Online-Journals.org

iJIM International Journal of Interactive Mobile Technologies

iJIM | eISSN: 1865-7923 | Vol. 18 No. 7 (2024) | OPEN ACCESS

https://doi.org/10.3991/ijim.v18i07.48337

PAPER

Design and Optimization of Human-Computer Interaction System for Education Management Based on Artificial Intelligence

Yaqing Liu(✉)

Academic Administration, Nanjing Vocational Institute of Railway Technology, Nanjing, China

liuyaqing17@outlook.com

ABSTRACT

The continuous improvement and refinement of artificial intelligence (AI) technology has facilitated the broader application of human-computer interaction in the field of education management. The construction of an educational management human-computer interaction system based on AI technology can optimize and improve key parameters of educational management human-computer interaction scenarios, thereby creating a more comprehensive mobile learning (m-learning) application system. This paper is based on AI technology, analyzing gesture semantics and speech semantics, and combining fusion algorithms to construct an education management human interaction system. The performance changes of the system were compared with real experimental operations and the NOBOOK platform analysis. The results show that the education management human-computer interaction system constructed in this article can enhance the m-learning experience of participants. It ensures high recognition accuracy, leading to higher scores in all dimensions of indicator evaluation. Therefore, as one of the crucial forms of m-learning, the human-computer interaction system for education management based on AI can establish a foundation for the further enhancement and development of education management.

KEYWORDS

artificial intelligence (AI), human-computer interaction, education management system, mobile learning (m-learning), fusion algorithm

1 INTRODUCTION

The rapid development of the economy and society has promoted the updating and iteration of artificial intelligence (AI) technology, leading to continuous transformation and improvement in the education industry. In this process, the mode of educational management is also changing accordingly [1–2]. In the gradual popularization of education, educated individuals aspire to enhance and supplement

Liu, Y. (2024). Design and Optimization of Human-Computer Interaction System for Education Management Based on Artificial Intelligence. *International Journal of Interactive Mobile Technologies (iJIM)*, 18(7), pp. 107–124. https://doi.org/10.3991/ijim.v18i07.48337

Article submitted 2024-01-02. Revision uploaded 2024-02-09. Final acceptance 2024-02-12.

their professional knowledge and broaden their horizons. They also hope to further improve their knowledge structure and professional skills. The continuous improvement and innovation of education management models can help address the structural deviation in China's labor demand and supply. It can also optimize and enhance the deepening application of AI in education management. Not only can it optimize educational models, but it can also enrich educational forms and content and improve the quality of education [3]. With the rapid development of information technology and the continuous application of AI, new mobile learning (m-learning) models, such as human-computer interaction systems, provide a fresh platform for education. Participants in education management can utilize the human-computer interaction system developed using AI technology to select the suitable time and environment for engaging deeply with relevant courses in education management. They can more conveniently experience the educational form of human-computer interaction, including audio, video, text, and other forms of fusion.

In the field of educational management, there is a wide array of learning courses available, each varying in quality. As a result, participants often find themselves investing significant time and energy in selecting the most suitable content. They cannot make appropriate adjustments based on the individual differences of participants, and the process is a simple information transmission process, which cannot achieve deep learning in human-computer interaction scenarios [4–5]. Therefore, the situational deep learning method based on human-computer interaction in educational management has yet to be widely popularized. In the context of human-computer interaction in education management based on AI, gesture and speech interaction have become a new form of non-contact human-computer interaction, following the mouse, keyboard, and touch screen. Its intuitive, natural, and human-computer harmony have been widely applied in various fields. The intelligent control model, based on gesture and speech interaction, has effectively promoted the rapid development and updating of human-computer interaction in education management. It also provides a more efficient means for educational management [6–7].

In the development of human-computer interaction technology, traditional contact-based interaction devices are unable to fulfill the profound perception and experiential human-computer interaction requirements of diverse participants in educational management. The emergence of new interactive devices has significantly advanced gesture-based interaction. The interaction based on gesture and speech mainly involves techniques such as gesture tracking, gesture segmentation, and gesture and speech recognition [8]. In vision-based gesture interaction, gesture information can be captured by devices such as Kinect and Leap Motion. After processing, gesture recognition results can be obtained and combined with speech recognition technology. Then, based on the recognition results, 3D models, robots, and virtual objects can be manipulated. The data obtained from wearable devices is very accurate, but their popularity is limited due to the high cost and price issues associated with these devices. The method proposed by researchers, based on gesture and speech information for collaborative input, and has optimized the real-time performance of human-computer interaction [9–10]. This lays the foundation for the design and optimization of human-computer interaction systems in education management based on AI technology, further enriching the form of m-learning.

With the continuous development and maturity of human-computer interaction technology, the interaction methods in education management have gradually evolved from two-dimensional web page interaction to three-dimensional interaction. The education management human-computer interaction system platform designed using virtual reality technology allows participants to deeply engage with virtual objects in the scene through handheld devices, thereby enhancing the

participants' sense of experience [11]. In addition, the human-computer interaction system enables multi-dimensional channel feedback in the virtual scene to ensure convenient, timely, and accurate transmission of interactive information in education management. Operators can acquire information through visual channels and transmit information through various channels, including touch, speech, and more. Avoiding the disadvantage of excessive reliance on recognition accuracy for interaction accuracy in a single interaction mode [12].

A multimodal human-computer interaction system can be implemented by utilizing a multi-dimensional fusion of gestures, speech, and other forms of human-computer interaction, which can construct interaction modes that align with educational management in AI. Therefore, this system is also considered a more natural and efficient method of human-computer interaction in education management. Multimodal human-computer interaction integrates various new interaction methods such as gestures, speech, eye movements, and touch. It can simultaneously process information from multiple modalities and integrate information through multimodal fusion algorithms, providing participants with feedback on various information sources. This technology can effectively solve the problems faced in educational management, such as the need for increased interaction intelligence and higher accuracy. It can also assist participants in the more natural operation of virtual scenes and educational management. The multi-channel interaction method enables participants to interact naturally and efficiently based on real-world behavioral habits [13–14]. Most of the existing multimodal research focuses on fusion methods and technologies. However, there are still areas for improvement in designing and optimizing human-computer interaction systems for educational management within the realm of AI. This can lead to subpar performance of fusion methods in educational management scenarios. Moreover, most fusion models are trained using machine learning (ML) algorithms with a large amount of data, which consumes significant human resources and energy. Often, it can also lead to overfitting issues [15].

Based on this, the paper proposes an AI-based education management human-machine interaction system that can integrate the multimodal inputs of participants. By comprehensively analyzing participant intentions, we achieve human-machine interaction navigation in the education management process. This enhances the efficiency of m-learning and lays the groundwork for designing and optimizing education management human-machine interaction systems.

