



# Emotion-Aware Ambient Intelligence: Changing Smart Environment Interaction Paradigms Through Affective Computing

Alex Altieri<sup>(✉)</sup>, Silvia Ceccacci, and Maura Mengoni

Department of Industrial Engineering and Mathematical Sciences,  
Università Politecnica delle Marche, Ancona, Italy  
a.altieri@pm.univpm.it,  
{s.ceccacci,m.mengoni}@univpm.it

**Abstract.** This paper describes the conceptual model and the implementation of an emotion aware system able to manage multimedia contents (i.e., music tracks) and lighting scenarios, based on the user's emotion, detected from facial expressions. The system captures the emotions from the user's face expressions, mapping them into a 2D valence-arousal space where the multimedia content is mapped and matches them with lighting color. A preliminary experimentation involved a total of 26 subjects has been carried out with the purpose of assess the system emotion recognition effectiveness and its ability to manage the environment appropriately. Results evidenced several limits of emotion recognition through face expressions detection and opens to several research challenges.

**Keywords:** Emotion-aware system · Ambient intelligence · Emotion recognition · Smart environment · Face expression recognition · Affective computing · Emotion detection

## 1 Introduction

Ambient intelligence (AmI) refers to the concept of a smart environment sensitive to its inhabitants, able to support them in an unobtrusive, interconnected, adaptable, dynamic, embedded, and intelligent way, and even capable of anticipating their needs and behaviors [1].

To be reliable and robust, such systems should have the ability to recognize the affective state of the person communicating with, so as to manage their behavior according to this information. Consequently, AmI systems should be equipped with emotional skills and must be able to adapt their users' emotional mood expressing feelings [2].

The recent increase in interest, concerning embedding emotion recognition to HCI systems, gave birth to a newly introduced facet of human intelligence, named as "emotional intelligence" within the more general field of ambient intelligence (AmI) [3]. However, despite for several years, affective computing and emotion analysis have been the focus of researchers in the field of Human-Computer Interaction,

the development of effective AmI systems applications still represents a great challenge today.

A lot of basic research has been carried out in the last years, which has determined some important advances regarding emotion recognition systems. However, despite several studies proposed framework that embedded emotion recognition systems with the purpose to enable AmI systems with emotion regulation capabilities, the majority of them are described only at a conceptual level. By our knowledge, no studies reported experimental results with the aim to assess the effectiveness of the proposed systems. Only [4] reported the results of an experimentation involved users, with the aim to assess user's satisfaction about the emotion regulation capability of the proposed system.

In this context, this paper describes the design of an emotion aware system able to manage multimedia contents (i.e., music tracks) and lighting scenarios, based on the user's emotion, detected from facial expressions. The system conceptual modelling and its implementation and validation in a real scenario are the main contribution of this research work.

## 2 Related Works

Affective computing, emotion analysis and study of human behaviors have been the focus of researchers in the field of Human-Computer Interaction over a long time [5].

Today several methods and technologies allow the recognition of human emotions, which differ in level of intrusiveness. The majority of such techniques, methods and tools refer to three research areas: facial emotion analysis, speech recognition analysis and biofeedback emotion analysis. Obviously, the use of invasive instruments (e.g., ECG or EEG, biometric sensors) can affect the subjects' behavior and, in particular, it may adulterate his/her spontaneity and accordingly the emotions experienced by them.

In the last years several effort have been made to develop reliable non-intrusive emotion recognition systems, in particular based on facial expression recognition. Facial emotion analysis aims to recognize patterns from facial expressions and to connect them to emotions, based on a certain theoretical model.

Nowadays, the majority of facial expression recognition system implements Deep Learning algorithms, in particular based on Convolutional Neural Networks (CNN), a Deep Learning mathematic model that takes in input different kind of pictures and make predictions on the trained model basis. There have been several models that have been proposed for analyzing emotions from facial expressions. However, the majority of facial expression databases currently available are based on the Ekman and Fresner's primary emotions (i.e., anger, fear, disgust, surprise, joy and sadness) [6]. Consequently, it is not surprising that most of the algorithms developed until now allow to recognize only these emotions. Among the main commercial tools currently available for visual emotional analysis, there is Affdex by Affectiva [7] and the Microsoft Cognitive Services based on the Azure platform [8].

