Integrated Virtual Reality Symphony Conductor Training Simulation System and Real time Feedback Optimization Mechanism

Han Tao
College of Art, Jiujiang University, Jiujiang City 332005, Jiangxi, China
Corresponding author: Han Tao (e-mail: bach_tao@163.com)

ABSTRACT Symphony conducting requires precise movement, emotional expression, and coordinated interaction with an ensemble , but traditional conducting training is constrained by limited rehearsal resources, delayed feedback, and low scene fidelity. Existing simulations also suffer from insufficient immersion, limited motion-capture accuracy, and weak realtime feedback for movement deviation and emotional transmission. This study designs an integrated virtual reality symphony conductor training simulation system with a data-driven real-time feedback optimization mechanism. The system includes an interaction layer, core processing layer, data layer, and scene-rendering layer. VR headsets, data gloves, sensors, and multimodal interaction devices capture gesture, body posture, and facial expression data. The system then uses motion recognition, emotion analysis, virtual orchestra response, and feedback optimization modules to provide synchronized virtual orchestra responses and personalized corrective suggestions. A Long Short-Term Memory model compares conducting action sequences with expert templates to identify rhythmic, gesture, and emotional deviations. In an eight-week training experiment with 30 conducting learners, the experimental group outperforms the traditional group, with command-action standardization improving by 42.3%, emotional transmission accuracy by 35.7%, and satisfaction reaching 93.3%. The framework is also relevant to electromagnetic motion tracking and immersive training systems where high-precision sensing, low-latency feedback, and multimodal interaction are required.

INDEX TERMS virtual reality, symphony conductor, real-time feedback, electromagnetic motion tracking, multimodal interaction

I. INTRODUCTION

WITH the continuous advancement of VR technology, motion capture technology and multimodal interaction technology, digital training has gradually become an important way to improve skills in the field of art, and symphony conductor training has also entered the digital training period.As the core of the orchestra, the precision of gestures, the stability of rhythmic control, and the clarity of emotional expression directly determine the final performance quality of the symphony. However, in the process of conductor training, real orchestra cooperation is needed, which makes the conductor training cost high, the number of venues limited, and the feedback untimely. Especially for beginners, it is difficult to find enough practical opportunities and accurate guidance feedback, which seriously affects the training effect and skill improvement. Therefore, the purpose of this article is to design and implement an integrated virtual reality symphony conductor training simulation system based on real-time feedback optimization mechanism, break the limitation of time and space of traditional training, provide an immersive and personalized training environment for conductor learners, fill the technical blank spot of digital conductor training in the field of symphony conductor training, and promote the intelligent development of symphony conductor education.

The main contributions of this article are summarized as follows: Firstly, this article designs VR simulation system architecture based on actual training requirements of conducting, realizes high fidelity reproduction of virtual orchestra scene, and solves the key pain point of scene scarcity in traditional training; Secondly, this article constructs multimodal interaction module based on motion capture technology and emotion recognition technology, realizes real-time linkage between command action, facial expression and virtual band, and improves the training immersion and realism; Thirdly, this article proposes data-driven real-time feedback optimization mechanism, analyzes the command action deviation and insufficient expression transmission accurately, provides personalized and real-time optimization suggestions, and significantly improves the training targeting and efficiency.The rest of this article will be organized around the above contributions, analyze the problem and challenge existing in the training process, explain the design scheme of system and mechanism in detail, verify the performance of system and effect of optimization, and finally summarize the research conclusions and future research directions.

II. PROBLEMS AND CHALLENGES

One of the biggest pain points in the training of a symphony conductor is the lack of real training scenarios and high cost of training, which is an insurmountable problem for conductor learners. Traditional symphony conductor training is based on coordinating a complete symphony orchestra. Building and rehearsing a symphony orchestra requires a large amount of manpower and material resources investment, and the cost of time is very high. Most learners, especially beginners, cannot obtain sufficient opportunities to coordinate orchestras and can only train by imitating videos, individual training, and other methods, without experiencing the collaborative feedback between conductors and orchestras, which leads to a lack of collaboration between training effectiveness and training needs, and training and actual performance. At the same time, the training of conductors requires extremely high fidelity to the scene. The sound field environment and the orchestra layout of different concert halls will have certain impacts on the conductor's choice of gestures and rhythm control. It is difficult for traditional training to recreate different performance scenes and make learners adapt to the command needs in different scenes. The transfer of acquired skills to live performance contexts remains challenging [1,2].