2 MODELING OF EDUCATION MANAGEMENT ENVIRONMENT BASED ON ARTIFICIAL INTELLIGENCE

Designing and optimizing an AI-based human-computer interaction system for education management requires modeling and processing interaction scenarios. A support vector machine (SVM) classifier can be used to develop gesture and speech recognition models for educational management environments. SVMs transform linearly inseparable samples in a low-dimensional input space into a high-dimensional feature space to enable linear separability through the selection of various kernel functions. Based on institutional risk minimization, the optimal hyperplane is constructed in the feature space to obtain a structural description of data distribution, thereby reducing data size and distribution requirements. In environmental modeling, SVMs are developed to find from the optimal it surfaces under linearly separable conditions. The requirement for the optimal classification surface is that classification can not only correctly separate two classes but also maximize the classification interval.

Assuming training sample $(\vec{x}_1, y_1),...,(\vec{x}_i, y_i), \vec{x}_i \in R^m, i = \{1,2,...,l\}$, determine a hyperplane with the largest interval to make the training set linearly separable, where $\vec{x}_i$ is the eigenvector and y_i is the corresponding label, then it can be converted into the problem described in formulas (1) and (2):

$$P(\vec{w}, b, \vec{\xi}) = \frac{1}{2}\vec{w}^T C \sum_i^l \xi_i \tag{1}$$

$$\begin{cases} y_i\left[\vec{w}^T, \phi(\vec{x}_i) + b\right] \geq 1 - \xi_i \\ \xi_i \geq 0 \quad i = 1,2,...,l \end{cases} \tag{2}$$

Among them, $\vec{w}$ represents a vector in m dimension; b represents a scalar; ξ_i represents a relaxation variable; C is a penalty factor that controls the trade-off between maximizing edges and minimizing classification errors, mapping training data $\vec{x}_i$ to a higher-dimensional space through a function of $\phi(\vec{x}_i)$.

The maximum win algorithm is used in classification decision-making, where each classifier votes for the class it determines. The final classification result depends on the class that receives the most votes. By adopting this method, the classification results can be predicted, and probability information about classification can be provided for each test sample. For a k classification problem, the goal is to estimate the probability that the sample $\vec{x}_i$ belongs to each class, as described by formula (3) [16]:

$$P_i = P\left(y = i \middle| \vec{x}\right), \quad i = 1,2,...,k \tag{3}$$

Among them, P_i can be obtained by solving the following optimization problems, as shown in formulas (4) and (5):

$$W(\vec{p}) = \frac{\sum_1^k \sum_{j:j\neq i} (r_{ij} p_i - r_{ij} p_j)^2}{2} \tag{4}$$

$$\sum_{i=1}^k p_i = 1 \quad p_i \geq 0, \forall i \tag{5}$$

Among them, r_{ij} represents a paired probability.

In addition, virtual agents have been a crucial component of human-computer interaction system design in education management and have been a prominent research topic since the 1990s. Virtual agents are graphical entities that can simulate the behavior and actions of humans or other living organisms. Their appearance enhances the authenticity of virtual reality scenes and improves the experience of human-computer interaction [17–18]. Virtual agents have demonstrated significant value in various fields, including education, entertainment, training, and military applications. They are even used as substitutes for real people in virtual scenes to assist individuals in completing hazardous operations. Human behavior is divided into three stages: environmental perception, cognitive decision-making, and motion control, forming a recurrent network. Before engaging in a behavior, individuals first perceive changes in their environment, then make decisions based on their cognitive abilities, and finally execute appropriate actions through their organs. Therefore, virtual agents must also perceive their environment and make cognitive decisions before taking final actions to simulate natural human behavior when designing human-computer interaction systems for education management.

At the same time, the behavior of virtual agents can also affect the surrounding environment, causing other objects to react and trigger additional actions by the agent. In the education management human-machine interaction system, combined with m-learning, the state changes of participants can be analyzed and studied using virtual agents. As shown in Figure 1, a schematic diagram of the human-computer interaction architecture in education management is provided.

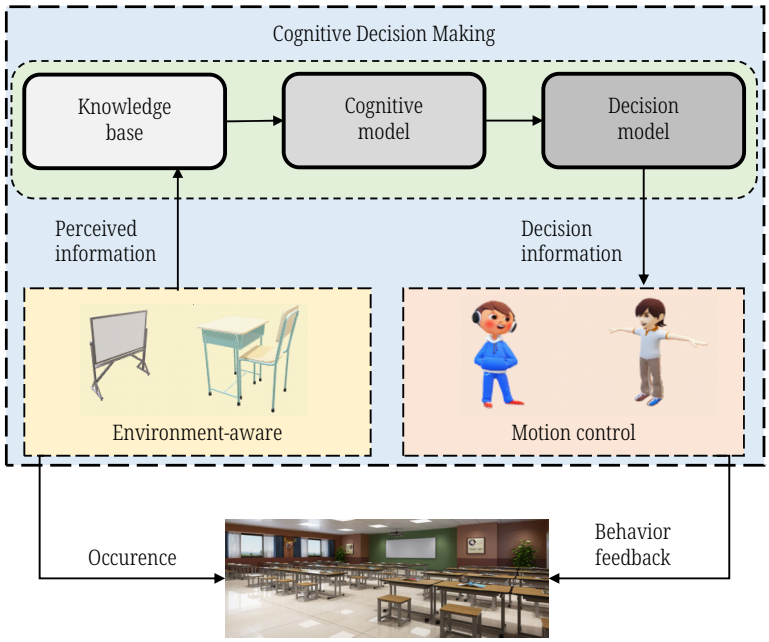
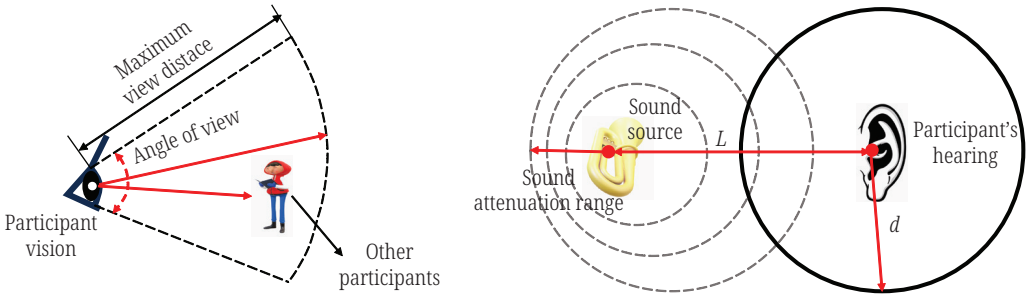


Fig. 1. Schematic diagram of human-computer interaction structure for education management system