Several studies have been conducted with the aim to develop emotion-aware systems, which embed cognitive agents, to regulate human emotions, for example by managing music and lighting color [2, 9–12].

The proposed framework implements the results of a lot of basic research regarding for example the link between human emotion and color association [13, 14] and the associations between music and colors [15]. Moreover, several methods have been proposed in literature with the aim to classify music tracks according to human emotions [16, 17].

Nowadays, all this knowledge together with the technology available may lead to the effective implementation of new human-computer interaction paradigm, e.g. symbiotic interaction [18]. However, by our knowledge, most of the proposed system are only at a conceptual level. No studies report experimental results able to evidence in a clear manner the real effectiveness of the proposed emotion-aware systems.

### 3 The Proposed System

A schematic layout of the proposed system architecture is reported in Fig. 1. Five processing nodes characterize it: the face detection, the facial expression recognition, the mood identification, the emotion/music track matching and the emotion/lighting color matching.

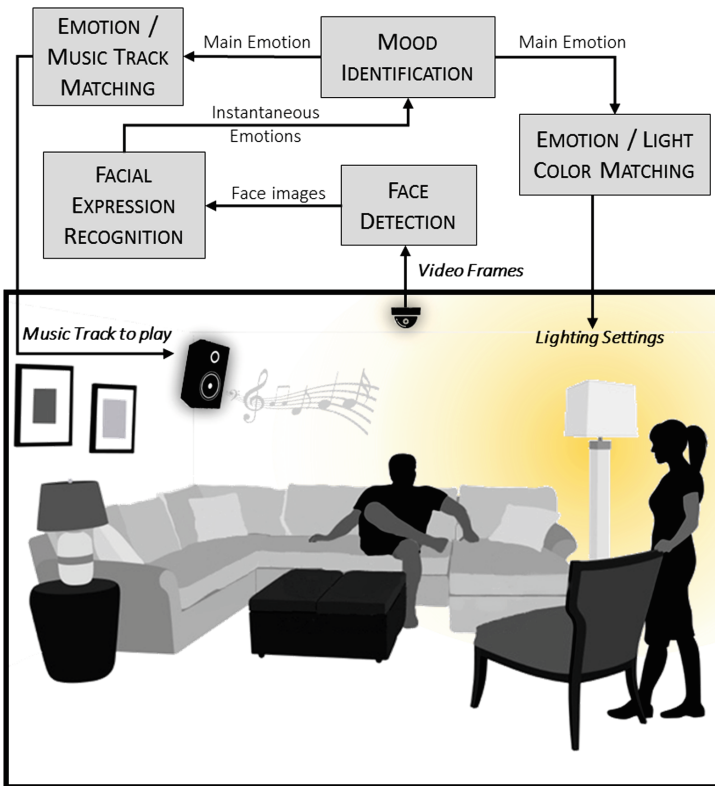


Fig. 1. General architecture of the proposed emotion-aware environment

The face detection node processes the video streams captured by a video camera (e.g., an IP camera). It detects, crops and aligns, frame by frame, the user's face by the original image. The resulted image is resized to  $64 \times 64$  and converted in grayscale.

The facial expression recognition node (FERN) implements a CNN, based on a revised version of the VGG13 [19], which consists of 10 convolution layers spaced with max pooling and dropout layers. It was trained from scratch on the FER+ dataset [20], so that it is able to recognize the six basic Ekman's emotions (i.e., joy, surprise, sadness, anger, disgust and fear), plus the neutral emotion condition. The CNN takes in input the aligned  $64 \times 64$  grayscale images provided by the face detection node and gives in output an array of seven values (Eq. 1), each representing the respective emotion probability.

$$Emotion(\bar{t}) = (a_t, b_t, c_t, d_t, e_t, f_t, g_t), \quad a_t + b_t + c_t + d_t + e_t + f_t + g_t = 100 \quad (1)$$

The FERN was already tested in a real environment to determine its emotion detection effectiveness compared to a traditional video analysis. Results of experimentation are reported in [21].

