Research on Optimization of Stage Presentation Effects and Intelligent Assisted Decision-Making for Music Performances Based on Multimodal Data Fusion

Kai Yao, Yan Shi
College of Fine Arts, Jingchu University of Technology, Jingmen 448000, Hubei, China
Corresponding author: Yan Shi (e-mail: hsh6666688888@126.com)

ABSTRACT The integration of digital technologies into performing arts has created opportunities for quantitative evaluation and intelligent optimization of stage presentation effects. This study proposes a multimodal-data-fusion-based framework for stage-effect evaluation and intelligent decision support in music performance scenarios. A multimodal acquisition platform is developed to synchronously collect electromyography signals, inertial measurement unit data, audio spectral information, spatial trajectories, and audience physiological responses. To address temporal asynchrony and heterogeneous feature distributions, an attention-based feature-fusion strategy is employed to construct a multidimensional stage-effect evaluation model. Furthermore, a reinforcement-learning-based decision-support mechanism dynamically coordinates lighting, acoustic, and visual presentation elements. Experimental results demonstrate that the multimodal framework significantly improves effect-prediction accuracy and enhances audience immersion and emotional engagement. The proposed approach provides methodological references for multimodal signal fusion, intelligent sensing, human-centered interaction, and adaptive communication environments.

INDEX TERMS multimodal data fusion, intelligent decision support, electromyography signals, intelligent sensing, adaptive communication environments

I. INTRODUCTION

AS a whole art, the efficacy of a musical stage presentation extends beyond the technical proficiency and interpretive depth of the individual performer. It is a synergistic manifestation of multiple variables, including dynamic illumination, acoustic engineering, spatial orchestration, vestiary design, and real-time spectator engagement. Traditional stage presentation often relies on the subjective experience of directors and designers, lacking a systematic understanding of the non-linear interactions between complex elements such as material morphology and material spatial distribution within the scenographic environment. This research addresses that gap by establishing a quantifiable technical path that integrates industrial manufacturing excellence into the intelligent creation of digital stage art. With the rapid development of sensor technology, artificial intelligence and Internet of things, the introduction of multimodal data acquisition and analysis in the performing arts is achieving the quantitative description and dynamic optimization of stage presentation effects has become a current hot topic in interdisciplinary research. During a musical performance, the movements of the performer's body, the micro expressions on his face, the audio signals from the musical instruments, the dynamic lighting and shadows on the stage, as well as the real time feedback from the audience is a highly coupled and complex system. How to extract good features from this heterogeneous, multi-source data and reveal its intrinsic connection to the final artistic impact are the aspect of the scientific questions for reaching intelligent stage control. Existing research mainly considers the single-modal analysis such as the evaluation of performance quality only from audio or the evaluation of dance postures only from motion capture without sufficient in-depth exploration of the synergistic effects of multiple information channels. Furthermore, optimization decisions for stage effects are often behind the performance process, and real-time intervention and adjustment are very difficult. To address these challenges, this paper proposes an intelligent auxiliary decision-making framework that leverages multimodal data fusion, incorporating material morphology parameters and material spatial distribution models to optimize the synergistic presentation of musical performance and stage aesthetics. By integrating these technical attributes into the digital environment, the system provides a quantifiable path for achieving high-fidelity interaction between material textures and performing arts. The research aims to construct an analytical framework with such functions as acquisition, processing, and fusion of the various performance-related data in real time and, on this basis, establish an effect evaluation model. Finally, intelligent algorithms make dynamic presentation optimizations for the stage. The structure of the paper is as follows: Part II reviews relevant research progress at home and abroad, clarifying the starting point of this research, Part III explains multimodal data acquisition, feature extraction, fusion strategy and intelligent decision-making method in detail, Part IV verifies the effectiveness of the proposed method through experiments, discusses the research results in detail, Part V summarizes the research results and looks forward to the future.