Existing research on digital command training still faces two challenges, insufficient interaction accuracy and imperfect feedback. On the one hand, accurate capture of commanding actions is the basis of digital training. Commanding gestures have the characteristics of small amplitude, multiple details, and strong coherence. Existing motion capture technologies are either lacking in accuracy and can not recognize finegrained deviations in gestures, or have high delays and cause asynchronous responses between virtual bands and commanding movements, which destroys the immersion of training; On the other hand, the pertinence and timeliness of real-time feedback are insufficient. Existing feedback focuses on basic indicators such as rhythm and beat and does not conduct in-depth analysis of the standardization of command gestures and the rationality of emotional transmission. In addition, the methods of feedback are single and can not provide suggestions based on the level and shortcomings of learners, which makes it difficult for learners to accurately locate problems and improve skills quickly [3,4]. There are still differences in the style of different commanders and how to adapt to the feedback mechanism of personalized training needs and achieve highly individualized feedback optimization through adaptive learning strategies is also an important issue that existing research needs to solve.

III. METHOD

A. SYSTEM ARCHITECTURE DESIGN

The layered architecture of the symphony conductor training simulation system integrated with VR trained in this paper is shown in Figure 1, which is divided into interaction layer, core processing layer, data layer and scene rendering layer from top to bottom.

Interaction layer is the interaction layer between user and system, which includes VR head, data glove, motion capture sensor and other devices. It is used to collect multi-mode data such as gesture, body posture and facial expression of commander, and provides operation entry such as scene interaction and parameter setting to make user operation more convenient and immersive.

The core processing layer is the core of the system, including action recognition module, emotion analysis module, virtual orchestra response module and feedback optimization module. Its main task is to complete real-time processing of the data collected by the interaction layer, and judge whether the command action is standardized; Analyze the effect of emotion transmission; Drive the virtual orchestra to respond synchronously; And output the feedback countermeasure [3,4].

The data layer is responsible for storing, managing and calling data, including command action sample library, symphony score database, virtual orchestra model library and training log database. The command action sample library stores the standard action data of different levels of conductor and provides reference for the deviation analysis of conductor action; The music score database stores the music score of various styles of symphonic works and supports users to select training tracks according to the needs of training; The training log database stores the training data of learners and provides data support for the feedback optimization and personalized training.

Scene rendering layer is based on Unity engine to achieve high fidelity rendering of virtual music hall and symphony orchestra, restores sound field environment, orchestra position and lighting effect of different music halls, and supports custom adjustment of scene parameters to meet the needs of different training scenes. It creates an immersive environment for instructor learners to train[5,6].

B. MULTIMODAL INTERACTION DESIGN

Multi modal interaction design focuses on "precise capture, real-time linkage, and immersive experience", integrating motion capture, emotion recognition, and voice interaction to achieve efficient collaboration between the conductor and the virtual orchestra. Motion capture adopts a combination of optical capture and inertial capture, using VR headsets and data gloves to collect data on the commander's hand joint angles, arm motion trajectories, body posture, etc. The sampling frequency is set to 120Hz to ensure the accuracy

