ingegneria sismica
International Journal of Earthquake Engineering
Anno XLIII - Num. 2 - 2026

Implementation of Multimodal Learning Model in Personalized Teaching of English Listening and Speaking Courses

Dandan Yang[1,*]

[1] School of Science and Engineering, Xi'an Kedagaoxin University Xi'an 710000, Shaanxi, China

SUMMARY: *For the problems of same content, not enough individual character and low interaction in traditional English listening and speaking teaching, this research builds a multi-modal learning model that combines speech identification, natural language handling, computer vision and big data technique. This model has realized the fusion of multimodal information, the accurate portraying of learner features, the adaptive carrying out of teaching and the diversified conducting of evaluation, hence it forms a closed-loop teaching system. The teaching experiments which we carry out on 200 college students display that this model can significantly promote students' listening and speaking ability, learning activeness and degree of satisfaction. This research offers a standardized and repeatable scheme for the intellectual and personalized transformation of English listening and speaking classes.*

KEYWORDS: *multimodal learning model; English listening and speaking courses; personalized teaching; speech recognition; learning analytics*

1 Introduction

1.1 Research Background

Under the background that global communication happens again and again, English listening and speaking ability has become the core measurement for students' comprehensive language ability. Our country has pushed forward reforms in English listening and speaking teaching work, changing the teaching that is for passing examinations into education that is oriented toward actual use.

Even so, the present instruction work still possesses obvious shortcomings. The traditional teacher-centered unified teaching method cannot adapt to students' differences in language basis and learning speed. The practice of listening and speaking has shortage of real immersive situations, hence making students that they can understand knowledge but cannot use it in flexible expression. Furthermore, the one-time final assessment is not able to achieve real-time study tracing and pointed individual-customized response.

Along with the progress of artificial intelligence, big data and multimedia technology, multimodal study has got wide usage in language teaching. This thing makes many different feeling roads and many-mode messages join together. Putting this mode into use for individualized English listening and speaking teaching breaks the restrictions of time and space, fits the personal learning needs of students, hence promotes the digitized and individualized development of English instruction.

*y15202952790@163.com
https://doi.org/10.65102/is2026996
Published: 30 April, 2026

1.2 Research Significance

1.2.1 Theoretical Significance

This research discusses the combination roads between multi-mode study models and individual-oriented English listen-speak teaching, therefore increasing the applied study of multi-mode study theory in language education and hence making the theory frame for individual-oriented teaching methods more precise. Through the building of an execution frame for multi-modal learning models, and through making clear their core functions and working mechanisms, we hence offer theoretical bases for follow-up usages of multi-modal technologies in language education. At the same time, when we bring in personalized teaching theories, we are to study the application logic of multimodal information in learner demand analysis, teaching content adjustment, and teaching flow optimization. These works expand the realization roads for individualized teaching, injecting new life force into the creative development of language education theories.

1.2.2 Practical Significance

The multi-modal learning model which this study has put forward offers actionable teaching references for frontline English education workers, therefore enabling them to go beyond the limits of traditional teaching methods, while accurately solving the individual differences of learners, hence realizing differentiated teaching. At the same time, this model, which is based on learners' language proficiency and their needs, provides personalized learning experiences through recommending content that is specially made, thus creating immersive practice situations for listening and speaking, hence effectively raising English communication abilities and capabilities for self-directed learning. In addition, the carrying out of this work makes the optimization and integration of English listening and speaking teaching resources become easier, makes the digital and intelligent management of teaching processes more convenient, and thus gives practical support for the reform of English language education practices.

1.3 Research Content and Methods

1.3.1 Research Content

This research carries out discussion on the putting into practice of multimodal study models in personalized instruction for English listening and speaking classes. This research framework contains five important constituent parts: Firstly, we carry out systematic inspection on the core concepts and theoretical bases of multimodal learning and personalized teaching, hence we make clear the integration logic between multimodal learning models and custom-made English language education. Second, we carry out analysis on current teaching activities in English listening and speaking classes, find out existing problems and their basic root reasons, and thus prove the necessity and feasibility that multimodal learning models are applied to personalized teaching activities. Third, we have established an execution framework for multi-modal study models in individualized instruction, which defines core modules, technical support systems and working procedures. Fourth, we have carried out teaching experiments for the purpose of verifying the model’s effectiveness in promoting teaching results and satisfying individual study requirements. At last, we make summarization of experimental results and bring up optimization strategies for the usage of multimodal learning models in individualized English language education, hence providing actionable references for the wider carrying out.

