A social robot's friendly appearance will for sure help to be accepted by the users and emotional bonds, similar to those existing with a pet or a friend, will arise. The inclusion of head, face and arms in the robot benefits perception of a living entity or an animal. In addition, all sensors are located on board, avoiding the need of a physical set-up and the modification of the environment. Moreover, this implies that the robot is able to move around the environment to approach the user, or to move to another room. Therefore, its is mandatory that the social robot integrates a mobile platform. Most of the human-robot interactions with social robots are usually conducted is short distances (around few meters), similar to human-human interactions. This proximity between the user and the robot ensures the acquisition of a high quality data from the embarked sensors.

Another important aspect is the robot's size. This will depend on the users it is intended to interact with. In usual daily living environments, the robot would live with adults who are able to easily communicate with the robot in a natural manner. A small robot would cause users to bend when communicating, and a big robot could be overwhelming. In consequence, considering the average height of adult men and women, we believe that the robot's height should range between 1.5 and 1.7 m.

In the next subsections, the required hardware (sensors and actuators) that are needed by the robotic platform to successfully operate in a smart environment is described.

10.2.1 A Mobile Social Robot

As already stated, the social robot needs to autonomously move around to approach, follow, or search for users. To date, the most reliable technology to achieve this functionality is a mobile platform based on wheels. In order to allow a safe navigation, the robot needs a map of the environment for detecting obstacles and calculating the trajectory from one point to another. These tasks require that the robot gets data to build the map and perform obstacle detection.

A clever combination is to install a 2D LIDAR in the base and a 3D depth camera. The LIDAR device provides accurate data to build the map and to navigate. Nevertheless, it only provides 2D data that could be insufficient to detect obstacles like a table or a stretcher. This type of barriers are easily detected by a 3D camera, like the well-known RGB-D Kinect sensor. The merged data from these two types of devices, together with the precise control of the motorised wheels, results in a safe and reliable navigation of our mobile platform.

10.2.2 A Social Robot with Ears and Eyes

Based on previous works (e.g. [12, 13]), we want to use a mobile, social robot as a smart sensor to estimate the user's emotional state or mood. We will use two sources of information mainly, the user's voice and image. This is why our social robot needs "ears" and "eyes".

In order to detect users utterance, we suggest to endow the robot with an array of directional microphones that provide high quality sound, even if the source is not very close, but at the cost of a narrow operational angle. This is why we need a ring of micros around the robot. As an extra benefit, the ring of microphones allows the localisation of the audio source, which results very interesting during human-robot interaction.

In the audio domain, the emotional content of the signal can provide helpful information to asses the user's state—for instance, the tone of voice of a person about to cry is completely different to a happy person's one. In relation with the robot's "eyes", the user's face is analysed to assess the current emotion [16, 17].

10.2.3 A Social Robot with Expressive Capabilities

In this paper, our intention is to present a system that combines a smart environment with a social robot to modify the user's emotional state. Therefore, the robot needs to have expressive capabilities to impact on the user's affective state. The way the robot moves could affect the user's perception of the robot. In case the robot moves with abrupt movements, the robot could be perceived as dangerous and the user would feel agitated. On the contrary, if the robot performs soft, smooth movements, it could contribute to calm the user.

Many social robots are endowed with screens (for example in the eyes, the face, or the body [14, 15]). The dominant colour displayed on the screens can change according to the desired effect on the user, as described later on in the section dedicated to emotion regulation. Also the multimedia content shown can help to communicate an emotion: high-valence and low-arousal images similar to those labelled in the International Affective Picture System database [18] help to calm down a person, or emoticons help to perceive the robot's artificial emotional state (like smiling or sad faces). Similarly to the intention followed with screens, some recent robots are endowed with a projector to display images on the environment. Due to the impact of the projections on the surroundings, this sounds like a promising method to guide the user's mood. Besides, colour LEDs can be placed in different parts of the robot, e.g. chicks, mouth, or base. The colour displayed in these elements would depend on the user requirements.

Finally, other crucial robotic elements capable of altering the human mood are non-verbal sounds. Different sounds full of emotional content can be synthesized or reproduced by the robot. Most of these elements, both sensors and actuators, are

Fig. 10.1 Maggie (*left*) and MBot (*right*), the mobile social robots from the Robotics Lab at Universidad Carlos III de Madrid

present in the social robots from the Robotics Lab at Carlos III University of Madrid (Fig. 10.1). Additionally, these robots are a good example of mobile platforms with a friendly external appearance and an appropriate size.

10.3 The Multi-modal Emotion Detection Module

The emotion detection module combines analysis in two different domains. On the one hand, it uses voice analysis techniques to assess emotions in the user's speech. Computer vision complement the audio-based approach by adding face detection and recognition capabilities that are also able to detect emotions in video sequences. Finally, a bayesian decision rule is in charge of merging the outputs from both analyses, generating a unique result. The combination of both approaches is described in Fig. 10.2.

