

A Study on the Aesthetic Characteristics and Transformation Mechanisms of Cross-media performers' Stage and Screen Performances under Media Differences

He Xu

Xi'an International University, Xi'an 710077, Shaanxi, China

Corresponding author: He Xu (e-mail: hexudedaxue@163.com)

ABSTRACT Cross-media performers face challenges in rhythm control, emotional delivery, and spatial deployment when shifting between stage and screen performance, resulting in aesthetic inconsistency across media. This paper constructs a performance transformation mechanism model for cross-media performers under media differences. The study integrates literature review, experimental comparison, and interviews. Sixty video clips of stage and screen performances from 30 dual-media actors are collected, and four variables are quantified: camera distance, performance rhythm, facial micro-expression, and spatial movement. Audience responses and attentional focus are measured through eye-tracking and emotion recognition systems. Structural equation modeling is used to validate the performance transformation mechanism. The results show that performance adaptability across media can be optimized through emotional accuracy training, camera awareness, rhythm reconstruction, and spatial perception training. Emotional expression accuracy and camera awareness adaptability influence audience perception consistency, while spatial rhythm coordination plays a mediating role. The study provides a quantitative framework for analyzing performance aesthetics and offers practical guidance for actor training across stage, film, and television media.

INDEX TERMS cross-media performance, media difference, performance aesthetics, eye tracking, structural equation modeling

I. INTRODUCTION

THE two mediums of performance, stage and screen, possess fundamental differences in space, rhythm, and audience interaction, mirroring the distinct technical requirements and yarn spatial distributions found across diverse machinery and industrial manufacturing processes. Navigating these performance variations requires a systematic understanding of structural layers, comparable to the digital twin-knowledge graph fusion used to bridge the semantic gap in the digital preservation of industrial heritage. Those who perform in both mediums must learn to dynamically change their approach between the outwardness of the performance energy and the intangibility of the transmission of emotions in switching between media. With the distance of the camera compressing the openness of the stage, the rhythms of body expressions, the variety of emotional projection and the spatial control strategies are altered and significantly tense aesthetically. This disparity not only influences the style of performance of the actor in question, but also it has the immediate effect of influencing the

perceptual experience and emotional resonance of the audience. Expressive physical gestures and spatial expansion are commonly used in stage performances to evoke an aesthetic atmosphere, whereas subtle facial expression, vocal rhythm, and control of visual attention dominate in screen performances [1,2]. Cross-media performers often experience challenges such as rhythmic imbalance, emotional latency, and limited camera awareness when transitioning between the two media systems. The existing studies are primarily concerned with the pattern of performance of a particular medium without systematic analysis of the sophisticated adjustment mechanisms generated by the differences in the media and an explanation of how actors can find aesthetic continuity in their transitions.

This paper explores the performance transformation mechanism of Cross-media performers across media differences, integrating the same digital twin-knowledge graph fusion used to map the complex fiber morphology and yarn spatial distribution in modern manufacturing. By bridging the semantic gap between abstract artistic adaptation and the

technical precision of industrial heritage, the study provides a systematic framework for understanding structural shifts in both performance art and high-performance engineering. It exposes the cause and effect links of performance variables through an experimental analysis and structural modeling of cross-media adaptability through three dimensions; facial expression, spatial rhythm and audience perception. The study will also seek to address the question on how a coordinated and unified performance style can be achieved through camera awareness training and rhythm reconstruction that will provide an operating quantitative foundation on the cross-media performance of actors and broaden the usefulness of performance aesthetics on the convergence of film, television and stage.

II. RELATED WORK

The study of body language, rhythm control, and empathy with the audience has long been studied as one of the mechanisms of stage performing, and it was found that actors create energy fields of emotions via movement rhythm and space organization. Scholars have since mainly studied stage physical expression through the concept of psychological reality and motivational drive, focusing on the imaginative extension and magic discharge of the actors in open spaces since the Stanislavski system. Researchers have started applying motion capture and facial recognition technology to measure the physical rhythmic properties of stage performances in recent years e.g. the relationship between rhythm and audience attention of a stage performance is synchronized. These findings, however, can mainly be applied to the theater area, and the complexity of the facial expression and camera flexibility is not thoroughly analyzed [3,4]. Moreover, the research on the stage aesthetics is more concerned with the symbolization and shared emotions, and not much is said about the strategies of individual control in the media transformation.