In human-computer interaction systems, participants acquire information about the surrounding environment through environmental perception, which is a prerequisite for their movement. The level of perception directly influences participants' behavioral decision-making and motion control. Currently, the most common method for constructing environmental perception is to simulate the perception process of living organisms using visual and auditory cues. As shown in Figure 2, a schematic diagram illustrating visual and auditory perception is used in the design of an education management system based on artificial intelligence.



a) Visual perception diagram **b) Auditory perception diagram**

Fig. 2. Visual perception and auditory perception in the human-computer interaction system for educational management system

In the design process of the human-computer interaction system for educational management, the simulation of participants' hearing mainly considers distance and

sound intensity. When the distance between the sound source and the participants exceeds the defined hearing radius, or the sound intensity transmitted from the sound source to the participants does not reach the sound threshold set by the participants, the participants cannot receive sound information and will not be able to make decisions. On the contrary, a judgment decision will be made based on the received sound information [19–20]. In the design of a human-computer interaction system, the perceptron identifies the trigger carried by the object and extracts the decision information to carry out the following action. This illustrates the implementation process of the auditory perception algorithm in the education management human-computer interaction system. A specific auditory perception pseudocode is provided in Algorithm 1. (The pseudocode of visual perception can also be analyzed regarding auditory perception-related logic.)

Algorithm 1: Auditory Perception

Input: Type of trigger t_{target}, position information of trigger pos_{target}, *decision* information carried by trigger, position information of perceptron *pos*, hearing radius of perceptron r_{sound}, speech value of perceptron $threshold_{sound}$.
Output: Auditory decision $result_{sound}$
Step:
1. **if** (t_{target}=″*sound*″) **then**
2. dis=*Euclidean* (*pos*, pos_{target})
3. *power*=*SoundFade*(*pos*, pos_{target})
4. **if** (*dis*<r_{sound}) & (*power*≥$threshold_{sound}$) **then**
5. $result_{sound}$=*decision*
6. **return** $result_{sound}$
7. **end**

Cognitive decision-making plays an essential role in the behavioral mechanisms of participants. It combines the environmental information obtained from the environmental perception module with its cognitive ability to make corresponding decisions. The system transmits decision information to the motion control module, which then initiates corresponding actions. The finite-state machine model is a standard cognitive decision model, a mathematical model composed of states and transitions. Anything can be abstracted into a finite number of state sets composed of different states at every moment. A state transition occurs when stimulated or influenced by the external environment, as depicted in Figure 3, illustrating the state transition process.

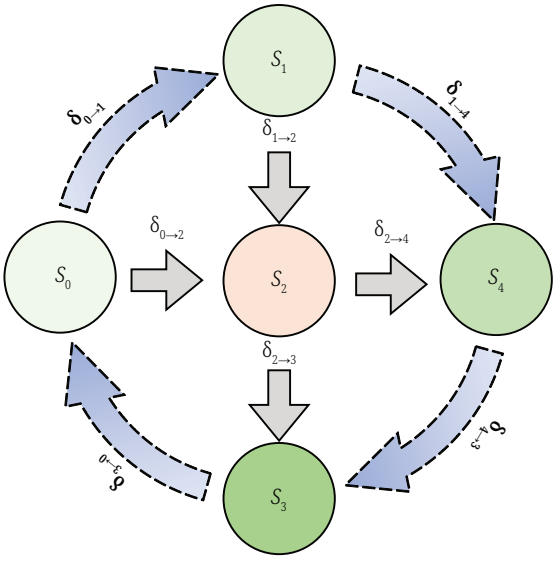


Fig. 3. Schematic diagram of state transition

In the human-computer interaction system for educational management, the state switching between each participant's state points can be represented by a directed weighted edge. The weight of the edge signifies the information input from the external world. The finite-state machine can be expressed by the formula (6):

$$M = (Q, s_0, X, Y, \delta, F) \tag{6}$$

Among them, Q represents a finite set of states; S_0 represents the initial state in the finite state, which is one of the finite set Q; X represents the influence of external stimuli and a limited set of inputs; Y represents the state transition caused by input and is also one of the finite state sets Q; δ is a parameter that controls state transitions; F is the final set of states, also a subset of Q.

Also, we can describe how participants move within the education management system using the linear quadratic form optimal control algorithm. Let's assume we represent the equation of state of a linear system with formula (7).

$$\begin{cases} \dot{x}(t) = Ax(t) + Bu(t) \\ y(t) = Cx(t) + Du(t) \end{cases} \tag{7}$$

The quadratic performance index function is generally denoted as J, as shown in formula (8).

$$J = \frac{\int_0^{\infty} [x^T(t)Qx(t) + u^T Ru(t)]dt}{2} \tag{8}$$

Among them, $x(t)$ represents the state variable, Q is the semi-definite weighted matrix of the state variable, $u(t)$ denotes the control input variable, R stands for the positive definite weighted matrix of the control input, and J represents the performance indicator function. The goal of optimal control is to minimize the performance index function J. According to the minimum principle, the optimal control input formulas (9) and (10) of the system can be derived as follows:

$$\overset{*}{u}(t) = -Kx(t) = -R^{-1}B^T P(t)x(t) \tag{9}$$

$$\dot{P}(t) = -P(t)A - A^T P(t) + P(t)BR^{-1}B^T P(t) - Q = 0 \tag{10}$$

Among them, K is the feedback gain matrix, and $P(t)$ is the solution of the *Riccati* differential equation formula.

3 HUMAN-COMPUTER INTERACTION SYSTEM MODEL FOR EDUCATION MANAGEMENT BASED ON ARTIFICIAL INTELLIGENCE

3.1 Semantic algorithm for human-computer interaction for educational management

Semantic analysis of gestures in human-computer interaction. Based on the construction of gesture semantics in the design of human-computer interaction systems for education management, the goal is to eliminate the semantic gap between gesture and speech. To achieve this, both modalities are uniformly expressed at the

semantic level. This involves determining the active object, passive object, and interactive action referred to by each modality. Firstly, it is necessary to determine the active object (*GA*) that the gesture refers to. In human-computer interaction systems, participants use virtual hands to manipulate virtual objects. By default, the active object of gesture semantics is the virtual hand. When the virtual hand manipulates other virtual objects, the active object is transformed into the manipulated object. Assuming there are l possible active objects in the virtual scene, the probability that the active object referred to by the gesture is the *i-th* virtual object is shown in formula (11) [21].

$$GA = \begin{cases} 0 & active\ object\ is\ i \\ 1 & active\ object\ isn't\ i \end{cases} \tag{11}$$

In gesture semantic analysis, the virtual object that the participant plans to interact with can be determined based on the direction of motion of the active object and the distance from other virtual objects. In a virtual environment with *m* passive objects, the probability of the participant intending to interact with a virtual object, GP_j, can be represented by formulas (12) and (13):

$$GP_j = \frac{1}{\sum_{t=1}^{m} I_t} \tag{12}$$

$$I_j = \frac{1}{\theta_j + d_j} \tag{13}$$

Among them, *j* is the angle between the motion direction of the active object and the vector between the active object and the *j-th* virtual object; D_j is the distance between the active object and the *j-th* virtual object; GP_j represents the probability that the active object wants to operate the *j-th* virtual object, which is the passive object indicated by the gesture.