In the figure, GEVA is the component in charge of emotion detection through voice analysis while GEFA is in charge of emotion detection through video (facial expressions) analysis. Therefore, the input to GEVA is the user's voice, where features are extracted and a couple of classifiers obtain one of the basic emotions "neutral", "happiness", "sadness" and "surprise". Complementary, GEFA uses a couple of tools (as described in Sect. 10.3.2) to obtain the same basic emotions. Lastly, there is a decision rule that uses the output of both components to get a final consensus basic emotion as detailed in Sect. 10.3.3.

Four basic emotions, "neutral", "happiness", "sadness" and "surprise", have been chosen for two main reasons: The first one is that these basic emotions can be represented very easily in a classical circumplex model of affect [19]. This model suggests that emotions are distributed in a two-dimensional circular space, containing arousal (activation) and valence (pleasantness) dimensions. Arousal represents the vertical axis and valence expresses the horizontal axis, while the centre of the circle means a

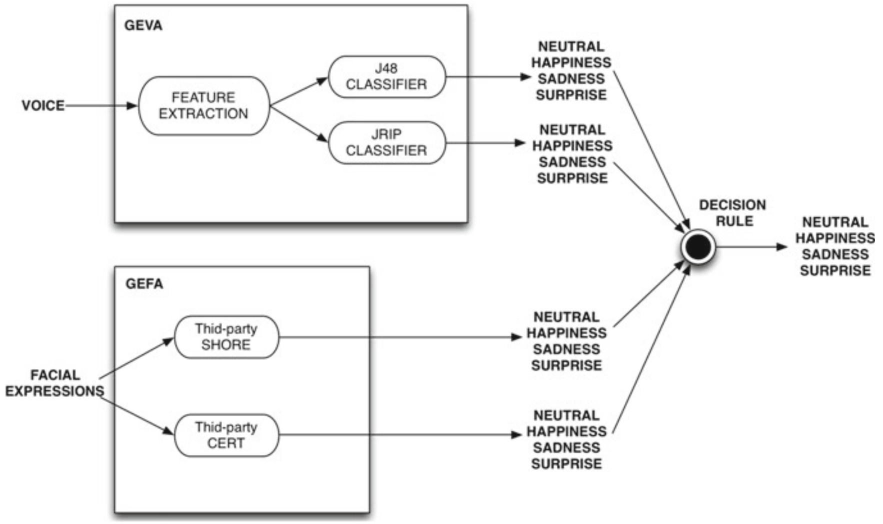

Fig. 10.2 General scheme of the multimodal emotion detection module

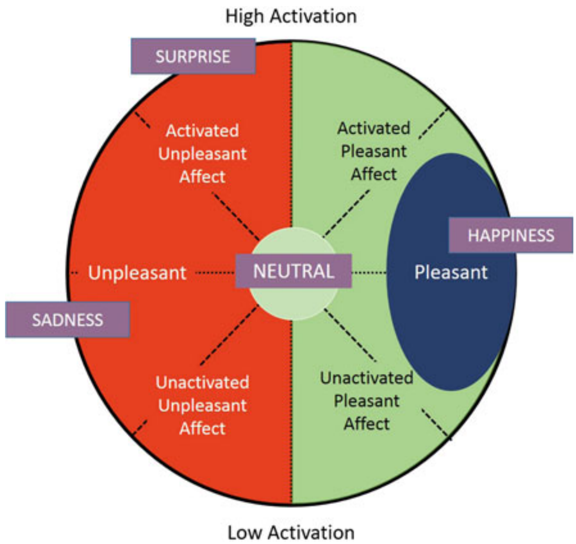

Fig. 10.3 Circumplex model of affect including the four basic emotions considered

neutral valence and a medium level of arousal. In this model, emotional states can be represented at any level of valence and arousal, or at a neutral level of one or both of these factors. Figure 10.3 offers a circumplex model of affect particularised for the four emotions considered in this proposal.

In second place, the pathways to regulate the chosen emotions are also well-known in Psychology. In fact, empirical evidence shows that the two emotional dimensions of arousal and valence are not independent of each other [20]. Rather, they form a V ("boomerang")-shaped function, with the unpleasant stimuli tending to be more highly arousing than pleasant stimuli, and both pleasant and unpleasant stimuli being more arousing than neutral stimuli. Considerable evidence supports this "boomerang" shape in averaged data on arousal and valence ratings of people's reactions to affective visual stimuli. In Fig. 10.3 you may also observe that the two axis split the circumference into four parts denominated *Activated Unpleasant Affect*, *Unactivated Unpleasant Affect*, *Unactivated Pleasant Affect* and *Activated Pleasant Affect*.

