Methodology for Stage Lighting Control Based on Music Emotions

Article *in* Information Sciences · May 2017

DOI: 10.1016/j.ins.2017.05.026

Shih-Wen Hsiao
National Cheng Kung University

Chu-Hsuan Lee
National Cheng Kung University

Information Sciences 412–413 (2017) 14–35

Contents lists available at ScienceDirect

Information Sciences

journal homepage: www.elsevier.com/locate/ins

Methodology for stage lighting control based on music emotions

Shih-Wen Hsiao*, Shih-Kai Chen, Chu-Hsuan Lee

Department of Industrial Design, National Cheng Kung University, No. 1, University Road, Tainan 701, Taiwan

ARTICLE INFO

Article history:
Received 20 June 2016
Revised 12 May 2017
Accepted 16 May 2017
Available online 18 May 2017

Keywords:
Automatic stage-lighting regulation
Music emotion recognition
Lighting color regulation based on music emotions and genre
Support vector regression (SVR)
Automatic music segment detection

ABSTRACT

Traditionally, stage lighting regulations have required that professionally trained technicians operate the lighting equipment; however, contemporary demands for higher-quality performances require more preparation before a performance. Thus, technicians or club DJs now spend double to triple the time previously required before a show on matching the lighting control sequence musical instrument digital interface (MIDI) with the music, which is very time consuming. Thus, a methodology for automatic stage-lighting regulation would be very useful. Recently, the development of music emotion recognition (MER) and neural network algorithms has progressed significantly. Feelings related to music can be recognized and are even quantifiable using a supervised machine learning approach. In this study, a variety of music signal features from 2,087 song clips were captured, and then, a cross-validation test based on the support vector machine's (SVM) accuracy of classifying them into Thayer's emotion plane was applied to the main features related to music emotions, in order to produce linear quantitative values for describing music emotions. Music emotions and color preferences for stage lighting were subsequently studied. Using the experimental results, a support vector regression (SVR) model was trained to construct simulations. To increase the realism of the simulations, we developed an automatic music segment detection methodology based on music signal intensity to capture the different music strengths and feelings in each segment. Furthermore, music genres were studied as a factor for developing a comprehensive automatic stage lighting system based on feelings, genre, and the intensity of each segment of music.

1. Introduction

Traditionally, stage lighting regulations have required that professionally trained technicians operate the lighting equipment. However, with rising labor costs, most technicians are now required to control both light and sound. Existing systems can use quantifiable signal features to adjust the stage lighting parameters. However, music feelings are abstract, as are lighting and color feelings, and therefore, it is difficult to build a computer simulation and automatically generate the lighting and color that reflect these feelings. The development of music emotion recognition (MER) and neural network algorithms has recently progressed significantly. Feelings evoked by music can now be recognized in a more scientific manner; they have even become quantifiable by means of a supervised machine learning approach. In addition, studies on color emotions have indicated that light and color can influence people's perceptions and emotions. Overall, it seems that the development

* Corresponding author.
E-mail addresses: swhsiao@mail.ncku.edu.tw, swhsiao2002@yahoo.com.tw (S.-W. Hsiao).

http://dx.doi.org/10.1016/j.ins.2017.05.026

of an automatic lighting regulation system based on feelings, genre, and the intensity of each segment of music is no longer impossible.

1.1. Music emotions

Research on music and emotions has a long history. In recent years, MER has undergone significant development. A typical approach is to choose several adjectives that express emotions and use a property classification algorithm to recognize different music emotions [7,20]. Although the use of this approach has become widespread, researchers face a problem: it is difficult to use a small number of adjectives to cover all possible music emotion expressions. Furthermore, people having different cultural backgrounds interpret adjectives differently. Therefore, researchers have started to seek a general and efficient means of expressing music emotions. An emotion model, known as the Thayer model, which is based on linear and continuous emotion expressions, is now frequently used. This model, which is described in detail below, uses two consecutive values: valence (indicating positive or negative emotions) and arousal (the energy degree of an emotion) to define a plane covering the expressions of human emotion. For example, many studies have used a support vector machine (SVM) to classify music into the four quadrants of the Thayer model [24,49]. When these types of emotion expressions are used, it is not necessary to strictly define each music emotion by a particular adjectival term. Therefore, even people from different cultural backgrounds can easily understand the terms. Meanwhile, the Thayer model allows emotions to be expressed as two linear values, which is very convenient for computer applications. However, it is very difficult to analyze the impact of music emotion classification or to construct a computer simulation regressively.

Since neural network-supervised learning machines have reached a mature phase of development, the use of a regression approach has gradually become a trend in current research on music emotion recognition. In one study, 253 subjects were invited to listen to 10 music clips randomly selected from 195 pop song clips and to give them emotional values [49]. Yang subsequently used the experimental results to train several supervised learning regression algorithms; the validation results showed that the trained support vector regression (SVR) model output had the highest statistical correlation coefficient [49]. However, the validation results were not outstanding, an outcome that could have been caused by the small amount of data derived from the experiment.

In the SVR approach, a supervised learning algorithm is used to simulate music emotion regressively. Instead of a small amount of experimental data, a large amount of music information derived from social network Web data is used as the training data. In order to reduce the errors caused by relying solely on the subjective personal perceptions of a few subjects, the current trend is to conduct experiments using a large number of users. This approach should enhance the accuracy of music emotion simulations, allowing the relationship between music emotions and lighting regulation according to its properties to be studied in depth.

1.2. Lighting colors and emotional feelings

Colors carry both messages and images. They have different significance for people having different cultural backgrounds. Currently, researchers are using the Kansei Engineering measurement approach to describe or measure "the amount" of emotions people feel [29]. This approach has enabled researchers to interpret abstract color imagery feelings numerically. Recently published studies on color emotions have focused on the selection of emotional scales and on investigating the manner in which these scales are related by using factor analysis. Regression analysis is usually applied to reveal the relationships of human responses; herein, the scales describe the underlying color appearance attributes, such as lightness, chroma, and hue. In many studies, researchers have advanced to applying their methods [11], and with the mature development of computer-aided design, in these studies computer-aided design color plans for products or clothes based on eliciting specific emotions have been developed [9,10,11]. Light and color can influence people's perception of the characteristics of the area around them, such as its comfort and atmosphere; they can even cause the pulse and endocrine activities to accelerate [28]. These studies indicated that the regulation of light can trigger specific emotions and affect the excitement level of an audience, and a specific color can be associated with emotional feelings in a specific atmosphere. Thus, it is reasonable to arrange an appropriate lighting regulation based on music emotions.

2. Related work

2.1. Emotion model

MER has recently been significantly developed [24,26,49]. In almost every related study, emotions were defined differently. Scholars have used a number of adjectives to describe some of the basic emotions [16,41,38]. In addition, Li and Ogihara [22] presented 13 classifications for different types of emotions, including 11 taken from Farnsworth's research [6]; they added the remaining two types. They sought a definition that would cover all the expressions of music emotions. However, their results were not very convincing; the small sample of people who labeled the music under consideration reduced the reliability of their study results. Therefore, in many studies on music emotion recognition, psychological research on music, as a theoretical background, is now employed to find relevant expressions to define emotions. Some adjectives, such as “delicate,” “charming,” and “gloomy,” are used frequently. However, the total number of adjectives used to describe

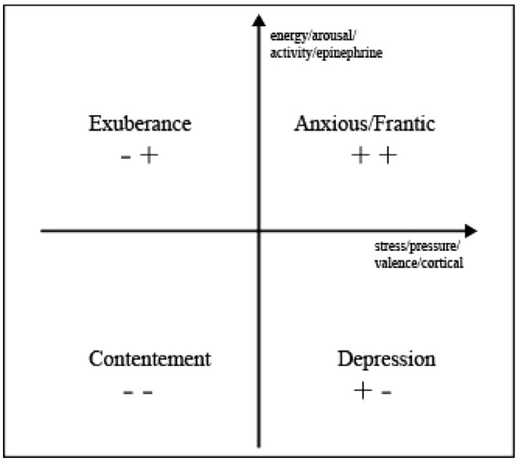


Fig. 1. Thayer's emotion model.

emotions has grown significantly, as has the number of words in the descriptions [31]. Although different interpretations of adjectives have been used in various studies, no set of adjectives has been widely accepted thus far.

Although emotions can be described as a set of dimensions, as above, there is additional evidence that these emotional dimensions are interrelated in a highly systematic fashion. Russell [34] proposed a circumflex model of affect based on two bipolar dimensions, instead of on several mono-polar states. The two dimensions are termed pleasant-unpleasant and arousal-sleep. Thus, each affective word can be defined as some combination of the pleasure and arousal components. Thayer, using Russell's music emotion model, constructed a two-dimensional energy-stress mood model [43], as shown in Fig. 1, where the energy dimension corresponds to arousal and stress corresponds to the pleasure included in Russell's model. In fact, the precise name of each dimension varies significantly in the literature. For example, in Juslin's work [13], the dimensions are called valence and activity level, whereas in affective computing [31], they are called valence and arousal. However, the basic meaning of each dimension is very similar. The dimension of arousal, energy, or activity level ranges from quiet to energetic (or in the reverse direction), while the dimension of stress, pleasure, or valence represents a range from unpleasant to pleasant or from negative to positive (or in the reverse direction). Based on the level of stress and energy, Thayer's model divides music moods into four clusters (each in one quadrant), contentment, depression, exuberance, and anxious/frantic, where the horizontal dimension (from left to right) represents the range from pleasant to unpleasant and the vertical dimension (from bottom to top) illustrates the range from quiet to energetic.

These four emotion clusters almost cover the basic mood responses to music, and they are usually among the most highly rated emotions [14,23]. Moreover, these four clusters are explicit and discriminative, and the two-dimensional structure gives important cues for computational modeling while avoiding the problems associated with using a small number of adjectives to cover all possible music emotions and the cultural biases that cause people to interpret adjectives differently. Thus, Thayer's emotion model was adopted in the present study to describe music emotions.

2.2. Music feature extraction

Many features for describing music have been proposed, such as linear prediction coefficients (LPC), linear prediction cepstrum coefficients (LPCC) [16], mel-frequency cepstrum coefficients (MFCC) [3,17,37,48], entropy and dynamism [1], timbre [1,3,24,37,50], intensity [24], rhythm [24,44], pitch [44], amplitude envelope [48], and Daubechies wavelet coefficient histograms [37].

The classification of continuous general audio data for content-based retrieval, where the audio segments were classified based on MFCC and LPC, was also addressed. Li et al. [19] showed that cepstral-based features provide better classification accuracy. The method described for content-based audio classification and retrieval using joint time–frequency analysis exploits the non-stationary behavior of music signals and extracts features that characterize their spectral changes over time [5]. Audio signals were decomposed in Umapathy's study by using an adaptive time frequency decomposition algorithm; signal decomposition parameters based on octaves (scaling) were used to generate a set of 42 features over three frequency bands within the auditory range [45]. These features were analyzed using linear discriminant functions and then were classified into six music groups.

A related music feature capture approach includes a time domain basis and a frequency domain basis, which provides different physical characteristics. Many studies have indicated that the intensity feature is necessary for emotion detection [12,24]. This feature is simple and easy to compute. Timbre refers to audio signal information. Timbre features are usually described by centroid, bandwidth, roll-off, flux, zero crossing rate, and octave-based spectral contrast [24,37,44,48]; they can be computed based on a short-time Fourier transform. Music features were classified into intensity- and timbre-related features, depending on the model on which they were based; they were then divided into the time-domain basis and frequency-domain basis for the purposes of this research. A total of 21 music features were investigated in this study.

