# Modeling of Performance Creative Evaluation Driven by Multimodal Affective Data

Yufeng Wu[1], Longfei Zhang[1]*, Gangyi Ding[1], Tong Xue[1], Fuquan Zhang[2]

[1] Key Laboratory of Digital Performance and Simulation Technology, Beijing Institute of Technology, Beijing,100081 (China)
[2] Fujian Provincial Key Laboratory of Information Processing and Intelligent Control Minjiang University (China)

Received 10 October 2020 | Accepted 2 July 2021 | Published 4 August 2021



## Abstract

Performance creative evaluation can be achieved through affective data, and the use of affective features to evaluate performance creative is a new research trend. This paper proposes a "Performance Creative—Multimodal Affective (PC-MulAff)" model based on the multimodal affective features for performance creative evaluation. The multimedia data acquisition equipment is used to collect the physiological data of the audience, including the multimodal affective data such as the facial expression, heart rate and eye movement. Calculate affective features of multimodal data combined with director annotation, and defined "Performance Creative—Affective Acceptance (PC-Acc)" based on multimodal affective features to evaluate the quality of performance creative. This paper verifies the PC-MulAff model on different performance data sets. The experimental results show that the PC-MulAff model shows high evaluation quality in different performance forms. In the creative evaluation of dance performance, the accuracy of the model is 7.44% and 13.95% higher than that of the single textual and single video evaluation.

## Keywords

Performance Creative Evaluation, Multimodal Affective Feature, Multimedia Acquisition, Data-driven, Affective Acceptance.

DOI: 10.9781/ijimai.2021.08.005

## I. Introduction

THE cultural industry is prospering, and the performing arts, as an important branch of the cultural industry, highlights the core aesthetic values Lee [1]. Performance evaluation research has very important academic and application value, which helps to promote people's cognition and exploration of performing arts. After a long period of development, performing arts have evolved into many different forms of performance, such as stage performance, dramatic performance, large-scale event performance (opening and closing ceremonies of sports games, etc.) and multi-media interactive performance. Performance creativity includes the director or choreographer's novel ideas and clever designs in stage design, content structure, audience interaction, and the use of multimedia technology. A good performance creative can not only improve the performance quality, but also go deep into the hearts of the audience and affect the emotional and aesthetic cognition of the audience. So, how do we determine whether the creative of a performance is success? At present, the performance creative evaluation is mainly guided by professional aesthetic experience. In this context, it is difficult to make a true and objective evaluation of the performance creative, and it even brings a negative impact on the performance itself. In the field of performing arts, performance creative evaluation has begun to show its research value and practical significance.

* Corresponding author.
E-mail address: longfeizhang@bit.edu.cn

Performance creativity evaluation has its particularity, which is mainly reflected in the three types of people involved in the performance: director, actor and audience. From the perspective of time and space, performances can be divided into linear performances and non-linear performances. Linear performances mainly refer to film, television, media and other linearly edited video content. Non-linear performances refer to performances dominated by live performances, such as stage performances and square performances. Performance creators can design performance content through stage design, scenery, costume props and other performance elements. The actors will be affected by the performance environment and audience feedback, and even the possibility of on-site improvisation may occur. This is also the essential difference from linear performance, and it is also the main elements that affect performance creative. The performance creative evaluation proposed in this paper is mainly for non-linear performance.

In recent years, with the improvement of computer hardware and computing power, powerful and automatic feature extraction capabilities can effectively determine better features, thereby improving the efficiency of the model recognition system. Significant progress has been made in the research of computer vision [2]-[4], speech recognition [5]-[6], and natural language processing [7]-[9]. For example, an image recognition system based on deep learning has been described by [2], which uses various image preprocessing algorithms to perform grayscale processing and banalization on training data to enhance the training data, and then uses the GoogLeNet model to train these preprocessed images and test its recognition result. The method proposed in this study provides a new idea for future medical image-

assisted diagnosis. An attention segmental recurrent neural network (ASRNN) that relies on a hierarchical attention neural semi-Markov conditional random fields (semi-CRF) model has been proposed by [7] for the task of sequence labeling.

The improvement of computing power has promoted the arrival of the era of sensing intelligence [10]. Digital audio-visual technology, video stage design, and interactive performance content provide performance creators with broad ideas and promote the diversified development of performance creative. Using digital and intelligent methods to evaluate performance creative is the current main research direction. Performance evaluation uses these digital methods to promote the realization of new evaluation ideas and methods. Cross-topic research integrating machine learning, affective computing, artificial intelligence and other fields is gradually forming. In the complex performance environment, many factors such as performance behavior, stage setting, viewing angle, audience perception, etc. have had a great impact on performance evaluation. Only considering a single modal feature as a research object cannot meet the needs of performance evaluation. The use of multi-modal data features for modeling to solve complex systemic problems such as performance creative has become the focus of attention of researchers.

In order to solve the above-mentioned problems in the evaluation of performance creative, this article starts from the perspective of intelligent multimedia analysis, and integrates the physiological signal data such as audience evaluation text, audience facial expressions, audience heart rate, and eye movement data, to extract multimodal affective features and to evaluate performance creative. This paper argues that the core issue of performance creativity evaluation is to obtain the true emotions of performance works and the true evaluation of performance content, which poses new challenges to the methods and means of performance creativity evaluation.

The contributions of this work are presented as follows:

- This article proposes a "Performance Creative— Affective Acceptance (PC-Acc)" to evaluate the quality of performance creative, trains and builds a "Performance Creative—Multimodal Affective (PC-MulAff)" model, which can evaluate creative for different performance forms.
- This paper proposes a new "Performance Creativity-Multimodal Evaluation Data Set", which is composed of performance video data, audience evaluation text data and audience physiological data, which makes up for the problem of insufficient description of affective features by a single data type.
- Based on the establishment of a multi-modal evaluation data set, the correlation analysis between the audience's multi-modal physiological signals and the emotional dimension of the "Director Label" is realized. This work plays a decisive role in the evaluation of performance creativity.

The structure of this document is as follows: section II reviews the related works, section III details the methods proposed in our research, section IV presents the experiment results, section V details the discussion, and section VI conclusions and look forward to the future work.

## II. Related Works

### A. Performance Creative Evaluation (PCE)

In 1994, Abbé Decarroux put forward the importance of performing arts quality assessment and the challenges of analyzing it [11]. Frieder [12] proposed the concept of introducing computer technology into the field of fine arts, transforming the creative process into a computable process as early as 2007. **Evaluation from the perspective of the creative process:** Yamada [13] proposed the use of weighted sum of wavelet coefficients to generate creative dance movements, although implemented on the algorithm, it is difficult to describe an action that does not exist in words, and it is difficult to empathize with the result of this creative at the emotional level. Chang [14] proposed a calculation on integrating creative operation into music creation process. Although calculations and extensions have been made on musical characteristics such as timbre, there is still the problem of how to evaluate the calculated results. Gove [15] proposed that the coordination and precise interaction of players in musical performance should be regarded as the main constituent parameters of the creative process.

**Evaluation from the perspective of the creative methods:** Cabral [16] used interactive digital media to visualize the ideas of choreography, this paper proposed to use video annotator to annotate, analyze and evaluate the creative process of dance performance. However, this kind of interactive method will decompose the actors' attention and produce irreparable interference to the performance taking place. Therefore, the accuracy of the evaluation conclusion obtained from this method is difficult to be guaranteed. Cisneros [17] put forward the creative potential of VR and how to provide creative process for choreographers and dancers. This kind of creative method relying on new interactive tools may have a certain impact on the surface form, but this article believed that this has little ability to solve the structural problems of artistic creative and the problem of creative evaluation.

**Evaluation from the perspective of the creative cognition:** Christensen [18] analyzed the creative process and the creative evaluation process from the cognitive dimension. Pegah [19] proposed and evaluated a new method for stimulating creative in a common design system. The semantic similarity of creative expression and the structural similarity of hand-drawn sketch were calculated respectively. The creative level of the volunteers ware divided by the three intervals of high similarity, existing similarity and low similarity, so as to stimulate and transform the creativity of the volunteers, but the thresholds of these three intervals were not quantified and set. Tiffany [20] and Richardson [21] put forward the idea of solving the creative strategy and evaluation problems from the perspective of cognition, but it also needs to solve the quantitative analysis and modeling of the attention of creative works, which cannot be well realized in a short time.

