AN EFFICIENT DEEP LEARNING APPROACH TO STAGE LIGHTNING CONTROL: PART 1 - SYSTEM ARCHITECTURE

Amerena Matteo, Cantoni Corrado, Piccirillo Jacopo, Stucchi Gabriele**

Dipartimento di Elettronica, Informazione e Bioingegneria (DEIB), Politecnico di Milano
Piazza Leonardo Da Vinci 32, 20122 Milano, Italy
[matteo.amerena, corrado.cantoni,
jacopo.piccirillo, gabriele.stucchi]@mail.polimi.it,

ABSTRACT

Stage lighting control requires professionally trained technicians to operate with the lighting system and contemporary demands for higher quality performances requires more time and effort for preparation. We propose a system for automatic stage lighting control based on the music emotion of what is being played. The system features a Music Emotion Recognition deep learning model to dynamically predict valence and arousal, that are used to determine the color of the light based on a mapping between points in the Russell's model of affect and colors; while the brightness of the lights is controlled through Root Mean Square. The validity of the system is checked through different experiments that show the good performances of the model (average RMSE around 0.21), the validity of the adopted color model (in most cases the majority of tested subjects confirmed our decisions) and the correctness of the mapping between valence and arousal and colors.

***Index Terms*—** Stage lighting control, music emotion recognition, color emotions, deep learning, music information retrieval

1. INTRODUCTION

Stage lighting control traditionally have required professionally trained technicians to operate with the lighting system, and with rising labor costs, most technicians are now required to control both light and sound [1]. With the recent fast development of new technologies and of Music Information Retrieval (MIR) automated systems based on music features have become more and more present.

Music Emotion Recognition (MER) is a branch of MIR that aims at detecting the mood of a musical piece as a whole (static MER) or in smaller segments that composes it (dynamic MER) and has brought benefits to emotion-related music applications as well as personalized experiences such as recommendation systems and music psychology [2]. A further differentiation in the approaches to MER is generally done in terms of: categorical approach, in which emotions are described through discrete labels (such as happy, sad, angry and fear) [3] or clusters [4], treating MER as a classification problem; and dimensional approach, in which the emotions are described through a continuous N-dimensional plane, treating MER as a regression problem. The typical plane used in this approach is the 2-dimensional circumplex model of affect proposed by Russel [5] where the horizontal axis (*valence*) indicates the positivity of an emotion and the vertical axis (*arousal*) indicates the intensity of an emotion.

Moreover, studies on color emotions reports that light and color can influence people's perception of the characteristics of the area around them, such as its comfort and atmosphere, and they can even cause the pulse and endocrine activities to accelerate [6]. These studies suggested a reasonableness in developing an automatic stage lighting control system based on music emotions.

2. RELATED WORKS

Among the literature concerning automatic stage lighting control the only approach based on music emotion is the one of Hsiao, Lee, Chen [1] in which they propose a methodology for controlling both light brightness and color. For the brightness regulation they use the intensity of the music while for the color they conducted a study to analyze the relationship between lighting color regulation preferences and music emotions. They called five professional stage lighting technicians with more than two years' stage experience and asked them to to regulate lighting color for 988 selected music clips. The results were first used to analyze the influences of music emotions on lighting regulation preference using a Multivariate analysis of variance (MANOVA) and then to train a Support Vector Regression model employed to construct the color maps of the Russell's plane. In the experiment, based on the feedback from the subjects, they also considered the music genre as a factor of influence ending up with eighteen (one for considered genre) different color maps.

Given the scarcity of works in the specific subject of automatic stage lighting control based on music emotions we consulted the literature concerning the two main areas of the system: MER and color emotions. Studies related to MER are deeply presented in the second part of this report [7].

Regarding color emotions, different studies [8] [9] rely on the approach presented by Ståhl, Sundström and Höök [10]. In their work they propose eMoto: an emotional text messaging service that allow users to adjust the affective background of a message with the use of colors, shapes and animations thanks to specific gestures. Regarding colors they considered the Russell's model of affect and associate a color to the points of the plane starting from the Itten's circular color model [11] and adapting it following the Ryberg's theory in which red represents the most powerful and strong emotions while blue, the color at the other end of the color scale, represents emotions with less energy; and the Goethe's theory [12] in which colors are considered as positive or negative, where the negative colors are blue, red-blue and blue-red, and the positive colors are yellow, red-yellow and yellow-red.