2.3. Supervised learning machine

In the field of automatic pattern recognition, a supervised learning algorithm for classifying properties is indispensable. In the field of music emotion recognition research, some algorithms have been used for learning, such as SVM [3], SVR [48,49], fuzzy C-Mean [30], Gaussian mixture models (GMMs) [24], multi-layer perceptrons (MLPs), hidden Markov model (HMM) [1], K-nearest neighbor (KNN) [21,44], AdaBoost [50], and radial function basis neural networks (RBFNNs) [3]. In this study, music emotion recognition was considered a regression problem, and regression techniques were appropriately used to predict music emotions from extracted features. Some regression algorithms were tested for training a music emotion recognition simulation, including multiple linear regression (MLR), SVR [39], and AdaBoost.RT (BoostR) [40]. The SVR approach was used as the supervised learning algorithm to simulate music emotion regressively. Instead of a small amount of experimental data, a large amount of music information from social network Web data was used as the training data.

2.4. Connection between lighting color and emotions

In the color vision domain, a significant number of studies have addressed the correspondences between auditory dimensions (mainly pitch and loudness) and color dimensions (such as hue, lightness, and saturation) [2,32,35]. These studies were based on correspondences that may exist between the physical characteristics of sound and color. For example, in Caivano's study [2], hue was associated with pitch, since these two dimensions are closely related to the dominant wavelengths in color and sound spectra, respectively. Finally, the lightness of color has been associated with loudness (black and white represent silence and maximum loudness, respectively, with the grayscale representing the intermediate levels of loudness). Sebba [35] took a different approach: in a series of experiments, she investigated the structural correspondences that may exist between color and music elements, as opposed to the direct comparisons of the elements themselves. Her experimental results suggest that such structural correspondences between color and music do exist (e.g., emotional expression, hierarchical organization, and contrast). Mark's report [27] shows associations between pitch and light intensity, as well as between loudness and light intensity. Marks' overall conclusion was that pitch and loudness are related not to light intensity, but rather to auditory brightness. This conclusion is based on the assumption that auditory brightness is the same as auditory density, a dimension that increases when both pitch and loudness increase [27]. The lighting intensity regulation presented in this paper is also based on this conclusion. Furthermore, as mentioned earlier, light and color can influence the way people perceive their level of comfort and the atmosphere of the space around them, and may even affect their physiological functions, such as heart rate [28]. These studies indicated that lighting regulation can trigger specific emotions and affect the excitement level of the audience, with specific colors being associated with specific emotional feelings; thus, it is reasonable to regulate lighting appropriately based on music emotions.

Five professional stage-lighting technicians with more than two years' experience were invited to execute the lighting color regulation according to its properties for 988 music clips, in order to study lighting color regulation preferences and music emotions. A MANOVA was used to investigate the relationships revealed by the experimental results. The analysis showed that music emotion and lighting color are significantly related. Subsequently, the results were used to train an SVR to construct the simulations.

3. Theoretical framework

The lighting regulation was divided into two parts in this study: brightness and lighting color. Lighting brightness regulation was conducted according to the music intensity. The deviations in the signal peak were used to detect music segments, and music onset was determined using a flashing light corresponding to the rhythm of the music. Lighting color regulation was conducted to study the relationship between lighting color regulation preferences and music emotions. Several professional stage-lighting technicians were invited to execute lighting color regulation according to its properties for 988 music clips. Therefore, the lighting color was dependent on the value of the music emotion recognition results. The selected music was used to extract several features that were related to music emotions. After a principal component analysis (PCA) dimension reduction approach was applied, these music features were reflected in the Thayer model so that the trained SVR would recognize direct emotional value (arousal and valence).

3.1. Music emotion recognition

A total of 2087 music clips of 20-sec duration were captured from the most popular songs on the Musicovery Website, and several musical features were extracted from these music clips using PCA to reduce the dimension of the feature space. These dimension-reduced features were used for a cross-validation (CV) test based on the SVM accuracy of classifying them into the four quadrants of Thayer's model, in which the parameters are chosen by an optimized parameter selection methodology. The objective was to choose the feature combination with the highest recognition accuracy rate as the music features that best express emotions on which to base the regression simulation approach. An SVR approach was used as the regression simulation approach, in which the parameters were selected using a 5-fold CV technique. The correlation coefficients of the testing data were found to surpass 80%, which was suitable for the construction of a music emotion recognition system.

3.1.1. Audio analysis and processing

In our experiments, all the music pieces in the dataset were converted into 44.1 kHz mono WAV files. The audio data were segmented into fixed-length frames and overlapping samples (in this study, 2048 signal samples with 25% overlap were used). Before features were acquired in the frequency domain, each signal sample was implemented via a pre-emphasizing filter, and defined as

$$x_i'(n) = x_i(n) \times h(n), \tag{1}$$

where $x_i(n)$ represents the input signals of the i-th sample [24,32,35]

$$h(n) = 0.54 - 0.46 \times \cos\left(\frac{2\pi n}{N-1}\right), \tag{2}$$

where $0 \le n \le N-1$ and N signifies the number of points in each frame.

3.1.2. Music feature extraction

Researchers have noted that intensity and timbre are two major features in music emotion recognition [12,24]. The intensity, average, variance, and maximum were extracted as the music intensity features in this study, and the centroid, bandwidth, roll-off, flux, zero crossing rate, linear predictive cepstral coefficients, and the mel-cepstral coefficients were used as the music timbre features. These features were divided into the time domain and frequency domain to produce a total of 21 music features. In addition, LPCC and MFCC are general features used in audio signal recognition and classification; they are robust and reliable [3,17,25,36,37,46,48].

Each frame was multiplied by a Hamming window. The fast Fourier transform (FFT) was then applied on each frame to obtain the corresponding spectrum. Since the FFT transforms signals in the time domain to the frequency domain, some features were preserved in the time domain.

The details of the operation are as follows.

Intensity:

$$I_i = \sum_{n=0}^{N-1} |x_i'(n)| \tag{3}$$

where $x_i'(n)$ is the n-th sample point in the i-th frame.

Average:

$$A_i = \frac{1}{N}\sum_{n=0}^{N-1} x_i'(n) \tag{4}$$

Variance:

$$V_i = \frac{1}{N}\sum_{n=0}^{N-1} (x_i'(n) - A_i)^2 \tag{5}$$

Maximum:

$$M_i = \max\left\{x_i'(n) | 0 \le n < N\right\} \tag{6}$$

Centroid:

$$C_i = \sum_{n=0}^{N-1} x_i'(n) \times n \Bigg/ \sum_{n=0}^{N-1} x_i'(n) \tag{7}$$

Bandwidth:

$$B_i = \sum_{n=0}^{N-1} \left(\left|x_i'(n)\right| - (n - C_i)\right) \Bigg/ \sum_{n=0}^{N-1} \left|x_i'(n)\right|^2 \tag{8}$$

Roll-off:

$$R_i = \sqrt{0.95 \times \sum_{n=0}^{N-1} \left|x_i'(n)\right|^2} \tag{9}$$

where R_i is less than 95% of the power distribution.

Table 1
Music signal features.

Time Domain		Frequency Domain	
Intensity	Average (TA)	Intensity	Average (FA)
	Variance (TV)		Variance (FV)
	Intensity (TI)		Maximum(FM)
Timbre	Centroid (TC)		Intensity (FI)
	Bandwidth(TB)	Timbre	Centroid (FC)
	Roll-off (TR)		Bandwidth(FB)
	Flux (TF)		Roll-off (FR)
			Flux (FF)
			Peak (FP)
			Valley (FV)
			Contrast (FC)
			Zero Crossing Rate (FZ)
		LPCC (FL)	
		MFCC (FM)	

Flux:

$$F_i = \sum_{n=0}^{N-1} \left(\left| x_i'(n) \right| - \left| x_{i-1}'(n) \right| \right)^2 \tag{10}$$

Zero Crossing Rate:

$$ZCR_i = \sum_{n=0}^{N-1} \left| sgn\left(x_i'(n)\right) - sgn\left(x_{i-1}'(n)\right) \right| \tag{11}$$

where the sgn function is 1 for a positive argument and 0 for a negative argument (sgn(0)= 0).

A total of 21 types of music features were examined in this study, as shown in Table 1.

3.1.3. Feature dimensions reduction

The 21 types of music features may not be of equal importance or quality from the perspective of machine learning. Although information facilitates machine learning training, irrelevant or redundant features may lead to inaccurate conclusions. Furthermore, the time-series data length changes with time. Therefore, a suitable feature dimension reduction approach should be used to reduce features as much as possible to obtain high-accuracy recognition. One solution for solving this problem is to utilize PCA, which is normally used to increase efficiency and improve accuracy by eliminating irrelevant data [4]. In the area of music emotion recognition, Yang et al. [49] also used this approach to increase calculation speed and accuracy by analyzing the principal data using dimensional reduction.

PCA is a statistical procedure that uses an orthogonal transformation to convert a set of observations of possibly correlated variables into a set of values of linear uncorrelated variables. The PCA number is less than or equal to the number of original variables. This transformation is defined such that the first principal component has the largest possible variance (i.e., it accounts for as much of the variability in the data as possible), and each succeeding component in turn has the highest variance possible under the constraint that it is orthogonal (i.e., uncorrelated with) to the preceding components. The principal components are orthogonal because they are the eigenvectors of the covariance matrix, which is a symmetric matrix. PCA is sensitive to the relative scaling of the original variables and is a frequently used method for reducing the correlation between variables, which entails the computation of a loading matrix L to transform the original data Y to principal components U such that

$$U = L(Y - mean(Y)) \tag{12}$$

$$Y = L^{-1}U + mean(Y), \tag{13}$$

where U is the representation of Y in the principal component space.

3.1.4. Support vector machine

Before training a regression simulation, features that are related to emotions must be selected to ensure that the learning process is effective and robust. A classifier simulation was used as a pre-test for selecting the best group of music features for music emotion recognition. Laurier et al. [15] evaluated the implementation of multiple classifiers using 10 runs of 10-fold CV. The best results were achieved by an SVM algorithm with linearly normalized features between 0 and 1, default parameters, and a polynomial kernel. In addition, an SVR model produced by support vector classification depends only on a subset of the training data, because the cost function for building the model ignores the training points that lie beyond the margin. SVMs and SVR define the range of data error by applying the structural risk minimization principle. If the SVM classification results perform well, this can be considered indirect supporting evidence that good simulation results were achieved by SVR training. Training a classifier is faster and requires less computation than conducting a regression simulation, and therefore, an SVM is suitable for pre-testing a large number of CVs, and for selecting music emotion features as regression simulation training data. These extracted music features are then input into the SVM trained to classify arousal and the SVM trained to classify valence. The results of the arousal SVM may be high (+) or low (-), whereas the results of the valence SVM may be positive (+) or negative (-). If the arousal SVM classifies a music sequence as high (+) and the valence SVM classifies it as negative (-), the final result is located in the second quadrant of the Thayer model. Similarly, if the arousal SVM classifies a music sequence as low (-) and the valence SVM classifies it as negative (-), the final result is located in the third quadrant of the Thayer model. According to the recognition accuracy, pre-testing identifies the most relevant group of music features for training input for the SVR simulation.

The algorithm used to determine the best hyperplane as a reasonable choice is that which represents the widest separation, or margin, between two classes. This approach is aimed to minimize the error bound (structural risk) rather than the mean squared error (empirical risk minimization principle). Thus, the SVM completes an optimized network structure by applying a structural risk minimization principle, the Vapnik-Chervonenkis dimension trade-offs of a confidence interval. The final results are better than those obtained using other neural network algorithms. Thus, an SVM can solve a nonlinear classification problem.