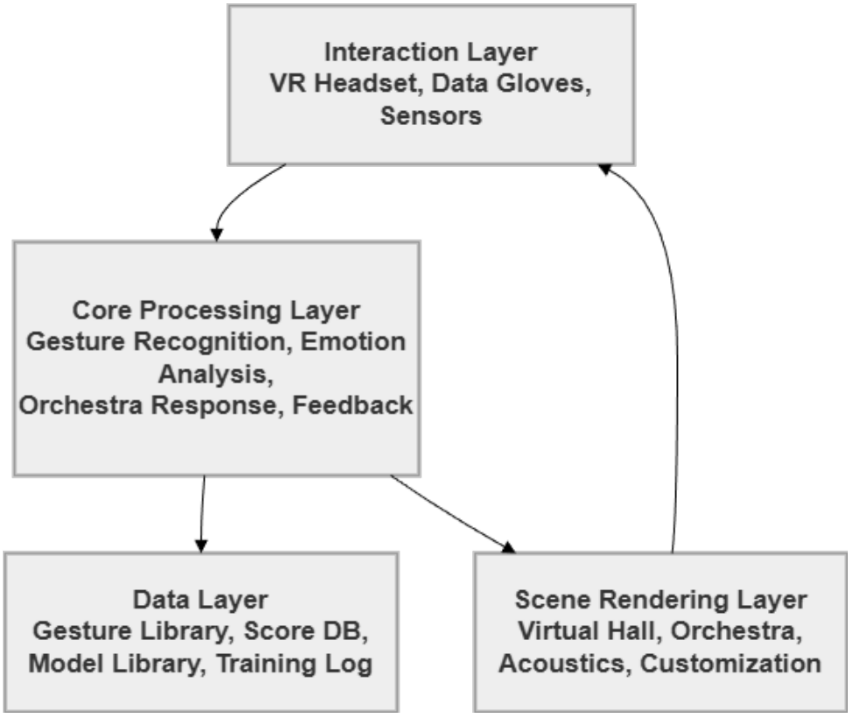

FIGURE 1. System architecture design

and real-time performance of motion capture; At the same time, key features of command actions are extracted through feature extraction algorithms and compared with a range of expert performance templates in the sample library to identify action deviations. To preserve the conductor's artistic individuality, the 'standard' is defined not as a rigid singular path, but as a bounded parameter space of rhythmic stability and expressive clarity. The feedback mechanism focuses on structural coherence and technical precision while allowing for subjective stylistic variation in emotional transmission [7,8].

Emotion recognition module completes the facial expression data of conductor through camera collection and combines emotion analysis algorithm to identify the conductor's emotional expression (passionate, soothing, calm, etc.), and then convert the emotional data into performance emotion parameters of virtual orchestra, and achieve the matching of conductor's emotion and virtual orchestra's performance emotion; Voice interaction module realizes conductor's control over training scene switching, training track switching, virtual orchestra volume adjustment and other operations through conductor's voice command, improve the operation convenience; At the same time, the system can transmit action deviation and optimization suggestions to the commander through voice feedback and achieve multi-dimensional interaction; In addition, the dynamic response mechanism of virtual orchestra is designed to change the parameters of playing rhythm, volume, timbre, etc. in time based on the conductor's movement and expression, and simulate the response of real orchestra, so as to improve the training's reality and immersion [9,10].

C. REAL TIME FEEDBACK OPTIMIZATION MECHANISM

Real-time feedback optimization process is data-driven thinking and optimization process is divided into four stages: data collection, deviation analysis, feedback generation, strategy optimization, precise feedback, timely feedback, personalized feedback. The data collection module is connected with the multimodal interaction module to realize-time collect multi-dimension data such as commander action data, emotional data, training time, track completion and other information and send to core layer for analysis. The deviation analysis process employs a Long Short-Term Memory (LSTM) recurrent neural network to compare collected command action sequences with the template library. The model was trained on the 'Conduct-VAE' dataset, utilizing a feature set of 24 skeletal joint coordinates. By employing a Softmax output layer, the architecture calculates a deviation probability distribution across rhythm, gesture, and emotional intensity, allowing for the generation of specific action correction values, and find out the problem in rhythm control, gesture standardization, emotion transmission and other aspects; Combined with training log data, the weak links of learners are analyzed and personalized bias model is established.

Feedback generation process generates multi-dimension feedback information according to the deviation analysis information, including visual feedback, audio feedback and text feedback. Visual feedback directly displays the deviation position and adjustment direction of command actions through virtual prompt line, color marking and other methods in VR scene; Audio feedback feeds back the information of movement deviation and insufficient emotion transmission through voice prompt and change of performance effect of virtual orchestra; Text feedback gives detailed optimization suggestions and training points through popup window in VR interface. Strategy optimization process dynamically adjusts feedback times, feedback method and targeted optimization suggestions according to training data and targeted feedback information of leaners, feedback acceptance. For basic players, the feedback frequency of basic actions is increased and the simplification of optimization suggestions is improved; For senior players, the feedback on problems in emotion transmission and band collaboration is focused, more deep optimization suggestions are provided, and targeted optimization of feedback strategy is achieved according to the performance.