II. RELATED WORK

Multimodal research on music and performing arts has gradually attracted the attention of the academic community. Researchers have explored the complex interaction mechanism of elements such as music, movement, space, costume and audience perception in stage presentation from different dimensions. Moura et al. [1] explored the role of music style and movement quality in audience perception, and how music expertise regulates this audiovisual integration perception process. Doğantan-Dack [2] analyzed how performance as a research medium generates new knowledge that cannot be obtained through traditional music score analysis, and explored the potential impact of this interdisciplinary integration on the development of music theory, music performance practice and music education. Borgianni et al. [3] proposed and implemented a software-defined wide area network architecture based on reinforcement learning, aiming to provide high-quality, low-latency network connections for online music performances. Li et al. [4] found that specific musical melodies and rhythms often correspond to specific dance movements and body postures, while the performer's emotional investment and stylized expression integrate these elements into an infectious artistic whole. Ollora-Triana et al. [5] believed that applying semiotic processes to dance performance analysis can profoundly reveal the internal mechanism of dance as a complex expressive system. Wahid et al. [6] revealed how dancers perceive space, how they transform the physical and abstract qualities of space into physical movements, and how this "embodiment" process shapes the final performance and the audience's perception. Kongshi et al. [7] proposed a method that effectively enhances physical interaction in dance performance. Le Goff [8] explored how performance costume design transcends mere aesthetic decoration. Wiese et al. [9] extracted key findings on the role of attention focus in specific dance movements and overall performance quality. Engler et al. [10] found that the presence of the audience has a significant impact on dance performance. Existing research lacks a systematic exploration of deep integration and collaborative optimization of data from different modalities. This paper aims to construct a comprehensive analysis framework that integrates multi-source data, explore data-driven optimization paths for stage presentation, and provide theoretical support and practical tools for the artistic expression and technical realization of music performance.

III. METHODS

A. DESIGN AND SYNCHRONIZATION PROTOCOL OF MULTIMODAL DATA ACQUISITION PLATFORM

To achieve comprehensive capture of diverse information during musical performances, this study designed an integrated multimodal data acquisition platform. This platform includes a motion capture module, an audio acquisition module, a physiological feedback module, and an environmental perception module. The motion capture module uses a nine-axis inertial measurement unit to record the three-dimensional acceleration, angular velocity, and geomagnetic orientation data of the performer's major joints at a sampling rate of 100 frames per second. A total of 15 sensor nodes are positioned at key locations such as the head, shoulders, elbows, wrists, hips, knees, and ankles. The audio acquisition module consists of a distributed microphone array deployed around the stage and in the audience area, simultaneously picking up the performance's sound field and spatial reverberation information at a sampling rate of 44.1 kHz. The physiological feedback module collects the performer's electromyography (EMG) signals and the audience's heart rate variability and electrodermal response using wearable sensors, with a sampling frequency of 256 Hz. The environmental perception module is responsible for recording the temporal parameters of stage lighting intensity, color temperature changes, and projected visual elements. To ensure the consistency of time sequence of multisource data, all acquisition devices are connected to a unified time synchronization server, and microsecond-level synchronization is achieved through a precise time protocol. Hardware trigger calibration is performed before each acquisition. The data stream is aggregated to the central storage system via the fieldbus, providing a high-quality raw data foundation for subsequent analysis [11-12] .

B. DATA PREPROCESSING AND MULTIMODAL FEATURE EXTRACTION

The raw data contains a large amount of noise and redundant information, requiring preprocessing to improve the signal-to-noise ratio and feature effectiveness. For inertial measurement unit (IMU) data, a Kalman filter algorithm is first used to fuse accelerometer and gyroscope information to calculate the precise posture angles of each joint of the performer, and then calculate joint angular velocity, angular acceleration, and intersegment coordinated motion index. Audio data is converted into a time-spectrum graph through short-time Fourier transform, and acoustic features such as Mel frequency cepstral coefficients, spectral centroid, bandwidth, and harmonic noise ratio are extracted. Simultaneously, a peak detection algorithm is used to identify the note initiation point and changes in playing speed. Electromyographic (EMG) signals are bandpass filtered and full-wave rectified to extract the root mean square value in the time domain and the median frequency in the frequency domain to characterize muscle activation and fatigue state. Audience skin conductance data is analyzed using continuous decomposition to separate phase-dependent skin conductance response and tension-dependent skin conductance level; the former reflects instantaneous emotional arousal events. All features are normalized after extraction to eliminate dimensional differences. To address the issue of inconsistent sampling rates for different modal data, a multi-rate synchronization strategy is employed. High-frequency audio and physiological signals are first processed in their native resolution to extract spectral and temporal features (e.g., MFCCs and HRV indices). Subsequently, these pre-computed feature vectors—rather than the raw waveforms—are resampled to a consistent 100 Hz frame rate using cubic spline interpolation. This approach prevents aliasing and preserves the high-frequency components essential for technical completion analysis while ensuring temporal alignment across modalities.