3.1.5. Optimal kernel function parameter selection

SVMs constitute one of the most powerful techniques for supervised classification. However, their performance depends on choosing the appropriate kernel functions or the appropriate parameters of a kernel function. The application of k-fold CV to choose the “almost best” parameters is extremely time consuming. Nevertheless, the search range and the fineness of the grid in the grid method should be determined in advance. Li [18] proposed an automatic method for selecting the parameters of the RBF kernel function. The parameter selection when using our proposed method required very little time as compared to k-fold CV. Moreover, the corresponding SVMs were found to achieve a performance that is more accurate than, or at least equal to, that of SVMs applying k-fold CV to determine the parameters. The implemented SVM algorithm was based on [18].

3.1.6. Support vector regression

In this study, the features combination having the highest recognition accuracy rate was chosen as the music feature that best expresses emotions for the regression simulation approach. SVR performed better than other regression simulation algorithms [49], signifying that the simulation output shows the highest statistical correlation coefficient. As compared to other neural network regressions, when SVR is used to estimate the regression function, it has three distinct characteristics. First, SVR estimates the regression using a set of linear functions that are defined in a high dimensional space. Second, SVR uses a risk function consisting of the empirical error, and third, it uses a regularization term that is derived from the structural risk minimization principle. The implemented SVR algorithm was based on [42].

3.2. Multivariate analysis of variance

Multivariate analysis of variance (MANOVA) is simply an ANOVA with several dependent variables; i.e., ANOVA tests for the difference in the means of two or more groups, while MANOVA tests for the difference in two or more vectors. MANOVA is one of the most frequently used multivariate statistical procedures in studies on means in the social science literature. All current MANOVA tests are based on $A = E^{-1}H$, where H denotes the hypothesis sums of squares and the cross products matrix and E denotes the error sums of squares and the cross products matrix. The multivariate equivalent of the A statistic is matrix A. Four different multivariate tests are applied to $E^{-1}H$; each test's statistics has its own associated F ratio. In some cases, the four tests yield an exact F ratio for testing the null hypothesis, while in others the F ratio is approximated. The reason for four different statistics and for approximations is that the mathematics of MANOVA becomes so complicated in some cases that it has never been solved. In this study, Wilks' lambda was used.

Wilks' lambda was the first MANOVA test statistic to be developed and is very important for several multivariate procedures, in addition to MANOVA.

$$\text{Wilks lambda}(\Lambda) = \frac{|E|}{|H + E|} = \prod_{i=1}^{n} \frac{1}{1 + \lambda_i} \tag{14}$$

The quantity (1 - Λ) is frequently interpreted as the proportion of variance in the dependent variables explained by the model effect. When the statistics have been obtained, they are translated into F statistics in order to test the null hypothesis. The reason for this translation is identical to the reason for converting Hotelling's T^2, i.e., the easy availability of published tables of the F distribution. It should be noted that, in some cases, the F statistic is exact, while in other cases it is approximate [47].

3.3. Automatic lighting regulation methodology

Several professional stage-lighting technicians were invited to execute lighting color regulation according to its properties for 988 music clips in order to study the relationship between lighting color regulation preference and music emotions.

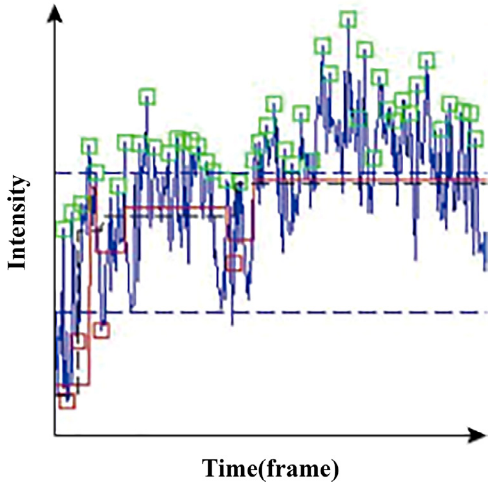


Fig. 2. The green box shows the peak and the red box shows the valley of the wave. The red square wave line shows the mean intensity for each music segment. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Then, using the experimental results, an SVR was trained to construct simulations. Music segmental differences were considered, as different emotional feelings and intensity are caused by specific music segments. Therefore, an automatic music segment detection methodology, based on music signal intensity, intended to represent different music strengths and the feelings associated with each segment, was developed in this research study. The music was segmented according to this methodology, and then, the music emotion recognition and automatic lighting regulation approach were executed separately for each segment.

3.3.1. Automatic music segment detection

The emotional feelings related to music are likely to change one or more times within a single piece of music. Since it is not appropriate to detect an exclusive emotion for an entire piece of music, we detect a unique emotion in each segment correspondingly. The main subject of this section is our proposed approach to music segment detection in a music piece using a two-step mood tracking scheme. Subsequently, the mood tracking in an entire piece of music can be easily performed by identifying the emotion in each independent segment by using the music emotion recognition approach. In this approach, an intensity outline is implemented to coarsely detect a music segment. The intensity outline is detected to temporally divide a music piece into several segments, where each segment contains an almost constant intensity or corresponds to a single energy contour that grows from low to high and then returns to low energy. These segments can roughly represent the mood development along a timeline. Two thresholds are defined to preliminarily quantize the energy envelope into three levels instead of binary levels, thereby generating a larger number of potential segment boundaries whenever the energy envelope crosses either threshold. The thresholds are adaptively set, based on the statistics of the energy envelope of the entire piece of music, as $Th1=\mu-0.5\sigma$, $Th2=\mu+0.5\sigma$, where μ and σ are the corresponding mean and standard deviation. The next section describes the peak and valley detection methodology in detail. However, the boundaries of some segments may be biased. In order to derive more reliable results, the boundary of each segment is aligned to the nearest valley of the energy envelope, where a valley is defined as a point of local minimum, as shown in Fig. 2.

In the first part of the segmentation approach, the length of each segment is too small to be considered a musical paragraph, but is suitable as a lighting brightness indicator. The average intensity of each segment is normalized to grayscale as the lighting brightness index. The next step is to make the segment concur more realistically with the music paragraph. The average intensity data segmented in the previous step are segmented again using the aforementioned approach to define the final paragraph of the music. In this stage, a 4-min song is divided into 10 to 14 paragraphs, which fits most of the arrangement of the music's original paragraphs.

3.3.2. Audio peak and valley detection

A Gaussian function convolved to the original audio signal is set to reduce the noise peaks, as shown in Eq. (15), where f represents the original audio data, and g represents a Gaussian function:

$$y(t) = f(t) * g(t), \tag{15}$$

$$where\ g(t) = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{(t-\mu)^2}{2\sigma^2}\right) \tag{16}$$

where y(t) represents the new data derived after the convolving approach. If the original k sub-band is represented as $\{y_{k,\,1}, y_{k,\,2}, L, y_{k,\,n}\}$, the new data after convolving are represented as $\{y'_{k,\,1}, y'_{k,\,2}, L, y'_{k,\,n}\}$. The peaks and valleys

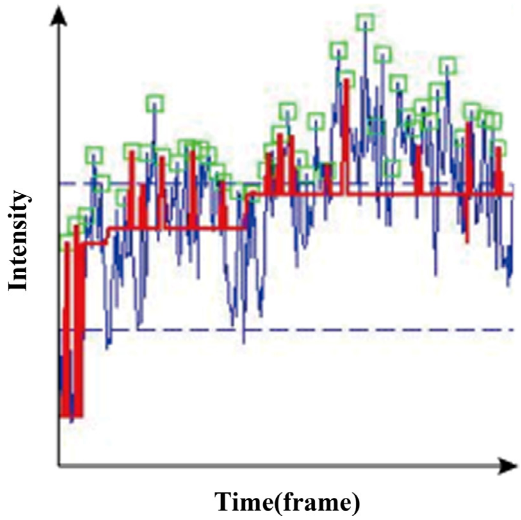


Fig. 3. The thick red line shows the average music intensity, which was used as the signal of the corresponding light intensity. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

are defined as

$$Peak:\ if\ sign(y'_{n-1}) \geq sign(y'_{n+1})\&\&x_n \geq mean(y) \tag{17}$$

$$Valley:\ if\ sign(y'_{n+1}) > sign(y'_{n-1})\&\&x_n < mean(y) \tag{18}$$

3.3.3. Audio onset detection

Audio onset detection processing is widely used in music signal applications. The signal onset is considered a flashing regulation indicator. In this study, the Roberts [33] cross operator was used for signal onset detection. The Roberts cross operator performs a simple, quickly computed, 2-D spatial gradient measurement of an image. In the onset detection, the Roberts cross operator is used to generalize the waveform intensity and quickly find rising edges that denote onsets. According to Roberts [33], in its most common usage, the input to the operator is a grayscale image, as is the output. The pixel values at each point in the output represent the estimated absolute magnitude of the spatial gradient of the input image at that point. In this proposed approach, the pre-emphasizing audio intensity signals are first calculated using the equations

$$y_n = \sqrt{x_n} \tag{19}$$

$$z_n = \sqrt{2(y_n - y_{n+1})^2}, \tag{20}$$

where x is the initial intensity value in the matrix, z is the computed derivative, and n represents the location in the matrix. Then, a threshold is used to identify the edge; the calculations are

$$\begin{aligned} &\mathrm{Th} = \sqrt{\mathrm{a} \times \mathrm{Mean(z)}} \\ &\begin{cases} \mathrm{z_n} \geq Th, & true \\ \mathrm{z_n} < Th, & false \end{cases} \end{aligned} \tag{21}$$

where Th is the threshold, z is the result of the Roberts cross operator, and a is parameter with a value of 6.

The results of this operation highlight changes in intensity in a diagonal direction. The rising edge is detected as the onset. The foregoing segmentation method divides music into several segments, where the average intensity value is calculated as the lighting brightness regulation indicator. Then, the onsets of audio signals are detected using the aforementioned approach, and these values are added into the segmental average intensity data, as shown in Fig. 3. The final data are normalized to grayscale as lighting brightness indicators.

4. Research procedures

Various music signal features from 2087 song clips were captured, and the main features related to music emotions were selected, following Thayer's emotion plane, in order to produce linear quantitative values describing music emotions. Then, the music emotions and color preferences for stage lighting were examined. Using the experimental results, an SVR was trained to construct simulations. As mentioned, for greater realism, an automatic music segment detection methodology

Table 2
Reduced dimensions of each music feature.

Time Domain		Frequency Domain	
Feature	Dim.	Feature	Dim.
Average (TA)	6	Average (FA)	7
Variance (TV)	40	Variance (FV)	45
Intensity (TI)	7	Maximum (FM)	7
Centroid (TC)	1	Intensity (FI)	7
Bandwidth (TB)	1	Centroid (FC)	1
Roll-off (TR)	1	Bandwidth (FB)	2
Flux (TF)	42	Roll-off (FR)	1
		Flux (FF)	42
		Peak (FP)	7
		Valley (FV)	2
		Contrast (FC)	1
		Zero Crossing Rate (FZ)	1
		LPCC (FL)	2
		MFCC (FM)	2

based on music signal intensity was developed to present different musical strengths and the feelings associated with each segment. In this study, MATLAB was used to analyze data, and the JAVA language was used to construct the experimental user interface and firmware communication; in addition, a MATLAB model was imported for computations.

4.1. Music emotion recognition

Music clips were captured from the most popular songs on the Musicovery Website, and several musical features were extracted from these music clips using PCA to reduce the dimensions of the feature space. These dimensionally reduced features were used for a CV test based on the SVM accuracy of classifying into the four quadrants of Thayer's model, in which the parameters are chosen using an optimized parameter selection methodology. The combination of the highest recognition accuracy rate features was then chosen as the music features that best reflect emotions in order to implement the regression simulation approach. An SVR algorithm was used as the regression simulation approach, where the parameters were selected using 5-fold CV techniques.