**Evaluation from the perspective of the creative modeling:** Kyu [22] put forward the Computational Thinking Pattern Analysis (CTPA). It used CTPA to calculate differences in three different learning conditions as an index for measuring creativity. The differences were calculated by CTPA, and the creativity itself was mapped to 9 kinds of high-dimensional cosine space as divergent elements, which were used as indicators to measure the creativity. However, these indicators were only tested on the creative ideas of the two game designs, and there was no reasonable explanation for the on-site or off-site creative behaviors. Ajit [23] constructed a novel computing model combining visual and conceptual features to quantitatively represent and analyze feedback on creativity. Jonas [24] proposed a strategy of evaluating ideas through crowdsourcing feedback. Although directors can obtain feedback information from online audience groups, the behavior of group feedback is easy to have a guiding influence on individual subjective judgment.

**Evaluation from the perspective of the management:** Abfalter [25] recognized the importance of leading creative teams and creative environments in the performing arts context. The influence of leadership and organizational structure on performance and evaluation is elaborated, which can be regarded as a new perspective to solve problems, but there is no practical verification.

Above, we can see that the researchers to try various **perspectives (creative process, creative methods**, **creative cognition**, **creative modeling**, **management)** to evaluate the performance creative. However, the existing methods of performance creative evaluation have subjective limitations, and the feedback strategy of evaluation is complicated, which have different effects on the accuracy and effectiveness of the digital evaluation of creative performance.

### *B. Multimedia Computing in PCE*

From the perspective of sentiment analysis and machine learning, the core problem of performance evaluation is to solve audience emotion detection and how to build an evaluation model.

#### *1. Affective Computing Based on Semantic Features*

It is a common research method to obtain and analyze audience emotion through semantic network. The Lee [26] proposed a solution to build big data for dance performance and develop a big data creative analysis model system suitable for dance research, 25 kinds of high-frequency words were classified according to dance theme, characters and themes. Min [27] analyzed 20,776 dance research data texts accumulated in South Korea from 1958 to 2016 by using text mining and degree concentration based on semantic network, and obtained the core creative ideas and emotional themes of Korean dance performances in different historical periods. Ryeon [28] determined the determinants of Korean dance performance through tree analysis based on data mining. Choi [29] aimed to systematically investigate the knowledge structure of modern dance research through text mining, and establish the emotional cognitive system of modern dance performance in the future. Kyung [30] Choihyojin [31]and Kimhayeon [32] analyzed the thematic emotional trend of Korean dance performances in the past 20 to 30 years through text mining. Zhou [33] proposed a method to integrate acoustic features and text features to calculate emotions from large-scale Internet voice data. Liang [34] proposed a UAM proposed a universal affective Model (UAM) to calculate the potential emotion of short text in social media. Hung [35] introduced and discussed the classification methods for MuSe-Topic sub-challenges, as well as the data and results. For topic classification, Hung integrated two language models, ALBERT and RoBERTa, to predict 10 topic categories. In order to classify valence and arousal, SVM and random forest are combined with feature selection to enhance performance.

Although this method can quickly establish the semantic network of the text, it cannot guarantee whether the semantic of the text truly reflects the psychological and emotional of the audience. This method relies heavily on the performance cognition of the audience.

#### *2. Affective Computing Based on Physiological Signal Features*

Sowden [36] analyzed the influence of different emotions on creative. Corness [37] described the research of extracting the audience's experience in the performance scene to evaluate the audience's empathy experience in the performance process. Coursaris [38] developed and tested a cognitive model of cognitive user satisfaction with high explanatory power, which was used to assess the direct impact of cognitive and emotional dimensions on satisfaction. Altuwairqi [39] proposed an emotional model and a new process to test students' engagement in learning, six core silver factors affecting the model were analyzed by statistical methods (strong, high, medium, low, disengagement). Rahdari [40] introduced a multi-modal emotion recognition system based on two different modalities, namely emotional speech and facial expressions. For emotional speech, common low-level descriptors including prosodic and spectral audio features are extracted. Loprinzi [41] proposed a cognitive emotional model to assess physical activity based on experience, to evaluate the exercise habit and intensity of adults. The emotional changes can be seen after the acute exercise, it is related on the exercise intensity and exercise cycle of the volunteers. At the same time, personal health will also have an impact on emotional changes, which shows that this measurement method has inevitable flaws in its universality. A method of using emotional calculation to evaluate football players is proposed by Liu [42], which combines the text information of the post-match report and emotional calculation to measure the performance quality of the players. The author established a player performance evaluation model based on LSTM, However, this method still has some problems to be solved. For example, it is difficult to achieve accurate statistics and quantification of specific behaviors in the game, which has a greater impact on the collection and evaluation of key information. Wei [43] proposed a new method for extracting emotional features from facial expression images using multi-modal strategies is proposed. The basic idea is to combine low-level experience features and high-level self-learning features into multi-modal features. The convolutional neural network was used to extract the two-dimensional coordinates of the key points of the face as low-level experience features, and the convolutional neural network is used to extract high-level self-learning features. In order to reduce the free parameters of CNNs, small filters are used in all convolutional layers. Chen [44] used sLORETA to analyze the significant difference between the active source area and frequency band of the EEG reconstruction source based on emotion, and selected 26 Brodmann regions as regions of interest (ROI). On this basis, the support vector machine was used to extract the time-frequency domain features of 6 important activity regions and frequency bands, and to classify different emotions. Choi [45] proposed an emotional response generation model based on emotional feature extraction is proposed. Deepika [46] identified three bases for speech emotion recognition system: database, feature extraction and various classification methods. The performance of speech emotion recognition system was discussed. Features were divided into basic, prosodic and spectral features. Wei [47] proposed a new method for extracting emotional features from facial expression images using multi-modal strategies. The basic idea was to combine low-level experience features and high-level self-learning features into multi-modal features. The convolutional neural network was used to extract the two-dimensional coordinates of the key points of the face as low-level experience features, and the convolutional neural network was used to extract high-level self-learning features.

It can be seen that although researchers have proposed various affective models, there are still many problems in the data collection and input of the models. It is a relatively superior research method to establish an emotional evaluation mechanism by obtaining the physiological information and cognitive status of the audience. Usually this method is called implicit sentiment measurement. Radbourne [48] proposed to monitor the emotional experience of the audience in live performances, and emphasized the importance of interaction firstly. Radbourne [49] proposed that the ability to stimulate the audience's emotions is the key to accurate performance evaluation, and the role of the audience in the performance is gradually changing. In 2011, Latulipe [50] used GSR to monitor the emotional arousal of 6 spectators in dance performances. The research divided the performance attributes into two categories: LH scale and ER scale, which represent the audience's degree of affection for the performance and the degree of arousal of the audience by the performance content. The results showed that the GSR value is positively correlated with the degree of ER arousal. Wang [51] used GSR signals to monitor the emotional state of 15 volunteers in live performances. The research conducted a cluster analysis on the audiences through the GSR values, and found that the data of 10 volunteers were closely related. Martella [52] analyzed the audience's feedback in live performances, and used a three-axis accelerometer to calculate the acceleration of the audience's body movements, which integrated complex emotional experiences

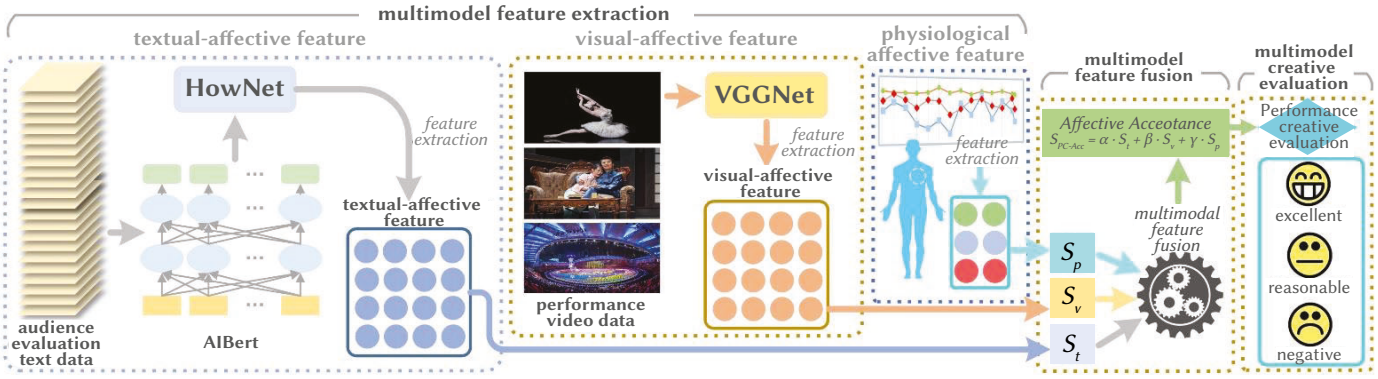


Fig. 1. Research framework.

such as "Enjoyment" or "Immersion" for quantification. By calculating the dynamic changes of acceleration, the research has reached nearly 90% accuracy in predicting whether the audience is in a "enjoying state". Psychological research has confirmed that the relationship between emotion and EEG signals has an important impact on human cognitive processes [53] [54].