There is also research of screen performance based on the grammar of shots, visual composition, and emotional encoding of facial micro-actions. The theories of film and television performance training support the concepts of emotional micro-amplification and visualization of internal expression, when camera language is employed in order to increase the emotional concentration of the audience. Experimental analyses on accuracy of recognition of facial expressions, changes in speech frequency and influence of camera distance have been done on relevant literature although most studies are done on film and television actors without any longitudinal tracking of the groups of actors who appear on both the stage and the screen [5,6]. Furthermore, even though media communication studies have examined the perceptual logic of image narrative, they never amalgamated it with the performance psychology and hence it has been hard to uncover the dynamic trajectory through which the performers arrive at aesthetic equilibrium amidst media diversities. This is the gap that is filled up in this research.

III. METHODS

A. RESEARCH SUBJECTS AND DATA SOURCES

A sample consisting of 30 professional actors who had experience of performing with both media was chosen in the study with a male to female ratio of 1:1 and an average of 32.4 years. The experience of all in stage plays and film and television shooting had been over five years. A sample was selected consisting of three kinds of works namely, experimental plays, TV series and short films to be certain that the variability in the performance style and media context was quantifiable and analyzable. The materials of performance on the stage were gathered as the result of multi-cam recording on the scene, in which the camera angle was positioned above the chest line of the actor to capture the connection between body movement and facial expression. Materials of screen performance were acquired on the basis of the official shots of film and television works, which were shot in the same time and did not contain editing interference. Each type of media sample was then taken out of each actor in two segments and the 120 segments added together, which consisted of about 360 minutes in total [7,8]. Images were handled in Premiere Pro timeline and frame rate of 30fps and audio-visual sync within 0.02 seconds. ELAN software was used to perform data annotation and time-locked annotations of facial micro-expressions, gaze direction, movement trajectory, and speech rhythm were made, producing a total of 8742 time-series labels. There were 18 non-performance observers who gathered the audience reaction data. Gaze hotspots, dwell times were successfully captured using eye-tracking devices and timecodes of the respective segments were automatically synchronized to create a bimodal material of the analysis. Informed consent was obtained from all participants. The information data is shown in Table 1.

Table 1 shows the differences in sample type, duration, and number of annotations, which facilitates the establishment of media comparison relationships between performance rhythm, spatial use, and emotional density.

B. SETTING PERFORMANCE VARIABLES AND CONSTRUCTING THE INDICATOR SYSTEM

The performance variables were set based on the perceptual characteristics presented by the differences between stage and screen media, and the distribution of variables followed the principle of coupling the quantifiable dimensions of the actor's expression with the dimensions of the audience's perception. The camera distance variable measures the difference in the actor's body projection through changes in spatial scale and perspective, and is expressed as the percentage of the body in the picture, with the range controlled between 0.25 and 0.85. The performance rhythm variable consists of the duration of speech and the cycle of movement, and the rhythm fluctuation per unit time is identified by ΔT . Time series data is generated through Praat acoustic analysis. The facial micro-expression variable is coded according to the action unit (AU) defined by the

TABLE 1. Statistics on the Sample and Material Information of Cross-media performers

Serial number	Performance type	Number of materials (segments)	Duration (minutes)	Number of labels	Sampling equipment
A01–A05	Stage play + short film	8	24	712	SONY FX6 camera
A06–A10	Musical + TV series	8	30	686	Canon C300 Mark II
A11–A15	Experimental drama + film	8	27	745	Blackmagic URSA Mini
A16–A20	Drama + Web Series	8	28	712	Panasonic EVA1
A21–A25	Stage play + movie	8	26	701	RED Komodo 6K
A26–A30	Small theater play + short film	8	25	672	Nikon Z9

Facial Action Coding System (FACS), focusing on four action intervals that are highly related to emotion recognition: AU1, AU6, AU12, and AU15. The movement amplitude is calculated as the trajectory offset in a two-dimensional coordinate system, with the unit being pixel displacement [9,10]. The spatial motion variable is reconstructed based on the displacement path of the actor on the stage or in the composition of the picture. Key nodes are extracted through the OpenPose skeleton recognition system, and the mean of horizontal movement and the variance of vertical displacement are calculated. Audience perception metrics measure visual attention by the overlap rate between eye movement dwell time and gaze trajectory hotspots, and combine this with the matching degree of emotional response output by the emotion recognition system to form a “performance-perception” collaborative matrix.