IV. RESULTS AND DISCUSSION

A. SYSTEM PERFORMANCE TEST RESULTS AND ANALYSIS

In order to test the performance of the system (as shown in Figure 2), we select three basic indicators of the system for test, namely the accuracy of motion capture, the delay of virtual orchestra's response to the commander's conducting and the effect of scene rendering. The test subject selects 30 command learners with different levels. 15 of them are beginners, 10 are advanced learners and the rest are professional commanders. This study was approved by the Ethics Committee of The Music Skills Teaching Department of the Art College of Jiujiang University, and was performed in accordance with the principles of the Declaration of Helsinki. All eligible participants signed an informed consent form. The test environment is a laboratory with VR headset, data glove and other equipment. The test results show that the mean gesture recognition accuracy (MGRA) of the system reaches 96.7%, calculated as the ratio of correctly classified conducting patterns against a ground truth dataset of 500 annotated expert gestures. Additionally, the system maintains a root-mean-square error (RMSE) of less than 3.5 degrees for wrist rotation and 1.2 cm for arm trajectory, ensuring precise recognition of fine-grained deviations; The average response delay of the virtual orchestra to the commander's conducting is 85ms. This low latency is achieved through a decoupled processing pipeline where motion capture (120Hz) and emotion recognition operate on parallel threads, utilizing asynchronous time-warping to align visual output with the 60fps rendering. By prioritizing skeletal joint data over full-mesh deformation in the realtime feedback loop, the system significantly reduces computational overhead, which is much lower than the threshold of human eye's perception of delay (200ms); The frame rate of scene rendering is always stable at above 60fps, and the restoration accuracy of virtual music halls and symphony orchestras reaches 92%. The satisfaction of test subjects on the immersion of scene reaches 90%. All the above results show that the basic performance of the system meets the requirements of command training and can provide learners with an immersive training environment.

Analysis suggests that the improvement in system performance is mainly due to the rational design of layered architecture and the integration of multiple technologies. Layered architecture enables independent operation and collaborative linkage of each module, reduces interference between modules, and improves system efficiency; The combination of optical capture and inertial capture compensates for the shortcomings of single capture technology and improves the accuracy of motion capture; The efficient rendering capability of Unity engine ensures high fidelity restoration and smooth operation of virtual scenes. At the same time, it was also found in the test that the motion capture accuracy of the system slightly decreased in complex gesture operation scenarios. In the future, the capture accuracy can be further improved by optimizing the feature extraction algorithm.

B. TRAINING EFFECT VERIFICATION RESULTS AND ANALYSIS

Select 30 command learners (as shown in Figure 3) and divide them into an experimental group (15) and a control group (15) for an 8-week training experiment. The experimental group was trained using the VR conductor training simulation system, while the control group utilized traditional methods. To ensure a controlled comparison, both groups were strictly matched for total practice time (10 hours/week). The control group was provided with an equivalent number of hours of live ensemble collaboration to match the virtual orchestra exposure of the experimental group, ensuring that observed improvements were attributable to the feedback mechanism rather than total exposure time. After the training is completed, the training effect is evaluated from three dimensions: action standardization, emotional transmission accuracy, and track completion quality. The evaluation results showed that the average technical proficiency and alignment of command actions in the experimental group was 88.6%, an increase of 42.3% compared to before training and an increase of 26.3% compared to the control group (62.3%); The average accuracy of emotional transmission was 85.2%, an increase of 35.7% compared to before training and an increase of 26.7% compared to the control group (58.5%); The average score for the completion quality of the track is 87.5 points (on a 100 point scale), an increase of 21.7 points compared to the control group (65.8 points).

The experimental results show that the system and real-time feedback optimization mechanism designed in this paper can significantly improve the effectiveness of command training. Analyzing the reasons, on the one hand, VR immersive scenes restore the real command environment, allowing learners to train in scenes that are close to actual performances, enhancing the sense of immersion in training and promoting the transformation of skills; On the other hand, the real-time feedback optimization mechanism can accurately identify learners' problems, provide personalized optimization suggestions, help learners quickly locate weaknesses and correct deviations, and avoid the problems