It has confirmed that when users watch highly aroused content, the EEG coherence between the hemispheres has increased significantly [55]. The interaction of paired EEG electrode amplitudes has been confirmed to be correlated with positive arousal in emotional states [56]. According to the four-quadrant theory of emotions proposed by Koelstra [57], 32 performance emotion-inducing materials have been divided into "high arousal high valence" (HAHV), "high arousal low valence" (HALV), and "low arousal high valence" (LAHV) and "Low Excitation Low Valence" (LALV) four emotional level categories. Pinto [58] processed electrocardiogram, electromyography and dermal electrical activity to find a physiological model of emotion. Using samples of 55 healthy subjects, Pinto used single-peak and multi-peak methods to analyze which signals or combinations of signals can better describe emotional responses.

Zhang [59] proposed a multi-modal emotion recognition method using deep autoencoders for facial expressions and EEG interactions. The decision tree is used as the target feature selection method. Then, based on the facial expression features recognized by the sparse representation, the solution vector coefficients are analyzed to determine the facial expression category of the test sample. After that, the bimodal depth autoencoder was used to fuse EEG signals and facial expression signals.

We can clearly see that the use of human physiological data characteristics to calculate audience emotions, thereby realizing the evaluation of performing arts, has become a new research focus. Based on the affective computing method of physiological signals, this paper designed and collected the physiological signal data of the audience while watching the performance. For the first time, the method of "Director Label" was used to label the performance videos and physiological signals. Based on the above mentioned, "Performance Creative - Multimodal Evaluation Dataset" was developed for performance evaluation.

## III. Method

### A. Research Framework

The architecture of the PC-MulAff model is proposed in this paper, as in Fig. 1. The model consists of three parts: multimodal feature extraction module, multimodal feature fusion and multimodal creative evaluation. The multimodal feature extraction includes textual-affective feature, visual-affective feature and physiological affective feature unit. First of all, the audience evaluation text data in textual-affective feature unit get score $S_t$, and then, the performance video data get score $S_v$, and similarly the physiological data get score $S_p$. Finally, the $S_{PC-Acc}$ of was calculated, and the performance creativity was evaluated according to the value of $S_{PC\text{-}Acc}$.

### B. Multimodal Feature Extraction

This paper extracts the affective features of the audience and the visual affective features of the performance video, and conducts analysis and evaluation of performance creative through the quantified scores after the fusion of multimodal features.

#### 1. Textual-Affective Feature Extraction

In the depth of the traditional learning method, usually using Word2Vec extracted feature, such as the Skip-gram, Continuous Bag of Words (CBOW), but they catch the embedded part of the training, this part of the parameter is less, if continue to downstream text processing tasks you will need to add a lot of parameters, and from the training, increase a lot of training data and the training time, and Word2Vec emotional analysis of the context is limited by the length of the context, with the result of the classification of emotional impact. In order to reduce training data to complete the downstream tasks of natural language processing, researchers began to learn the embedding of general text through a large corpus: two-way LSTM M. Peters [60] combined embedding for forward and backward propagation, and BERT using Transformer model for two-way encoding, decoding and pre-training A.Radford [61] J. Devlin [62]. AlBert Lan [63], which adopted the full-network pre-training algorithm, adopted the parameter sharing mechanism, which could share most of the parameters with subsequent processing tasks, not only saving calculation time, but also avoiding the problem of limited context length. Therefore, in this study, AlBert is selected to extract the emotional features of the text.

$$Attention(Q,\ K,\ V) = \text{softmax}\left(\frac{QK^{\mathrm{T}}}{\sqrt{d_k}}\right)V \tag{1}$$

The core algorithm of AlBert's feature extraction is Transformer. Transformer USES multi-head self-attention mechanism. It projects H different Queries(Q), Keys(K) and Values(V) respectively, and the mapped dimensions are ALL $D_k$ and $D_v$ as the Equation.1 Ashish [64].

#### 2. Visual-Affective Feature Extraction

In the classification of visual emotions, the description of features is particularly important. A new spatial representation is proposed to solve the limitation of image compression domain, which provides a fast quasi-spatial transformation and effectively integrates the histograph-based method to solve the shortcomings of image enhancement in compression domain [65]. This method not only

achieves excellent visual quality, but also provides new possibilities in the acquisition of image features. Zhao [66] proposed Principles of Arts-based emotion features, and Xu [67] proposed three emotions for deep convolutional neural network to predict images: Positive emotion, negative emotion and neutral emotion, Jou [68] cross residual neural network is proposed for image emotion classification, Gao [69] visual attention and circulation is put forward combining the neural network to capture the key content, Zha [70] put forward the methods of extracting video features of visual attention, combined with the research in recent years, we will introduce the concept of visual attention to enhance the visual characteristics of expression ability, we will VGGNet as visual emotional feature extraction model as the Equation.2 Simonyan [71]. After obtaining the feature representation we input it into Softmax's full connection layer for visual emotion classification.

$$F_I = CNN_{VGGNet}(I) \tag{2}$$

### 3. Physiological-Affective Feature Extraction

We refer to Zhang [59] physiological feature extraction method to collect and calculate the physiological data of the audience. Instead of a single mode, this method uses a fusion dual mode depth autoencoder (BDAE) to integrate EEG and facial expression data to obtain an emotional model. In the process of multi-modal emotion recognition, Zhang adopts Restricted Boltzmann Machine (RBM) model. All collected data correspond to the visible layer of the model, and extracted features correspond to the hidden layer of the model. There is no connection between the nodes in the layer but the edge of the layer has. The variables $v \in \{0, 1\}^M$ in the visible layer and $h \in \{0, 1\}^N$ in the hidden layer are defined as follows:

$$E(v, h; \theta) = -\sum_{i=1}^{M}\sum_{j=1}^{N} W_{i,j}\, v_i h_j - \sum_{i=1}^{M} b_i v_i - \sum_{j=1}^{N} a_j h_j \tag{3}$$

## C. Multimodal Feature Fusion

### 1. Score of Textual-Affective Feature

HowNet is mainly divided into Chinese and English parts. There are 3730 Chinese positive evaluation words, 3116 Chinese negative evaluation words, 836 Chinese positive emotion words and 1254 Chinese negative emotion words. There are 3594 positive evaluation words, 3563 positive evaluation words, 769 positive emotion words and 1011 negative emotion words in English. We define the emotional characteristic score of the text as $S_t$, as the Equation.4.

$$S_t = \sum_{i}^{n} HowNet\,(T_i) \tag{4}$$

Where, n is the number of affective words, $T_i$ is the score value of the $i$ th word in the affective word set in HoeNet, when $S_t > 1$ is positive affective, and when $S_t < -1$ is negative affective.

### 2. Score of Visual-Affective Feature

After extracting the affective features through VGGNet visual affective feature extraction model, we obtained the affective classification result $S_v$ in the Softmax classifier, as the Equation.5. Where, $V_{-1}$represents the probability of negative affective, $V_0$ represents the probability of neutral affective, and $V_1$ represents the probability of positive affective.

$$S_v = \begin{cases} V_{-1} \\ V_0 \\ V_1 \end{cases} \tag{5}$$

### 3. Multimodal Feature Fusion

After obtaining the textual—affective feature score $S_t$, the visual—affective feature score $S_v$, and the physiological—affective feature score $S_p$. The weighted sum method is used to fuse the three features to obtain the final emotional score $S_{PC-Acc}$, which is defined as the affective acceptance, as the Equation.6.

$$S_{PC-\mathrm{Acc}} = \alpha \cdot S_t + \beta \cdot S_v + \gamma S_p \quad (\alpha + \beta + \gamma = 1) \\ \alpha \in [0,1]\ \beta \in [0,1]\ \gamma \in [0,1] \tag{6}$$

In this paper, principal component analysis (PCA) is used to determine the weights of α, β and γ. When $S_{PC-Acc} > 0$, it means that the audience acceptance is positive, so the performance creative is excellent. When $S_{PC-Acc} < 0$, it means that the audience acceptance is negative, so the performance creative is a failure. When $S_{PC-Acc} = 0$, it means that the audience acceptance is neutral, so the performance creative is reasonable.