4.1.1. Experiment music sample selection

A large number of user experiments can be assumed to reveal a general trend, which can be utilized to enhance the accuracy of music emotion simulations. This allows the provision of emotion retrieval services that enable users to select music according to an emotion plane that is the same as the Thayer emotion plane. The Musicovery Website offers an application programming interface (API) that provides data to generate music recommendations and playlists of all types of music, based on a mood, artist, track, genre/style, theme, period/year, and so on. The top 100 most popular songs according to the information provided in the Musicovery database were selected as the music samples for the experiment. A total of 2087 20-sec music clips were selected. Meanwhile, music emotion values that include arousal and valence values were used as the training target for the supervised learning algorithm.

4.1.2. Music features extraction

The intensity features included the intensity, average, variance, and maximum. In addition, the centroid, bandwidth, roll-off, flux, zero crossing rate, LPCC, and MFCC were used as the music timbre features. Music features were classified into intensity- and timbre-related features, and then were accordingly divided into the time-domain basis and frequency-domain basis. A total of 21 music features were examined in this study.

4.1.3. Music feature dimensions reduction

All the samples were 20-sec clips in the 44.1 kHz standard mono WAV format. The data were segmented into fixed-length and overlapping frames, using 2048 signal samples per frames with 25% overlapping. In order to preserve the characteristics of the different features of music signals at different time points, in this study the feature values for each frame were captured, except for LPCC and MFCC. Each 20-sec music clip had 574 frames, and each feature constituted the data of a feature matrix having 574 dimensions. The 2087 music clips were collected as a set of 2087×574 matrix data based on the same features. After extracting all the features mentioned above, a PCA approach was employed to reduce the feature matrix data. The dimensions of these feature matrixes were determined by the matrix in the PCA approach, with eigenvalues greater than one. The features of each dimension are shown in Table 2.

4.1.4. Emotion features selection

To test which features are most suitable for emotion recognition, the 21 music features mentioned above were used to train a music recognition classification system. This system in fact included two SVMs, one of which was trained for arousal,

Table 3
Best classification accuracy performance feature groups.

Arousal		**Valence**	
Feature combination	Accuracy	Feature combination	Accuracy
FA, FM, TB	100%	FA, FC, FF, FM, TC, TF	80%
FF, FM, TA, TC, TF, TV	100%		
FA, FF, FM, FV, TB, TC, TR	100%		
FA, FC, FF, FV, TB, TC, TF, TR, TV	100%		

which classified the music into either a high or low degree arousal and the second was trained for valence, which classified the music into the positive or negative direction. There were 50 music clips for each emotion quadrant, with a total of 200 music clips randomly selected from the 2087 music clips. These samples were randomly separated into two parts; 10% were used as validation data, and the remaining 90% for training. All the parameters were selected using optimal kernel function methodology with 10% of the training data, in order to elevate the classification accuracy rate. The classification accuracy, after the validation stage, is shown in Table 3.

4.1.5. Music emotion recognition support vector regression construction

In this study, two SVRs with a Gaussian kernel function, arousal and valence, were used for each axis of the Thayer emotion plane. The 2087 music samples were randomly separated according to a 7:3 ratio for training and validation; that is, 1461 music samples were used for training and 626 for validation. As the effectiveness of an SVR depends on the appropriateness of the parameters of the kernel functions, half of the training samples were used to determine the appropriate parameters using a 5-fold CV grid search technique; the root mean square error value was determined as an indicator. The search range was based on the recommendations in Camps-Valls et al.'s paper [8]. The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 99%. The results of the arousal value simulation reached 99.96% accuracy, and those of the valence value simulation reached 99.93%. The statistical correlation coefficients between the SVR simulation output using the testing sample and validated music emotion value results surpassed 80% accuracy. The results of the arousal value simulation reached 89.93% accuracy, and the results of the valence value simulation reached 80.76% accuracy.

4.2. Lighting color regulation experiment

Using the experimental results, an SVR was trained to construct simulations. In this experiment, a CIE HSI color space system was used. The subjects in the experiment regulated lighting color according to its saturation and hue, while taking into consideration that color intensity can be confused with dimmer lighting and directly associated with music intensity; in this experiment, this factor was eliminated. The subjects watched an LED Par Light, 2 m high and $2 \times 2\ \mathrm{m}^2$, projected on painted white walls; they were also able to watch the color displayed on the computer screen. A MANOVA was used to investigate the relationships revealed by the experimental results; it was found that music emotion and lighting color are significantly related. Then, using the experimental results an SVR was trained to construct simulations with lighting regulation and music emotions.

4.2.1. Experimental system

First, 988 music clips were used for the experiment. So that the operation of this experiment would be more intuitive, the CIE HSI color space system was used. The HSI color system describes a color intuitively, and it is frequently used in image processing. Hue is generally understood to signify the color itself, such as yellow, orange, cyan, or magenta, while saturation refers to the purity of the color. Furthermore, the addition of gray reduces the purity and the saturation of a color. If a user wants to perform a color conversion, the hue value can be adjusted for the image using the HSI color system. To convert hue, saturation, and intensity to a set of red, green, and blue values, one must first note the value of H. If $\mathrm{H} = 0$, then R, G, and B are given by

$$\begin{cases} R = I + 2IS \\ G = I - IS \\ B = I - IS \end{cases} \tag{22}$$

If $0 < \mathrm{H} < 120$, then

$$\begin{cases} R = I + IS \times \cos \mathrm{H} / \cos(60^\circ - \mathrm{H}) \\ G = I + IS \times [1 - \cos \mathrm{H} / \cos(60^\circ - \mathrm{H})] \\ B = I - IS \end{cases} \tag{23}$$

If $\mathrm{H} = 120$, then the red, green, and blue values are

$$\begin{cases} R = I - IS \\ G = I + 2IS \\ B = I - IS \end{cases} \tag{24}$$

Table 4
Multivariate analysis of experimental results.

Variables	Wilks Lambda	F test between dependent variables				
		Hue	Saturation	Red	Green	Blue
Arousal	2.277***	2.065*	5.968***	2.717**	1.104	1.801
Valence	1.106	2.173*	0.935	1.049	1.294	1.728
Arousal * Valence	1.017	.932	.907	.941	.767	1.084

*** $p < 0.001$;
** $p < 0.01$;
* $p < 0.05$

If $120 < H < 240$, then

$$\begin{cases} R = I - IS \\ G = I + IS \times \cos(H - 120^\circ)/\cos(180^\circ - H) \\ B = I + IS \times [1 - \cos(H - 120^\circ)/\cos(180^\circ - H)] \end{cases} \tag{25}$$

If $H = 240$, then

$$\begin{cases} R = I - IS \\ G = I - IS \\ B = I + 2IS \end{cases} \tag{26}$$

and, if $240 < H < 360$, we have

$$\begin{cases} R = I + IS \times [1 - \cos(H - 240^\circ)/\cos(300^\circ - H)] \\ G = I - IS \\ B = I + IS \times \cos(H - 240^\circ)/\cos(300^\circ - H) \end{cases}. \tag{27}$$

In this experiment, color intensity was fixed, since previous studies have demonstrated the association of music intensity with color brightness [27], and it was taken into consideration that color intensity can be confused with dimmer lighting. The subjects therefore regulated lighting color according to color saturation and hue, while the intensity value was fixed at 100, in order to ensure that the color was rendered accurately.

In this experiment, an experimental interface written in JAVA was used for executing the lighting regulation and the subsequent recording. Freestyle lighting control software was used as the program regulating stage lighting, and a computer was connected through the MCSWE USB-DMX, which converts a digital signal to a DMX signal, and finally produces the full color LED Stage Par Light.

4.2.2. Experimental operation

Five professional stage lighting technicians with more than two years' stage experience were invited to participate in the experiment. They were asked to regulate lighting color using the computer interface described above for the 988 music emotion clips that had been randomly sorted. The subjects regulated lighting color with two sliders, one for hue regulation and one for saturation regulation, using a music playback/control button for repeating a music sample. If the subjects were not satisfied with a specific regulation, they could enter a direct count order and execute it again. As mentioned previously, they watched an LED Par Light projection 2 m high and 2×2 m^2 on painted white walls; they were also able to watch the color displayed on the computer screen to regulate the colors until they were satisfied with their results.

4.2.3. Experimental results analysis

To analyze the influences of music emotions on lighting regulation preference, a MANOVA was conducted with the hue and saturation designated as the within-participant factors. Moreover, RGB light parameters were added in order to understand how RGB light color mixing is affected by music emotions. Furthermore, the influence of noise in the data was reduced in order to make it easier to determine the trend of the results. Music emotion values were divided into 11 orders by 0.1 intervals. The results are shown in Table 4. In general, if the value of lambda is smaller, the group disparity will be larger and the difference for factors in one group will be smaller. It is assumed that the p-value, which represents the statistical error of the samples, is checked. A smaller p-value indicates that the results are less likely to include errors. Three values are commonly adopted for determining whether or not two factors are significant. It can be seen that arousal has a statistically significant influence (Wilks' lambda $= 2.277$, $p = 0.000 < 0.01$) and the correlation of valence and hue is just above the statistical correlation threshold (Wilks' lambda $= 2.173$, $p = 0.018 < 0.05$). The interactions of the two emotion value influences were less significant (Wilks' lambda $= 1.017$, $p = 0.408 < 0.5$).

According to the factor analysis results for the dependent variables, the influences of arousal on hue and saturation were significant ($F = 2.065$, $p = 0.025 < 0.05$; $F = 5.968$, $p = 0.000 < 0.001$), especially the influence of arousal on saturation. The influence of arousal on hue can be verified from the test results for red light ($F = 2.717$, $p = 0.003 < 0.05$). The valences also influenced hue ($F = 2.173$, $p = 0.018 < 0.05$).