C. FEATURE-LEVEL FUSION STRATEGY BASED ON ATTENTION MECHANISM

Heterogeneous modal features exhibit complex nonlinear complementarity and redundancy relationships, making it difficult to fully exploit their synergistic effects through simple concatenation or weighted averaging. This study introduces a multi-head attention mechanism to construct a feature fusion module, mapping feature vectors from different modalities to the same semantic space and strengthening the contribution of key modalities through adaptive weight allocation. Specifically, preprocessed motion, audio, physiological, and environmental features are encoded into fixed-dimensional embedding vectors through independent fully connected networks, forming a feature sequence. The attention module uses the embedding of one modality as a query and the embeddings of other modalities as keys and values to calculate cross-modal attention weights, generating an enhanced representation that incorporates contextual information. By stacking multiple attention heads, the model can capture interaction patterns between modalities from different subspaces. The fused feature vector is then input to the downstream task network. The advantage of this strategy lies in its ability to adjust the weights of each modality in real time according to the dynamic changes in the performance process. For example, in the climax of a piece of music, the weights of audio features and the performer's large-scale movements will increase accordingly; while in lyrical passages, the audience's physiological feedback features may receive higher attention to reflect the depth of emotional resonance. The fusion module is optimized end-to-end during training, and the loss function is backpropagated in conjunction with the evaluation task objective.

D. CONSTRUCTION OF A MULTI-DIMENSIONAL EVALUATION MODEL FOR STAGE PRESENTATION EFFECTS

Stage presentation effect is an abstract, subjective concept, difficult to quantify with a single indicator. This study constructs an evaluation model from four dimensions: performer expressiveness, technical completion, audience immersion, and visual aesthetics. Each dimension includes several calculable indicators. Performer expressiveness is comprehensively measured by the fluidity of movement, the intensity of emotional expression, and musical synchronicity. Movement fluidity is calculated based on joint angle range and smoothness of motion; the intensity of emotional expression is based on audio dynamic range and electromyographic activity levels; and musical synchronicity is measured by the cross-correlation peak between the movement beat and the audio beat. Technical completion focuses on performance accuracy and movement stability, including indicators such as pitch deviation, rhythm deviation, and joint trajectory repeatability. Audience immersion is characterized by the synchronicity index of group skin conductance response and the low-frequency to high-frequency ratio change in heart rate variability, reflecting the degree to which the audience is attracted to the performance. Visual aesthetics integrates the spatial entropy indicators of lighting color temperature and stage composition, combined with the contrast of costume color and background for scoring. Each dimension's indicators are first dimensionality reduced using principal component analysis, and then the multidimensional features are mapped to a comprehensive effect score using a support vector regression model. The training data uses a weighted average of expert scores and audience questionnaire scores as labels to ensure that the evaluation results are both professional and acceptable to the general public. The expert panel consisted of five professionals: two senior stage directors with over 15 years of experience in performing arts, two musicology professors specializing in performance theory, and one digital media artist. Scoring was conducted based on a standardized rubric evaluating "Dynamic Expressiveness" (weighted 40%), "Technical Accuracy" (30%), and "Emotional Impact" (30%) to minimize subjectivity.

E. INTELLIGENT DECISION SUPPORT SYSTEM BASED ON REINFORCEMENT LEARNING

To achieve the stage presentation dynamic optimization, this study builds an intelligent auxiliary decision-making system. This system takes the multimodal fusion features of the present performance state as input and makes an output as control commands (lighting parameters, sound equalization and switching of visual content of the stage) aiming to maximize the accumulated evaluation score of the stage effect. The decision problem is modeled as a Markov decision process whereby the state space is fusion features and historical actions, the action space is a discrete set of more control options, and the reward function is the change in effect score calculated in real time by the evaluation model. A deep deterministic policy gradient algorithm is used to learn the policy network. This algorithm can be used for continuous action spaces and is good for fine-tuning continuous variables such as lighting brightness and color temperature. The training process is carried out in a simulation environment which includes state transition models and reward models for standard scenarios of music performance. Through many iterations, the optimal control strategy is learned by the agent under various performance styles and emotional trajectories. In practical applications, the system operates online and generates predictive control commands with a temporal resolution of 100 milliseconds. While major artistic shifts require manual confirmation via a pre-set 'confidence threshold' interface, the system executes low-latency micro-adjustments (such as subtle lighting intensity and sound equalization) autonomously. This hybrid architecture reduces the decision-path latency to sub-second levels, ensuring the stage response remains dynamic and synchronized with the live performance. However, recognizing that heart rate variability (HRV) requires longer windows for statistical validity, the physiological feedback module utilizes a sliding 60-second window, updated every two seconds, to ensure that the HRV metrics used in the decision-making process are physiologically meaningful and stable. These suggestions are then carried out, after manual confirmation, a balance of intelligence and safety. The system also has self-learning capabilities where the data generated during actual performances is included in the training, and decision-making strategy is continuously optimized.