Each face image with the associate related emotion data are saved time by time in the background. The mood identification node allows evaluating the main emotion experienced by the user in a certain period. To this end, second by second, it gets and aggregates emotion data related to the video frames collected during a certain acquisition window (e.g., 30 s) and it processes them through an algorithm, in order to determine the resulting "average" emotion.

Emotion data, related to the considered acquisition window, are aggregated within 7 arrays (one per each emotion). To eliminate non-significant emotion data, values lower than 1 are discarded. Then, for each acquisition window, the emotion resultant, representing the user's mood, is determined as the maximum values of the sums of each emotional array elements (Eq. 2).

$$EmotionResultant = \text{MAX}(Joy, Surprise, Fear, Sadness, Anger) \quad (2)$$

Where:

$$\begin{aligned} Joy &= \sum_{i=1}^n a_i, \quad a_i > 1 \\ Surprise &= \sum_{i=1}^n b_i, \quad b_i > 1 \\ Fear &= \sum_{i=1}^n c_i, \quad c_i > 1 \\ Sadness &= \sum_{i=1}^n d_i, \quad d_i > 1 \\ Anger &= \sum_{i=1}^n e_i, \quad e_i > 1 \\ Disgust &= \sum_{i=1}^n f_i, \quad f_i > 1 \\ Neutral &= \sum_{i=1}^n g_i, \quad g_i > 1 \end{aligned}$$

In order to manage ambient lighting and music playlist according to the user's emotion feelings, it is then necessary to match the detected user's mood with the music tracks and the lighting color that best suit it. To this end, proper algorithms have been defined, which are described in the following paragraphs, with the aim to match

lighting color and music with five basic emotions: Joy, Surprise, Fear, Sadness and Anger.

Emotion feelings related to the Neutral mood, have not be considered to actuate ambient modification: the system does not perform any action when it detects a “Neutral” mood. Otherwise, because it is not considered useful to create experiences that arouse disgust, the system immediately stops any music stimulus and provides white lighting when it detects “Disgust”.

### 3.1 Managing Lights Based on Emotions

In order to manage ambient lighting according to users’ emotion feelings, by projecting the most appropriate chromatic light through a RGB led lighting system, it has been necessary to define a total of 5 color transitions. To this purpose, a survey was carried out involving about 300 people (58.4% females and 41.6% males), to determine the most suitable color-emotion associations. The questionnaire was managed through Google Surveys and anonymously administered. The user, at first must indicate his generalities (e.g., sex and age). Then he is asked to associate a color with each of the considered basic emotions (i.e., Joy, Surprise, Fear, Anger, Sadness). For each emotion the user is asked to answer to the question: “According to the following color palette, which color would you associate with the emotion X?”

The user must choice a color among a color palette made up of 8 colors (Fig. 2), which can be easy reproduced through an RGB led lighting system.

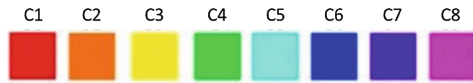


Fig. 2. The color palette

Table 1. Predominant emotions-color associations

Emotion	Color	Result (%)
Joy	C3 (yellow)	46,6
Surprise	C2 (orange)	41,1
Fear	C7 (purple)	42,7
Anger	C1 (red)	71,5
Sadness	C6 (blue)	62,2

Table 1 shows the colors that users have predominantly associated with the five considered emotions. Results revealed in almost all cases a prevalence of a color among the 8 available. At this point, it was possible to associate a color with each emotion and reproduce them through a RGB lighting system.

### 3.2 Managing Playlist Based on Emotions

In order to classify music tracks according to the emotion feelings they arouse, it is possible to map them on the bi-dimensional valence-energy space, according to the model proposed by Russell in [22]. In fact, based on the results discussed in [23], music characterized by “high-valence/high-arousal” seems to be more related with exciting sensations, while music with “low-valence/low-arousal” is more sad, melancholic and boring and music with “low-valence/high-arousal” are generally associated with tension. Accordingly, we identify on the space valence-arousal a total of 5 areas, which can be respectively associate with five Ekman’s basic emotions: surprise, joy, fear, anger and sadness (see the image on the top left corner of Fig. 3). The barycenters of these areas have been considered as starting centroids in a k-means clustering process, which has been applied to subdivide a generic music playlist in a total 5 cluster, respectively related with the five considered emotions.