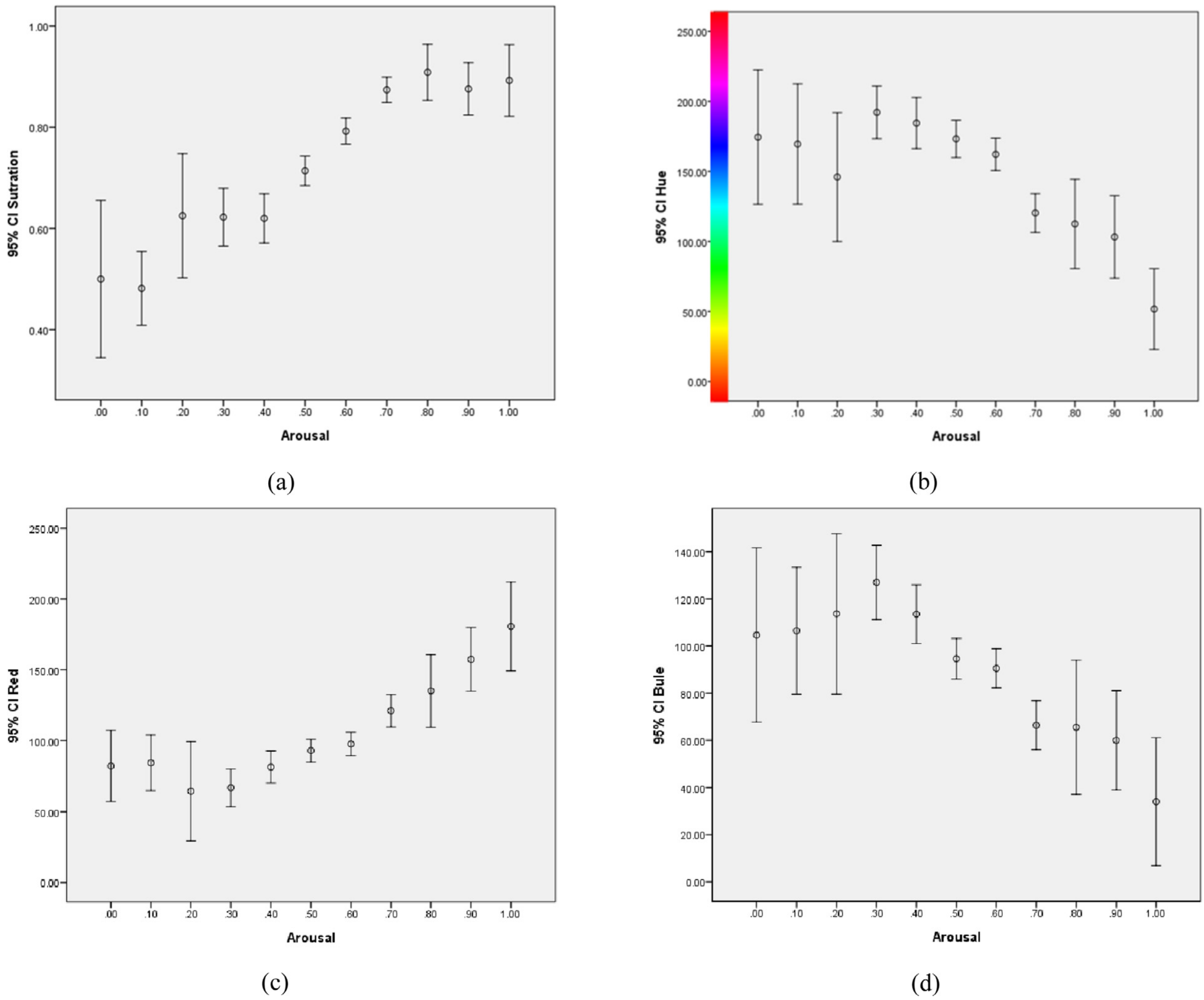


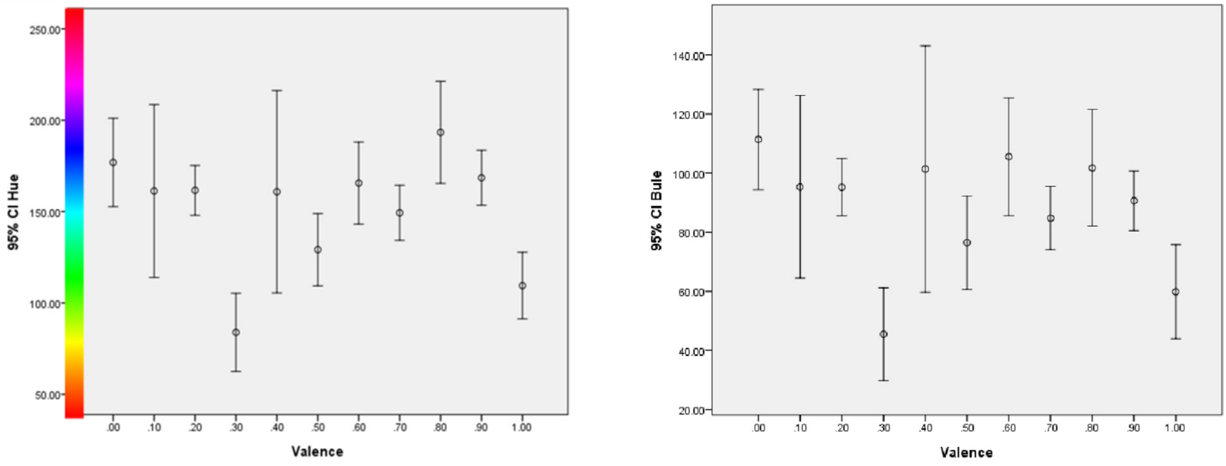
Fig. 4. Error bar chart between the arousal value and dependent variables. (a) Arousal value and lighting color saturation; (b) arousal value and lighting color hue; (c) arousal value and red light; (d) arousal value and blue light. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

The arousal value caused the subjects to make specific light color regulations of hue and saturation, and the influence on the hue can be seen in the blending of a specific amount of red and blue light. The error bar chart between the arousal and dependent variables data at a 95% trust level is shown in Fig. 4, in which the association can be observed more easily. The arousal value shows a positive correlation with the lighting color saturation value. In other words, a higher arousal value is given a higher saturation lighting color. The arousal value shows a slightly negative correlation with the lighting color hue value, and the association can also be observed in the RGB value result, which shows a positive correlation with the amount of red light and a negative correlation with the amount of blue light. In other words, the arousal value is more strongly associated with the red color phase and less with the blue color phase.

The valence value also causes a subject to perform a specific light color regulation of hue and saturation; the effect on the hue can be seen in the blending of specific amounts of red and blue light. The error bar chart between the valence and dependent variable data at a 95% trust level is shown in Fig. 5, in which the association can be observed more easily. A specific valence value is given a specific range of lighting color saturation hue, which is clearly observed at the minimum and maximum. Furthermore, the results show a slightly negative correlation between valence and blue light; in other words, the valence value is associated less with a greater amount of blue light.

4.2.4. Lighting color regulation support vector regression construction

The MANOVA results indicate that music emotions have a statistically significant influence on the lighting color, and the saturation can be seen as equivalent to the RGB blending result; thus, the saturation and the hue were taken as the simulation targets. Two SVRs with a Gaussian kernel function were used for each color parameter. A total of 988 samples



(a) (b)

Fig. 5. Error bar chart between the valence value and dependent variables. (a) Valence value and lighting color hue; (b) valence value and blue light. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

Table 5
Experiment music genre.

Genre	Classical	Country	Disco	Electro	Folk	Funk	Metal	Jazz	Latin
Data num.	41	1	17	40	6	9	9	22	13
Genre	Pop	R&B	Rap	Reggae	Rock	Soul	Soundtrack	Vocal Pop	World
Data num.	154	42	38	10	180	40	9	111	6

were randomly separated according to a 7:3 ratio for training and validation; that is, 691 samples were used for training and 297 samples for validation. As mentioned previously, half of the training samples were used to determine the parameters using a 5-fold CV grid search technique, and the root mean square error value was determined as an indicator. The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 82%. The results of the arousal value simulation reached 82.75% accuracy, and those of the valence value simulation reached 84.78%. The statistical correlation coefficients surpassed 71%. The results of the arousal value simulation reached 71.6% accuracy, and those of the valence value simulation reached 75.06%.

4.2.5. Lighting color emotion map

An input matrix representing the arousal and valence values was constructed across data elements ranging from 0 to 1, with 0.01 intervals. The lighting color regulation simulation SVR was used to simulate the output color and to draw a 100 million-pixel music emotion map. This emotion map can be used as a quick retrieval database for lighting color regulation application programs.

4.3. Lighting color regulation with music genre

In this section, the music genre as a factor of lighting color regulation is addressed, although there are no direct studies indicating an association between them. However, the feedback of the subjects in the lighting color regulation experiment indicated that some technicians correctly consider music genre as naturally generating a certain type of musical feeling and performed their lighting color regulation accordingly. Therefore, the relationship between music genre and lighting color regulation was analyzed. Eighteen types of music genres were obtained from the Musicovery Website, as shown in Table 5. The experimental results were then used to train an SVR to construct simulations with lighting regulation, with the addition of the music genre factor.

4.3.1. Experimental results analysis

To show the influences of music emotions on lighting regulation preferences, a MANOVA was conducted with hue and saturation as the within-participant factors. Moreover, the music genre factor was added. The same approach as that previously mentioned was applied, where the normalized data for music emotion values, including the arousal and valence values, were divided into 11 ranks, with 0.1 intervals. The results are shown in Table 6, where it can be seen that music genre has a statistically significant influence (Wilks' lambda $= 1.920$, $p = 0.000 < 0.001$). The interaction of music emotion and

Table 6
Experimental results for the multivariate analysis when the music genre factor is added.

Variables	Wilks Lambda	F test between dependent variables				
		Hue	Saturation	Red	Green	Blue
Genre	1.920***	2.015**	.830	1.955*	2.846***	2.131**
Arousal	2.277***	2.065**	5.968***	2.717**	1.104	1.801
Valence	1.106	2.173*	0.935	1.049	1.294	1.728
Genre *Arousal	.995	.927	.983	.730	.691	1.203
Genre *Valence	.980	.910	1.007*	.624	1.140	.938
Arousal*Valence	1.017	.932	.907	.941	.767	1.084
Genre * Arousal*Valence	0.923	.948	1.068	1.111	.941	1.134

*** $p < 0.001$
** $p < 0.01$
* $p < 0.05$

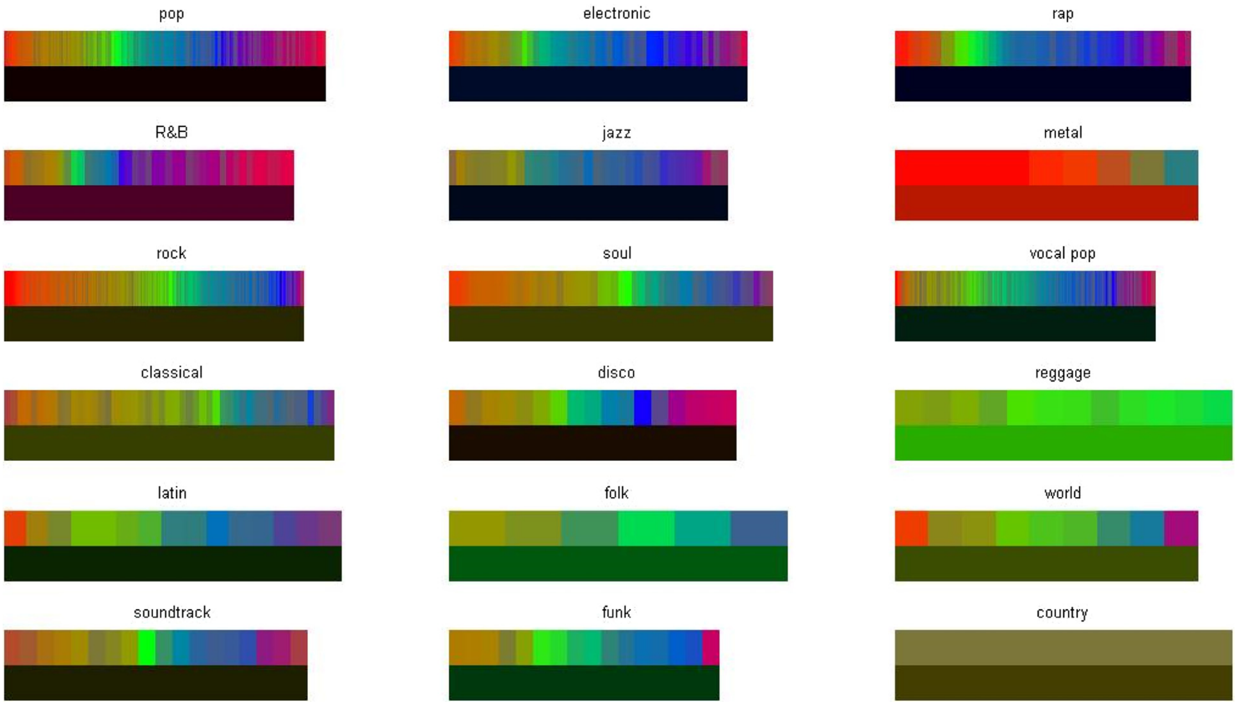


Fig. 6. All the regulated colors and the average color for each music genre.

music genre showed a weaker influence (Wilks' lambda $=0.923$, $p=0.408>0.05$). The results empirically show that music genre affects lighting color regulation preferences. The interactions of music emotion and music genre are not mutually independent, but the influence is not significant.