## IV. Result

This paper evaluates the creative of three different performance forms and verifies the validity and rationality of the PC-MulAff model. In this paper, dance performance, drama performance and square performance were selected for the experiment. All the experiments were carried out in the Linux environment on a PC computer with core I7 processor and 512GBRAM. All methods are implemented in Python.

## A. Data Description

Lisetti [72] used the audience's physiological signals to recognize the emotion of the movie and divide the emotion into six discrete categories. Sun Kai [73] used the IMDB scoring system based on bayesian statistical algorithm to select the movie emotional content data set and establish the movie emotional space model. MAHNOB HCI [74] movie emotion database was established. Currently there is still a lack of emotional data sets for performances. Our research builds performance evaluation data set with emotional features based on the audience's evaluation text and physiological signals annotated in three common representative performance types (dance performances, drama performances, and square performances).

This study recruited 50 volunteers (25 men and women), age range 22-45 years old, their professional background is computer science (39%), digital media art (46%), dance performance (15%). EEG signal acquisition uses the whole brain induction head-mounted device Emotiv, which consists of 14 channel sensors and 2 bipolar reference electrodes, samples at a rate of 128 Hz, and is directly connected to the computer via Bluetooth. It has better wearing comfort for the audience, which is conducive to the collection of EEG signals. The data collection of the audience's facial expressions uses a conventional high-definition camera with an image resolution of 1920*1080. Considering that the use of wearable eye tracking devices will adversely affect the audience's viewing experience and may cause noise to the EEG signal. Therefore, we add a single camera for the collection of eye movement data and the collection of facial expressions. Eye movement frequency mainly includes eye tracking and saccade. Because eye tracking can clearly describe the audience's interest points in the performance, the saccade movement can well reflect the audience's continuous attention. These two key parameters provide important for the subsequent "Director Label". While the audience is watching the performance videos, we use a smart bracelet to collect the audience's heart rate. The device uses a PPG heart rate sensor, a three-axis acceleration sensor and a three-axis gyroscope, which can more accurately sample and record the audience's heart rate. In this article, we set the heart rate collection interval with 1 minute, and the heart rate detection can assist in verifying the audience's excitement. Each data sample includes 1

piece of performance video, 50 pieces of audience evaluation text and corresponding length of audience physiological signal data (facial expression, EEG, heart rate and eye movement frequency). The form of the data set as in Fig. 2.

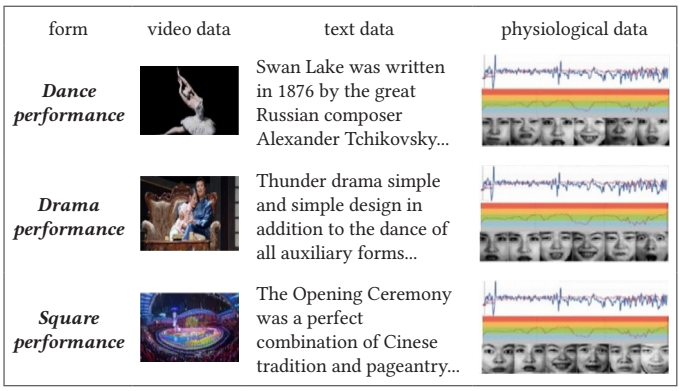


Fig. 2. Performance creative evaluation data set.

The performance creativity evaluation data set including 215 Chinese and Foreign famous dance works, 182 drama works and 213 large-scale square performances (such as opening and closing ceremony performances). The total amount of data is 610 pieces of performance video data. We have performed manual editing for each performance content to remove noise images that are not related to the performance content, and the time for intercepting the performance content is controlled within 30-50 minutes. 30,500 pieces of audience evaluation data are collected. We divide the training set and the test set in a ratio of 7:3, (see Table I).

TABLE I. Experimental Data Set

| Dataset | Mode | Amount | TrainSet | TestSet |
|---|---|---|---|---|
| **Dance** | performance video | 215 | 150 | 65 |
| | evaluation text | 10750 | 7500 | 3250 |
| **Drama** | performance video | 182 | 127 | 55 |
| | evaluation text | 9100 | 6350 | 2750 |
| **Square** | performance video | 213 | 149 | 64 |
| | evaluation text | 10650 | 7450 | 3200 |
| **Total** | performance video | 610 | 426 | 184 |
| | evaluation text | 30500 | 21300 | 9200 |

## B. Results Analysis

### 1. Comparative Experiment of Singlemodal and Multimodal Affective Features

This article uses a supervised model training method to label the training set and the test set. In view of the professionalism and particularity of the performance data in this article, the research uses the "director label" method on the data set to achieve the most effective training result. Based on the creative intention and aesthetic experience of the performance work, the director defined and marked the high-point and low-point of the performance creative. Among them, high-point corresponds to positive and wonderful ideas, and low-point corresponds to negative failed creative, the others corresponds to general reasonable creative, and the same annotation mode is used for the same text data set evaluated by the audience. The director label as in Fig. 3.

In this paper, the collected physiological data is associated with the director's annotations to form a complete annotated multimodal emotional data set. Annotation system for multimodal affective data sets as in Fig. 4. This part of the experiment compares the creative evaluation methods of singlemodal and multimodal features. First, we only use a single textual feature for creative evaluation. The accuracy of the creative evaluation model training results is shown in Fig. 5.

The x-axis in the figure represents the number of the test set, and the y-axis represents the accuracy rate. It can be seen from the figure that the accuracy rate obtained by the evaluation method of a single textual feature can improve faster. It can be seen from the figure that the accuracy rate obtained by the evaluation method of a single text feature can reach more than 80.56%, indicating that the audience's evaluation quality is relatively objective, but the degree of the audience's emotional fluctuations in the creative is relatively stable, which reflects the evaluation has a greater impact on the audience's artistic background and aesthetic experience. Then, we use a single visual feature for creative evaluation, and the accuracy of the creative evaluation model training results in Fig. 6.

It can be seen from the figure that the accuracy rate obtained by the evaluation method of a single visual feature improve slowly. It can be seen that the evaluation method of visual features has great emotional fluctuations, and the visual effects including the change of light and shade, the richness of color and whether the picture is in a wonderful moment have a greater impact on the evaluation results. Finally, we adopt the creative evaluation method of multimodal feature fusion, and the accuracy of the creative evaluation model training results in Fig. 7.

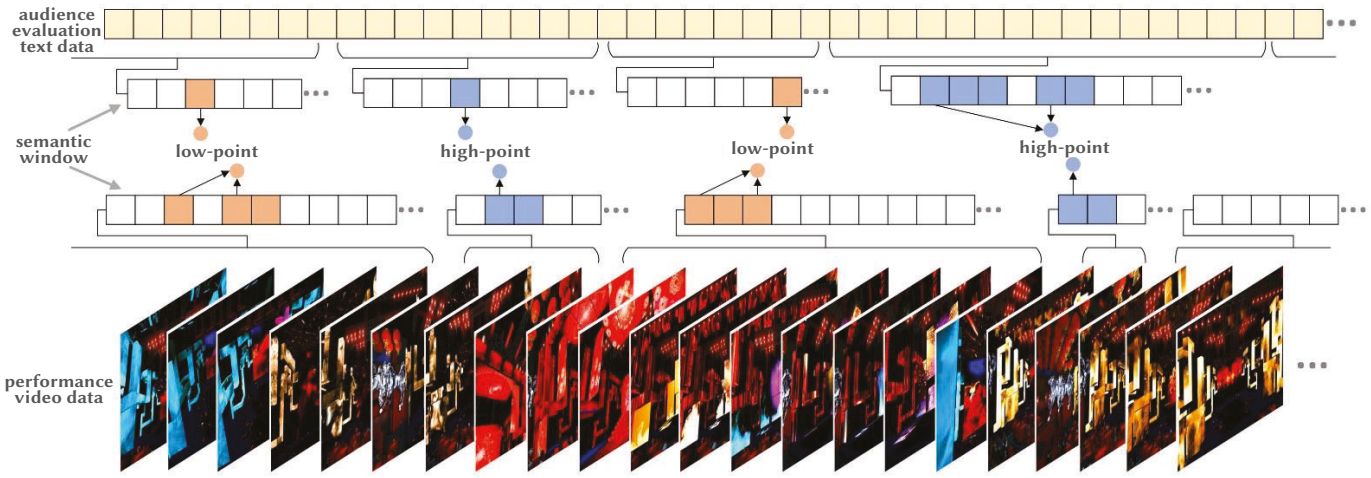


Fig. 3. Director label.