10.3.1 Emotion Detection Through Voice Analysis

Changes in emotions can be detected by analysing the interlocutor's voice tone [21]. This proposal exploits the capabilities of a multi-domain audio-based emotion detection approach named GEVA [22, 23]. It extracts some relevant features from the input signals that are sent to a classifier that obtains the emotion as an output.

10.3.1.1 Feature Extraction

In this first phase of the GEVA component, the analysis is performed in three domains: (i) time domain, in which the analysis is directly performed in the analogue signal; (ii) frequency domain, in which the Fast Fourier Transform is applied to the input signal; and (iii) time-frequency domain, in which the analysis is performed after applying the Discrete Haar Wavelet transform to the input data. Features are extracted in these three domains using Chuck,[1] a programming language for real-time sound synthesis and sound wave analysis. The features extracted are the following ones:

- **Pitch**: Sound frequency as perceived by the human ear.
- **Flux**: How big the sound variations are in the frequency domain. Values close to zero indicate small differences whilst values near one imply that there are important variations in the frequency domain.
- **Rolloff-95**: Frequency value at which 95% of the signal energy is already contained.
- **Centroid**: Median of the signal spectrum. That is, the frequency the signal approaches the most. This frequency is related to the tone of a sound (timber).
- **Zero-crossing rate**: Number of times that the signal crosses the zero (x axis). This is useful to distinguish between background noise and voice since the former tend to cross the axis more frequently than the latter.

[1]Chuck website: http://chuck.cs.princeton.edu/uana/.

- **Signal-to-noise ratio (SNR)**: Voice signal volume with respect to the background noise.
- **The communicative rhythm**: This feature is defined as the number of words pronounced per minute. This is useful when trying to distinguish emotions since each of them have characteristic communication rhythms.

10.3.1.2 Classification

The features described in the previous phase characterise the main parameters of the acquired sound wave. Next, these are used to train a classifier that obtains the detected emotion. GEVA implements an off-line universal classifier, trained with no constraint regarding the type of user. Therefore, the dataset for training contains sentences from a wide range of users of different age, language and gender. The dataset contains samples of tagged locutions for the four target emotions: “happiness”, “sadness”, “surprise” and “neutral”. The training dataset is composed from several sources:

- Voice examples from the developers simulating emotions.
- Interviews with colleagues asked to fake emotions.
- Real or spontaneous interactions between the robot and colleagues.
- Interviews obtained from the Internet.
- TV shows from the Internet.
- Audiobooks from the Internet.
- Databases with a tagged voice corpus:
  - Emotional Prosody Speech Database (LDC): with 15 types of emotions [24].
  - Berlin Emotional Speech Database (EmoDB): with seven types of emotions [25].
  - Danish Emotional Speech Corpus (HUMANAINE): with five types of emotions [26].
  - FAU Aibo Emotion Corpus: 8.9 h of spontaneous voice recordings from 51 children, with 11 types of emotions[2] [27].

In order to find the best-suited classifier for emotion detection we have used the software library Weka [28], which integrates more than one hundred machine learning techniques.

Using our linguistic corpus and the voice feature extraction module, a file containing about 500 emotion-tagged locutions with training patterns for Weka is built. Using cross-validation over the dataset, two algorithms are selected: a decision tree-based algorithm, J48, and a decision rule-based one, JRIP, with a 80.51% and 81, 15% of success rate, respectively (see [23] for more details).

[2] Fau website: http://www5.cs.fau.de/de/mitarbeiter/steidl-stefan/fau-aibo-emotion-corpus.

10.3.2 Emotion Detection Through Video Analysis

The literature offers a number of techniques for emotion detection from facial expressions [16, 17]. Most of these techniques follow the same steps: (i) face detection in the image flow; (ii) facial feature extraction, such as distance between eyes or mouth shape, among others; and (iii) emotion classification using the previous features and several learning techniques.

Currently, there is a number of algorithms for user face detection and tracking in multiple applications. Usually, these techniques focus on the detection of frontal faces, identifying candidate regions for characteristic components of a face such as eyes, nose or mouth. Among the most widespread techniques we may find the Viola and Jones algorithm [29], although there are other trends such as the ones based on machine learning, e.g. neural networks [30] or support vector machines (SVM) [31]. Besides, well-known image processing libraries such as OpenCV[3] also include other wide-spread algorithms for face detection.

After locating a face in the image, it is necessary to extract features to simplify the classifiers' work. For this purpose, there are approaches based on interest points and their geometric relationships (local approximation) [32, 33]. Also, some works deal with this problem through representing the face as a whole unit, e.g. placing a 3D mesh over the detected face, calculating the differences between the current detection and a target one (holistic approach) [34, 36]. These two approaches can be mixed together by using the interest points to determine the initial position of the mesh [37].