The factor analysis results for the dependent variables show that the weaker influence of music genre on hue was significant ($F=2.015$, $p=0.008<0.01$). This weaker influence on hue can be verified through the RGB lighting value test results ($F=1.955$, $p=0.01<0.05$; $F=2.846$, $p=0.000<0.001$; $F=2.131$, $p=0.004<0.01$). The music genre caused the subjects to execute a specific light color regulation of hue, and the weaker influence on the hue can be seen in the blending of specific amounts of red, blue, and green light. Specific types of music genre show a specific range of lighting color hues. For example, metal music is associated with a range of red and yellow, which can be observed in the significant relationship between the music genre and the amount of red light; electronic music, jazz, and rap are associated with a range of blue, which can be observed in the significant relationship between these genres and the amount of blue light; reggae, folk, funk, and Latin music (as influenced by the Latin American culture) are associated with a range of green, which can be observed in the significant relationship between these genres and the amount of green light. Fig. 6 shows all the regulated colors and the average color for each music genre in the experiment. It is easier to observe the color preference for each music genre, such as metal, electronic, jazz, rap, reggae, folk, funk, and Latin, individually. In contrast, pop music, rock, and soundtrack music include a wide variety of elements, and therefore show trends toward a wider variety of lighting colors.

Table 7
Comparison of support vector regression simulation performance with and without music genre.

Variables	Training result		Testing result	
	Hue	Saturation	Hue	Saturation
Music emotion	0.8275	0.8478	0.716	0.7506
Music emotion*Music genre	0.9855	0.9975	0.7597	0.8286
Improvement	19%	17.66%	6.1%	10.39%

4.3.2. Lighting color regulation adding the music genre factor for support vector regression building

The MANOVA results show that music emotion and music genre have statistically significant weaker influence on lighting color. Thus, music genre was added to the simulation input. As aforementioned, two SVRs with a Gaussian kernel function were used for each color parameter. A total of 988 samples were randomly separated according to a 7:3 ratio for training and validation; that is, 691 samples were used for training, and 297 samples for validation. As mentioned previously, half of the training samples were used to determine parameters using a 5-fold CV grid search technique, and the root mean square error value was determined as an indicator. The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 98%. The results of the arousal value simulation reached 98.54% accuracy and those of the valence value simulation reached 99.75%. The statistical correlation coefficients between the SVR simulation output using the test sample and the validation music emotion value results surpassed 75%. The results of the arousal value simulation reached 75.97% accuracy and those of the valence value simulation reached 82.86%. After adding the music genre factor, the effectiveness of the simulation was significantly improved, regardless of whether it was for the training or the validation stage. As shown in Table 7, the correlation coefficient of the hue simulation was enhanced by 19% in the training stage and by 6.1% in the testing stage; the correlation coefficient of the saturation simulation was enhanced by 17.66% in the training stage and by 10.39% in the testing stage. The significant enhancement also proved the significant influence of music genre on the lighting color regulation, which was the same as the results of the MANOVA analysis.

4.3.3. Lighting color emotion map for each music genre

Each music genre was constructed as mentioned above, where an input matrix represents the arousal value and valence value constructed across data elements in a range from 0 to 1, with 0.01 intervals. The lighting color regulation simulation SVR was used to simulate the output color and to draw a 10,000-pixel music emotion map of each music genre, as shown in Fig. 7. These emotion maps of 18 different music genres can be used as a quick retrieval database for lighting color regulation application programs.

5. Discussion of modes

This section discusses the instructions for the automatic lighting regulation system, and the operation of this program is explained, as well as the operation of the functions, the connection of hardware and software, and the operation of the lighting demonstration.

5.1. Automatic lighting regulation program structure

The structure of the automatic lighting regulation program is shown in Fig. 8. The graphic user interface (GUI) was built using JAVA and MATLAB was used for the core operation. If a chosen music file is being imported for the first time and has not been analyzed for producing light regulation sequence data, the system analyzes the music file. The audio signal data are extracted by the operational part of the program built using MATLAB. In the analysis stage, if the music genre factor has been considered and clearly identified, this music genre is applied to the light color regulation SVR simulation; if not, the system searches for the most similar music segment in the database and labels the music with the genre of that segment.

As the first step of the analysis process, the system detects the segments of the chosen music and separates the audio signal according to this segmentation. These signal segments are the indicators of the lighting brightness regulation, and each indicator extracts a music feature that reflects the specific music emotion and lighting color. The extracted features were mentioned previously in the description of the music emotion-featuring experiment using an SVM classifier. These features were used for music emotion recognition, as well as to find the most suitable lighting color regulation, whereas the lighting brightness was associated with the mean intensity value of each music segment. The music onset value was then added to the mean intensity value to obtain a flash effect. The lighting regulation sequence data that were produced were used to build the demonstration video, which was utilized for simulating the lighting regulation results during a series of image processing and time line fitting. After the MATLAB analysis process, the user is able to play the music and execute the lighting regulation by using the JAVA GUI. The program reads the lighting regulation sequence data, which were produced by the analysis, and plays the music and the lighting regulation simulation video synchronously. At the same time, the

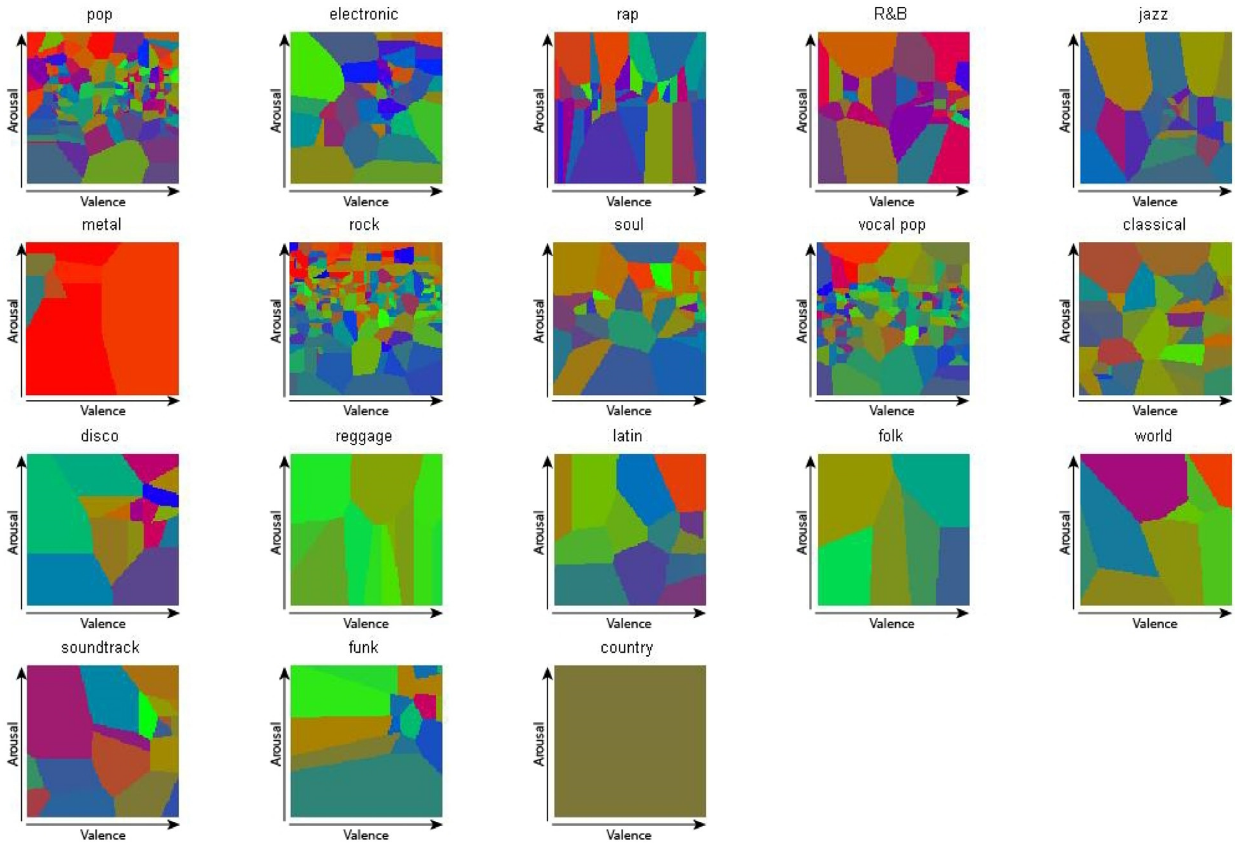


Fig. 7. Lighting color emotion map of each music genre.

Table 8
Multivariate analysis between music genre and music features.

Variable	Music genre							
Dependent variables	TC	TR	TF	FA	FC	FV	FF	FM
F test	2.397	8.94	2.086	6.495	6.922	2.324	2.0855	18.389
Distinctiveness (p)	0.001	0	0.045	0	0	0.0217	0.04525	0

lighting regulation sequence data are transferred to Freestyler programs and converted into DMX signals to control the LED stage lighting by MCSWE USB-DMX.

5.2. Music genre recognition

Music genre is to be distinguished from musical form and musical style. In practice, these terms are sometimes used interchangeably and are subject to different interpretations, and therefore, it is difficult to build a highly accurate music recognition classifier using a computer algorithm. Therefore, this program finds the music most similar to that of the user-selected music in the database as an indicator for determining the musical genre.

5.2.1. Music genres related to feature selecting

As aforementioned, PCA was used to reduce the dimensions of the feature space. A total of 2087 pieces of music collected from the Musicovery Website were analyzed according to music genre using a MONOVA test. According to the results, the features that were followed significantly outperformed ($p < 0.05$) music similarity comparative indicators. The factor analysis results for the dependent variables are shown in Table 8.

5.2.2. Correlation testing

The music genre-related features were used to calculate the correlation coefficients between the detected music and music clips in the database to find the top 10 most similar clips. Then, the Euclidean distance between these 10 music clips and the detected music was calculated for further identification. This approach is described in detail in the next section.