Speech semantic analysis in human-computer interaction. In the human-computer interaction system of education management, analyzing gesture semantics and related algorithms for speech semantics are necessary. In semantic understanding, most algorithms adopt the method of dividing the semantics of participants into limited, discrete categories. This method first defines the semantics of participants as multiple categories. Then, it collects various related datasets to utilize deep neural networks (DNNs) for distinguishing the input speech semantics [22]. This method requires a lengthy data collection process and network training, which limits its application in practical engineering fields. To address the above challenges, this paper utilizes the language technology platform (LTP) to perform syntactic analysis on the identified sentences. This process helps in obtaining active objects, passive objects, and interactive actions for interactive semantics. Then, word vectors for each component are generated using relevant models such as A_v, P_v, and I_v, corresponding to gesture interaction semantics. All interaction semantics in the human-computer interaction environment of the education management system are represented in the form of word vectors. Specifically, the *i-th* active object word vector in the human-computer interaction scene is represented as A_i, the *j-th* passive object word vector as P_j, and the interaction action as I_k. Finally, the similarity between each component is calculated. The calculation process is shown in formula (14), where the interactive semantics of speech are $\{VA_i, VP_j, VI_k\}$.

$$\begin{cases} VA_i = Sim(A_V, A_i) \\ VP_j = Sim(P_V, P_j) \\ VI_k = Sim(I_V, I_k) \end{cases} \tag{14}$$

Based on word similarity, the structural characteristics of sentences are determined by segmenting sentence components using the system. Then, they are added to the similarity calculation of sentences and combined with the application scenario of virtual scenarios in the education management human-computer interaction system. The method is then simplified to obtain formulas (15) and (16).

$$Sim(S_1, S_2) = \lambda \times \beta \tag{15}$$

$$\beta = a_1 Sim_1(B_1, B_2) + a_2 \left| Sim_2(B_1, B_2) \right| + a_3 Sim_3(B_1, B_2) \tag{16}$$

Among them, the calculation result of $Sim(S_1, S_2)$ is co-determination by two factors. S1 represents the speech provided sentence input by S_1 participants, while S_2 denotes the specific semantics in the semantic database. β represents the semantic similarity value, while λ is a negative coefficient. If there are apparent antonyms in two sentences, the value λ is set to –1, indicating opposite semantics.

Participants aim to identify the sentence with the highest similarity to the constructed intention set when they engage in speech input. At this point, the semantic similarity between the recognition statement and the semantic database can be expressed using Formulas (17) and (18):

$$I_i = Sim(S_1, S_2^i) \quad i = 1,2,3 \tag{17}$$

$$I_0 = 1 - max(I_1, I_2, I_3) \tag{18}$$

Among them, S_1 represents the sentence that the participant input speech through speech recognition, S_2^i corresponding to the non-empty intention in the intention set. I_0 represents the strength of the current empty intention, and the largest among I_0, I_1, I_2, and I_3 represents the current participant's speech semantics.

3.2 Design of human-computer interaction for education management system based on fusion algorithm

A comprehensive framework for multimodal interaction based on AI has been developed to create an educational management human-computer interaction system with multimodal interaction at its core. In the multimodal interaction framework, the input modes of human-computer interaction are set to tactile and speech input [23]. In human-computer interaction, gesture information is selected as the primary input method, supplemented by speech input. This combination of virtual and real interfaces helps streamline the interaction process within the education management system. As shown in Figure 4, the basic framework structure of multimodal interaction is presented. It mainly consists of an input layer, a perception and recognition layer, a fusion layer, and an application layer. In the design of educational management human-computer interaction systems, the overall framework of multimodal human-computer interaction is divided into the input of multimodal information, the perception and recognition of multimodal information, the fusion of multimodal information, and the application of fusion results.