3. SYSTEM ARCHITECTURE

A scheme of the proposed architecture is shown in Figure 1. Two paths are present: one aims at controlling the light color based on music emotion and the other aims at controlling the light brightness based on the Root Mean Square (RMS). In the following subsections we will give an explanation of the main components.

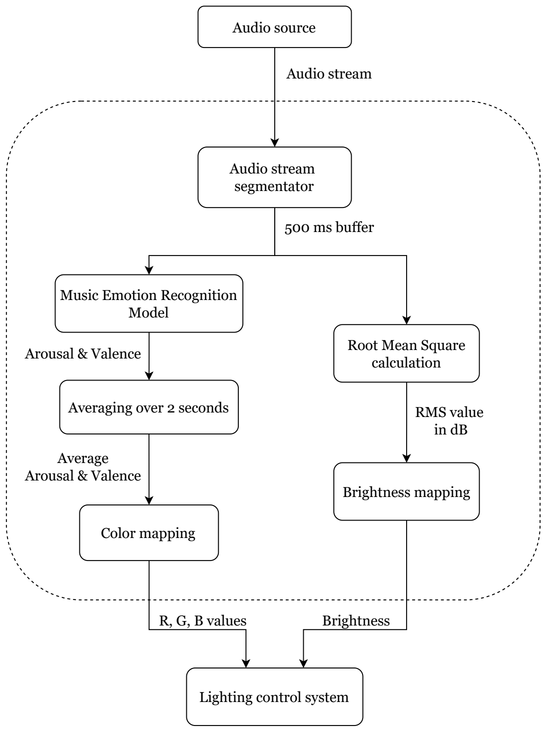

Figure 1: System Architecture

3.1. Music Emotion Recognition Model

To solve the Music Emotion Recognition task we adopted a dynamic and categorical (based on the Russell's model) approach.
The database we used is the MediaEval Database for Emotional Analysis of Music [13] that contains both static and dynamic (over 500 milliseconds) annotations. For this reason, our model takes as input raw audio files of 500 milliseconds and gives as output the predicted valence and arousal. The duration of the input segments sets a threshold for the time the model takes to perform the predictions, highlighting the need of a model that focuses on being as fast as possible while maintaining good performances. A detailed description of the model is given in the second part of this report [7].

The predicted valence and arousal of 4 subsequent buffers (corresponding to 2 seconds) are then averaged in order to not have sudden and unpleasant changes.

3.2. Color model

The adoption of the Russell's model of affect made it necessary to find an appropriate color model that reflects the corrispondence between points in the valence/arousal plane and colors. Following the already mentioned approach of Ståhl et al. [10] we started from the Itten's circular color model [11] and then adapted it. This meant reflecting it with respect to the first and third quadrant's bisector in order to have the negative colors placed on the negative valence side and the positive colors placed on the positive valence side. Moreover, colors are smoothed into one-another to avoid sharp boundaries and get weaker towards the center where valence and arousal are neutral (Figure 2). In the resulting model the colors that are most powerful, like red, are placed at the higher arousal zone, while more peaceful colors, like blue and green, end up at the lower arousal zone.

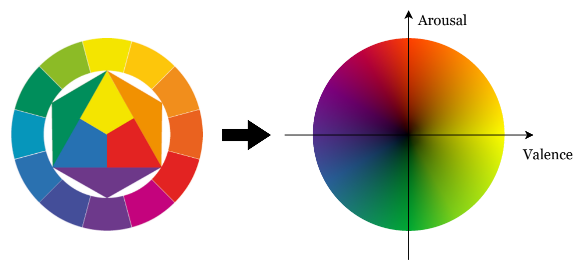

Figure 2: Itten's color model (left) adapted to the Russell's model (right)