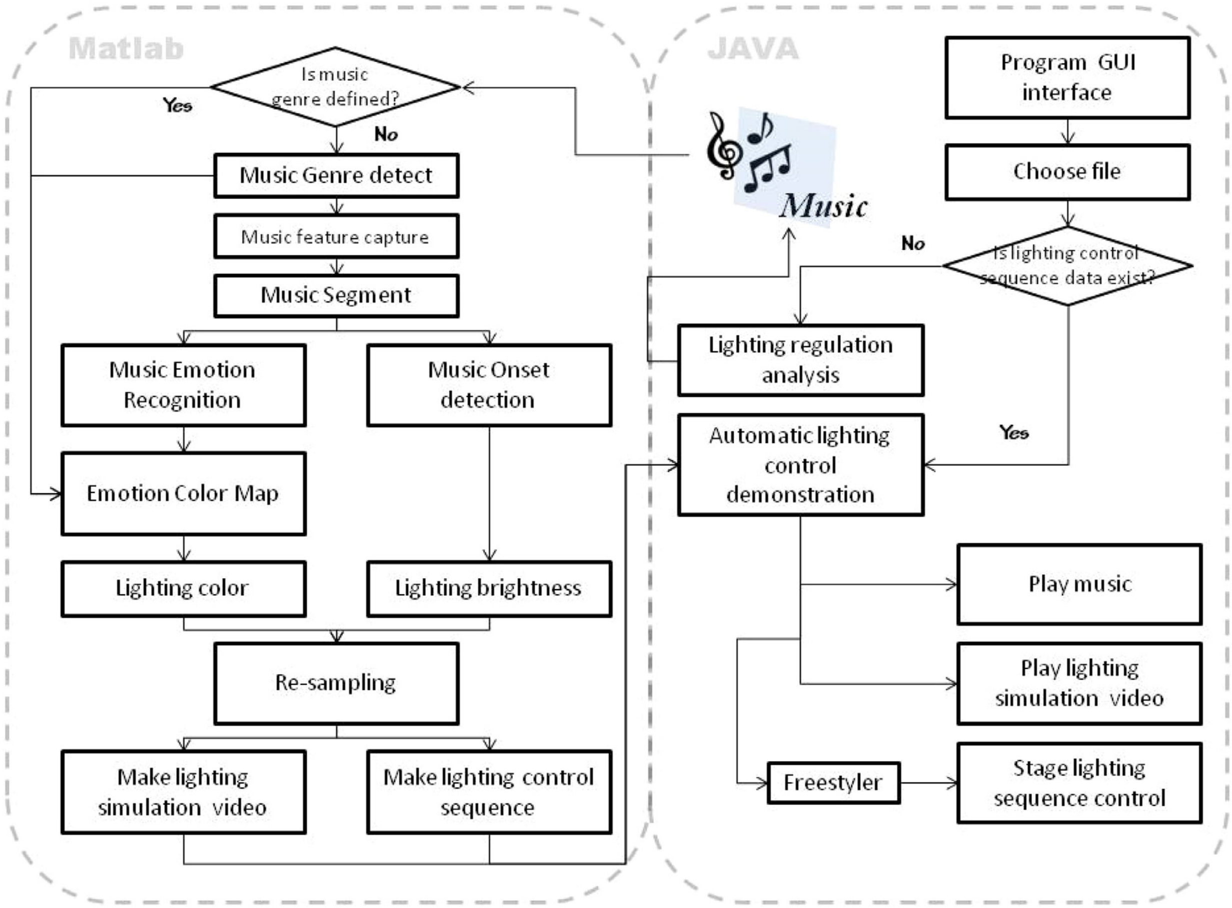


Fig. 8. Automatic lighting regulation program structure.

5.2.3. Similarity testing

The top 10 most similar music clips in the database selected using the correlation coefficient approach were used to calculate the Euclidean distance from the detected music:

$$E = \sqrt{\sum_{i=0}^{n} \left(q_{i,j} - p_{i,j}\right)^2} \tag{28}$$

where E is the calculated error value, q represents the features of the detected music data, and p represents the features of the music clips in the database. On basis of the minimum error value between the features, the music genre is determined according to the music genre that is most similar to that in the music clips.

5.3. Automatic music segment detection and lighting brightness regulation approach

The lighting brightness regulation is assessed according to the average intensity of each music segment and onset value. The program calculates the average intensity of each music segment, which is determined by the automatic music segment detection algorithm, and the music onset values are then added. The lighting brightness regulation sequence data are then normalized and converted to grayscale. For example, visualized analysis data of 2 Unlimited's song "No Limit" using this approach are shown in Fig. 9. The figure shows the intensity of the original audio intensity data, the threshold used for segmentation, and the lighting brightness regulation grayscale value.

5.4. Music emotion and lighting color regulation

This section describes the lighting color regulation approach. Music features were extracted from each of the music segments, reflecting the specific music emotion and lighting color. These features were used for music emotion recognition and to determine the most suitable lighting color regulation. The analysis data for 2 Unlimited's song "No Limit," for example, can be visualized, as shown in Fig. 10. If the music genre factor has been considered and the genre clearly identified, this music genre is applied to the light color regulation SVR simulation; if not, the system searches the database for the most

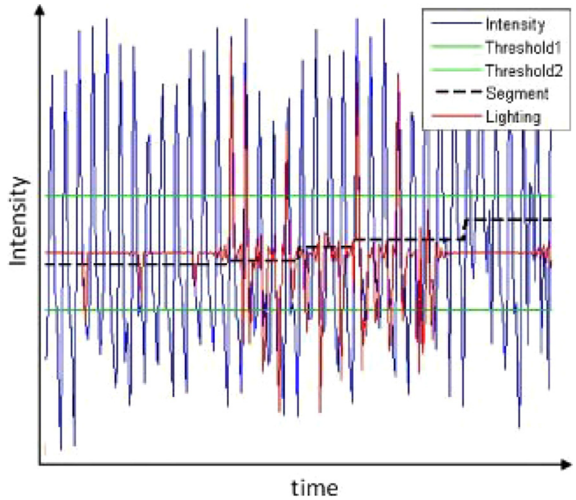


Fig. 9. Visualization of the music segmentation and lighting brightness regulation approach using 2 Unlimited's song "No Limit," as an example. The solid blue line represents the intensity of the original audio data, the solid green line represents the threshold for music segmentation, the solid red line represents the lighting brightness regulation data, and the black dotted line represents the music segment average intensity value. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

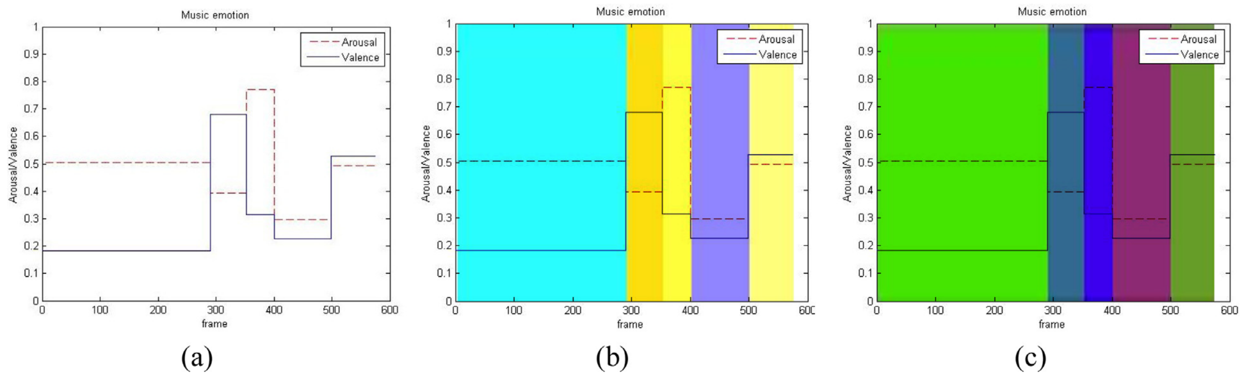


Fig. 10. Lighting color analysis result. (a) Music emotion recognition results, in which the solid blue line represents the valence value and the red dotted line represents the arousal value; (b) lighting color results without the music genre factor; (c) lighting color results with the music genre factor. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

similar music segment and labels it with this music genre. For example, the music genre classification of 2 Unlimited's song "No Limit" is electronic. Without considering the music genre factor, the lighting color regulation shows a pale blue color tone, which is more technological in character and suitable for electronic music.

5.5. Lighting regulation simulation video construction

The lighting brightness regulation sequence data were converted to a series of grayscale 200 pixel × 200 pixel pictures and blended with a band picture in the "darken blending" mode to simulate the light-off effect. Then, the same series of pictures was highlight blended with the blended image to simulate the lighting brightness and flash; finally, the lighting color regulation sequence data were also converted to a series of 200 pixel × 200 pixel pictures with color and overlay blended with the abovementioned blended image to simulate the lighting color effect. These images were re-sampled and converted to a 15 sample rate AVI file format video, as shown in Fig. 11. The algorithm of each blending mode is described in detail in the following.

5.5.1. Darken blending

Darken blending creates a pixel that retains the smallest components of the foreground and background pixels. If the foreground pixel has the components R1, G1, and B1 and the background has R2, G2, and B2, the resultant pixel is expressed as

$$[min(R_1, R_2), min(G_1, G_2), min(B_1, B_2)]. \tag{29}$$

(a) (b) (c)

Fig. 11. Lighting regulation simulation results. (a) Original image; (b) image after darkening and hard light blending; (c) image after overlay blending.

Table 9
System verification questionnaire results.

Song	(1)	(2)	(3)	(4)
Elton John & Leon Russell–In The Hands Of Angels	○	3.2	4.6	4.0
Orianthi–Bad News	○	4.6	4.6	4.2
O'Funk'Illo–O'Funk'IlloGroove	○	4.8	4.8	3.8
Kevin Kern–Sundial dreams	○	3.8	4.4	3.9
Average	○	4.1	4.6	3.98

5.5.2. Overlay blending and hard light blending

The parts of the top layer where the base layer is light become lighter, and the parts where the base layer is dark become darker. An overlay with the same picture resembles an S-curve:

$$f(a,b) = \begin{cases} 2ab, & if\ a < 0.5 \\ 1-2(1-a)(1-b), & otherwise \end{cases}, \tag{30}$$

where a is the base layer value, and b is the top layer value.

Hard light blending is equivalent to overlay blending, but with the bottom and top images swapped.

5.6. Case studies

To verify the automatic lighting regulation system developed in this study, 10 subjects evenly divided between males and females were invited to participate in the verification experiment. The subjects were asked to listen to the music and watch the stage lighting regulation effect, including a simulation video displayed on a screen and stage lighting on the stage. The test samples were four different pieces of music chosen from each of the four quadrants of the Thayer emotion plane. After each demonstration, the subjects were asked to fill a questionnaire consisting of four questions: (1) Does this automatic lighting regulation system help you enjoy the music and add extra entertainment? (2) How well does the automatic segmentation algorithm result fit the music? (3) How well does the automatic lighting regulation result fit the music genre? (4) How well does the automatic lighting regulation result fit the music emotion?

The questionnaire results are shown in Table 9, which shows that in the case of music with a harder rhythm, it was considerably easier for the subjects to feel the association between music and lighting, and songs with obvious segmentation were more likely to make subjects feel the association between music and lighting color changes, as in O'Funk' illo's "O'Funk' illoGroove," for example. Rock music also received high scores (for example, Orianthi's "Bad News"). However, music that includes contemplative feelings and monotonous rhythms obtained lower scores (for example, Elton John and Leon Russell's "In the Hands of Angels"). However, as compared with the aforementioned music, in terms of the lighting regulation effect scores of music genre and music emotion, the gap between the scores of the pieces of music was not very great, and therefore, this was still considered a good response. The scores of the automatic segmentation algorithm and lighting regulation for music genre in this verification were high. However, the lighting regulation scores for music emotion were lower. This may indicate that delicate emotions are not easy to observe.

6. Concluding remarks

An automatic lighting regulation methodology was developed in this study. A total of 2087 20-sec music clips were captured from the most popular songs on the Musicovery Website, and 21 musical features were extracted from these music clips and PCA was applied to reduce the dimensions of the feature space. These dimension-reduced features were used in a CV test based on the support vector machine (SVM) accuracy of classifying into the four quadrants of Thayer's model. The SVM classifier accuracy results show four groups of music features that are suitable for music arousal recognition; their classification accuracy was found to be 100%, and one group of music features was shown to be suitable for music valence recognition, having a classification accuracy of 80%. These feature combinations are considered the music features that best express emotions, using which the regression simulation approach can be implemented. An SVR approach was used as the regression simulation approach, where the parameters were selected using a 5-fold CV technique. The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 99%. The results of the arousal value simulation reached 99.96% and those of the valence value simulation reached 99.93% accuracy. The statistical correlation coefficients between the SVR simulation output using the testing sample and the validation music emotion value results surpassed 80%. The results of the arousal value simulation reached 89.93% and those of the valence value simulation reached 80.76% accuracy. The significant results indicate that these SVR simulations are suitable for music emotion recognition.