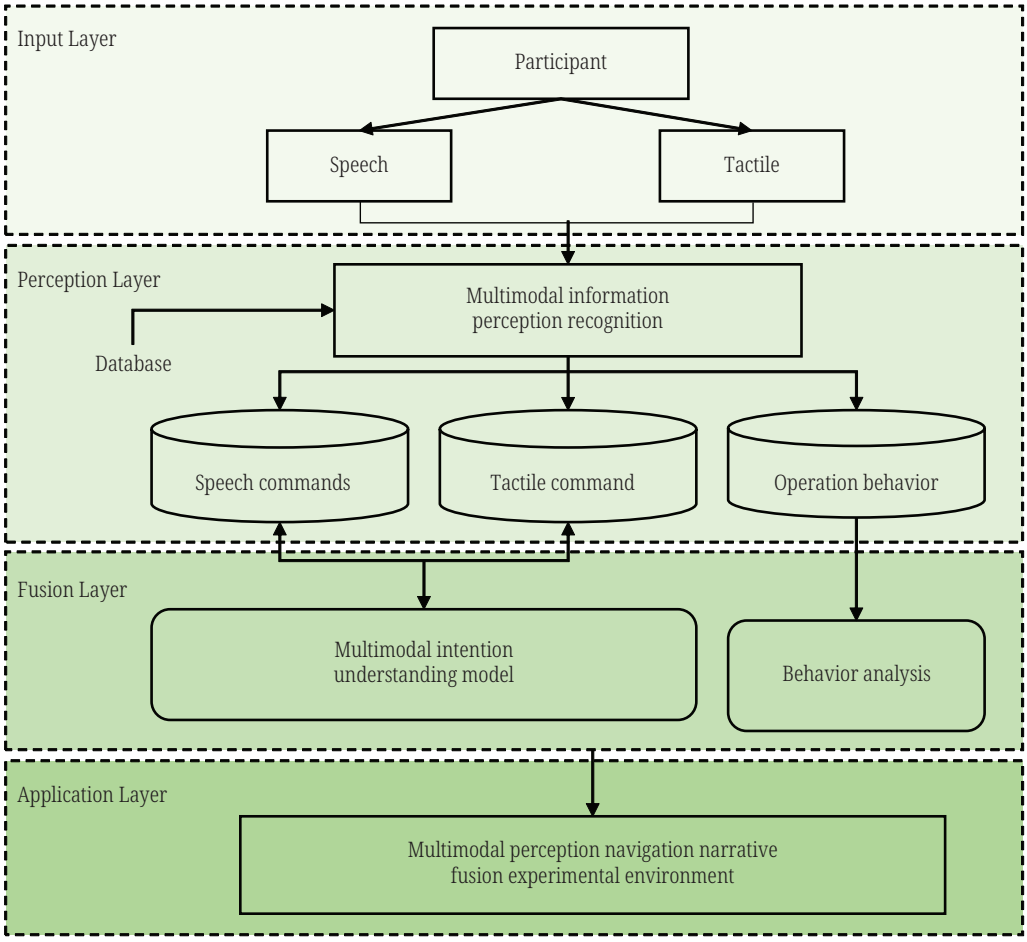


Fig. 4. Basic framework structure of multimodal interaction

Based on the analysis of gesture semantics and speech semantics algorithms in designing human-computer interaction systems for education management, a unified semantic expression for gesture and speech modalities has been achieved, addressing the issue of heterogeneous data from non-homologous sources. The derivation of participants' true intentions through the above two semantics is a vital link. Due to the significant difference in frequency between the two types of semantics, gesture semantics are generated in every frame of the educational management human-computer interaction system. In contrast, speech semantics are only generated after participants produce speech input and recognize segmentation. This also leads to the asynchrony of the two modalities in time, and it is necessary to determine the correspondence between gesture semantics and the generated speech semantics.

Because gesture semantics are generated frame by frame based on video images, there may be frame loss or noise data during gesture operations, which can lead to errors in recognizing gesture intentions. Simply selecting the speech intention to generate the previous or subsequent frame for calculating gesture semantics cannot accurately represent the participants' actual operational intentions. The semantics of participants' gestures and speech are not generated simultaneously; their order of generation is random. However, the generation time of gesture semantics is always within 1 second of the generation of speech semantics, and gesture semantics are concentrated near the time point of speech semantics generation. Therefore, temporal constraints can be applied to gestures and speech. As shown in Figure 5, a time threshold T is set to determine whether the gesture of each frame within the

period is related to speech semantics. If the gesture frame is not within the range, it does not affect the actual operational intention of the participants during this period. Conversely, the gesture frames within the T period are all related to the actual operational intention of the participants.

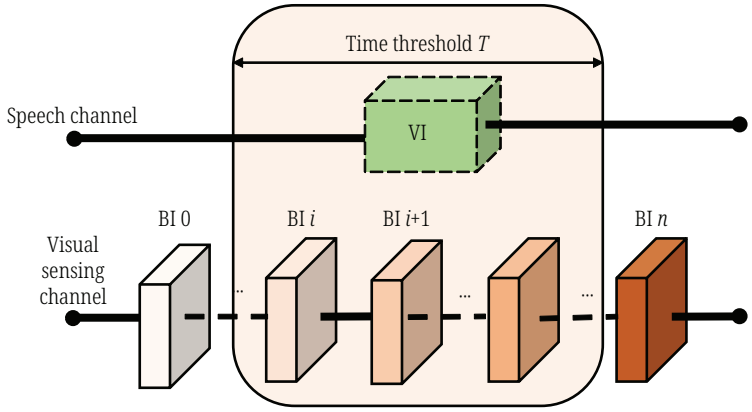


Fig. 5. Time constraints of gestures and speech

In the educational management human-computer interaction system scene, there is one active object for analyzing gesture interactions. Each frame of gesture-activated objects is encoded, and a one-dimensional vector represents each frame of gesture-activated objects. The correlation λ between gesture and speech at *t*, the active object *GA* referred to by the gesture, can be represented by formulas (19) and (20):

$$\lambda = \frac{1}{\sigma\sqrt{2\pi}} exp\left(-\frac{\left(\frac{t}{30} - \mu\right)^2}{2\sigma^2} \right) \tag{19}$$

$$GA = \lambda \bullet M \tag{20}$$

In complex scenarios within education management human-computer interaction systems, relying solely on intelligent devices or speech for interaction can often result in the distortion of input information due to external or internal factors. This can impact the efficiency of the human-computer interaction system. Therefore, to construct a fundamental problem based on gesture and speech intention understanding models and algorithms, a multimodal fusion perception algorithm is proposed. This algorithm aims to perceive the true intentions of participants by mutually supplementing gesture and speech information. In the process of multimodal interaction, the multimodal fusion perception algorithm is used to integrate participants' gesture and speech information. The basic flowchart of the multimodal fusion perception algorithm is presented in Figure 6.

1. Participants' gesture recognition results and speech recognition data are input into the multimodal intention understanding model, and the fusion results are obtained through the intention understanding model.
2. Determine the accuracy of the fusion result's intention, divided into accurate and fuzzy intentions.
3. Input the accurate intentions of the participants into the interactive application layer and check and correct the steps in the current navigation according to the participants' intentions.

If there is a vague intention, ask the participants to confirm the accuracy of the currently obtained intention to facilitate the education management human-computer interaction system to judge the participants' intentions more accurately.