## 4 System Implementation and Experimental Test

A system prototype has been deployed in order to support experimentation. A preliminary test has been carried out to assess the system effectiveness. The experiment involved 26 participants: 13 males, 13 females, aged between 23 and 47.

### 4.1 The Experimental Set-Up

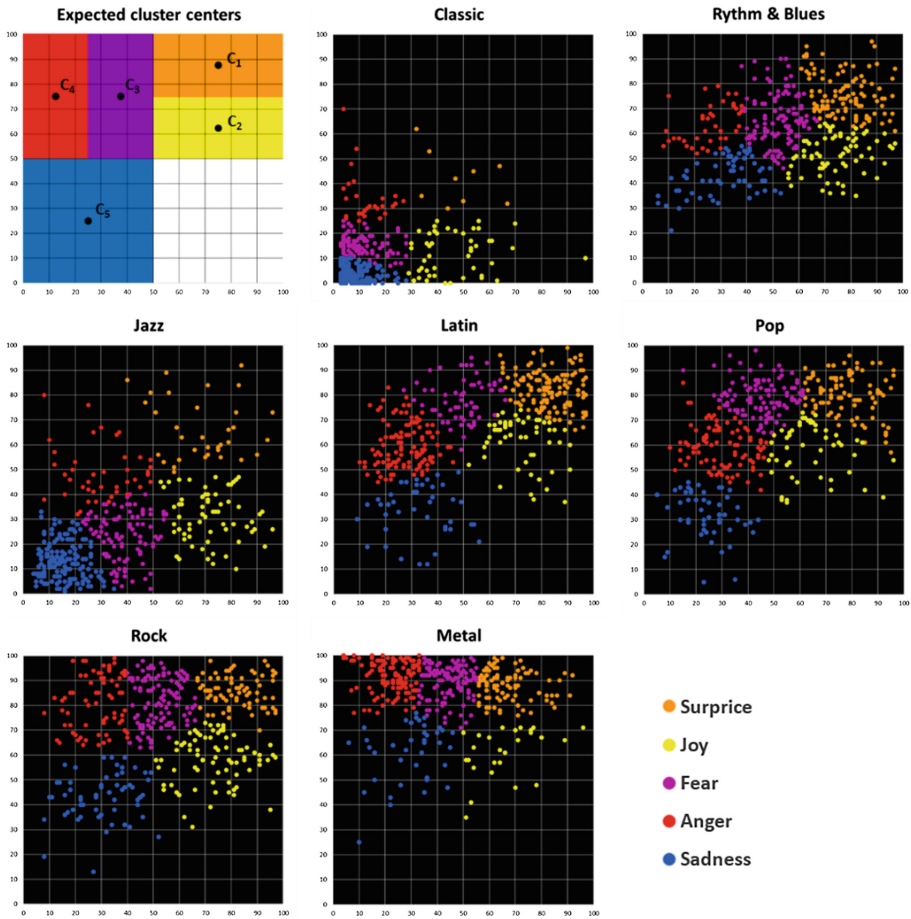
The experimental set-up consists of the following hardware systems:

- A 49” Samsung 4K TV;
- A webcam Logitech Brio 4K;
- Two stereo speakers Logitech Z130;
- An iPad 2018, 9.7”, wi-fi;
- A Brightsign Media Player XD233;
- An Intel NUC mini-PC;
- A Crestron DIN-DALI-2 controller;
- Two LINEARdrive 720D eldoLED;
- Two strip led RGBW;
- A router NETGEAR Nighthawk XR500.

The system is based on an architecture of clients/servers positioned within a local network to guarantee the security of the system and not have problems of interference and loss of data between the client and the server and between the various modules and physical peripherals that they are activated by the server (Fig. 4).

A Netgear Nighthawk XR500 router is used for communication within the network and routing packets.

The ipad is connected to the local wifi network and represents the Client of the architecture: the user sends the selection of his favorite kind of music through a web page developed and realized in.Net language.