1.3.2 Research Methods

The present research utilizes the combination of a number of research methods to guarantee scientific strictness and practice effectiveness.

(1) Literature summary method: Systematically carry out analysis on research results which are related to multimodal study, personalized teaching activity, and English listening and speaking teaching, in order to make clear theoretical bases and current research situation, thus this can provide inspiration and reference for this research of ours.

(2) Investigation Research Technique: Through questionnaire investigations and interview researches, this approach has the goal to know the present situation of English listening and speaking course teaching, learners' individual demands, and the difficulties which teachers meet in teaching practices, hence therefore providing experience-based proof for the building of model.

(3) Method of Model Building: Through combining the features of multi-modal technology with the requirements of individual English listening and speaking teaching, this method builds an execution frame for multi-modal learning models, hence clearly stipulating their core modules, technical support systems, and operation mechanisms.

(4) Experiment Research Method: Concrete teaching objects were chosen and split into one experiment group and one control group. The experiment group used a multimode study pattern to carry out personalized teaching, meanwhile, the control group applied traditional teaching ways. Through the comparison and the analysis of the teaching results between these two groups, the effect of this model has been verified.

(5) Data analysis method: It is to carry out statistical analysis on the learning data, test scores and questionnaire data which have been collected in the experimental process. We use the combination of quantitative method and qualitative method to sum up the effect of model using, find out current problems, hence provide data support for putting forward optimization strategies.

1.4 Research Innovations and Limitations

1.4.1 Research Innovations

The core innovation of this study is in the construction of a course-special multimodal learning model which integrates audio, text, vision and behavior data to realize accurate perception and adaptive adjustment of the states of learners. It has broken through the restrictions that single-modal input brings, and therefore realizes the deep integration of multimodal technology with the whole process that personalized English listening and speaking teaching follows.

1.4.2 Research Limitations

This research still has some restrictions: Firstly, the choice of experiment sample was comparatively narrow, it only focuses on learners of specific stages and from particular groups, hence this may limit the popularization of the results. Secondly, some functions of the multi-mode study model (for instance, identification of emotional condition and dynamic follow of study interest) need further optimization, hence technical compatibility and stability need promotion. Thirdly, the application of this model has a high requirement for educators' grasping ability of multimodal technology. How the enhancement of teachers’ technical abilities can be done effectively and the wider promotion of this model’s application can be achieved still is a domain that needs extra research work.

2 Relevant Theoretical Foundations

2.1 Multimodal Learning Theory

2.1.1 Core Connotation of Multimodality

Multimodality means the combining of two or more modalities (e.g., text, speech, images, videos, gestures, facial expressions) in information passing and meaning building. By the cooperating participation of many sense channels (vision, hearing, touch, and movement feeling), this therefore enables effective information conveying and deep understanding. In the domain of education, multimodal learning thus transcends the limitations of single-modal methods through synthesizing manifold information sources. This method enables learners to take part through multiple senses, therefore it can cultivate active knowledge building and deep comprehension.

In the study of multiple mode learning, "mode things" have both difference and mutual help together. Diversity shows itself via the abundant sorts of modal kinds, including both language modalities (text and speech) and non-language modalities (pictures, videos, body motions, and face expressions). Synergy is the name for the complementary and mutually strengthening connections between different modalities, which together give support to learning goals. For example, in English listening and speaking teaching, the cooperation between speech pattern ways (listening materials and spoken expressions) and vision pattern ways (situation videos and face movements) helps learners more understand language backgrounds and promote language using abilities.

2.1.2 Core Features of Multimodal Learning Models

The multi-modal study model is a study system which is constructed on the basis of multi-modal study theory, it integrates technologies including artificial intelligence and big data. It allows for the gathering, analysis, handling, and utilization of multi-modal information. Its core characteristics mainly include the below aspects:

(1) Multimodal information gathering union: The model has the ability to collect and unite many kinds of modalities which include text, speech, images and actions, hence it gets over the restrictions of single-modal information, hence comprehensively catches learners' study conditions and actions, therefore provides firm data holding for individual-based teaching.