There are works that verse about the possibility of having a universal facial expressions classification for emotion recognition, taking into account gender, age and culture [38, 39]. This is related to the different intensities when expressing emotions which make the recognition problem more challenging when the intensity is low, as facial variations are more subtle.

In conjunction with the advances in the literature, some works propose off-the-shelf tools for face detection, feature extraction and emotion classification. An example is CERT [40], a visual tool that allows classifying six emotions in real time as well as 16 action units (AU) from the Facial Action Coding System [41]. Feature extraction follows a local approximation-based approach and AUs are classified using a SVM with results ranging between 80 and 90% for on-line emotion recognition. CERT outputs the intensity value for each detected emotion. In this case, the outputs are: fun, joy, smile detector (these three are grouped in one set as happiness), disgust and sadness (grouped as sadness), surprise, neutral, fear and anger (these last two are not considered in our approach).

There is another interesting tool, SHORE [42, 43], for face detection and emotion recognition in challenging environments (e.g. lightning changes). This software implements a holistic approach able to track the position and orientation of user faces although the detection performance is better when dealing with front views. SHORE provides the intensity values of the following emotional states: happiness, sadness,

[3] OpenCV website: http://opencv.org.

surprise and anger. Again, this last emotion is not considered in our proposal. Moreover, if none of these emotional states has an intensity greater than 50%, then it is assumed that the emotional state of the user is neutral.

These two tools, CERT and SHORE have been used to build the GEFA module. Using these two tools, GEFA is able to operate at a maximum interaction distance of 4.5 m.

10.3.3 Integration of GEVA and GEFA

The outputs of the GEVA and GEFA emotion detection components need to be merged in order to get a unified outcome. Figure 10.4 shows a scheme of the process to determine the user emotion. A previous work [23] demonstrates that the accuracy of the visual emotion detection system drops when users are talking, as the image libraries for training are mainly composed of non-speaking face images. Therefore, when a user is quiet, the output of GEFA is considered and, otherwise, the information coming from GEVA is the one taken into account. These outputs may be seen as independent, since they occur at different time instants. Nevertheless, emotions do not occur instantaneously, so a decision rule can be useful to combine the information from both emotion detection modules, generating a single output with lower temporal resolution.

Once the success rate of each module (with two classifiers each) is estimated, a decision rule is applied in order to create a unified output. Prior to that, it is necessary to determine the *confidence degree* of the output (the detected emotion) for each

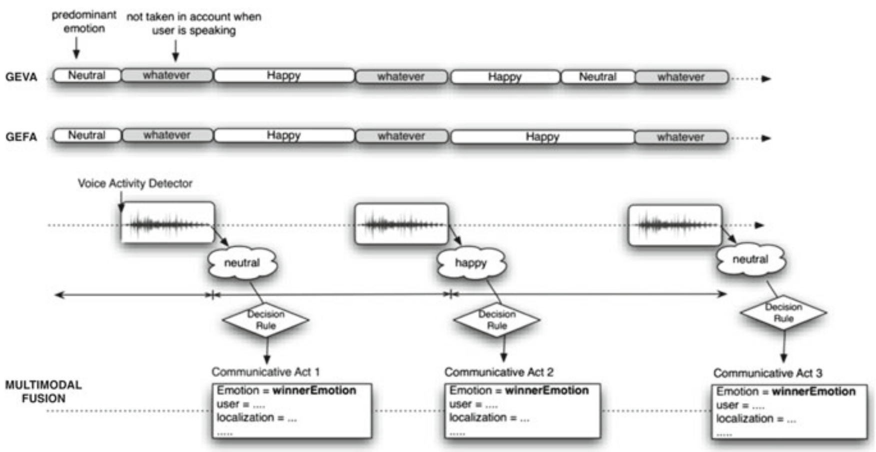

Fig. 10.4 Graphical representation of the proposed system. Outputs from GEVA and GEFA are fused using the decision rule

classifier. In this work, the *confidence degree* of each output is calculated using the Bayes theorem and the confusion matrices. The decision rule used is as follows:

Decision rule: The easiest way of determining the most probable state of the system is to find the state $\hat{s}$ that has the highest probability among all classifiers given by:

$$\hat{s} \in \mathbf{S}, \hat{C} \in \mathbf{C} \mid \forall s_i \in \mathbf{S}, \forall C \in \mathbf{C}\ ,\ p(S = s_i|S_C) \leq p(S = \hat{s}|S_{\hat{C}}). \tag{10.1}$$

where the current state of the system is S and the list of states is $\mathbf{S} = \left\{s_1, \ldots, s_n\right\}$ where $n \in \mathbb{N}$, and $\mathbf{C}$ is a collection of m classifiers $\mathbf{C} = C_1, \ldots, C_m, m \in \mathbb{N}$.