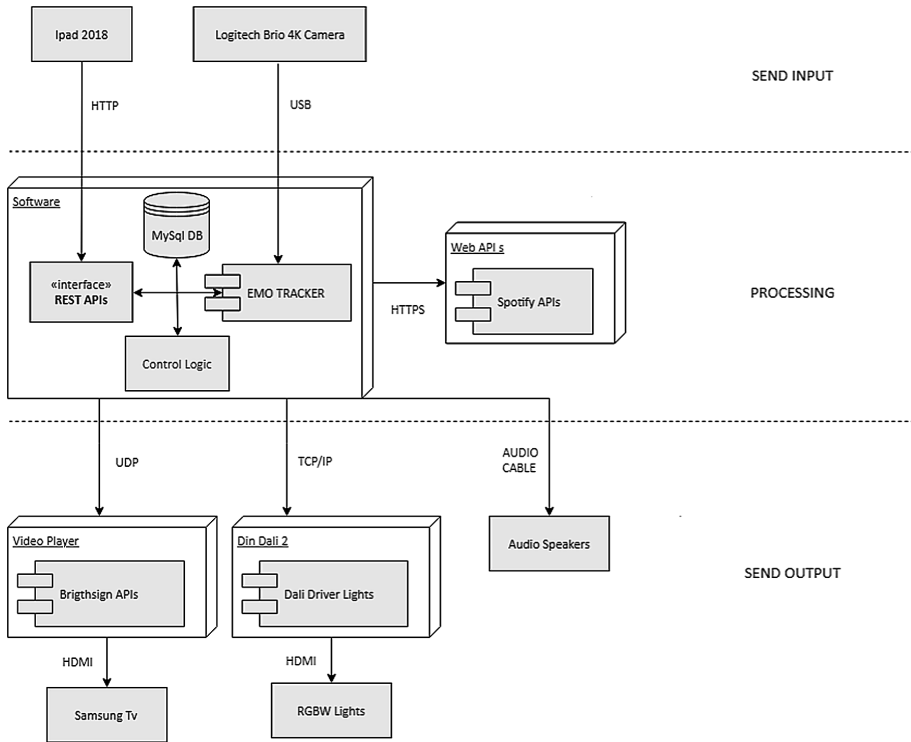


**Fig. 3.** Hypothetical areas most related with the five basic emotions in the valence-arousal space (on the top left) and results of the k-means process applied to various music playlists

The Samsung TV is positioned on a table 90 cm high and 150 cm distant from the user. The NUC acts as a central server and it is connected via USB to the Logitech webcam, positioned in front of the user on the top of the Samsung TV, which captures 8 frames per second (equal to a frame-rate of 0.125 f/s) and streams the video that will be processed frame by frame by the FERN. The NUC embeds software and REST services that allows to receive the client requests. Such software consists of various modules:

- The FERN CNN developed in Python, which allows to detect and evaluate the facial expressions of the subject and convert them into emotions;
- The data storage in a MySQL database;
- The management of data logic developed in C # code.

In particular, the NUC central unit processes the client requests and creates an information package containing:



**Fig. 4.** The experimental system architecture

- The favorite music genre, selected by the user, which is stored in the database;
- The video selected by the user, among a pre-loaded video playlist, which is inserted into an Udp packet and routed through the network to reach a Brighsign xd233 Video player device connected to the LAN.

While the user watching the video on the TV, the system acquires and processes the video streamed by the USB camera, detect the user's face, and recognizes the user's emotions. At the end of the video the decision-making process is activated, and the system provides two outputs, based on the result of the mood identification algorithm, to proper manage respectively the music playlist and lighting.

In particular, the NUC, through an http call to the Spotify web APIs [24], select music tracks belonging to the music genre selected by the user and according on the results of the music management algorithm, it selects the song most related to the detected user's mood.

At the same time, the NUC, based on the result of the color management algorithm builds a DALI package and sends it to the DIN-DALI 2 device via TCP/IP in order to properly manage the two RGBW strips-led through the Eldoled drivers.



## 4.2 Experimental Procedure

The experimental tests took place in a dimly illuminated room. The experience proposed to the participant was characterized by two phases: a “stimulus phase” and a “reaction phase”.