C. MULTIMODAL ANALYSIS PROCESS

The principle of multimodal analysis is founded on the simultaneous matching of visual, auditory and physiological responses to make stage and screen performance material as accurate as possible in time and features in their complementarity. On the visual dimension, a joint skeleton recognition algorithm based on OpenPose and MediaPipe is used to produce 33 key nodes of the joints, resulting in a motion trajectory JSON file that is automatically correlated with the audio track frame rate through timestamps. RetinaFace identifies facial areas and feeds them into the DeepFace model to derive 128-dimensional vectors of facial features as five basic confidence levels of emotions per frame. Audio signals are decomposed acoustically in Praat and the speech intensity, rate and frequency distribution are calculated and the results are an energy matrix of signal at 0.1-second intervals.

Tobii Pro Fusion device was used to record the eye-tracking data of the audience with the sampling rate of 120Hz, which captured the coordinates of the fixation points, the fixation duration, and the changes in pupil diameter. The synchronization module took frame numbers as a point of reference to synchronize multi-source data, whereby the error could be under 8 milliseconds. The Emotion Recognition System processed emotional response data, by frame recognizing face video of the audience and providing output of emotion category and confidence through a convolutional neural network and comparing the results to the time series of the changes in the facial expressions of the actor through a sliding window model.

D. STRUCTURAL EQUATION MODELING (SEM) CONSTRUCTION AND PATH VALIDATION

The multidimensional characteristics of amphibious actor performance variables were inputted into structural equation modeling in order to bring out both the causal pathways of emotions expression, spatial control, and aesthetic perception in media transition. Four latent constructs that were determined following an exploratory factor analysis of the imported data into the AMOS 26.0 platform included: “Accuracy of Emotional Expression,” “Spatial Rhythm Coordination,” Audience Perception Consistency,” and Camera Awareness Adaptability. Latent variables have 3-5 observational indicators in each (confidence level of features of facial micro-expressions, standard deviation of movement amplitude, total gaze duration and rhythm control deviation). The theoretical assumptions form the model path, which is defined as follows: accuracy of emotional expression and camera awareness adaptability both affect audience perception consistency with spatial rhythm coordination mediating the effect of both. Maximum likelihood estimation was used to estimate the parameters and the convergence criteria was 0.001.

IV. RESULTS AND DISCUSSION

A. EXPERIMENTAL DESIGN

The within-subjects design structure was used to carry the experiment, as it is based on the principle of comparability of dual-media performance. A control group was the performance sections of the same actor in the stage and screen. Every actor acted twice with the same passage, keeping the script, dialogue and emotional purpose unchanged, however, the scene was shifted slightly to the performance area, the distance of the camera, and the position of the audience. The stage performances were held in an open theater environment with the audience being about 12 meters away to the actors and the filming was done by multi-cameras and fixed position, so that the body movement and sound would propagate naturally. Single-lens tracking shot with 1.5 meters distance was utilized in the screen performances to bring out facial expression and vocal characteristics. All performances underwent one rehearsal and one formal recording, with lighting controlled at $650 \text{ lx} \pm 10 \text{ lx}$ and an audio sampling rate of 48 kHz. Audience reaction experiments were conducted simultaneously, with 18 observers watching video stimuli in each session. Eye trackers and facial recognition systems simultaneously recorded gaze trajectories and emotional responses. Data collection lasted

TABLE 2. Statistical Results of Facial Expression Range in Stage and Screen Performances

Performance type	Average mood fluctuation amplitude ΔE	Average peak frequency (Hz)	Variance σ^2
Stage performance	0.84	1.72	0.063
Screen performance	0.48	1.05	0.041
Difference value	-0.36	-0.67	-0.022
Change percentage (%)	-42.6	-38.9	-34.9

approximately 12 minutes, with actor performance and audience reaction time series aligned for interactive analysis of performance variables and perceptual indicators, ensuring the quantifiability and statistical comparability of media difference effects.