(2) Intelligence analysis ability: The model uses technologies for example natural language processing, speech identification, and computer vision to carry out intelligence analysis on collected multi-mode information. It correctly ascertains learners’ language basis, study demands, study habits, and study problems, hence therefore allowing accurate learner character description.

(3) Individualized Adaptive Ability: The model is able to provide custom-made teaching content, speed, and teaching methods on the basis of learners' accurate personal files, hence realizing "one-person-one-method" individualized teaching to satisfy each person's study demands.

(4) Dynamic mutual function: The model supports many mode interactions between learners and the teaching system, for example speech dialogue, movement operations, face expression feedback, it creates a sinking type learning environment for strengthen learners' participation and active nature.

(5) Real-time feedback ability: The model is able to follow the learning processes of learners in real time, gather learning data, assess learning results dynamically, and give directional feedback and suggestions quickly to help learners make adjustments to their

learning strategies.

2.1.3 Application Logic of Multimodal Learning in Language Teaching

The study of language is a process that has multiple senses, multiple modes and cooperation among people. The progress of language abilities including listening, speaking, reading, writing needs the support from many different modal information. The core tenet of utilizing multimodal study in language teaching resides in the combination of diverse modal inputs for the constructing of authentic and attractive language learning situations. This method effectively causes learners to have interest in language study, therefore at the same time it promotes their understanding and actual usage of language knowledge.

The multi-mode study method has the built-in superiority for English listening and speaking class. The obtaining of listening ability depends on auditory modalities (audio materials), which are combined with visual modalities (scenario videos) and textual modalities (listening scripts), for helping learners comprehend contexts and identify key information. Speech practice needs phonetic modes (natural utterance and standard pronunciation) which combine with kinesthetic modes (hand movements and face expressions) and visual modes (situation simulations) to make pronunciation standard and improve fluency and accuracy. Multimodal learning models through seamless integration of these modalities, build immersive practice environments that realize the integrated "listening, speaking, viewing, practicing" approaches, hence thus effectively promote English listening and speaking ability level.

2.2 Personalized Teaching Theory

2.2.1 Core Connotation of Personalized Teaching

Individualized teaching is a method which takes learner as center, and it fully respects the differences between each person. Through the consideration of learners' language level, learning capabilities, interests and speed, it makes designed teaching content, methods and evaluation systems to realize the educational thought of "teaching in accordance with each person's natural ability". In the innermost core, individualized instruction places priority on taking the learner as center, laying stress on custom-tailored teaching processes, content connection, and multiple evaluation approaches. Its final objective is to cultivate all-round development and skill promotion for each individual learner.

Individualized teaching in English listening and speaking classes is specifically expressed through: providing teaching materials of different difficulty degrees that are made suitable for learners' language ability; establishing personally customized study arrangements on the basis of each person's own learning rhythm; to make topic-type listening and speaking tasks that match what learners are interested in; and thus providing pointed guidance tutoring and practice courses which aim at specific study difficulties. This method can guarantee each learner can promote their listening and speaking abilities in a study environment which fits their own demands.

2.2.2 Core Principles of Personalized Teaching

(1) Principle of Subjectivity: We must respect the primary position that learners hold, fully bring their learning initiative and enthusiasm into play, and let learners participate in the design and execution of the teaching course, therefore permitting them to independently choose learning content, speed and ways.

(2) Principle of Individual Difference: Fully acknowledge the individual differences that exist among learners in the aspects of language level, learning ability, and learning interest.

Make the teaching schemes which have difference to suit the features of learners which are not same, to avoid the teaching which all are same.

(3) Adaptability Principle: The teaching content, the instruction methods, and the speed of progress should be in accordance with the individual needs of learners, to guarantee that the difficulty degree and the advancement of the content fit their cognitive ability, while instruction methods are in correspondence with their learning inclinations.

(4) Principle of Dynamism: The learning states and needs of learners undergo the dynamic changes. Personalized teaching schemes ought to be adjusted and optimized in good time according to learners' progress and improvement speeds, hence guaranteeing the relevance and effect of teaching activities.