During the “stimulus phase”, the subject underwent the viewing of a 30-s video clip (stimulus) selected from the FilmStim database validated by Schaefer’s studio and collaborators [25], in order to arouse a particular emotional state. While watching the videos, the subject’s facial expressions were analyzed by the system. Once the video has ended, the reaction phase starts, and the system plays a song excerpt of 30 s, through the Spotify Web APIs, according to the results of the facial expression analysis. Moreover, the light system will adapt its color and intensity to reflect the emotion felt by the user during the video.

Before starting the experience, the subjects was asked to select, through the tablet, a music genre between Classic, Rock, R and B, Jazz, Pop, Latin and Metal. Then they were asked to start the video-player.

At the end of the experience, the subject was asked to fill out a questionnaire to assess the reliability of the system. In particular, the subjects were asked to report:

- The main emotion aroused by the video (among fear, joy, sadness, anger and surprise);
- Whereas the proposed music was coherent with the emotion experienced by them. Otherwise, they had to indicate the emotion most associated with music;
- Whereas the proposed light was coherent with the emotion experienced by them. Otherwise, they had to indicate the emotion most associated with the proposed lighting.

Overall, the experience lasts 1 min plus the time needed to fill the questionnaire. The subject were asked to repeat the experience four time: four videos has been proposed to each subject, in order to arouse them: fear, anger, sadness and joy (i.e., amusement). The order of the video clips was previously defined in order to ensure counterbalance across subjects.

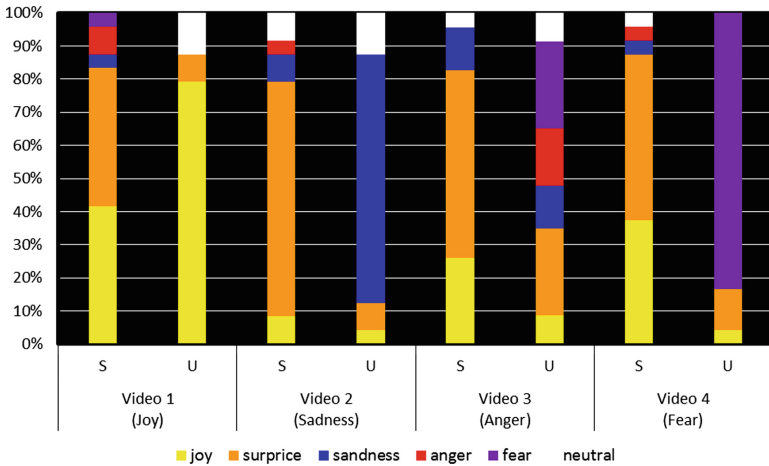
In order to avoid any distraction from the tasks, the room has been organized with as fewer objects as possible. To limit the researcher intervention, the experiment was supervised from a different room.

## 4.3 Results

Experimental results evidenced several issues and limitations that should be addressed to ensure system effectiveness.

In particular, by comparing the system outputs in term of detected mood and emotions the users reported they had experienced (Fig. 5), it is possible to infer that the system is slightly able to effective detect the main emotion the user experiences in a certain time period (i.e., within 30 s).

Despite the CNN, used to detect the facial expressions, as demonstrated by the experimental results reported in [21], is characterized by a good level of reliability, the percentage of mood recognition resulted by the experiment is very low. This may be



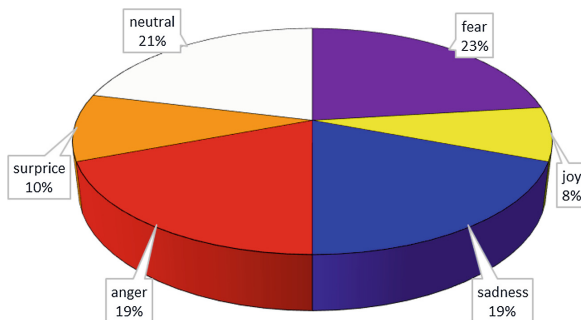
**Fig. 5.** System effectiveness in terms of mood detection: comparison between mood percentages detected by the system (S) during each video playing and the percentages of emotions the users reported they had experienced.

due to the limits of the algorithms used to determine the prevailing emotion in a time window.

However, although the scarce effectiveness of the system in mood recognition has proved to be poor, it is surprising most of the subjects anyhow have positively evaluated the experience.

In particular, the 73% of the subjects found the color of the proposed lighting coherent with the emotional state they experienced.