B. DIFFERENCES IN AESTHETIC CHARACTERISTICS BETWEEN STAGE AND SCREEN PERFORMANCE

The DeepFace feature vector mean squared error was employed to determine the change in emotion intensity by the use of the facial expression amplitude test. In the case of each performance video, facial regions were cropped at a rate of 0.04 seconds and following automatic registration based on facial key points, six different emotion confidence curves were generated. The amplitude of facial expression of every actor using both the media was measured using emotional feature change range whereby the range of emotion intensity change and frequency of distribution of peaks were compared. The 120 video clips were sampled and added up to 74,000 valid data frames. Once the data had been smoothed and standardized over time, an amplitude comparison matrix was created that revealed that the average amplitude of the face expression was 42.6 percent smaller when the experiment was on the screen compared to when the performance was on the stage.

Table 2 indicates that the intensity and the tempo of the emotional swings in onscreen performances have largely reduced, and this implies that the actors are more likely to express emotions via expression of low intensity, however, with more delicate timing in the cases of a compressed camera distance. The reduction in the number of changes in frequency implies the more consistent change of emotions, so allowing the audience to create a stable vision under the close-up conditions; the reduction of the variance means the accuracy of the facial expression and the tendency of the focused detail expression development.

C. SPATIAL RHYTHM CONTROL SCORE

The scores of spatial rhythm control are determined in terms of stability of body displacement, movement cycle, and rhythm of an actor. The coordinates of 33 important nodes were extracted with the help of OpenPose skeleton recognition, which measures the horizontal and vertical change of displacement in time series form. The stability of the movement cycle was analyzed with the help of Fourier transform and it was combined with the data about the speech rhythm that is extracted with the help of Praat and made R the movement-Speech coupling Index. The overall score is obtained through weighting three aspects, including

TABLE 3. Comparison Results of Spatial Rhythm Control Scores

Group	Mean	Standard deviation	Stability coefficient (Cv)	Action-speech coupling index R
pre-stage test	56.3	8.4	0.149	0.62
Post-stage testing	71.8	6.9	0.096	0.75
Screen test	58.7	7.6	0.129	0.68
Behind the screen	79.2	5.3	0.067	0.83
average improvement	+20.5	-	-	+0.15

displacement regularity, beat accuracy of movement, and spatial usage density using standardized Z-scores. The results are based on comparative tests of 30 actors prior to and following media conversion training, amounting to 180 video clips being tested, with more or less 126, 000 frames of frame level motion analysis. The results are shown in Table 3.

The scores of the control of the spatial rhythm before and after the media conversion training showed an increase of about 20.5 points after the conversion, whereas the coefficient of the stability declined to 55, which indicated that the movement rhythm was more likely to stabilize. The fact that the R value has risen shows the augmented coordination of the physical and vocal rhythms, which is due to augmented spatial control awareness through camera distance and the limitation of the framing. Screen training made actors

change the rate and amplitude of movements, creating a more balanced rhythmic hierarchy, shifting the loose expression of stage to the camera, showing a systematic advancement in the logic of movement organization and a greater level of visual specificity of rhythmic aesthetics.

V. CONCLUSION

The spatial environment, camera distance, and emotional delivery methods are intrinsically linked, determining the cross-media performance of Cross-media performers in a manner consistent with how fiber morphology and yarn spatial distribution dictate the structural integrity of industrial products. Mastering these variables allows for the same level of precision and adaptability required in the digital twin-knowledge graph fusion of modern machinery and automated production systems. Stage performances are focused on externalization and extension whereas, screen performances are focused on details and emotional emphasis. The restrictions of the camera will result in a smaller area of control of the body movements and the facial expression, and the correspondence of the rhythm and emotion is rebalanced in the process of training. Adaptability of the actors is expressed through increased accuracy of body control and absorption of the emotional levels. Aesthetic difference is concretized in multimodal analysis, which creates consistent performance mechanisms to various media constructions. The research reveals that performers need to develop a dynamic media awareness, viewing performance rhythm and spatial application as a process of mutual shaping with audience perception. Model analysis further indicates that the aesthetic coherence of crossmedia performances is built upon actors' controllable adjustment of emotional rhythm and spatial density. Current research, limited by sample

