An Adaptive System to Manage Playlists and Lighting Scenarios Based on the User's Emotions

A. Altieri[a], S. Ceccacci[a], L. Ciabattoni[b], A. Generosi[a], A. Talipu[a], G. Turri[a], M. Mengoni[a]
[a]Dipartimento di Ingegneria Industriale e Scienze Matematiche,
[b]Dipartimento di Ingegneria dell'Informazione,
Università Politecnica delle Marche, Ancona, Italy
Email: {a.altieri, a.generosi, t.abudukaiyoumu}@pm.univpm.it,
{s.ceccacci, l.ciabattoni, g.turri, m.mengoni}@univpm.it

***Abstract*—This paper introduces a new system capable of adaptively managing multimedia contents (e.g. music, video clips, etc.) and lighting scenarios based on the detected user's emotional state. The system captures the emotion from the user's face expression mapping it into a 2D valence-arousal space where the multimedia content is mapped and matches them with lighting color. Results of preliminary tests suggest that the proposed system is able to detect the user's emotional state and manage proper music and light colors in a symbiotic way.**

I. Introduction

In the last years, research in the context of smart home automation systems has focused on how automation systems can be controlled based on the user's physiological and behavioral characteristics or facial cues [1]. In particular, several studies have argued how emotion recognition and regulation can help to reduce various stress-related health problems [2] and how music and colored lights can be handled to support it [3]. However, to the best of our knowledge, no studies introduce and test systems able to manage music playlists and ambient lighting scenarios, in a symbiotic way, according to the user's emotions. Anyway, a lot of basic research has been conducted, which can be useful to achieve this goal, by using lighting and music stimuli. Researchers tried to investigate the link between human emotion and color association [4], [5] and the association between music and colors [6], and several methods have been proposed in literature with the aim to classify music tracks according to human emotions [7], [8]. Fernandez et al. [3] proposed a theoretic approach to perform emotion regulation based on music and colored lights management. All this knowledge together with the technology available nowadays leads to the effective implementation of new human-computer interaction paradigm, e.g. symbiotic interaction [9]. In this context, the proposed system aims to detect facial expressions of a person in front of a camera, recognize his/her emotions, and match them, in a symbiotic way, to a specific ambient lighting scenario and music track.

II. The Proposed System

The basic idea is to make an environment sensitive and responsive to the person's emotions. In particular, the proposed system detects the mood of a person within the environment, and manage the ambient lighting color and music, in order to provide a symbiotic experience. To this end, it integrates different technologies and implements an algorithm able to manage music tracks and lightning colors based on human emotions. It detects the person's facial expressions and associate them to the six basic Ekman's emotions [10] (happiness, surprise, sadness, anger, disgust and fear) plus the neutral expression. Regarding ambient lighting management, we considered a total 7 color transitions. Emotion-color association has been defined based on the findings reported in [2] and [3]. To validate the assumed color-emotion association a survey has been carried out, which involved about 300 people. In order to associate music track to the basic Ekman's emotions in an objective way, we considered the two-dimensional space valence-arousal introduced in [10]. Using the valence and energy (i.e. arousal) values it is possible to map music track within the two-dimensional space valence-arousal [11]. To extrapolate valence and energy (i.e., arousal) for each music track, the Spotify Web APIs [12] can be used, which provides metadata for each song (i.e., genre, bpm, energy, loudness, valence, popularity, etc.). Since a determined medium range of valence and arousal characterizes every musical genre, we classified music tracks by considering 7 music genres: Pop, Rock, Classical, Latin, R&B, Jazz, and Metal. Then, for each genre, we associated valence-arousal space coordinates to the basic Ekman's emotions. In this way, once selected a certain music genre, the algorithm is able to determine the best match between emotions, color lighting and music tracks.

III. Prototype Architecture

To support test with user a system architecture 1 has been defined and implemented. The hardware architecture requires the use of an Intel NUC mini-PC, an iPad, a Logitech Brio 4K webcam, a Brightsign Media Player for 4K multimedia content, a 49" Samsung 4K TV, a Crestron DIN-DALI-2 controller to drive two tubes and a plate of LED lights and a high-performance speaker. The NUC should act as a central server: the Logitech webcam has been connected via USB to stream the video that will be processed frame by frame via a Convolutional Neural Network (CNN, EmoTracker) trained on the FER+ dataset [13] able to recognize the aforementioned 7 different facial expressions. The iPad is connected to the NUC via a Wi-Fi connection: on it, a web interface that

Authorized licensed use limited to: Universita' Politecnica delle Marche. Downloaded on January 04,2023 at 12:28:10 UTC from IEEE Xplore. Restrictions apply.

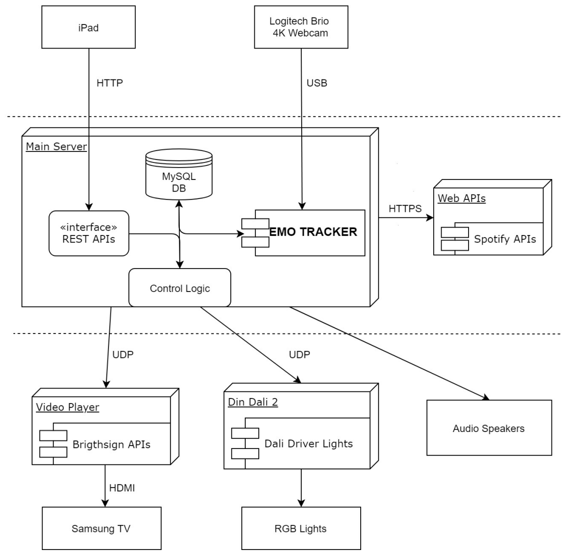

Fig. 1: *Hardware and software architecture of the proposed system.*

allows the user to select his/her preferred music genre; the choice will be after sent to the server that will expose a REST interface awaiting for instructions. On the basis of the received command, the NUC have to manage the control logic. Once the UDP packets are sent to the Brightsign Media Player to activate the videos in the predefined sequence, the NUC will be able to contact the Spotify APIs to match the mood detected by the EmoTracker and the previously chosen music genre, to the most suitable song in the Spotify database. Such database includes the top 100 most popular songs of Spotify for each of the considered music genres. Moreover, the main server will pilot (always via UDP) the Crestron light system to provide the most appropriate light with the average mood detected by the EmoTracker while the user is watching the video.