IV. RESULTS AND DISCUSSION

A. RESULTS OF MULTIMODAL DATA ACQUISITION AND FEATURE ANALYSIS

To verify the effectiveness of the data acquisition platform and the accuracy of feature extraction, this study deployed the aforementioned system in an experimental piano recital. The repertoire covered three styles: Classical, Romantic, and Modern, lasting approximately 45 minutes. A total of 270,000 effective inertial measurement unit (IMU) data frames, 119 million audio data sampling points, and 691,200 physiological feedback data frames were acquired. Through preprocessing and feature extraction, a total of 67,500 aligned multimodal feature sequences were obtained. Statistical analysis of key features showed significant differences in the performer's motion characteristics under different musical styles. Table 1 shows a comparison of the mean and standard deviation of typical joint motion parameters and audio features under the three styles. As can be seen from the data in Table 1, the Romantic style had the highest mean wrist joint angular velocity, reaching 142.6 degrees per second, accompanied by a large standard deviation, indicating that wrist movements in this style of performance are more free and varied. The Classical style showed a relatively convergent range of wrist joint motion, while the Modernist style exhibited the largest standard deviation of angular velocity, reflecting the unpredictability and irregularity of its movements. In terms of audio characteristics, the Romantic style exhibits the lowest spectral centroid, suggesting a more low-frequency, heavy timbre, while the Modernist style boasts the highest spectral centroid and the largest bandwidth, consistent with its pursuit of sonic complexity. Audience skin response data shows the highest frequency and amplitude in the Romantic style segments, indicating that this style more easily evokes emotional arousal in the audience. These differences provide the underlying basis for the subsequent fusion model to distinguish between different performance styles.

TABLE 1. Statistical Comparison of Typical Characteristics under Different Musical Styles

Musical Style	Mean angular velocity of the wrist joint ($^{\circ}$/s)	Standard deviation of wrist joint angular velocity ($^{\circ}$/s)	Mean centroid of audio spectrum (Hz)	Average audio bandwidth (Hz)	Mean amplitude of audience skin conductance response (μS)
Classicism	118.3	45.2	823.6	1562.4	0.32
romanticism	142.6	67.8	756.2	1783.9	0.58
Modernism	130.7	82.5	1047.5	2206.1	0.41

B. PERFORMANCE EVALUATION AND COMPARISON OF FUSION MODELS

To verify the effectiveness of the attention-based multimodal fusion strategy, it was compared with four baseline models: a single-modal model using only motion features, a single-modal model using only audio features, and a simple feature concatenation fusion model, and a Canonical Correlation Analysis (CCA)-based fusion model representing traditional state-of-the-art cross-modal alignment. All models aimed to predict the combined expert-audience score, and their coefficient of determination (COD) and root mean square error (RMSE) were evaluated using five-fold cross-validation. The experimental dataset consisted of 800 10-second segments from the original collected data. Each segment was scored by five experts and thirty audience members, and the weighted average was used as the ground truth label. During model training, the input was the complete multimodal feature sequence of the corresponding segment. The evaluation results are summarized in Table 2. The data shows that the attention-based fusion model achieved a COD of 0.81 and an RMS of 0.47, significantly outperforming all baseline models. The single motion modal model and the single audio modal model performed similarly, with CODs of 0.58 and 0.61, respectively, indicating that each carried some, but not all, predictive information, demonstrating information complementarity. The simple splicing and fusion model improved its coefficient of determination to 0.72, indicating that multimodal combination is indeed effective. However, its performance is still lower than that of the attention mechanism model, proving that simple splicing fails to fully explore the complex interactions between modalities. The attention mechanism, through dynamic weight allocation, can strengthen effective modalities and suppress noisy modalities according to the characteristics of different segments, thereby achieving better prediction results. Visual analysis of attention weights revealed that in technically demanding fast-paced segments, the weight of motion features increases; in lyrical segments, the weight of audio features and audience physiological features increases. This aligns with intuition from performance practice and verifies the interpretability of the model.