3.3. Color mapping

To map arousal and valence to our color model we started from the Hue Saturation Value (HSV) color space and considered as Hue the angle created by the point (with average valence and arousal as coordinates) in the Russell's plane and as Saturation the distance of the point from the center. Since the brightness of the light will be controlled independently we kept the Value constant at maximum to have the brightest colors. If we consider H as Hue (ranging from 0° to 360°), S as Saturation (ranging from 0 to 1), V as Value (ranging from 0 to 1), AR as Arousal (ranging from -1 to 1) and VA as valence (ranging from -1 to 1) we can define H, S, V as:

$$\begin{cases} H = \arctan(\frac{AR}{VA}) \\ S = \sqrt{AR^2 + VA^2} \\ V = 1 \end{cases} \tag{1}$$

HSV is a different representation of the Red Green Blue (RGB) color model, that is an additive model where red, green and blue are the primary colors and are added together to create all the other colors. The Itten's color model however is a Red Yellow Blue (RYB) model, that is a subtractive model where the primary colors are instead red, yellow and blue. Since the standard in the digital world is the RGB model we decided to consider the HSV approach applied to the RYB model and finally map back the color to RGB.

We started by considering the mapping from HSV to RGB, but instead of the green color we substituted it with the yellow one. Following the same notation as before we can define three variables that will be useful for the calculation:

$$C = V * S \tag{2}$$

$$X = C * (1 - |(H/60)mod2 - 1|) \tag{3}$$

$$m = V - C \tag{4}$$

and the mapping is done as follows:

$$(R', Y', B') = \begin{cases} (C, X, 0), & \text{if } 0 \leq H < 60 \\ (X, C, 0), & \text{if } 60 \leq H < 120 \\ (0, C, X), & \text{if } 120 \leq H < 180 \\ (0, X, C), & \text{if } 180 \leq H < 240 \\ (X, 0, C), & \text{if } 240 \leq H < 300 \\ (C, 0, X), & \text{if } 300 \leq H < 360 \end{cases} \quad (5)$$

Finally R, Y and B are calculated as:

$$(R, Y, B) = (R' + m, Y' + m, B' + m) \quad (6)$$

Once the color has been defined in the RYB space it has to be converted to the RGB space. For this purpose we followed the work of Gosset and Chen [14] in which they propose a reasonable approximation obtained by defining a cube with each axis representing either Red, Yellow or Blue (see Figure 3). By defining appropriate RGB values for each of the eight colors represented by the corners of the cube, we can use trilinear interpolation to obtain suitable RGB values for any color defined in RYB. For instance,if a color were to be defined as 100% Red, 50% Yellow and 25% Blue, it could be represented as (1.0, 0.5, 0.25) in RYB coordinates. By interpolating the RGB values defined at the 8 corners of the cube, this color would have RGB coordinates of (0.8375, 0.19925,0.0625), producing a slightly muddy orange color as expected.

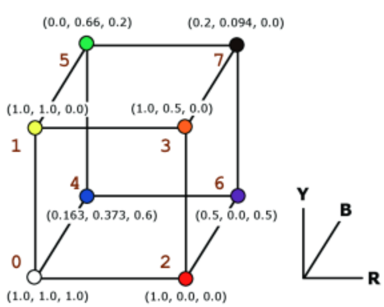

Figure 3: RYB interpolation cube

3.4. Brightness Mapping

Light brightness is controlled through Root Mean Square calculated over the 500 milliseconds segments. Differently from the color, the brightness control maintains a finer granularity, to obtain more dynamicity. Considering a discrete signal composed of n values $x_1, x_2, ..., x_n$ the RMS is defined as:

$$RMS = \sqrt{\frac{1}{n}\sum_{i=1}^{n} x_i^2} \quad (7)$$

The RMS value is then converted in decibels and added to the maximum brightness so that quieter segments will result in less brightness and viceversa.

4. IMPLEMENTATION DETAILS

The proposed system is though for a live concert environment and to be as easy to use as possible. For this reason we have implemented it as a Virtual Studio Technology (VST) plugin, in order to be able to use it inside a Digital Audio Workstation (DAW). The audio from the live stage is captured through an audio interface connected to a computer on which the DAW and the VST plugin are running; the plugin implements the procedure explained above and send messages with RGB and Brightness value through OSC protocol to the lighting control software. The lighting control software communicates with a control unit connected to the computer that controls the lighting system through Digital MultipleX (DMX) protocol (Figure 4).