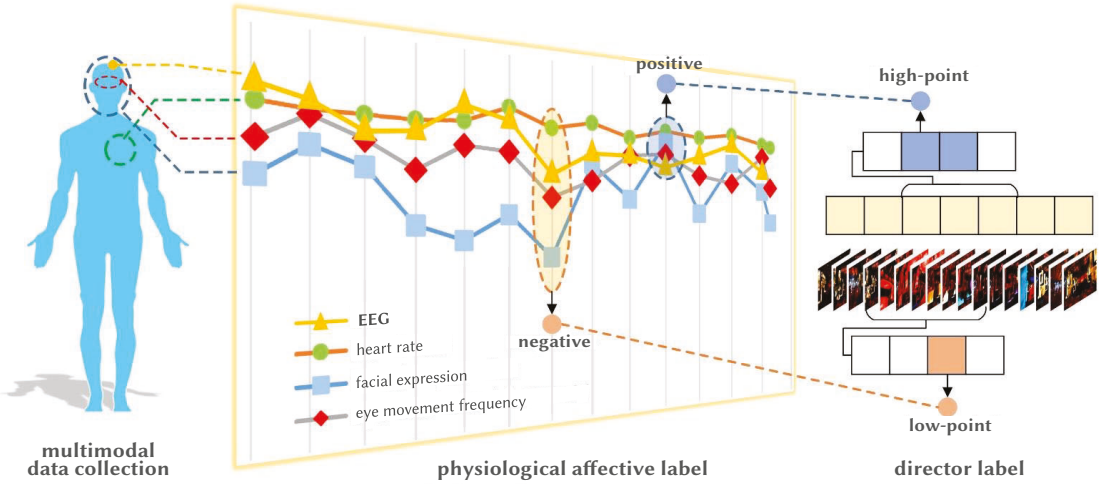


Fig. 4. Annotation system for multimodal affective data.

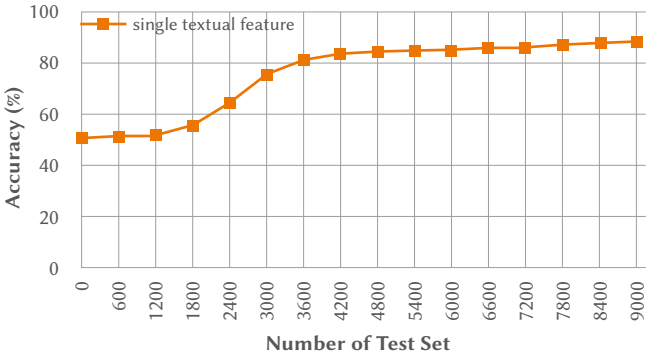


Fig. 5. Single mode evaluation accuracy in single textual feature.

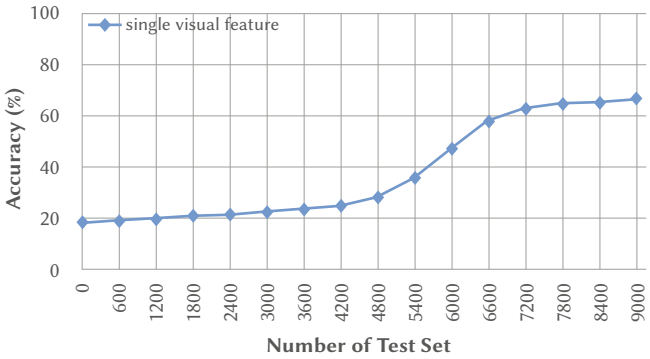


Fig. 6. Single mode evaluation accuracy in single visual feature.

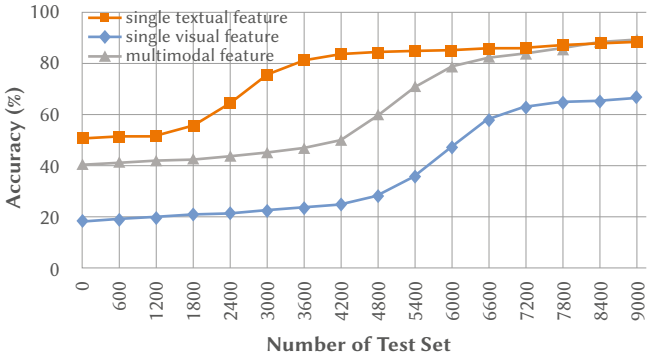


Fig. 7. Comparison of evaluation accuracy between single and multimodal.

From the figure, we can see that the multi-modal method makes up for the shortcomings of the single feature evaluation method. While detecting the objective evaluation of the audience, it also expresses the degree of emotional fluctuation well. Although the accuracy of the multi-modal method before the test data volume of 5400 has not increased faster than the accuracy of the text feature evaluation method, this precisely shows that the multi-modal evaluation method is more stable than the single modal. The multi-modal evaluation model minimizes the impact of noise in the preliminary calculations. After the test data volume is 5400, the accuracy of the multi-modal starts to rise rapidly and begins to exceed the single text feature evaluation when the test data volume is 8400. It conforms to the expectations of the study in this paper, and further verifies the stability and accuracy of the multimodal evaluation method.

## *2. Comparative Experiments in Three Performance Forms with Singlemodal and Multimodal Affective Features*

By experiment method is the textual—affective feature, visual—affective feature and PC-MulAff model performed in three different kinds of performance creative evaluation experiment. The validity of PC-MulAff model in creative performance evaluation is verified. Based on the Accuracy(A), Precision(P) and Recall(R) and F-measure (F) to evaluate the three different performance creative evaluation model. F-measure is the geometric average Charu [57] of accuracy and recall rate, as the Equation.6.

$$F1-score = 2 \times \frac{Precision \times Recall}{Precision + Recall} \tag{7}$$

The experimental comparison of three performance creative evaluation modes in “Dance Performance” data set (see Table Ⅱ).

TABLE Ⅱ. Experimental Comparison of Three Performance Creative Evaluation Modes in Dance Performance Data Set

| Dataset | Mode | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| **Dance** | $S_t$ | 0.7674 | 0.7545 | 0.7830 | 0.7684 |
| | $S_v$ | 0.7023 | 0.6909 | 0.7169 | 0.7036 |
| | $S_{PC-Acc}$ | 0.8418 | 0.8272 | 0.8584 | 0.8386 |

The experimental comparison of three performance creative evaluation modes in “Dance Performance” data set (see Table Ⅲ).

TABLE III. EXPERIMENTAL COMPARISON OF THREE PERFORMANCE CREATIVE EVALUATION MODES IN DRAMA PERFORMANCE DATA SET

| Dataset | Mode | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| **Drama** | $S_t$ | 0.6428 | 0.6272 | 0.7419 | 0.6797 |
| | $S_v$ | 0.5659 | 0.5636 | 0.6666 | 0.6107 |
| | $S_{PC-Acc}$ | 0.6978 | 0.6727 | 0.7956 | 0.7290 |

The experimental comparison of three performance creative evaluation modes in "Dance Performance" data set (see Table IV).

TABLE IV. EXPERIMENTAL COMPARISON OF THREE PERFORMANCE CREATIVE EVALUATION MODES IN SQUARE PERFORMANCE DATA SET

| Dataset | Mode | Accuracy | Precision | Recall | F1-score |
|---|---|---|---|---|---|
| **Square** | $S_t$ | 0.6197 | 0.6090 | 0.6380 | 0.6231 |
| | $S_v$ | 0.6760 | 0.6636 | 0.6952 | 0.6790 |
| | $S_{PC-Acc}$ | 0.7605 | 0.7454 | 0.7809 | 0.7627 |

It can be seen that the evaluation mode of PC-MulAff model has achieved good expected effects in different evaluation of performance creative in Fig. 8.

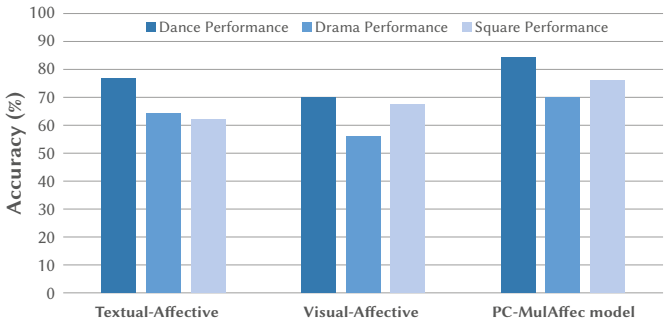


Fig. 8. Comparison of accuracy of three evaluation modes in different performance data.