This decision rule allows fusing the four outputs of the integrated classifiers in a single one with a higher accuracy with respect to the isolated original ones. Details about the mathematics to estimate the conditional probabilities that result in the decision rule are available in a previous paper [23].

10.4 The Emotion Regulation Module

The aim of the emotion regulation module is to provide the best-suited conditions of music performance, as well as colour/light to attain the desired affect in the person interacting with the personal assistant robot. We propose to simultaneously use multiple induction techniques to regulate the user' mood in his/her smart environments. These techniques are fully incorporated both in the social robot and in the smart ambient. Taking advantage of the actuators embarked in the robot's body as well as the ones in the ambient, music is performed by means of an box that plays background music. And, colours are automatically displayed in the robot's parts as well as in the environment in a smart manner.

The different emotions are detected/regulated through the combination of activation/pleasantness values and position in the orthogonal dimensions.

- Neutral: medium activation/medium pleasantness; undefined orthogonal dimension.
- Sadness: medium activation/low pleasantness; Unactivated Unpleasant Affect.
- Happiness: medium activation/high pleasantness; Activated Pleasant Affect.
- Surprise: high activation; medium pleasantness; Activated Unpleasant Affect.

10.4.1 Musical Emotion Regulation

A research work states that the most potent and frequently studied musical cues for emotion elicitation are mode, tempo, dynamics, articulation, timbre, and phrasing [44]. In fact, musical parameters such as tempo or mode are inherent properties of

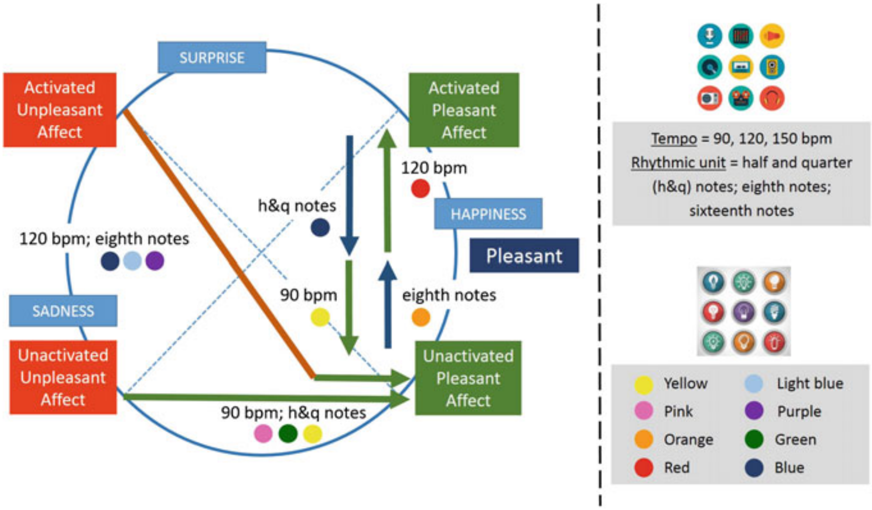

Fig. 10.5 Circumplex model for emotion regulation from personal assistant robot

musical structure [45]. These characteristics are important, as it has been shown that they influence listeners' emotions.

Musical emotion regulation focuses in our case on note value, an important musical cue related to rhythm. The two musical components of note value used are tempo and rhythmic unit. Beyond the effect of mode, tempo was found to modulate the emotional ratings, with faster tempo being more associated with the emotions of anger and happiness as opposed to slow tempo, which induces stronger feeling of sadness and serenity. This suggests that the tempo modulates the arousal value of the emotions [46].

Figure 10.5 shows the values of tempo and rhythmic unit that best provide the desired transitions from negative to positive affect in accordance with some of our prior experimental works [47, 48]. This figure is also based in the expert knowledge on the "boomerang" shape that psychologists and neuroscientists have described during the two last decades. At this point, it is important to clarify that the intervention proposed in this work is for situations where a negative affective state is unhealthy and a transition to a positive emotional state is necessary.

Thus, depending on the starting position of the emotion detected in the orthogonal dimensions, we have:

- *Activated Unpleasant Affect* is an unhealthy affective state, so that a transition to a healthy state is required. In accordance with psychological studies, firstly the activation level has to be dropped. In this case (see brown line in Fig. 10.5), a musical piece mostly based on eighth notes and a tempo of 120 beats per minute (bpm) is played. Once, the activation is below a given value, another musical piece is played (90 bpm, and half and quarter notes), which is shown with a green arrow. Notice that a complete and varied battery of musical pieces is available.