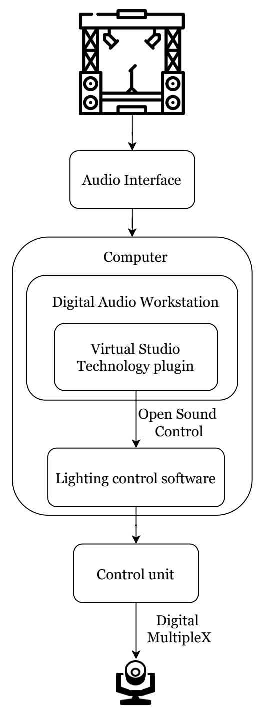

Figure 4: Implementation setup

5. EXPERIMENT AND RESULTS

The validation of the proposed system must take into account the different related aspects. For this reason we have conducted 3 different experimental tests explained in the subsections below.

5.1. Music Emotion Recognition model

We evaluated our model both on the perfomances (in terms of average RMSE on a 10-fold cross-validation) and, given the aforementioned time limit, on the time of execution. We obtained:

$$RMSE = .21 \pm .02 \quad (8)$$

and an Average Execution Time (AET) on 60 runs of:

$$AET = 4.1 \pm 0.4\text{ms} \quad (9)$$

Color	Category	Results			
		1	**2**	**3**	**4**
	2	28,9%	**55,5%**	7,2%	8,4%
	3	4,8%	35%	**56,6%**	3,6%
	1	**60,2%**	20,6%	9,6%	9,6%
	2	22,9%	**53%**	10,8%	13,3%
	4	31,4%	12%	14,5%	**42,1%**
	3	7,2%	34,9%	**50,6%**	7,3%
	1	**38,6%**	38,6%	15,4%	7,3%
	4	13,2%	8,4%	27,8%	**50,6%**

Table 1: Survey results

As always for a more detailed discussion we suggest reading the second part of this report [7].

5.2. Color model

To evaluate the validity of the adopted color model we conducted a perceptive test to 83 subjects. We divided the Russel's plane into eight different categories and selected a color belonging to each category accordingly to our color model. The test consisted in showing the selected colors and asking the subjects to associate each color to a category. The results we obtained were quite satisfactory, also considering the high subjectivity of the task, but we noted a tendency on making errors with nearby categories along one axis (for example categories with high arousal where often exchanged). For this reason we decided to aggregate the nearby categories as shown in Figure 5. The results of the test can be seen in the Table 1, where for each color is reported the belonging category and the percentage of votes for all the categories. The obtained results highlight that among all the categories the correct one has the highest percentage and this reinforces the validity of the color model.

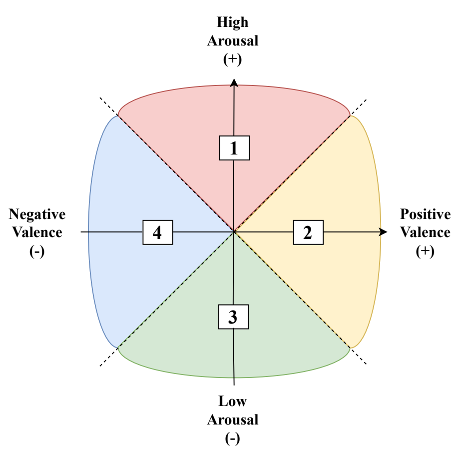

Figure 5: Categories considered for the test

Song	Artist	ID
Yemaya	Alfredo Rodriguez Trio	A
Dialoghi Sonori	Deb Rose	B
Damne	Fanfara Station	C
Abusey Junction	Kokoroko	D
Memphis	Lovesick Duo	E
Abstract Travels	Tommaso Cappellato	F
Stop Bajon	Tullio De Piscopo	G
True Love Waits	Radiohead	H