This is probably because the majority of users' moods detected by the system are related to the surprise, which has been associate with orange lighting, and a wide variability in user's judgement regarding the emotions associated to the orange color has been observed (Fig. 6).



**Fig. 6.** Emotions aroused to users by orange color. (Color figure online)

Furthermore the 54% of the subjects found a correct association between the emotional experience aroused by the video and the proposed song. However, the small sample of users involved in the experimentation does not allow to deep analyze these results, given the variability of proposed music. Probably, also this result may be due to the variability of user's opinions about the emotions aroused by a song.

## 5 Conclusion

The present paper described the design, the deployment and experimentation of a emotion-aware system for managing some features of a smart environment (i.e., lighting characteristics and background music).

Experimental results evidenced that the proposed system is still far away from being effective. In particular, several issues have emerged:

- The system resulted slightly able to correctly identify the main emotion perceived by the users during the viewing of the videos, although the FERN can be considered reliable, given the results of previous experimentation [21].

This can be due to the limits of the implemented mood identification algorithm. Several studies have to be carried out to improve the algorithm. For example, several experiments should be carried out to test its effectiveness by varying the threshold value (currently set to 1) defined to limit the effect of ambiguous facial expressions on the user's mood evaluation. However, the conceptual hypothesis on which the algorithm is based (i.e., "The total emotions experienced in a certain period of time is equal to the sum of all the emotions experience in that period") probable is too simplistic and must be revised. As the emotions are time-related, because they occur in response to certain events [26], the algorithm should consider also the temporal relationships between the various facial expressions. Moreover, it is also possible that facial expressions are not sufficient to determine the actual emotional state of the person, so that the collection of other contextual information are needed. In fact, Russell and Fehr [27] argued that context is the principal determinant in interpreting person's emotion from facial expressions, despite Ekman and O'Sullivan [28] stated that contextual information is only useful to interpret neutral or ambiguous face expressions.

- Even if the system has not been able to detect users' emotions correctly, most of the people involved in the test were satisfied with the music and lighting offered by the system.

In our opinion this result is surprising and can be motivated only if it is admitted that the personal preferences play a role so considerable in the emotion-color association and in the music-color association that it is not possible to hope that an environment management system, according on univocal rules, even if based on statistical findings, can satisfy all users. In this context, the system customization capability seems to be crucial and must be improved. In particular, the possibility to enhance the system adaptability by introducing self-learning functions should be considered.

**Acknowledgements.** This research has been co-funded by Videoworks spa and Marche Region within the POR Marche FSE 2014/2020 - Asse 1- P.I. 8.1- R.A. 8.5.