TABLE 2. Comparison of the predictive performance of different models for stage effect scores

Model	Coefficient of determination (R2)	Root Mean Square Error (RMSE)
Single motion mode model	0.58	0.76
Single audio modal model	0.61	0.72
Simple splicing and fusion model	0.72	0.61
CCA-based fusion model	0.75	0.56
Attention mechanism fusion model	0.81	0.47

C. VERIFICATION OF THE EFFECTIVENESS OF THE INTELLIGENT DECISION-MAKING AUXILIARY SYSTEM

To examine the application effect of the intelligent assisted decision-making system in real-world scenarios, a comparative experiment was designed. Two versions of the same Chopin Nocturne performed by the same pianist were selected, using a traditional manual control mode without systemic intervention and a control mode based on the intelligent assisted decision-making system of this study. Forty audience members were invited and divided into two groups of twenty each. Each group watched the two recorded video versions and rated their immersion and overall artistic appeal in real time using a continuous sliding bar from 1 to 10. Skin conductance data were also recorded for both groups as objective physiological indicators. In the intelligent system control mode, the brightness and color temperature of the lights, as well as the changes in the abstract graphics projected behind the stage, were dynamically adjusted based on real-time multimodal characteristics to enhance the emotional fluctuations of the music. Table 3 shows the comparison results of the average subjective scores and objective physiological indicators of the audience under the two modes. As can be seen from the data in Table 3 , the audience's immersion score was on average 1.3 points higher and their artistic appeal score was on average 1.5 points higher under the intelligent control mode. In terms of objective physiological indicators, the synchronicity index of the audience's skin conductance response significantly improved under the intelligent control mode, reaching 0.67, indicating that the audience's emotional fluctuations were more consistent and reflected a stronger collective resonance; while in the traditional mode, the index was only 0.42, indicating that the audience's response was more dispersed. The ratio of low-frequency to high-frequency heart rate variability also decreased in the intelligent mode, suggesting that sympathetic nervous system activity was relatively enhanced and the audience was in a higher state of arousal. These results indicate that the intelligent auxiliary decision-making system based on multimodal data fusion can effectively improve the overall effect of stage presentation and enhance the audience's sense of participation and emotional engagement.

TABLE 3. Comparison of audience reactions under traditional control and intelligent assisted decision-making control modes

Regulation mode	Average immersion rating (1-10)	Average artistic appeal score (1-10)	Audience skin response synchronicity index	Audience heart rate variability low-frequency/high-frequency ratio
Traditional manual regulation	6.8	6.5	0.42	1.83
Intelligent Assisted Decision-Making and Regulation	8.1	8.0	0.67	1.51

V. CONCLUSION

This work explores the application of multimodal data fusion in musical stage presentations, constructing a comprehensive technical framework that integrates material morphology data and material spatial distribution models into the processes of data acquisition, feature extraction, and intelligent decision-making. By leveraging industrial manufacturing excellence, this architecture ensures a seamless transition from material-level structural parameters to the high-fidelity synthesis of digital performance environments. By the design of a high precision synchronous acquisition platform, a variety of multi-source data containing performer motion information, sound data, physiological data, and environment-related data were acquired. An attention-based feature fusion strategy was proposed to make good use of heterogeneous modal information to improve the accuracy of the evaluation of stage effects greatly. Furthermore, an intelligent auxiliary decision-making system was developed using reinforcement learning, which made the dynamic optimization and control of stage elements, such as lighting and sound, possible. Experimental results indicate that the proposed method not only improves the evaluation performance of traditional models with single modes but also can effectively improve the audience's immersion and emotional resonance in practical applications. The key contribution of this research is to use data-driven optimization techniques in the field of artistic creation, which provides a feasible solution for quantitative analysis and intelligent enhancement of stage performance, and also promotes the information interaction among performance arts and artificial intelligence technologies. Current research is still subject to some limitations. First of all, data collection is limited to experimental performance environments, in terms of limited size and limited stylistic coverage. Importantly, while the current study utilizes a piano recital as a validation case, the system's reliance on stationary performance data may not immediately translate to high-mobility performing arts such as dance. Further research is required to adapt the motion capture and fusion algorithms to account for the rapid spatial displacement and complex limb trajectories characteristic of more dynamic stage arts. Second, the result of intelligent decision-making is now presented in form of suggestions, that would require manual confirmation before its actuation. Future research could investigate fully automated closed-loop control at certain stages to achieve better response speed. Furthermore, the real-time computational overhead of the attention mechanism in the fusion model has a certain effect on the system latency. Use cases in future research will focus on model lightweighting and edge computing deployment in order to achieve more stringent real-time requirements. With the continuous development of wearable sensors, spatial computing, and generative artificial intelligence, multimodal data fusion—incorporating morphology parameters and spatial distribution models—is on the cusp of producing innovative applications in the performing arts. These advancements leverage industrial manufacturing excellence to introduce a forward-thinking level of tactile expressiveness and material-driven interactivity to stage presentations.