Table 2: List of songs used for test

ID	Valence	Arousal	(R, G, B)	Color
A	-0.712	-0.214	(187, 156, 84)	
B	0.119	-1	(147, 160, 123)	
C	0.789	0.91	(255, 171, 71)	
D	-0.088	-0.845	(187, 184, 72)	
E	0.92	1	(251, 222, 121)	
F	1	0.517	(131, 209, 202)	
G	1	1	(252, 146, 67)	
H	-0.94	-0.73	(108, 139, 168)	

Table 3: Color mapping results

5.3. Color mapping

In the color mapping evaluation we were interested in evaluating the correctness of the mapping between predicted valence and arousal and color. For this reason we took 8 songs (to which we associated an ID as shown in table 2) and we made the color mapping for each of them. In Table 3 we can see the results for one segment belonging to each song[1].

The obtained results show that the point generated from the valence and arousal values is coherent with the color model confirming the validity of the adopted mapping.

6. CONCLUSION

In this paper we presented a system for automatic stage lighting control based on music emotion. The system features a deep learning Music Emotion Recognition model that predics valence and arousal of 500 milliseconds raw audio files and uses the predicted values to control the light color based on a color model derived from the Itten's circular one. The experimental results showed the effectiveness of the Music Emotion Recognition model both in terms of performance and execution time, confirming the validity for a live scenario. Moreover, the adopted color model seems to reflect the feelings associated to different colors by the tested population.

The system will be further evaluated during the Festivalle music festival organized at the "Valle dei Templi" in Agrigento, Sicily,

[1]Since the valence and arousal predicted by the model are very small we applied a multiplication factor of 3 in order to obtain more saturated colors and better check the correctness of the mapping

Italy where we will have the opportunity to test it in a real scenario and collect additional feedback.

7. REFERENCES

[1] S.-W. Hsiao, S.-K. Chen, and C.-H. Lee, “Methodology for stage lighting control based on music emotions,” *Information Sciences*, vol. 412, 05 2017.

[2] N. HE and S. Ferguson, “Multi-view neural networks for raw audio-based music emotion recognition,” in *2020 IEEE International Symposium on Multimedia (ISM)*, 2020, pp. 168–172.

[3] Y. Feng, Y. Zhuang, and Y. Pan, “Music information retrieval by detecting mood via computational media aesthetics,” in *Proceedings IEEE/WIC International Conference on Web Intelligence (WI 2003)*, 2003, pp. 235–241.

[4] A. Bhattacharya and K. V. Kadambari, “A multimodal approach towards emotion recognition of music using audio and lyrical content,” 2018.

[5] J. Russell, “A circumplex model of affect,” *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 12 1980.

[6] M. C. Miller, *Color for Interior Architecture*. John Wiley and Sons, New York, 1997.

[7] M. Amerena and J. Piccirillo, “A lightweight deep learning approach to stage lighting control: Part 2: Model description,” 2021.

[8] J. Dharmapriya, L. Dayarathne, T. Diasena, S. Arunathilake, N. Kodikara, and P. Wijesekera, “Music emotion visualization through colour,” in *2021 International Conference on Electronics, Information, and Communication (ICEIC)*, 2021, pp. 1–6.

[9] L. Bialoskorski, J. Westerink, and E. L. van den Broek, “Mood swings: Design and evaluation of affective interactive art,” *The New Review of Hypermedia and Multimedia*, vol. 15, pp. 173–191, 08 2009.

[10] A. Ståhl, P. Sundström, and K. Höök, “A foundation for emotional expressivity,” 10 2005.

[11] J. Itten, *The Art of Color: The Subjective Experience and Objective Rationale of Color*. Van Nostrand Reinhold Company, 1974.

[12] M. Schindler and J. W. Goethe, *Goethe’s theory of colour*. London: New Knowledge, 1970.

[13] A. Aljanaki, Y.-H. Yang, and M. Soleymani, “Developing a benchmark for emotional analysis of music,” *PLOS ONE*, vol. 12, no. 3, pp. 1–22, 03 2017. [Online]. Available: https://doi.org/10.1371/journal.pone.0173392

[14] N. Gossett and B. Chen, “Paint inspired color mixing and compositing for visualization,” in *IEEE Symposium on Information Visualization*, 2004, pp. 113–118.