- *Unactivated Unpleasant Affect* is also an unhealthy emotional state. In this case, it is not necessary to deactivate the patient. Just as in the previous case, a musical piece is played in 90 bpm, and half and quarter notes.
- *Unactivated Unpleasant Affect* is a health affective state. However, in this proposal we will be trying to maintain the patient as "Pleasant" as possible. This is why, it is foreseen that the affective state will balance between *Unactivated Unpleasant Affect* and *Activated Pleasant Affect*, depending on the reaction of the monitored person. In order to pass to *Activated Pleasant Affect*, two steps are necessary. In the first one, a musical piece based on eighth notes is played, and in the second one it is a musical piece played at 120 bpm.
- The *Activated Pleasant Affect* case is similar to the previous one. Here, firstly a musical piece based on half and quarter notes is played, and then a musical piece is played at 90 bpm.

10.4.2 Colour/Light-Based Emotion Regulation

Light is one of the principal environmental factors that influence on a user. Therefore, its affective impact has been widely researched and reported. Exposure to bright light and luminance distribution affect our mood. This is why, many researchers believe that light can be used to alter/improve emotional states. Colour is another factor that is constantly present in any environment. Moreover, the evidences of the influence of colour upon human emotions has been discussed in a number of previous publications from our research team (e.g. [49, 50]). Many authors highlight that certain colours are highly influencing for emotion elicitation and regulation.

Therefore, colour emotion regulation bases here on the combination of colour and light due to their double influence on emotion regulation. Also, in Fig. 10.5 you can appreciate the colours that best provide the desired transitions from negative to positive mood in accordance with a prior work [51]. Again, depending on the starting position of the emotion detected in the orthogonal dimension, we have:

- *Activated Unpleasant Affect* has to transit to *Unactivated Pleasant Affect*. This is achieved in two steps. Firstly, colour/light is changed to Blue, Light blue or Purple, and, then to Pink, Green or Yellow. The order in these lists corresponds to the velocity at which these colours provide the desired effect. Obviously, this has also to be personalised in accordance with the user's reactions.
- *Unactivated Unpleasant Affect* transits directly to *Unactivated Pleasant Affect* with colour/light Pink, Green or Yellow.
- Remember that *Unactivated Unpleasant Affect* is a healthy affective state, but we try to maintain the patient as "Pleasant" as possible. This is why, it is foreseen that the affective state will balance between *Unactivated Unpleasant Affect* and *Activated Pleasant Affect*. In order to transit to *Activated Pleasant Affect*, newly two steps are necessary. In the first one, colour/light is changed to Orange, and in the second one colour/light is changed to Red.

- The *Activated Pleasant Affect* case is similar to the previous one. Here, firstly colour/light is changed to Blue, and in the second one colour/light is changed to Yellow.

10.5 Conclusions

This work offers a proposal for a smart environment for emotion regulation including the capabilities of a personal assistant robot. Although traditionally sensors are placed throughout the environment, we propose an innovative approach where several sensing devices are embarked in a social robot. This robot is in charge of data acquisition and processing. Additionally, the environment and the robot are employed to influence the user's mood.

We propose the main features a robotic platform ought to incorporate such as a mobile base with sensors for navigation. The proposed robot needs to be endowed with screens, colour LEDs, audio devices to allow high expressibility to produce the desired effect on the user.

One important aspect of the proposal is the ability to detect emotions. This is achieved by equipping the robot with cameras and microphones which output is fused using a Bayes-based decision rule. Four basic emotions (neutral, happiness, sadness, and surprise) are detected using the analysis modules, which rely on machine learning techniques based on computer vision and audio features classification. On top of that, an approach for emotion regulation using music and colour has been presented, showing how tempo and rhythmic units have to be modulated, in the case of music, and what colours provide the transitions between moods.

Acknowledgements This work was partially supported by Spanish Ministerio de Economía y Competitividad/European Regional Development Fund under TIN2015-72931-EXP and DPI2016-80894-R grants. The research leading to this work has also received funding from the projects: Development of social robots to help seniors with cognitive impairment (ROBSEN), funded by the Ministerio de Economia y Competitividad/European Regional Development Fund; and RoboCity2030-III-CM, funded by Comunidad de Madrid and co-funded by Structural Funds of the EU.

References

1. Weiser, M., Gold, R., Brown, J.S.: The origins of ubiquitous computing research at PARC in the late 1980s. IBM Syst. J. **38**(4), 693–696 (1999)
2. European Commission.: Beyond Robotics (RO) Proactive Initiative (2006). https://cordis.europa.eu/ist/fet/ro-in.htm
3. Castillo, J.C., Castro-González, Á., Fernández-Caballero, A., Latorre, J.M., Pastor, J.M., Fernández-Sotos, A., Salichs, M.A.: Software architecture for smart emotion recognition and regulation of the ageing adult. Cogn. Comput. **8**(2), 357–367 (2016)