size and the range of repertoire types, does not cover the differences between cross-cultural contexts and virtual media performances. Future research could expand and validate the applicability and transferability of this transition mechanism within immersive video and virtual actor training systems by integrating the digital twin-knowledge graph fusion used to model complex fiber morphology and yarn spatial distribution. This cross-domain application ensures that the structural precision required for industrial heritage preservation and the technical monitoring of automated machinery provides a robust framework for enhancing the authenticity of virtual performance environments.

AUTHOR CONTRIBUTIONS

He Xu designed, collected and analyzed the data, and drafted the manuscript. He Xu conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. He Xu participated fully in the work, take public responsibility for appropriate portions of the content, and agreed to be accountable for all aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved.

CONFLICTS OF INTEREST

The author declares no conflict of interest.

FUNDING

This research was supported by,

Project Name: Research on the Training Mode of “Amphibious Actors” under the Background of New Media

Project ID: 20JK0334

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

- [1] H. Wang, “A Preliminary Discussion on the Performance Characteristics and Artistic Treatment of the Opera Film “Yi meng Mountain”,” *Western Radio and Television*, vol. 42, no. 19, pp. 137-139, 2021, doi: 10.3969/j.issn.1006-5628.2021.19.045.
- [2] Y. Zhu, “On the similarities and differences between film and television performance and stage performance,” *Western Radio and Television*, vol. 42, no. 1, pp. 147-149, 2021.
- [3] H. Xu, “Research on the Reform of College Course Teaching Based on OBE Concept-Taking “Piano Stage Performance” Course as an Example,” *Drama House*, no. 17, pp. 187-189, 2025, doi: 10.3969/j.issn.1007-0125.2025.17.062.
- [4] B. He, “An Exploration of the Characteristics of Opera Stage Performance Art,” *Art & Technology*, vol. 38, no. 19, pp. 4-6, 2025, doi: 10.3969/j.issn.1004-9436.2025.19.002.
- [5] P. Fan and J. Shen, “Development and testing of a scale for assessing stage performance anxiety among fashion performance students,” *Liaoning Silk*, no. 2, pp. 110-112+96, 2025, doi: 10.3969/j.issn.1671-3389.2025.02.048.
- [6] Y. Wang, Z. Lyu, M. Han, et al., “Initial-stage performance evolution of solid oxide fuel cells based on polarization analysis,” *Electrochemical Society Transactions* sof2021, vol. 103, no. 1, pp. 1261-1269, 2021, doi: 10.1149/MA2021-03145mtgabs.
- [7] H. Wang, Y. Yang, B. Xi, et al., “Inter-stage performance and energy characteristics analysis of electric submersible pump based on entropy production theory,” *Petroleum Science*, vol. 21, no. 2, pp. 1354-1368, 2024, doi: 10.1016/j.petsci.2023.10.032.
- [8] S. Jafar R M, Y. Sun, B. Niu, et al., “Revealing the secrets of metaverse technology adoption for sustainable performance via dual-stage SEM-ANN analysis,” *International Journal of Human-Computer Interaction*, vol. 41, no. 13, pp. 8296-8313, 2025, doi: 10.1080/10447318.2024.2407686.
- [9] M. Raouf A and G. Al-Ghamdi S, “Framework to evaluate quality performance of green building delivery: construction and operational stage,” *International Journal of Construction Management*, vol. 23, no. 2, pp. 253-267, 2023, doi: 10.1080/15623599.2020.1858539.
- [10] Q. Zhong, E. Xu, T. Jia, et al., “Dynamic performance and control accuracy of a novel proportional valve with a switching technology-controlled pilot stage,” *Journal of Zhejiang University-SCIENCE A*, vol. 23, no. 4, pp. 272-285, 2022, doi: 10.1631/jzus.A2100463.