## V. DISCUSSION

Through the comparative experiment 1, we can find the audience's evaluation text data can reach high accuracy in a short time, but the accuracy of performance video data has always been slower to improve for a long period of time. Neither of these two methods can objectively evaluate the accuracy of the model. The multimodal evaluation model makes up for the defects of single mode. Not only the accuracy rate is relatively stable, but also the high accuracy rate of text evaluation can be achieved. Through the comparative experiment 2, we also find the evaluation mode of PC-MulAff model is greatly improved compared with both textual—affective evaluation mode and visual—affective evaluation mode. Especially in accuracy and precision than visual—affective evaluation mode, the recall rate and F1 on increased 0.1395, 0.1363, 0.1415, 0.1350 respectively. It can also be seen from the table that the audience's evaluation effect of textual—affective on dance performance is better than visual—affective, indicating that the audience's evaluation text is richer than visual affective.

The evaluation mode of PC-MulAff model is significantly improved compared with the other two evaluation modes in drama performance. Especially than visual—affective evaluation mode in the accuracy, precision and recall rate and F1 on increased 0.1319, 0.1091, 0.1290, 0.1183 respectively. It can also be seen that the audience's evaluation text is richer than the visual affective features in the drama performance. The evaluation mode of PC-MulAff model is more advanced than the other two evaluation modes in square performance. Especially than textual—affective evaluation mode in accuracy, precision and recall rate and F1 on increased 0.1408, 0.1364, 0.1429, 0.1396 respectively. It can be seen from the table that visual—affective features are more abundant than textual—affective features in large square performances. This shows that in large square performances, the audience is limited by their visual field, so the overall effect of square performances is not strong. The creative effect of square performances is more suitable to be shown through video shots.

From the experimental results, the performance creative evaluation method proposed in this paper is effective and has reached the research expectations. In the perspective of the data set, we create a performance evaluation data set that integrates a variety of physiological signals from the audience and has a "Director Label" for the first time. This multi-modal data set corresponds the director's experience with the audience's physiological feedback for the first time. The establishment of this correspondence has played a key role in verifying the effectiveness and accuracy of the evaluation method. [76] present a Brain-Adaptive Digital Performance (BADP) was designed to measure and analysis of audience engagement level. Only to detect changes in the audience's engagement through the monitoring of EEG signals. First of all, in terms of the types of performances tested of stage performances are given, and there is no clear explanation of the types of performances and problem boundaries of participation monitoring. Second, the article only conducted monitoring experiments in a virtual performance environment. However, the audience can adapt more to the artistic perception ability and cognitive level of the real performance scene. The lack of experience in watching virtual performances and the freshness and cognition of visual perception will affect the accuracy of EEG signals. The authors did not give a method to remove the noise signal. The focus of our research is to use EEG signals as the attributes of the evaluation data set, and there are other physiological signal data as mutual verification of emotional features, which not only ensures the accuracy of emotional features, but also enhances the adaptability of the evaluation model. A multi-modal emotion recognition framework called EmotionMeter is proposed [77], which combined brain waves and eye movements, and verified that the modal fusion of multi-modal deep neural networks is higher than the performance of a single modal. However, the training data set mentioned in the article has differences and instabilities in cross-modal feature distribution. In the research of our paper, the method of "Director Label" is adopted to effectively avoid this defect. A Conditional Generative Adversarial Network (cGAN) is proposed to establish emotion-related EEG data [78], and two components that constitute an emotional EEG signal (YEEG) are defined: emotion-related (YE motion) and emotion-related (NOthers). While, the accuracy of labeling with facial expression images needs to be further improved. And the current coarse-grained labeling has limitations in emotion recognition and analysis.

We can also further conclude that in the evaluation model of dance performances the accuracy rate is the highest from Fig. 8, which can reflect that the audience's aesthetic feelings and perceptions of dance performances are higher than those of drama performances and square performances. It also confirms that the dance performances that the audience are exposed to are more than other performances. In the evaluation model of drama performance, the evaluation of text features is more accurate than the evaluation of visual features, which also shows that the literary of drama performance is stronger, and the audience's drama and literature accomplishment is higher than that of the square performance. In the evaluation model of square performances, visual feature evaluation is more accurate than text

feature evaluation, which shows that the creative core of large-scale square performances mainly revolves around the performance of visual effects. These viewpoints demonstrated from experiments not only show the effectiveness of the performance creative evaluation model proposed in this article, but also highlight the creative core of different performance forms. This helps directors to carry out effective and reasonable performance creative, and improves the performance creative efficiency and level. At the same time, the actor and audience's aesthetic perception of performance is improved. Therefore, it can be seen that our research has high potential value and practical benefits.

## VI. Conclusion

Experiments show that the PC-MulAff model is effective, especially in the comparative experiments for three different performance modes, PC-MulAff model has achieved good results. The purpose of this article is to evaluate performance creative through multimodal affective features. In order to achieve this goal, the affective features of the audience and the visual features of the performance video are extracted respectively, and the performance creative is analyzed through the quantified score after multimodal feature fusion. The main contributions of this article: 1). this paper proposes a PC-Acc to evaluate the quality of performance creative, trains and builds a PC-MulAff model, which can evaluate creative for different performance forms. And also 2). we propose a new "Performance Creativity-Multimodal Evaluation Data Set", which is composed of performance video data, audience evaluation text and audience physiological data. It not only makes up for the insufficient description of features by a single data type, but also provides a performance evaluation data set type with multiple physiological signal emotional features. This work provides a standardized verification basis for performance evaluation and fills the gap in the direction of the verification in performance evaluation data set. 3) Based on the establishment of multi-modal evaluation data, the correlation analysis between the audience's multi-modal physiological signals and different performance types has been realized. For the first time, the audience's emotional features and performance creative has been mapped through the method of "Director Label". This work plays a decisive role in the evaluation of performance creative. Digital and intelligent technical provide directors with scientific evaluation methods and verification basis.

Based on the limitations of currently collected performance works and experimental scenes, although the research methods proposed in this paper have been substantively verified, we believe that the existing evaluation framework can still be further optimized in future work , In particular, research on related algorithms for precise extraction of performance content based on emotion classification, and optimization of experimental scenes, to provide volunteers with a more immersive environment to extract more accurate physiological data and improve multimedia evaluation data sets . We hope that the next step will continue the ideas of this paper, and make corrections and adjustments to the score calculation including algorithm parameter adjustments of multi-modal special fusion.

The multimodal data-driven performance creative evaluation method proposed in this paper is effective. The model not only provides a multi-dimensional analysis for the performance creativity evaluation, but also proposes to solve the interpretability and data set support of creative evaluation of performing arts.

## References


[1] I. S. Lee, "Performing arts in the age of transmedia," *Journal of acting studies*, vol. 17, pp. 17–32, 2020.

[2] K.-W. Huang, C.-C. Lin, Y.-M. Lee, Z.-X. Wu, "A deep learning and image recognition system for image recognition," *Data Science and Pattern Recognition*, vol. 3, no. 2, pp. 1–11,2019.

[3] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[4] C. Szegedy, W. Liu, Y. Jia, P. Sermanet, S. Reed, D. Anguelov, D. Erhan, V. Vanhoucke, A. Rabinovich, "Going deeper with convolutions," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2015, pp. 1–9.

[5] J. S. Chung, B.-J. Lee, I. Han, "Who said that?: Audio-visual speaker diarisation of real-world meetings," *arXiv preprint arXiv:1906.10042*, 2019.

[6] J. Wu, Y. Xu, S.-X. Zhang, L.-W. Chen, M. Yu, L. Xie, D. Yu, "Time domain audio visual speech separation," in *2019 IEEE Automatic Speech Recognition and Understanding Workshop (ASRU)*, 2019, pp. 667–673, IEEE.

[7] J. C.-W. Lin, Y. Shao, Y. Djenouri, U. Yun, "Asrnn: a recurrent neural network with an attention model for sequence labeling," *Knowledge-Based Systems*, vol. 212, p. 106548, 2020.

[8] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "Bert: Pretraining of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[9] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[10] J. C.-W. Lin, G. Srivastava, Y. Zhang, Y. Djenouri, M. Aloqaily, "Privacy preserving multi-objective sanitization model in 6g iot environments," *IEEE Internet of Things Journal*, 2020.

[11] F. Abbé-Decarroux, "The perception of quality and the demand for services: Empirical application to the performing arts," *Journal of Economic Behavior & Organization*, vol. 23, no. 1, pp. 99–107, 1994.

[12] F. Nake, "Computer art: creativity and computability," in *Proceedings of the 6th ACM SIGCHI conference on Creativity & cognition*, 2007, pp. 305–306.