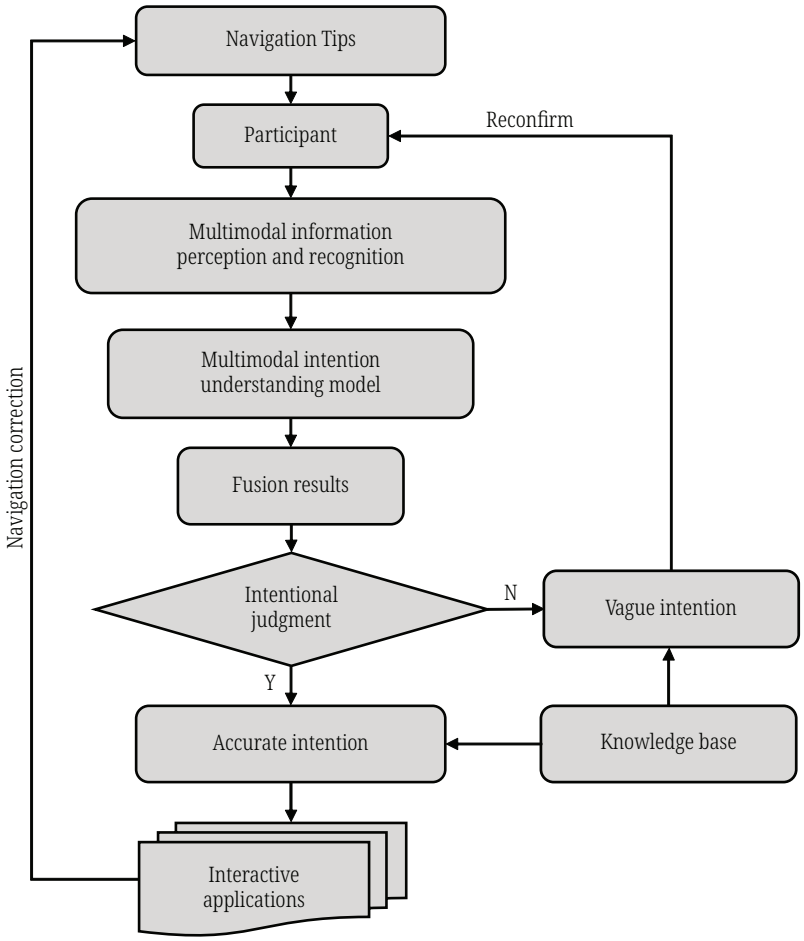


Fig. 6. Flow chart of multimodal fusion perception algorithm

4 APPLICATION OF HUMAN-COMPUTER INTERACTION SYSTEM FOR EDUCATION MANAGEMENT

4.1 System application conditions and processes

In the application process of the educational management human-computer interaction system, the system detects participants' gestures and speech information. It acquires participants' intentions and operational behaviors using multimodal intention understanding and operational interaction algorithms. The virtual experimental environment responds to users in real-time after receiving their intentions and operational behaviors. Participants can observe their behavior and actions in the virtual scene. At the same time, the system will detect the behavior of participants and provide reminders and corrections. Throughout the user experience process, the human-computer interaction environment conveys relevant information through speech broadcasting and screen display. Based on the developed education management human-machine interaction system, combined with the utilization of

m-learning devices, participants can engage in the relevant aspects of the education management process within the interactive system environment. When a participant inputs voice commands, the interaction system distinguishes and determines active objects, interactive actions, and passive objects.

4.2 System application results and analysis

Analysis of correlation between gesture semantics and speech semantics. In order to further explore the temporal constraints of gesture and speech semantics in AI-based education management human-computer interaction systems, 100 college students were randomly invited to participate in experiments. Each participant was asked to perform a gesture and verbally state the action name. The time for recognizing the gesture action and the action name were recorded separately. Each experimenter conducted ten experiments and statistically analyzed the relevant experimental results.

Figure 7 illustrates the temporal correlation between gesture semantics and speech semantics of participants in the education management human-computer interaction system. Among them, the horizontal axis represents relative time. The origin represents when speech is recognized, the negative coordinate represents the gesture recognized before speech is recognized, and the positive coordinate represents the gesture recognized after speech is recognized. The vertical axis represents the number of times a gesture has been recognized during a specific period. The correlation between gesture and speech follows a Gaussian distribution relationship. Through experiments, it can be determined that T = 2s, μ = 0.0136, δ = 0.3725.

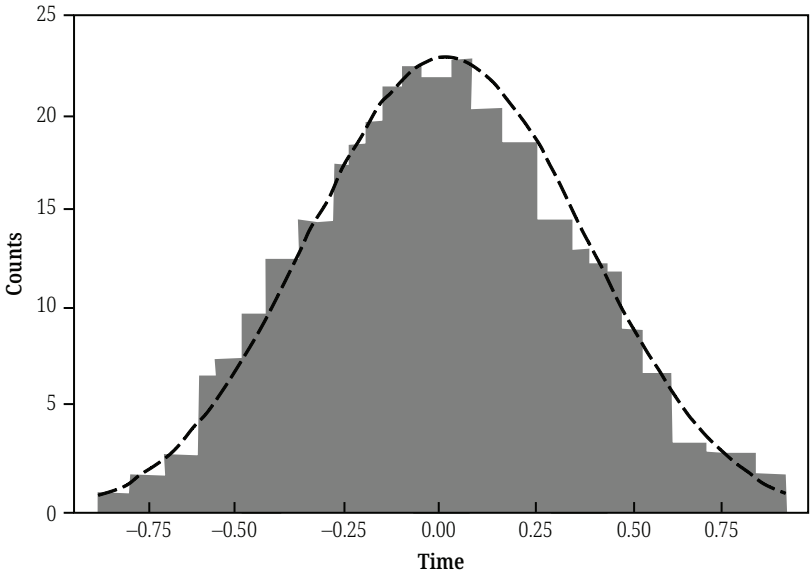


Fig. 7. Time correlation between gesture semantics and speech semantics

Comparison of operational efficiency and evaluation. In order to further evaluate the performance of the human-computer interaction system for educational management based on AI, a comparative analysis was conducted using the NOBOOK virtual platform. This analysis was complemented by fundamental experimental

operations and experimental operation scenarios in educational management. As shown in Figure 8, a comparison of the efficiency of participants in conducting experimental operations under three different scenarios is presented. To ensure the effectiveness of the comparative experiment, all three experimental methods were compared using the same scenario. We invited 15 participants to conduct the experiment, all of whom possess relevant basic knowledge and are proficient in completing real-world experiments. Each participant completed one experiment in three different ways, and the duration of the experiment was calculated for each participant.

The participants' experimental efficiency results show that the interaction efficiency of the education management system based on AI is significantly better than that of basic experimental operations and *NOBOOK* virtual experiments. On the one hand, experiments in virtual scenarios accelerate the occurrence of certain experimental phenomena. On the other hand, the utilization of human-computer interaction systems enables the understanding of participants' interaction intentions and guides them to execute correct operations, thereby enhancing their interaction methods. Among them, the longest duration during the experimental operation is approximately 13.6 minutes, while the longest duration for the *NOBOOK* virtual experiment is around 11.7 minutes. However, the longest duration for the education management human-computer interaction system constructed in this article is approximately 9.0 minutes, which is significantly superior to the efficiency of the actual experimental operation and the *NOBOOK* platform.

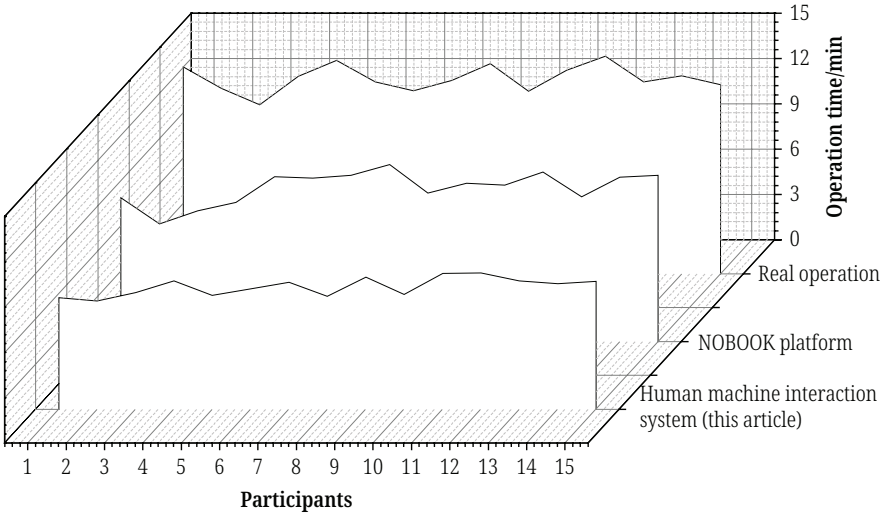


Fig. 8. Comparison and analysis of operation times under different experimental operations