AUTHOR CONTRIBUTIONS

Conceptualization –Kai Yao and Yan Shi; methodology – Kai Yao and Yan Shi; investigation – Kai Yao and Yan Shi; writing-original draft preparation – Kai Yao and Yan Shi. All authors have read and agreed to the published version of the manuscript.

CONFLICTS OF INTEREST

The authors declare no conflict of interest.

FUNDING

This research received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] N. Moura, P. Fonseca, P. Vilas-Boas J, et al., “Increased body movement equals better performance? Not always! Musical style determines motion degree perceived as optimal in music performance,” Psychological Research, vol. 88, no. 4, pp. 1314-1330, 2024, doi: 10.1007/s00426-024-01928-x.

[2] M. Doğantan-Dack, “Expanding the scope of music theory: Artistic research in music performance,” Zeitschrift der Gesellschaft für Musiktheorie [Journal of the German-speaking Society of Music Theory], vol. 19, no. 2, pp. 13-42, 2022, doi: 10.31751/1169.

[3] L. Borgianni, S. Giordano, and C. Chafe, “Democratizing networked music performance with a rl-based sd-wan,” IEEE Communications Magazine, vol. 62, no. 12, pp. 26-32, 2024, doi: 10.1109/MCOM.001.2400171.

[4] X. Li, P. Jirajarupat, and Q. Jin, “A Studying the Music, Dance, and Performance Styles in Shi Jing-Feng,” International Journal of Sociologies and Anthropologies Science Reviews, vol. 4, no. 5, pp. 349-358, 2024, doi: 10.60027/ijsasr.2024.4593.

[5] N. Ollora-Triana, M. Corbi, A. Jiménez-Eguizábal, et al., “Semiologic process in the performance of a dance performance: a critical review,” Arte, individuo y sociedad, vol. 36, no. 1, pp. 199-210, 2024, doi: 10.5209/aris.90332.

[6] A. Wahid N, N. Michael N, T. Saearani M F, et al., “Embodying Space through Improvisation in Site-Specific Dance Performance,” Harmonia: Journal of Arts Research and Education, vol. 25, no. 1, pp. 188-197, 2025, doi: 10.15294/ harmonia.v25i1.14728.

[7] M. Kongshi, D. Pan, and M. Wang, “Enhancing Dance Performance for Body Motion Interaction Through Swarm Intelligence and Deep Learning,” International Journal of Swarm Intelligence Research (IJSIR), vol. 16, no. 1, pp. 1-18, 2025, doi: 10.4018/ IJSIR.383942.

[8] L. Le Goff, “Costumes that tell multispecies stories: Critique of sympoiesis: A bio-inspired dance performance,” Studies in Costume & Performance, vol. 8, no. 2, pp. 241-262, 2023, doi: 10.1386/scp_00099_1.

[9] R. Wiese K, P. Ambegaonkar J, and J. Martin, “Impact of attentional focus on dance performance: A critically appraised topic,” Journal of Sport Rehabilitation, vol. 33, no. 2, pp. 128-134, 2023, doi: 10.1123/jsr.2023-0047.

[10] C. Engler, F. Pelzer, C. Kaczmarek, et al., “Effects of spectators on the performance of a dance routine,” German Journal of Exercise and Sport Research, vol. 54, no. 3, pp. 402-407, 2024, doi: 10.1007/s12662-023-00929-z.

[11] K. Maguire-Rosier, N. Abe, and F. Andreallo, “What other movement is there?: rethinking human-robot interaction through the lens of dance performance,” TDR, vol. 68, no. 1, pp. 87-103, 2024, doi: 10.1017/S1054204323000552.

[12] E. Herdiani, Y. Heriyawati, and A. Iskandar, “Back to Ritual: Virtual Indonesian Dance Performance on Social Media During Covid-19 Pandemic,” Journal Of Arts And Literature, vol. 3, no. 5, pp. 26-36, 2024, doi: 10.5281/zenodo.13998784.