## References

1. Chen, T.C.T., Chaovalitwongse, W.A., Hong, I.H.: Optimization of ambient intelligence systems. *Oper. Res. Int J.* **18**, 575 (2018)
2. Acampora, G., Vitiello, A.: Interoperable neuro-fuzzy services for emotion-aware ambient intelligence. *Neurocomputing* **122**, 3–12 (2013)
3. Frantzidis, C.A., Bratsas, C., Papadelis, C.L., Konstantinidis, E., Pappas, C., Bamidis, P.D.: Toward emotion aware computing: an integrated approach using multichannel neurophysiological recordings and affective visual stimuli. *IEEE Trans. Inf. Technol. Biomed.* **14**(3), 589–597 (2010)
4. Muñoz, S., Araque, O., Sánchez-Rada, J.F., Iglesias, C.A.: An emotion aware task automation architecture based on semantic technologies for smart offices. *Sensors* **18**(5), 1499 (2018)
5. Mengoni, M., Peruzzini, M., Mandorli, F., Bordegoni, M., Caruso, G.: Performing ergonomic analysis in virtual environments: a structured protocol to assess humans interaction. In: *ASME International Design Engineering Technical Conferences and Computers and Information in Engineering Conference*, pp. 1461–1472 (2008)
6. Ekman, P., Friesen, W.V.: *Facial Action Coding System: Investigator’s Guide*. Consulting Psychologists Press, Palo Alto (1978)
7. Affdex. <https://www.affectiva.com/product/affdex-for-market-research/>. Accessed 7 Feb 2019
8. Microsoft Azure: Cognitive Services. <https://azure.microsoft.com/en-us/services/cognitive-services/emotion/>. Accessed 7 Feb 2019
9. Fernandez-Caballero, A., et al.: Smart environment architecture for emotion detection and regulation. *J. Biomed. Inform.* **64**, 55–73 (2016)
10. Acampora, G., Loia, V., Vitiello, A.: A cognitive multi-agent system for emotion-aware ambient intelligence. In: *2011 IEEE Symposium on Intelligent Agent (IA)*, pp. 1–8 (2011)
11. Acampora, G., Loia, V., Vitiello, A.: Distributing emotional services in ambient intelligence through cognitive agents. *Serv. Oriented Comput. Appl.* **5**(1), 17–35 (2011)
12. Hsiao, S.W., Chen, S.K., Lee, C.H.: Methodology for stage lighting control based on music emotions. *Inf. Sci.* **412**, 14–35 (2017)
13. Hemphill, M.: A note on adults’ color–emotion associations. *J. Genet. Psychol.* **157**(3), 275–280 (1996)
14. Naz, K., Epps, H.: Relationship between color and emotion: a study of college students. *Coll. Student J.* **38**(3), 396 (2004)
15. Palmer, S.E., Schloss, K.B., Xu, Z., Prado-León, L.R.: Music – color associations are mediated by emotion. *Proc. Natl. Acad. Sci.* **110**(22), 8836–8841 (2013)
16. Kim, Y.E., et al.: Music emotion recognition: a state of the art review. In: *Proceedings ISMIR*, pp. 255–266. Citeseer (2010)
17. Chang, C.Y., Lo, C.Y., Wang, C.J., Chung, P.C.: A music recommendation system with consideration of personal emotion. In: *2010 International Computer Symposium (ICS)*, pp. 18–23. IEEE (2010)

18. Jacucci, G., Spagnoli, A., Freeman, J., Gamberini, L.: Symbiotic interaction: a critical definition and comparison to other human-computer paradigms. In: Jacucci, G., Gamberini, L., Freeman, J., Spagnoli, A. (eds.) *Symbiotic 2014*. LNCS, vol. 8820, pp. 3–20. Springer, Cham (2014). [https://doi.org/10.1007/978-3-319-13500-7\\_1](https://doi.org/10.1007/978-3-319-13500-7_1)
19. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint [arXiv:1409.1556](https://arxiv.org/abs/1409.1556) (2014)
20. Barsoum, E., Zhang, C., Ferrer, C.C., Zhang, Z.: Training deep networks for facial expression recognition with crowd-sourced label distribution. In: *Proceedings of the 18th ACM International Conference on Multimodal Interaction*, pp. 279–283 (2016)
21. Generosi, A., Ceccacci, S., Mengoni, M.: A deep learning-based system to track and analyze customer behavior in retail store. In: *2018 IEEE 8th International Conference on Consumer Electronics-Berlin (ICCE-Berlin)*, pp. 1–6. IEEE (2018)
22. Russell, J.A.: A circumflex model of affect. *J. Pers. Soc. Psychol.* **37**(3), 1161 (1980)
23. Kim, J., Lee, S., Kim, S., Yoo, W.Y.: Music mood classification model based on arousal-valence values. In: *13th International Conference on Advanced Communication Technology (ICACT)*, pp. 292–295 (2011)
24. Spotify web APIs. <https://developer.spotify.com/documentation/web-api/>. Accessed 7 Feb 2019
25. Schaefer, A., Nils, F., Sanchez, X., Philippot, P.: Assessing the effectiveness of a large database of emotion-eliciting films: a new tool for emotion researchers. *Cogn. Emot.* **24**(7), 1153–1172 (2010)
26. Ekman, P.: Facial expression and emotion. *Am. Psychol.* **48**(4), 384 (1993)
27. Russell, J.A., Fehr, B.: Relativity in the perception of emotion in facial expressions. *J. Exp. Psychol. Gen.* **116**(3), 223 (1987)
28. Ekman, P., O’Sullivan, M.: The role of context in interpreting facial expression: comment on Russell and Fehr (1987). *J. Exp. Psychol. Gen.* **117**(1), 86–88 (1988)