[13] K. Yamada, T. Taura, Y. Nagai, "Design and evaluation of creative and emotional motion," in *Proceedings of the 8th ACM conference on Creativity and cognition*, 2011, pp. 239–248.

[14] C.-y. Chang, Y.-p. Chen, "Fusing creative operations into evolutionary computation for composition: From a composer's perspective," in *2019 IEEE Congress on Evolutionary Computation (CEC)*, 2019, pp. 2113–2120, IEEE.

[15] L. Goves, "Multimodal performer coordination as a creative compositional parameter," *Tempo*, vol. 74, no. 293, pp. 32–53, 2020.

[16] D. Cabral, J. G. Valente, U. Aragão, C. Fernandes, N. Correia, "Evaluation of a multimodal video annotator for contemporary dance," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, 2012, pp. 572–579.

[17] R. E. Cisneros, K. Wood, S. Whatley, M. Buccoli, M. Zanoni, A. Sarti, "Virtual reality and choreographic practice: The potential for new creative methods," *Body, Space & Technology*, vol. 18, no. 1, 2019.

[18] B. T. Christensen, L. J. Ball, "Dimensions of creative evaluation: Distinct design and reasoning strategies for aesthetic, functional and originality judgments," *Design Studies*, vol. 45, pp. 116–136, 2016.

[19] P. Karimi, N. Davis, M. L. Maher, K. Grace, L. Lee, "Relating cognitive models of design creativity to the similarity of sketches generated by an ai partner," in *Proceedings of the 2019 on Creativity and Cognition*, 2019, pp. 259–270.

[20] T. Knearem, X. Wang, J. Wan, J. M. Carroll, "Crafting in a community of practice: Resource sharing as key in supporting creativity," in *Proceedings of the 2019 on Creativity and Cognition*, 2019, pp. 83–94.

[21] M. Richardson, F. Hernández-Hernández, M. Hiltunen, A. Moura, M. Fulková, F. King, F. M. Collins, "Creative connections: The power of contemporary art to explore european citizenship," *London Review of Education*, 2020.

[22] K. H. Koh, "Computing indicators of creativity," in *2011 IEEE Symposium on Visual Languages and Human-Centric Computing (VL/HCC)*, 2011, pp. 231–232, IEEE.

[23] A. Jain, "Measuring creativity: Multi-scale visual and conceptual design analysis," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, 2017, pp. 490–495.

[24] J. Oppenlaender, "Supporting creative workers with crowdsourced

feedback," in *Proceedings of the 2019 on Creativity and Cognition*, 2019, pp. 646–652.

[25] D. Abfalter, "Authenticity and respect: Leading creative teams in the performing arts," *Creativity and innovation management*, vol. 22, no. 3, pp. 295–306, 2013.

[26] J. Lee, E. Jun, J. Chae, "Big data analysis for dance studies using text mining," *The Journal of Dance Society for Documentation & History*, vol. 42, p. 191—212, 2016.

[27] L. J. Min, "An analysis of semantic relations in knowledge information in dance research data in korea from 1958 to 2016," *The Korean Journal of Arts Studies*, no. 16, p. 215—237, 2017.

[28] K. H. Ryeon, "Exploring the determinants of korean dance recognition and importance: Application of decision tree analysis based on data mining," *Dance Research Journal of Dance*, vol. 77, no. 1, p. 17—29, 2019.

[29] Choi, Hyo-jin, "Previous study research on korean contemporary dance using text mining," *The Korean Journal of Dance Studies*, vol. 76, no. 4, p. 97—111, 2019.

[30] K. Woo-Kyung, J.-Y. Yoo, "Analysis on the trends of research themes of the korean dance using text mining," *Journal of the Korea Entertainment Industry Association*, vol. 13, no. 5, p. 215—228, 2019.

[31] choihyojin, "Analysis of korean contemporary dance research trends using text mining," *Korean Journal of Arts Education*, vol. 17, no. 4, p. 103—118, 2019.

[32] Kimhayeon, "Analysis on the international contemporary dance research trend using text mining," *Korean Journal of Arts Education*, vol. 18, no. 1, p. 171—192, 2020.

[33] S. Zhou, J. Jia, Y. Wang, W. Chen, F. Meng, Y. Li, J. Tao, "Emotion inferring from large-scale internet voice data: A multimodal deep learning approach," in *2018 First Asian Conference on Affective Computing and Intelligent Interaction (ACII Asia)*, 2018, pp. 1–6, IEEE.

[34] W. Liang, H. Xie, Y. Rao, R. Y. Lau, F. L. Wang, "Universal affective model for readers' emotion classification over short texts," *Expert Systems with Applications*, vol. 114, pp. 322–333, 2018.

[35] H. M. Hung, H.-J. Yang, S.-H. Kim, G.-S. Lee, "Variants of bert, random forests and svm approach for multimodal emotion-target sub-challenge," arXiv preprint *arXiv:2007.13928*, 2020.

[36] P. T. Sowden, L. Dawson, "Creative feelings: the effect of mood on creative ideation and evaluation," in *Proceedings of the 8th ACM Conference on Creativity and Cognition*, 2011, pp. 393–394.

[37] G. Corness, K. Carlson, T. Schiphorst, "Audience empathy: a phenomenological method for mediated performance," in *Proceedings of the 8th ACM conference on Creativity and cognition*, 2011, pp. 127–136.

[38] C. K. Coursaris, W. Van Osch, "A cognitive-affective model of perceived user satisfaction (campus): The complementary effects and interdependence of usability and aesthetics in is design," *Information & Management*, vol. 53, no. 2, pp. 252–264, 2016.

[39] K. Altuwairqi, S. K. Jarraya, A. Allinjawi, M. Hammami, "A new emotion–based affective model to detect student's engagement," *Journal of King Saud University-Computer and Information Sciences*, 2018.

[40] F. Rahdari, E. Rashedi, M. Eftekhari, "A multimodal emotion recognition system using facial landmark analysis," *Iranian Journal of Science and Technology, Transactions of Electrical Engineering*, vol. 43, no. 1, pp. 171–189, 2019.

[41] P. D. Loprinzi, S. Pazirei, G. Robinson, B. Dickerson, M. Edwards, R. E. Rhodes, "Evaluation of a cognitive affective model of physical activity behavior," *Health Promotion Perspectives*, vol. 10, no. 1, p. 88, 2020.

[42] W. Liu, X. Xie, S. Ma, Y. Wang, "An improved evaluation method for soccer player performance using affective computing," in *2020 3rd International Conference on Artificial Intelligence and Big Data (ICAIBD)*, 2020, pp. 324–329, IEEE.

[43] W. Wei, Q. Jia, Y. Feng, G. Chen, M. Chu, "Multi-modal facial expression feature based on deep-neural networks," *Journal on Multimodal User Interfaces*, vol. 14, no. 1, pp. 17–23, 2020.

[44] G. Chen, X. Zhang, Y. Sun, J. Zhang, "Emotion feature analysis and recognition based on reconstructed eeg sources," *IEEE Access*, vol. 8, pp. 11907–11916, 2020.

[45] H.-J. Choi, Y.-J. Lee, "Deep learning based response generation using emotion feature extraction," in *2020 IEEE International Conference on Big Data and Smart Computing (Big- Comp)*, 2020, pp. 255–262, IEEE.

[46] C. Deepika, "Speech emotion recognition feature extraction and classification," *International Journal of Advanced Trends in Computer Science and Engineering*, vol. 9, no. 2, pp. 1257–1261, 2020.

[47] W. Wei, Q. Jia, Y. Feng, G. Chen, M. Chu, "Multi-modal facial expression feature based on deep-neural networks," *Journal on Multimodal User Interfaces*, vol. 14, no. 1, pp. 17–23, 2020.

[48] J. Radbourne, K. Johanson, H. Glow, T. White, "The audience experience: Measuring quality in the performing arts," *International journal of arts management*, pp. 16–29, 2009.

[49] Radbourne, Jennifer, "The quest for self actualization meeting new consumer needs in the cultural industries," *In ESRC Seminar Series Creative Futures-Driving the Cultural Industries Marketing Agenda*, vol. 6, 2007.

[50] C. Latulipe, E. A. Carroll, D. Lottridge, "Love, hate, arousal and engagement: exploring audience responses to performing arts," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2011, pp. 1845–1854.

[51] C. Wang, E. N. Geelhoed, P. P. Stenton, P. Cesar, "Sensing a live audience," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, 2014, pp. 1909–1912.