4. Fernández-Caballero, A., Martínez-Rodrigo, A., Pastor, J.M., Castillo, J.C., Lozano-Monasor, E., López, M.T., Zangróniz, R., Latorre, J.M., Fernández-Sotos, A.: Smart environment architecture for emotion recognition and regulation. J. Biomed. Inf. **64**, 55–73 (2016)
5. Castillo, J.C., Fernández-Caballero, A., Castro-González, Á., Salichs, M.A., López, M.T.: A framework for recognizing and regulating emotions in the elderly. Ambient Assisted Living and Daily Activities, pp. 320–327 (2014)
6. Fernández-Caballero, A., Latorre, J.M., Pastor, J.M., Fernández-Sotos, A.: Improvement of the elderly quality of life and care through smart emotion regulation. Ambient Assisted Living and Daily Activities, pp. 348–355 (2014)
7. Bartneck, C., Forlizzi, J.: A design-centred framework for social human-robot interaction. In: 13th IEEE International Workshop on Robot and Human Interactive Communication, pp. 591–594 (2004)
8. Moon, Y.E.: Sony AIBO: the world's first entertainment robot. Harvard Business School Case 502-010 (2001)
9. van Breemen, A., Yan, X., Meerbeek, B.: iCat: an animated user-interface robot with personality. In: The Fourth International Joint Conference on Autonomous Agents and Multiagent Systems, pp. 143–144 (2005)
10. Shibata, T., Inoue, k., Irie, R.: Emotional robot for intelligent system-artificial emotional creature project. In: 5th IEEE International Workshop on Robot and Human Communication, pp. 466–471 (1996)
11. Setapen, A., Breazeal, C.: DragonBot: a platform for longitudinal cloud-HRI. Human-Robot Interaction (2012)
12. Jiang, M., Zhang, L.: Big data analytics as a service for affective humanoid service robots. Proc. Comput. Sci. **53**, 141–148 (2015)
13. Alvarez, M., Galan, R., Matia, F., Rodriguez-Losada, D., Jimenez, A.: An emotional model for a guide robot. IEEE Trans. Syst. Man Cybern.? Part A Syst. Hum. **40**(5), 982–992 (2010)
14. Pérula-Martínez, R., Salichs, E., Encinar, I.P., Castro-González, Á., Salichs, M.A.: Improving the expressiveness of a social robot through luminous devices. In: 10th ACM/IEEE International Conference on Human-Robot Interaction. Extended Abstracts, pp. 5–6 (2015)
15. Mirnig, N., Tan, Y.K., Chang, T.W., Chua, Y.W., Dung, T.A., Li, H., Tscheligi, M.: Screen feedback in human-robot interaction: how to enhance robot expressiveness. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication, pp. 224–230 (2014)
16. Pantic, M., Rothkrantz, L.: Automatic analysis of facial expressions: the state of the art. IEEE Trans. Pattern Anal. Mach. Intell. **22**, 1424–1445 (2000)
17. Khatri, N.N., Shah, Z.H., Patel, S.A.: Facial expression recognition: a survey. Int. J. Comput. Sci. Inf. Technol. **5**(1), 149–152 (2014)
18. Lang, P.J.: The emotion probe: studies of motivation and attention. Am. Psychol. **50**(5), 372–385 (1995)
19. Russell, J.A.: A circumplex model of affect. J. Pers. Soc. Psychol. **39**(6), 1161–1178 (1980)
20. Libkuman, T.M., Otani, H., Kern, R., Viger, S.G., Novak, N.: Multidimensional normative ratings for the international affective picture system. Behav. Res. Methods **39**, 326–334 (2007)
21. Cowie, R., Douglas-Cowie, E., Romano, A.: Changing emotional tone in dialogue and its prosodic correlates. In: ESCA Tutorial and Research Workshop on Dialogue and Prosody, pp. 41–46 (1999)
22. Alonso-Martin, F., Castro-González, A., Gorostiza, J., Salichs, M.A.: Multidomain voice activity detection during human-robot interaction. In: International Conference on Social Robotics, pp. 64–73 (2013)
23. Alonso-Martin, F., Malfaz, M., Sequeira, J., Gorostiza, J., Salichs, M.A.: A multimodal emotion detection system during human-robot interaction. Sensors **13**(11), 15549–15581 (2013)
24. Liberman, M., Davis, K., Grossman, M., Martey, N., Bell, J.: Emotional Prosody Speech and Transcripts. Linguistic Data Consortium, Philadelphia (2002)
25. Vlasenko, B., Schuller, B.: Combining frame and turn-level information for robust recognition of emotions within speech. Interspeech, pp. 27–31 (2007)