The accuracy of gesture recognition and speech recognition is also essential in the education management human-computer interaction system. As shown in Figure 9, the accuracy of gesture recognition and speech recognition of participants in the interaction system is provided. It can be seen from the figure that the accuracy of gesture recognition ranges from 97.3% to 99.4%, while the accuracy of speech recognition ranges from 96.9% to 98.9%. Therefore, it indicates that the system can still achieve relatively high accuracy even in relatively complex experimental environments.

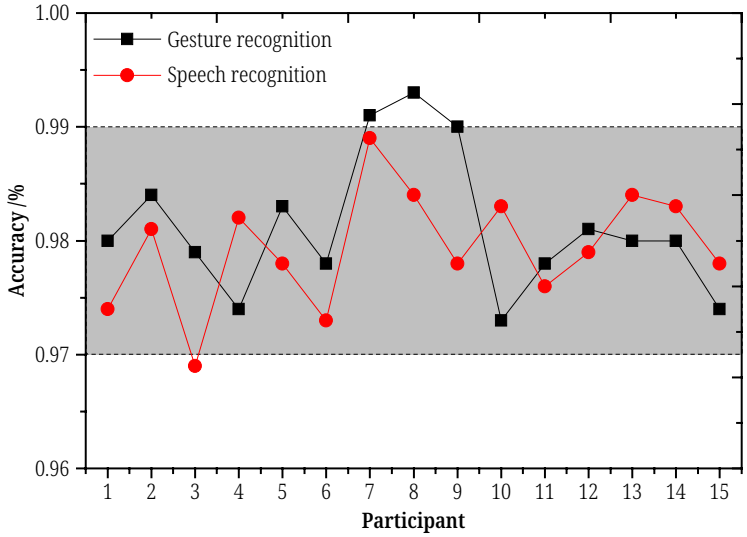


Fig. 9. Accuracy of gesture and speech recognition in the system (in this paper)

In addition, evaluation and analysis were conducted for three different operational scenarios. Based on the operational processes of the 15 participants mentioned above, evaluations were carried out on six dimensions: interactivity, intelligence, interest, effectiveness evaluation, convenience, and operational experience. Figure 10 displays the evaluation analysis of six dimensions across three distinct operating scenarios, with each item having a maximum score of 10 points. As depicted in the figure, the total score of the AI-based education management human-computer interaction system is approximately 51.5 points, the *NOBOOK* platform score is around 47 points, and the actual experimental operation score is about 44.5 points. The education management system, which incorporates a human-computer interaction system based on AI, has relatively better evaluation scores and performance regarding operational efficiency and comprehensive evaluation.

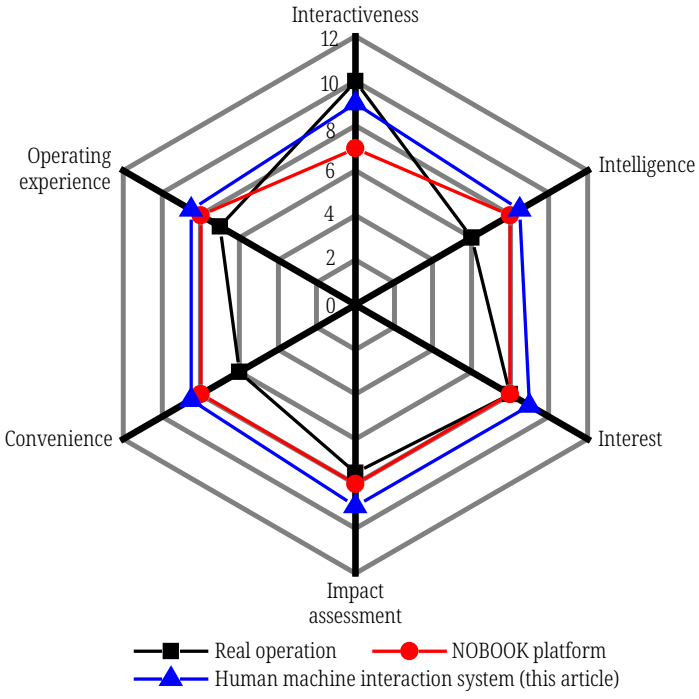


Fig. 10. Comparison of evaluation scores for various dimensions under different operating scenarios

5 CONCLUSIONS

The continuous development and application of AI technology and interaction technology have promoted ongoing progress and innovation in educational management methods from a technical perspective. This paper discusses the application of AI technology in education management, incorporating fusion algorithms to develop a human-computer interaction system for education management based on AI. This system aims to offer more efficient m-learning methods. The study examines the characteristics of the system, the variances among different *NOBOOK* platforms, and real operational scenarios. It also includes a comparison of operational efficiency and participant evaluation scores. The main conclusions are as follows:

1. The application of AI technology and fusion algorithms can integrate gesture semantics and speech semantics in human-computer interaction to create and construct virtual scenarios for educational management. Through auditory and visual perception, it is possible to achieve the perception and analysis of various scenarios in human-computer interaction systems for educational management. This can effectively enhance participants' multi-sensory visual experience in human-computer interaction scenarios and improve their overall experience.
2. In analyzing the temporal correlation between gesture semantics and speech semantics of participants in the human-computer interaction system for education management based on AI, a Gaussian distribution relationship exists between gesture semantics and speech semantics. For the analysis of related parameters, the time threshold $T = 2$s can be determined, and the corresponding distribution function parameters are $\mu = 0.0136$ and $\delta = 0.3725$. The highest accuracy rates for gesture and speech recognition that can be achieved simultaneously are 99.3% and 98.9%, respectively.
3. The system constructed in this article can significantly improve operational efficiency by comparing the performance parameters of the fundamental operation, the *NOBOOK* platform scenario, and the education management human-computer interaction system. In the same experimental scenario, the longest experimental operation time is about 9.0 minutes, which is reduced by 2.7 and 4.6 minutes compared to the accurate operation and *NOBOOK* platform, respectively. The education management human-computer interaction system has received higher scores in all dimensions for the comprehensive evaluation of participants, indicating superior overall performance advantages.