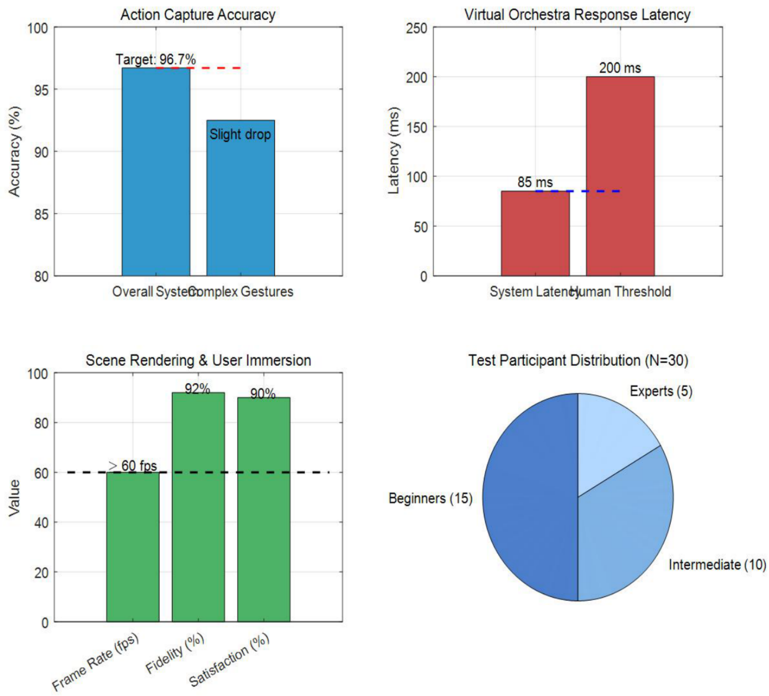

FIGURE 2. System performance test results

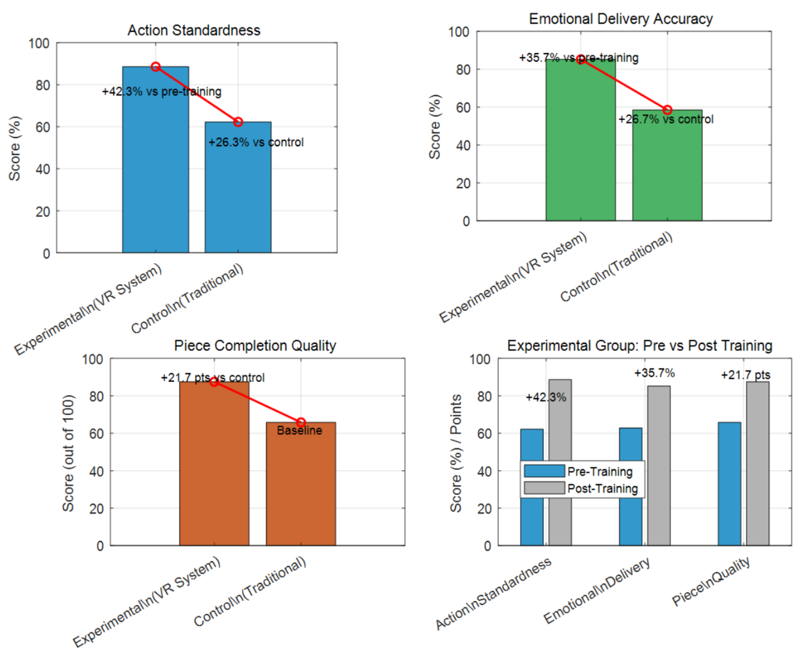

FIGURE 3. Training effectiveness verification results

of feedback lag and weak targeting in traditional training. In addition, the training enthusiasm of the experimental group learners was significantly higher than that of the control group, indicating that immersive training and precise feedback can effectively enhance learners' training interest and further improve training efficiency.

C. ANALYSIS OF THE EFFECTIVENESS OF FEEDBACK OPTIMIZATION MECHANISM

Take the correlation analysis of the feedback acceptance of experimental group learners in the real-time feedback optimization mechanism and skill improvement as the research object. Based on the training log data statistics, the acceptance rate of feedback information in experimental group training is 95.3%, and 86.7% of learners think that the feedback suggestions are positioning, and can help learners to correct action deviation, improve action standardization, and emotional transmission ability. The analysis results show that the adaptive adjustment function of feedback optimization mechanism can effectively improve the effectiveness of training: compared with the fixed feedback frequency group, the number of basic action feedback frequency of beginners increased 18.5% faster improvement in action standardization; for advanced learners, more emotional feedback strengthened the accuracy of emotional transmission, and 15.2% faster than the fixed feedback frequency group.

In addition, through the questionnaire survey, the experimental group learners were satisfied with the real-time feedback optimization mechanism accounted for 93.3%, thought that the feedback method was diversified and timely, and can give suggestions based on their own situation. But there are also some deficiencies: some learners feedback optimization suggestions for textual feedback report that professional content, not easy to understand; in the case of rapid command, feedback information presentation rhythm and command action connection not very smooth. These problems need to be further improved in the future research: simplify professional terms, improve the presentation rhythm of feedback information presentation, and optimize the usage of feedback mechanism.