26. Schuller, B., Arsic, D.: Emotion recognition in the noise applying large acoustic feature sets. Speech Prosody, 276–289 (2006)
27. Steidl, S.: Automatic Classification of Emotion Related User States in Spontaneous Children's Speech, pp. 1–250. University of Erlangen, Logos-Verlag (2009)
28. Holmes, G., Donkin, A., Witten, I.: WEKA: a machine learning workbench. The IEEE Australian New Zealand Intelligent Information Systems Conference, pp. 357–361 (1994)
29. Viola, P., Jones, M.J.: Robust real-time face detection. Int. J. Comput. Vis. **57**(2), 137–154 (2004)
30. Rowley, H.A., Baluja, S., Kanade, T.: Neural network-based face detection. IEEE Trans. Pattern Anal. Mach. Intell. **20**(1), 23–38 (1998)
31. Osuna, E., Freund, R., Girosit, F.: Training support vector machines: an application to face detection. In: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 130–136 (1997)
32. Kobayashi, H., Hara, F.: Facial interaction between animated 3d face robot and human beings. In: The IEEE International Conference on Systems, Man, and Cybernetics. Computational Cybernetics and Simulation 4, pp. 3732–3737
33. Padgett, C., Cottrell, G.: Representing face images for emotion classification. Advances in Neural Information Processing Systems 9 (1997)
34. Cootes, T., Edwards, G., Taylor, C.: Active appearance models. In: 5th European Conference on Computer Vision, pp. 484–498 (1998)
35. Terzopoulos, D., Waters, K.: Analysis and synthesis of facial image sequences using physical and anatomical models. IEEE Trans. Pattern Anal. Mach. Intell. **15**, 569–579 (1993)
36. Lucey, S., Matthews, I., Hu, C.: AAM derived face representations for robust facial action recognition. In: 7th International Conference on in Automatic Face and Gesture Recognition (2006)
37. Kearney, G., McKenzie, S.: Machine interpretation of emotion: design of a memory-based expert system for interpreting facial expressions in terms of signalled emotions. Cogn. Sci. **17**, 589–622 (1993)
38. Ekman, P., Friesen, W.: Constants across cultures in the face and emotion. J. Pers. Soc. Psychol. **17**(2), 124–129 (1971)
39. Russell, J., Dols, J.: The Psychology of Facial Expression. Cambridge University Press (1997)
40. Littlewort, G., Whitehill, J., Wu, T.-F., Butko, N., Ruvolo, P., Movellan, J., Bartlett, M.: The motion in emotion—a CERT based approach to the FERA emotion challenge. Face Gesture **2011**, 897–902 (2011)
41. Ekman, P., Friesen, W., Hager, J.: Facial action coding system: A technique for the measurement of facial movement. Number A Human Face. Consulting Psychologists Press, Palo Alto, USA (1978)
42. Wierzbicki, R.J., Tschoeppe, C., Ruf, T., Garbas, J.U.: EDIS-emotion-driven interactive systems. Int. SERIES Inf. Syst. Manag. Creat. Media **1**, 59–68 (2013)
43. Küblbeck, C., Ernst, A.: Face detection and tracking in video sequences using the modified census transformation. Image Vis. Comput. **24**, 564–572 (2006)
44. Gabrielsson, A., Lindstrom, E.: The role of structure in the musical expression of emotions. In: Theory, Research, and Applications, Handbook of Music and Emotion, pp. 367–400 (2010)
45. van der Zwaag, M.D., Westerink, J.L., van den Broek, E.L.: Emotional and psychophysiological responses to tempo, mode, and percussiveness. Musicae Sci. **15**(2), 250–269 (2011)
46. Trochidis, K., Bigand, E.: Investigation of the effect of mode and tempo on emotional responses to music using EEG power asymmetry. J. Psychophysiol. **27**(3), 142–147 (2013)
47. Fernández-Sotos, A., Fernández-Caballero, A., Latorre, J.M.: Influence of tempo and rhythmic unit in musical emotion regulation. Front. Comput. Neurosci. **10**, 80 (2016)
48. Fernández-Sotos, A., Fernández-Caballero, A., Latorre, J.M.: Elicitation of emotions through music: the influence of note value. In: Artificial Computation in Biology and Medicine, 488–497 (2014)
49. Sokolova, M.V., Fernández-Caballero, A., Ros, L., Fernández-Aguilar, L., Latorre, J.M.: Experimentation on emotion regulation with single-colored images. In: ICT-based Solutions in Real Life Situations, Ambient Assisted Living, pp. 265–276 (2015)

50. Sokolova, M.V., Fernández-Caballero, A.: A review on the role of color and light in affective computing. Appl. Sci. **5**(3), 275–293 (2015)
51. Ortiz-García-Cervigón, V., Sokolova, M.V., García-Muñoz, R., Fernández-Caballero, A.: LED strips for color- and illumination-based emotion regulation at home. In: ICT-based Solutions in Real Life Situations, Ambient Assisted Living, pp. 277–287 (2015)