Five professional stage-lighting technicians were invited to execute lighting color regulation according to its properties for 988 music clips. A MANOVA method was used to investigate the relationships revealed by the experimental results. The results showed that arousal has a statistically significant influence (Wilks' lambda $= 2.277$, $p = 0.000 < 0.01$). Valence has a less significant influence, but also reached a statistical correlation (Wilks' lambda $= 1.106$, $p = 0.284 < 0.5$). The arousal value shows a positive correlation with the lighting color saturation. The arousal value shows a slightly negative correlation with the lighting color hue value, and the association can also be observed via the RGB value, which shows a positive correlation with the amount of red light and a negative correlation with the amount of blue light. In contrast, the arousal value is less associated with the blue color phase. A specific valence value shows a specific range of lighting color saturation hue and is clearly observed at both the minimum and the maximum levels. Furthermore, the results show a slightly negative correlation between valence and blue light. In other words, the valence value is associated less with relatively large amounts of blue light. The results of the MANOVA test showed that the relationship between music emotion and lighting color is suitable for training a computerized simulation.

The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 82%. The simulated accuracy of the arousal value reached 82.75%, while that of the valence value reached 84.78%. The statistical correlation coefficients between the SVR simulation output using the training sample and the trained music emotion value results surpassed 71%. The results of the arousal value simulation reached 71.6% and those of the valence value simulation reached 75.06% accuracy. Moreover, the association between music genre and lighting regulation preference using the music clips from the Musicovery Website, where a total of 18 different music genres are tagged, was investigated. The MANOVA results showed that music genre has a statistically significant influence on lighting regulation preference (Wilks' lambda $= 1.920$, $p = 0.000 < 0.01$). A specific kind of music genre shows a specific range of lighting color hue; for example, metal music is associated with a range of red and yellow, which can also be observed in the significant relationship between this genre and the amount of red light. The correlation coefficient of the simulated hue was enhanced by 19% in the training stage and 6.1% in the testing stage by including the genre factor and the correlation coefficient of the simulated saturation was enhanced by 17.66% in the training stage and 10.39% in the testing stage. This significant enhancement also proved to have a significant influence on the lighting color regulation, which concurred with the results of the MANOVA analysis.

Based on the above findings, an automatic lighting regulation program was developed in this study. To validate this system, 10 subjects were invited to participate in a verification experiment. The results showed that in the case of music with harder rhythms the subjects can more easily to feel the association between music and lighting, and that in the case of songs with obvious segmentation the subjects are more likely to feel the association between music and lighting color changes. The scores for the automatic segmentation algorithm and lighting regulation for music genre in this verification were high. However, the lighting regulation score for music emotion was lower, albeit still high. This may be because extremely delicate emotions are not easily observed. In general, the automatic lighting regulation approach can be used to enhance a performance's entertainment value by regulating the lighting system based on the music emotion and genre feeling of each music segment.

Acknowledgment

The authors are grateful to the Ministry of Science and Technology, R.O.C for supporting this research under grant 102-2221-E-006-188-MY3.

References

[1] J. Ajmera, I. McCowan, H. Bourlard, Speech/music segmentation using entropy and dynamism features in a HMM classification framework, Speech Commun. 40 (3) (2003) 351–363.
[2] J.L. Caivano, Color and Sound:Physical and Psychophysical Relations, Color Res. Appl. 19 (2) (1994) 126–133.
[3] P. Dhanalakshmi, S. Palanivel, V. Ramalingam, Classification of audio signals using SVM and RBFNN, Expert Syst. Appl. 36 (3) (2009) 6069–6075.
[4] R.O. Duda, P.E. Hart, D.G. Stork, Pattern Recognition, Wiley, New York, 2000.
[5] S. Esmaili, S. Krishnan, K. Raahemifar, Content based audio classification and retrieval using joint time–frequency analysis, IEEE International Conference on Acoustics, Speech and Signal Processing, 2004.
[6] P.R. Farnsworth, A study of the hevner adjective list, J. Aesthet. Art. Crit. 13 (1) (1954) 97–103.

[7] Y. Feng, Y. Zhuang, Y. Pan, Popular music retrieval by detecting mood, in, in: Proceedings of the 26th annual international ACM SIGIR conference on Research and development in information retrieval, Toronto, ON, Canada, 2003.
[8] C.V. Gustavo, G.C. Luis, C.M. Javier, D.M.G. José, S.O. Emilio, A.C. Luis, M. José, Robust support vector method for hyperspectral data classification and knowledge discovery, IEEE Trans. Geosci. Remote Sens 42 (7) (2004) 1530–1542.
[9] S.W. Hsiao, C.F. Hsu, K.W. Tang, A consultation and simulation system for product color planning based on interactive genetic algorithms, Color Res. Appl. 38 (5) (2013) 375–390.
[10] S.W. Hsiao, F.Y. Chiu, H.Y. Hsu, A computer-assisted colour selection system based on aesthetic measure for colour harmony and fuzzy logic theory, Color Res. Appl. 33 (5) (2008) 411–423.
[11] S.W. Hsiao, Fuzzy set theory on car-color design, Color Res. Appl. 19 (3) (1994) 202–213.
[12] D. Huron, The Ramp Archetype and the Maintenance of Passive Auditory Attention, Music Percept. 10 (1) (1992) 83–92.
[13] P.N. Juslin, Cue utilization in communication of emotion in music performance: relating performance to perception, J. Exper. Psychol. 26 (6) (2000) 1797–1812.
[14] G. Kreurz, Basic emotions in music, The 6th International Conference Music Perception Cognition, 2000.
[15] C. Laurier, O. Meyers, J. Serr‘a, M. Blech, P. Herrera, Music Mood Annotator Design and Integration, 2009 Seventh International Workshop on Content-Based Multimedia Indexing,Chania, 2009.
[16] C. Laurier, P. Herrera, Audio music mood classification using support vector machine, in: Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, Austria, 2007.
[17] C.H. Lee, J.L. Shih, K.M. Yu, H.S. Lin, Automatic Music Genre Classification Based on Modulation Spectral Analysis of Spectral and Cepstral Features, IEEE Trans. Multimed. 11 (4) (2009) 670–682.
[18] C.H. Li, An automatic method for selecting the parameter of the RBF kernel function to support vector machines, Geoscience and Remote Sensing Symposium (IGARSS), Honolulu, 2010.
[19] D.G. Li, I.K. Sethi, N. Dimitrova, T. McGee, Classification of general audio data for content-based retrieval, Pattern Recognit. Lett 22 (5) (2001) 533–544.
[20] T. Li, M. Ogihara, Content-based music similarity search and emotion detection, in: Int. Conf. Acoust., Speech, Signal Process, Toulouse, France, 2004, pp. 705–708.
[21] T. Li, M. Ogihara, Toward intelligent music information retrieval, IEEE Trans. Multimed. 8 (3) (2006) 564–574.
[22] T. Li, M. Ogihara, Detecting emotion in music, The 4th International Conference on Music Information Retrieval, Baltimore, MD, USA, 2003.
[23] E. Lindstrom, P.N. Juslin, Expressivity Comes From Within Your Soul: A Questionnaire Study of Student's Perception on Musical Expressivity, Res. Stud. Music Educ. 20 (1) (2003) 23–47.
[24] L. Lu, D. Liu, H.J. Zhang, Automatic mood detection and tracking of music audio signals, IEEE Trans. Audio Speech Lang. Process 14 (2006) 5–18.
[25] R.J. Mammone, X. Zhang, R.P. Ramachandran, Robust speaker recognition: a feature-based approach, IEEE Signal Process. Mag. 13 (1996) 1053–5888.
[26] M.I. Mandel, G.E. Poliner, D.P.W. Ellis, Support vector machine active learning for music retrieval, Multimed. Syst. 12 (1) (2006) 3–13.
[27] L.E. Marks, On colored-hearing Synesthesia: Crossmodal Translations of Sensory Dimension, Psychol. Bull 82 (3) (1975) 303–331.
[28] M.C. Miller, Color For Interior Architecture, John Wiley & Sons, New York, 1997.
[29] M. Nagamachi, Kansei engineering: a new ergonomic consumer-oriented technology for product development, Int. J. Ind. Ergon. 15 (1995) 3–11.
[30] D.C. Park, Classification of audio signals using Fuzzy c-Means with divergence-based Kernel, Pattern Recognit. Lett 30 (9) (2009) 794–798.
[31] R.W. Picard, G. Cosier, Affective intelligence - the missing link? BT Technol. J. 15 (4) (1997) 150–161.
[32] R.W. Pridmore, Music and Color:Relations in the Psychophysical Perspective, Color Res. Appl 17 (1) (1992) 57–61.
[33] L.G. Roberts, Machine perception of three-dimensional solids, Garland Pub. (1980).
[34] J.A. Russell, Affective space is bipolar, J. Personal. Soc. Psychol 37 (3) (1980) 1161–1178.
[35] R. Sebba, Structural correspondence between music and color, Color Res. Appl 16 (2) (1991) 81–88.
[36] C. Seo, K.Y. Lee, J. Lee, GMM based on local PCA for speaker identification, Electron. Lett 37 (24) (2001) 1486–1488.
[37] B. Shao, D.D. Wang, T. Li, M. Ogihara, Music Recommendation Based on Acoustic Features and User Access Patterns, IEEE Trans. Audio Speech Lang. Process 17 (8) (2009) 1602–1611.
[38] Y.Y. Shi, X. Zhu, H.G. Kim, K.W. Eom, A tempo feature via modulation spectrum analysis and its application to music emotion classification, IEEE International Conference on Multimedia and Expo, Toronto, Canada, 2006.
[39] A.J. Smola, B. Schölkopf, A tutorial on support vector regression, Statist. Comput. 14 (3) (2004) 199–222.
[40] D.P. Solomatine, D.L. Shrestha, RT AdaBoost, A boosting algorithm for regression problems, in: Proceedings if the IEEE International Joint Conference on Neural Network, 2004.
[41] M. Sordo, C. Laurier, O. Celma, Annotating music collections: How content-based similarity helps to propagate labels, in, in: Proceedings of the 8th International Conference on Music Information Retrieval, Vienna, Austria, 2007.
[42] F.E.H. Tay, L. Cao, Application of support vector machines in financial time series forecasting, Int. J. Manag 29 (2001) 309–317.
[43] R.E. Thayer, The Origin of Everyday Moods: Managing Energy, Tension, and Stress, Oxford University Press, Oxford, 1996.
[44] G. Tzanetakis, P. Cook, Musical genre classification of audio signals, IEEE Trans. Audio Speech Lang. Process 10 (5) (2002) 293–302.
[45] K. Umapathy, S. Krishnan, S. Jimaa, Multigroup classification of audio signals using time-frequency parameters, IEEE Trans. Multimed 7 (2) (2005) 308–315.
[46] G. Velius, Variants of cepstrum based speaker identity verification, IEEE International Conference on Acoustics, Speech and Signal Processing, New York , 1988.
[47] R.T. Warne, A Primer on Multivariate Analysis of Variance (MANOVA) for Behavioral Scientists. Pract, Assess., Res. Eval. 19 (17) (2014) Available online: http://pareonline.net/getvn.asp?v=19&n=17.
[48] C.S. Xu, N.C. Maddage, X. Shao, Automatic music classification and summarization, IEEE Trans. Audio Speech Lang. Process 13 (3) (2005) 441–450.
[49] Y.H. Yang, Y.C. Lin, Y.F. Su, H.H. Chen, A regression approach to music emotion recognition, IEEE Trans. Audio Speech Lang. Process 16 (2008) 448–457.
[50] X. Zhu, Y.Y. Shi, H.G. Kim, K.W. Eom, An integrated music recommendation system, IEEE Trans. Consum. Electron 52 (3) (2006) 917–925.