6 REFERENCES

[1] Y. Li, "Innovation and enlightenment of 'Internet +' college education management mode," *Frontiers in Educational Research*, vol. 5, no. 22, pp. 152–164, 2022. https://doi.org/10.25236/FER.2022.052215

[2] G. Ying and X. Shanshan, "Analysis of different college music education management modes using big data platform and grey theoretical model," *Journal of Environmental and Public Health*, vol. 2022, pp. 1–10, 2022. https://doi.org/10.1155/2022/4522580

[3] M. Briceño Toledo, S. Correa Castillo, M. Valdés Montecinos, and M. Hadweh Briceño, "Educational management model for virtual learning programs [Modelo de gestión educativa para programas en modalidad virtual de aprendizaje]," *Revista de Ciencias Sociales*, vol. 26, no. 2, pp. 286–298, 2020.

[4] S. E. Wahyuningsih, N. A. Sugiyo, N. A. Samsudi, T. Widowati, and A. Kamis, "Model of local excellence-based on entrepreneurship education management for prospective vocational school teachers," *International Journal of Innovation and Learning*, vol. 24, no. 4, pp. 448–461, 2018. https://doi.org/10.1504/IJIL.2018.095383

[5] S. Yunpeng and C. Xingquan, "Human-computer interactive physical education teaching method based on speech recognition engine technology," *Frontiers in Public Health*, vol. 10, no. 9, pp. 256–269, 2022. https://doi.org/10.3389/fpubh.2022.941083

[6] M. Gori, S. Price, F. N. Newell, N. Bianchi-Berthouze, and G. Volpe, "Multisensory perception and learning: Linking pedagogy, psychophysics, and human-computer interaction," *Multisensory Research*, vol. 35, no. 4, pp. 335–366, 2022. https://doi.org/10.1163/22134808-bja10072

[7] F. Xiaoying and Z. Xianghu, "Artificial intelligence-based creative thinking skill analysis model using human–computer interaction in art design teaching," *Computers and Electrical Engineering*, vol. 100, p. 107957, 2022. https://doi.org/10.1016/j.compeleceng.2022.107957

[8] V. Chang, R. O. Eniola, L. Golightly, and Q. Xu, "An exploration into human–computer interaction: Hand gesture recognition management in a challenging environment," *SN Computer Science*, vol. 4, no. 5, pp. 441–456, 2023. https://doi.org/10.1007/s42979-023-01751-y

[9] S. Wang *et al.*, "Optical-nanofiber-enabled gesture-recognition wristband for human–machine interaction with the assistance of machine learning," *Advanced Intelligent Systems*, vol. 5, no. 7, pp. 678–695, 2023. https://doi.org/10.1002/aisy.202200412

[10] Md. A. A. Faisal, F. F. Abir, M. U. Ahmed, and Md. A. R. Ahad, "Exploiting domain transformation and deep learning for hand gesture recognition using a low-cost dataglove," *Scientific Reports*, vol. 12, no. 1, pp. 123–142, 2022. https://doi.org/10.1038/s41598-022-25108-2

[11] A. A. Yilmaz, "A novel hyperparameter optimization aided hand gesture recognition framework based on deep learning algorithms," *Traitement du Signal*, vol. 39, no. 3, pp. 396–406, 2022. https://doi.org/10.18280/ts.390307

[12] G. Peng, "Key technologies of human–computer interaction for immersive somatosensory interactive games using VR technology," *Soft Computing*, vol. 26, no. 20, pp. 10947–10956, 2022. https://doi.org/10.1007/s00500-022-07240-3

[13] S. Basak *et al.*, "Challenges and limitations in speech recognition technology: A critical review of speech signal processing algorithms, tools and systems," *Computer Modeling in Engineering & Sciences*, vol. 135, no. 2, pp. 1053–1089, 2022. https://doi.org/10.32604/cmes.2022.021755

[14] Z. Lv, F. Poiesi, Q. Dong, J. Lloret, and H. Song, "Deep learning for intelligent human–computer interaction," *Applied Sciences*, vol. 12, no. 22, p. 11457, 2022. https://doi.org/10.3390/app122211457

[15] C. Latella *et al.*, "Towards real-time whole-body human dynamics estimation through probabilistic sensor fusion algorithms – A physical human-robot interaction case study," *Auton. Robots*, vol. 43, no. 6, pp. 1591–1603, 2019. https://doi.org/10.1007/s10514-018-9808-4

[16] F. Cerutti, M. Giacomin, and M. Vallati, "How we designed winning algorithms for abstract argumentation and which insight we attained," *Artificial Intelligence*, vol. 276, pp. 1–40, 2019. https://doi.org/10.1016/j.artint.2019.08.001

[17] K. Li and X. Li, "AI driven human–computer interaction design framework of virtual environment based on comprehensive semantic data analysis with feature extraction," *International Journal of Speech Technology*, vol. 25, no. 4, pp. 863–877, 2022. https://doi.org/10.1007/s10772-021-09954-5

[18] I. U. Rehman, S. Ullah, and D. Khan, "Multi layered multi task marker based interaction in information rich virtual environments," *International Journal of Interactive Multimedia and Artificial Intelligence*, vol. 6, no. 4, pp. 57–67, 2020. https://doi.org/10.9781/ijimai.2020.11.002

[19] H. Liu *et al.*, "Real-time and efficient collision avoidance planning approach for safe human-robot interaction," *Journal of Intelligent & Robotic Systems*, vol. 105, no. 4, 2022. https://doi.org/10.1007/s10846-022-01687-0

[20] A. Naeimeh and U. Hiroyuki, "Effect of different listening behaviors of social robots on perceived trust in human-robot interactions," *International Journal of Social Robotics*, vol. 15, no. 6, pp. 931–951, 2023. https://doi.org/10.1007/s12369-023-01008-x

[21] Feng Quanzhi, Qiao Yu, Feng Sshichang *et al.*, "Research on flexible mapping algorithm of multi-gestures to one semantic," *Acta Electronica Sinica*, vol. 47, no. 8, pp. 1612–1617, 2019.

[22] Lu Zhenli, Jiang Ruixuan, Ma Zhipeng *et al.*, "Design of semantic training system for robot assisted rehabilitation of cerebral palsy," *High Technology Letters*, vol. 29, no. 2, pp. 183–188, 2019.

[23] H. Niu, C. Van Leeuwen, J. Hao, G. Wang, and T. Lachmann, "Multimodal natural human–computer interfaces for computer-aided design: A review paper," *Applied Sciences*, vol. 12, no. 13, pp. 345–358, 2022. https://doi.org/10.3390/app12136510

7 AUTHOR

Yaqing Liu, Academic administration, Nanjing Vocational Institute of Railway Technology, Nanjing 211800, China (E-mail: liuyaqing17@outlook.com).