[52] C. Martella, E. Gedik, L. Cabrera-Quiros, G. Englebienne, H. Hung, "How was it? exploiting smartphone sensing to measure implicit audience responses to live performances," in *Proceedings of the 23rd ACM international conference on Multimedia*, 2015, pp. 201–210.

[53] R. Adolphs, D. Tranel, A. R. Damasio, "Dissociable neural systems for recognizing emotions," *Brain & Cognition*, vol. 52, no. 1, pp. 61–69, 2003.

[54] A. R. Damasio, T. J. Grabowski, A. Bechara, H. Damasio, L. L. Ponto, J. Parvizi, R. D. Hichwa, "Subcortical and cortical brain activity during the feeling of self-generated emotions," *Nature neuroscience*, vol. 3, no. 10, pp. 1049–1056, 2000.

[55] J. Radbourne, K. Johanson, H. Glow, T. White, "The audience experience: Measuring quality in the performing arts," *International journal of arts management*, pp. 16–29, 2009.

[56] M. Wyczesany, S. J. Grzybowski, R. J. Barry, J. Kaiser, A. M. Coenen, A. Potoczek, "Covariation of eeg synchronization and emotional state as modified by anxiolytics," *Journal of Clinical Neurophysiology*, vol. 28, no. 3, pp. 289–296, 2011.

[57] S. Koelstra, C. Muhl, M. Soleymani, J.-S. Lee, A. Yazdani, T. Ebrahimi, T. Pun, A. Nijholt, I. Patras, "Deap: A database for emotion analysis; using physiological signals," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 18–31, 2011.

[58] G. Pinto, J. M. Carvalho, F. Barros, S. C. Soares, A. J. Pinho, S. Brás, "Multimodal emotion evaluation: A physiological model for cost-effective emotion classification," *Sensors*, vol. 20, no. 12, p. 3510, 2020.

[59] H. Zhang, "Expression-eeg based collaborative multimodal emotion recognition using deep autoencoder," *IEEE Access*, vol. 8, pp. 164130–164143, 2020.

[60] M. E. Peters, M. Neumann, M. Iyyer, M. Gardner, C. Clark, K. Lee, L. Zettlemoyer, "Deep contextualized word representations," *arXiv preprint arXiv:1802.05365*, 2018.

[61] A. Radford, K. Narasimhan, T. Salimans, I. Sutskever, "Improving language understanding by generative pretraining," 2018.

[62] J. Devlin, M.-W. Chang, K. Lee, K. Toutanova, "Bert: Pretraining of deep bidirectional transformers for language understanding," *arXiv preprint arXiv:1810.04805*, 2018.

[63] Z. Lan, M. Chen, S. Goodman, K. Gimpel, P. Sharma, R. Soricut, "Albert: A lite bert for self-supervised learning of language representations," *arXiv preprint arXiv:1909.11942,* 2019.

[64] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, Ł. Kaiser, I. Polosukhin, "Attention is all you need," in *Advances in neural information processing systems*, 2017, pp. 5998–6008.

[65] C.-M. Kuo, N.-C. Yang, J.-Y. Wu, S.-C. Chen, "Histogrambased image enhancement in quasi-spatial domain for compressed image,"

[66] S. Zhao, Y. Gao, X. Jiang, H. Yao, T.-S. Chua, X. Sun, "Exploring principles-of-art features for image emotion recognition," in *Proceedings of the 22nd ACM international conference on Multimedia*, 2014, pp. 47–56.

[67] C. Xu, S. Cetintas, K.-C. Lee, L.-J. Li, "Visual sentiment prediction with deep convolutional neural networks.," *arXiv preprint arXiv:1411.5731,* 2014.

[68] B. Jou, S.-F. Chang, "Deep cross residual learning for multitask visual recognition," in *Proceedings of the 24th ACM international conference on*

*Multimedia,* 2016, pp. 998–1007.

[69] L. Gao, Z. Guo, H. Zhang, X. Xu, H. T. Shen, "Video captioning with attention-based lstm and semantic consistency," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2045–2055, 2017.

[70] N. Zhao, H. Zhang, R. Hong, M. Wang, T.-S. Chua, "Videowhisper: Toward discriminative unsupervised video feature learning with attention-based recurrent neural networks," *IEEE Transactions on Multimedia*, vol. 19, no. 9, pp. 2080–2092, 2017.

[71] K. Simonyan, A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint arXiv:1409.1556*, 2014.

[72] C. L. Lisetti, F. Nasoz, "Using noninvasive wearable computers to recognize human emotions from physiological signals," *EURASIP Journal on Advances in Signal Processing* vol. 2004, no. 11, p. 929414, 2004.

[73] S. un Kai, Y. Junqing, "Audience oriented personalized movie affective content representation and recognition," *Journal of Computer-Aided Design &Computer Graphics*, vol. 22, no. 1, pp. 136–144, 2010.

[74] M. Soleymani, J. Lichtenauer, T. Pun, M. Pantic, "A multimodal database for affect recognition and implicit tagging," *IEEE transactions on affective computing*, vol. 3, no. 1, pp. 42–55, 2011.

[75] C. C. Aggarwal, C. Zhai, Mining text data. Springer Science & Business Media, 2012.

[76] S. Yan, G. Ding, H. Li, N. Sun, Z. Guan, Y. Wu, L. Zhang, T. Huang, "Exploring audience response in perfor ming arts with a brain-adaptive digital performance system," *ACM Transactions on Interactive Intelligent Systems (TiiS)*, vol. 7, no. 4, pp. 1–28, 2017.

[77] W.-L. Zheng, W. Liu, Y. Lu, B.-L. Lu, A. Cichocki, "Emotionmeter: A multimodal framework for recognizing human emotions," *IEEE transactions on cybernetics*, vol. 49, no. 3, pp. 1110–1122, 2018.

[78] B. Fu, F. Li, Y. Niu, H. Wu, Y. Li, G. Shi, "Conditional generative adversarial network for eeg-based emotion finegrained estimation and visualization," *Journal of Visual Communication and Image Representation*, vol. 74, p. 102982.

Yufeng Wu

Yufeng Wu obtained B.E. degree from Yantai University, China in 2011, and he is currently working toward the PhD in the School of Computer Science and Technology, Beijing Institute of Technology. His research interests are digital performance, computer simulation, deep learning and computer vision. His main research includes: intelligent generation methods for performance creativity, construction of the framework and theoretical system of performance creativity; and performance data analysis methods and creativity evaluation modeling based on graph neural networks and deep reinforcement learning. He is also working on developing a highly dynamic and intelligent creative evaluation platform based on the human-machine collaboration.

Longfei Zhang

Longfei Zhang obtained Ph.D. from School of Computer Science and Engineering, Beijing Institute of Technology, China in 2005. He is an associate professor in School of Computer Science and Technology at Beijing Institute of Technology. He went to Carnegie Mellon University as a visiting scientist from 2009 to 2011. His main research focuses on "Analysis, Prediction and Construction of 3D Behaviors of Video Personnel", "Cross Media Knowledge Base and Common Sense Library", "Intelligent Performance Creation and Evaluation", "Sports Content Understanding and Intelligent Directing".

Gangyi Ding

Gangyi Ding received the B.E. degree from Peking University, China in 1988, and Ph.D. at Beijing Institute of Technology, China in 1993. He is a professor with the Key Laboratory of Digital Performance and Simulation Technology of the Beijing Institute of Technology. He joined the faculty at the Beijing Institute of Technology in 1993. His research mainly involves training simulation, large-scale crowd simulation, environment simulation, digital performance and creative simulation. He provided simulation technical support for the arrangement of large-scale events such as the opening and closing ceremonies of the Beijing 2008 Olympic Games, and the Beijing 8-Minutes of the Pyeongchang Winter Olympics.

Tong Xue

Tong Xue received her B.E. degree from Communication University of China in 2016. She is currently working toward the PhD degree in School of Computer Science and Technology, Beijing Institute of Technology. She is a joint PhD student at Distributed and Interactive Systems, Centrum Wiskunde & Informatica (CWI). Her research interests lie in human-computer interaction and affective computing.

Fuquan Zhang

Fuquan Zhang received the PhD degree in School of Computer Science & Technology, Beijing Institute of Technology, China in 2019. Now he is a professor of Minjiang University, China. He is now a member of the National Computer Basic Education Research Association of the National Higher Education Institutions, a member of the Online Education Committee of the National Computer Basic Education Research Association of the National Institute of Higher Education, a member of the MOOC Alliance of the College of Education and Higher Education Teaching Guidance Committee, ACM SIGCSE, CCF member, CCF YOCSEF member, director of Fujian Artificial Intelligence Society.