IV. Experimental Assessment

An experiment that involved a total of 30 participants (15 male, 15 female, aged between 26 and 57) has been carried out. The experimental session took place in a dimly illuminated room predisposed with a 49" TV, a webcam, two speakers, a light system, and a tablet. In order to avoid any distraction from the tasks, the room has been organized with as fewer objects as possible and the researcher intervention is delivered and supervised from a different room. Before the experiment, the subject was asked to select, through the tablet, his favorite music genre. The execution of the whole experiment consisted in two separate sessions: the "emotion elicitation" and the "reaction session". In the first one, the subject underwent the viewing of a 30-second video clip (stimulus) selected from the FilmStim database validated by Schaefer's studio and collaborators [14], in order to arouse a particular emotional state. The order of the video clips was previously defined in order to ensure counterbalance across subjects. While watching the videos, the subject's facial expressions were analyzed by our system. Once the video has ended, the second session starts and the system plays a song excerpt of 30 seconds, 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 user emotion felt during the video according to the results of the survey conducted earlier. At the end of the experiment, the subject was asked to fill out a questionnaire to assess the reliability of the system. Overall, the experience lasts 1 minute plus the time needed to fill the questionnaire.

V. Conclusion

Experimental results arising from the questionnaire evidenced that 73% of the subjects believe that the matching between the mood experienced during the video and the multimedia scenario (music and lights) proposed by the system is accurate. Future studies are needed to deepen understand the correlations between emotions-music-lights and consequently improve the system effectiveness.

References

[1] S. A. Khowaja, K. Dahri, M. A. Kumbhar, and A. M. Soomro, "Facial expression recognition using two-tier classification and its application to smart home automation system," in *Emerging Technologies (ICET), 2015 International Conference on*. IEEE, 2015, pp. 1–6.

[2] P. R. Goldin and J. J. Gross, "Effects of mindfulness-based stress reduction (mbsr) on emotion regulation in social anxiety disorder." *Emotion*, vol. 10, no. 1, p. 83, 2010.

[3] A. Fernandez-Caballero, A. Martínez-Rodrigo, J. M. Pastor, J. C. Castillo, E. Lozano-Monasor, M. T. López, R. Zangróniz, J. M. Latorre, and A. Fernández-Sotos, "Smart environment architecture for emotion detection and regulation," *Journal of biomedical informatics*, vol. 64, pp. 55–73, 2016.

[4] M. Hemphill, "A note on adults' color–emotion associations," *The Journal of genetic psychology*, vol. 157, no. 3, pp. 275–280, 1996.

[5] K. Naz and H. Epps, "Relationship between color and emotion: A study of college students," *College Student J*, vol. 38, no. 3, p. 396, 2004.

[6] S. E. Palmer, K. B. Schloss, Z. Xu, and L. R. Prado-León, "Music–color associations are mediated by emotion," *Proceedings of the National Academy of Sciences*, vol. 110, no. 22, pp. 8836–8841, 2013.

[7] Y. E. Kim, E. M. Schmidt, R. Migneco, B. G. Morton, P. Richardson, J. Scott, J. A. Speck, and D. Turnbull, "Music emotion recognition: A state of the art review," in *Proc. ISMIR*. Citeseer, 2010, pp. 255–266.

[8] C.-Y. Chang, C.-Y. Lo, C.-J. Wang, and P.-C. Chung, "A music recommendation system with consideration of personal emotion," in *Computer Symposium (ICS), 2010 International*. IEEE, 2010, pp. 18–23.

[9] G. Jacucci, A. Spagnolli, J. Freeman, and L. Gamberini, "Symbiotic interaction: a critical definition and comparison to other human-computer paradigms," in *International Workshop on Symbiotic Interaction*. Springer, 2014, pp. 3–20.

[10] P. Ekman, "An argument for basic emotions," *Cognition & emotion*, vol. 6, no. 3-4, pp. 169–200, 1992.

[11] J. A. Russell and L. F. Barrett, "Core affect, prototypical emotional episodes, and other things called emotion: dissecting the elephant." *Journal of personality and social psychology*, vol. 76, no. 5, p. 805, 1999.

[12] https://developer.spotify.com/documentation/web-api//.

[13] E. Barsoum, C. Zhang, C. C. Ferrer, and Z. Zhang, "Training deep networks for facial expression recognition with crowd-sourced label distribution," in *Proceedings of the 18th ACM International Conference on Multimodal Interaction*. ACM, 2016, pp. 279–283.

[14] A. Schaefer, F. Nils, X. Sanchez, and P. Philippot, "Assessing the effectiveness of a large database of emotion-eliciting films: A new tool for emotion researchers," *Cognition and Emotion*, vol. 24, no. 7, pp. 1153–1172, 2010.

Authorized licensed use limited to: Universita' Politecnica delle Marche. Downloaded on January 04,2023 at 12:28:10 UTC from IEEE Xplore. Restrictions apply.