V. CONCLUSION

This article is based on the pain points of symphony conductor training. We carry out research on integrated virtual reality symphony conductor training simulation system and real-time feedback optimization technology. These technologies can solve the problems of few scenes, feedback delay and inaccurate interaction in conductor training. We have constructed a layered VR command training simulation system, which realizes high fidelity restoration of virtual scenes and accurate multimodal interaction; We have proposed data-driven real-time feedback optimization technology, which realizes precision, timeliness and personalized feedback. The experimental results verify that the whole system and optimization technology proposed in this article can improve the conductor trainees' action standards, accuracy of expression and emotion transmission and training efficiency greatly, and provide new technical route and method for conductor training. The theoretical value of this article lies in that it enriches the research results of digital art training and promotes the deep integration of VR technology and conductor education; The practical value of this article lies in that it provides a low cost and effective training platform for conductor trainees and supports the intelligent and popularized development of conductor education.There are still deficiencies in this article. The accuracy of complex gesture capture needs to be improved further and the usability of feedback suggestions should be enhanced. In the future, we will focus on the optimization of motion capture algorithm and feedback presentation method, and expand the training scene and track covered by the system. The practicality and applicability of the system will be further enhanced.

AUTHOR CONTRIBUTIONS

Han Tao designed, collected and analyzed the data, and drafted the manuscript. Han Tao conducted the study, critically revised the manuscript for important intellectual content, and gave final approval of the version to be published. Han Tao 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 received no external funding.

ACKNOWLEDGEMENTS

Not applicable.

REFERENCES

[1] J. P. C. D. Almeida, D. Russo, H. Iyer, et al., "Foramen Magnum Meningioma: An Example of Integrated Virtual Reality, Skull Base Laboratory, and Surgical Approach: 2-Dimensional Operative Video," Operative Neurosurgery, vol. 28, no. 3, pp. 455-455, 2025, doi: 10.1227/ons.0000000000001286.

[2] S. Olbina and S. Glick, "Using Integrated Hands-on and Virtual Reality (VR) or Augmented Reality (AR) Approaches in Construction Management Education," International Journal of Construction Education and Research, vol. 19, no. 3, pp. 341-360, 2023, doi: 10.1080/15578771.2022.2115173.

[3] B. M. Rajesh, R. Girishankar, and B. Sriram, "An Integrated Platform for Remote and Collaborative Augmented Reality and Virtual Reality of Artificial Intelligence Using AI Tools," Journal of Advancements in Robotics, vol. 10, no. 2, pp. 31-37, 2023.

[4] M. S. Le, M. Guenneguez, L. Macquet, N. Mangold, E. Le Menn, L. Lenta, H. Schmitt, G. Caravaca, and B. Seignovert, "Photogrammetric 3D reconstruction of Apollo 17 Station 6: From boulders to lunar rock samples integrated into virtual reality," Planetary and space science, vol. 240, pp. 1-15, 2024, doi: 10.1016/j.pss.2023.105813.

[5] T. I. Mohamed and A. Sicklinger, "An integrated curriculum of virtual/augmented reality for multiple design students," Education and Information Technologies, vol. 27, no. 8, pp. 11137-11159, 2022, doi: 10.1007/s10639-022-11069-6.

[6] L. E. L. Amrani and M. Moughit, "AN INTEGRATED ACCEPTANCE MODEL FOR VIRTUAL REALITY ADOPTION IN DISTANCE LEARNING: INVESTIGATING MOROCCAN STUDENTS' PERSPECTIVES," Journal of theoretical and applied information technology, vol. 102, no. 13, pp. 5136-5143, 2024.

[7] S. Slatman, T. Groenveld, R. Ostelo, et al., "Development of a Multimodal, Personalized Intervention of Virtual Reality Integrated Within Physiotherapy for Patients With Complex Chronic Low-Back Pain," Journal of Medical Extended Reality, vol. 1, no. 1, pp. 30-43, 2024.

[8] G. Sun, Y. Shang, K. Yuan, et al., "An Improved Whale Optimization Algorithm Based on Nonlinear Parameters and Feedback Mechanism," International Journal of Computational Intelligence Systems, vol. 15, no. 1, pp. 1-17, 2022.

[9] C. Pandey, "Performance Management System And Employee Feedback Mechanism," Poonam Shodh Rachna, vol. 3, no. 4, pp. 87-90, 2024.

[10] S. Mourya, S. D. Amuru, and K. K. Kuchi, "A Spatially Separable Attention Mechanism for Massive MIMO CSI Feedback," IEEE wireless communications letters, vol. 12, no. 1, pp. 40-44, 2023.