(5) Development Principle: The final aim of individualized teaching is to push forward the all-round development of learners' language ability and overall cultural attainment. It ought not only to pay attention to promoting learners' listening and speaking abilities but also lay stress on fostering their independent study abilities, creative capabilities, and cross-cultural communication abilities.

2.2.3 Implementation Conditions for Personalized Teaching in English Listening and Speaking Courses

The carrying out of individual-oriented English listening and speaking teaching depends upon three core preconditions. Firstly, carry out accurate study demand evaluation to identify students' language ability level, study difficulties and individual interests. Secondly, the teaching resources which have differences can provide the listening and speaking materials that are made-to-measure for learners who have various kinds. Thirdly, a perfect evaluation system which monitors the entire learning process, evaluates learning effects and guides timely teaching adjustment.

Traditional teaching does not have enough technical support, therefore it is difficult to realize accurate evaluation and adaptive resource matching, hence this restricts the development of personalized teaching. By comparison, the multi-mode learning model which is based on AI and big data can perfectly solve the above-mentioned problems, and thus provide solid technical guarantee for the effective carrying out of personalized teaching.

2.3 Integration Logic of Multimodal Learning and Personalized Instruction

The combination of multimodal study and individual teaching puts together advanced education technology with student-facing teaching ideas, therefore its core logic is reflected in three aspects.

First, multimodal study gathers learners' multi-dimensional data in order to achieve accurate judgment for individual study demands. Second, the various multi-mode teaching resources are matched to provide the custom-made learning content for the different students. Third, the real-time interactive feedback can continuously make teaching procedures get optimization, therefore promote the improvement of personalized teaching quality.

Say specifically, multimodal data which include learners' voice and learning behaviors enable the precise diagnosis of demands. Rich text, audio and video materials achieve self-adaptive individual content recommendation. Multiple types of immersion interaction methods can also make the entire learning process get optimization, enhance the degree of learning participation, and provide support for the teaching guidance which has clear targets.

3 Analysis of Teaching Issues in English Listening and Speaking Courses

3.1 Homogenization of teaching models with neglect of individual differences

Present English listening and speaking teaching mostly follows the traditional same-nature teaching modes. Education workers carry out standardized course programs, inflexible speed arrangement, and unified teaching methods while they ignore learners' personal differences in language ability, learning abilities, and participation degrees. For example, hearing activities often have same materials of same difficulty: beginner persons have big difficulty to follow and understand content, while high-level learners think the material is too simple to satisfy their demands. In the same way, oral practice tasks which are designed in a unified way do not have specific properties, hence leaving learners of all proficiency levels without pointed opportunities for promotion, hence effectively damaging the tenet of "teach students in accordance with their personal aptitude."

At the same time, traditional teaching modes are centered on teachers, which put learners in a passive position with limited independent ability when they choose learning content or study speed. This method cannot arouse learners' go-aheadism and participation. Classroom interactive activities mainly depend on teachers leading question asking and students giving answers, thus bring about dull forms which have no multimodal interactive experiences. These structures are not able to construct immersive environments for listening and speaking practice, therefore they ultimately hinder the development of learners' language application abilities.

3.2 Single teaching resources with insufficient adaptability

The quality and the variety of English listening and speaking teaching resources have the direct influence on teaching effect and the carrying out of individualized education. At present, these resources mainly depend on textbooks and matching audio materials, they have single types and are short of multimodal resources including texts, pictures, videos and scene simulations. In addition, the difficulty degrees of these resources still keep single, they cannot satisfy learners who have different language ability levels, therefore this leads to not good enough use of these resources.

In addition, the renewal of teaching resources is carried out with a slow speed, and it mainly depends on the traditional listening and dialogue exercises which are based on textbooks. Currently, there exists a striking insufficiency of practical listening and speaking materials which can reflect real life situations, for example daily communication, workplace interactions, and cross-cultural exchanges. This separation between theory knowledge and practice application therefore makes it have difficulty for learners to reach "learning for actual use." In addition, at present the teaching resources do not have the personalized recommendation functions, they cannot provide the suited materials according to the individual demands and preferences of learners, hence they cannot satisfy their customized learning demands.

3.3 Single teaching evaluation method with delayed feedback

At present, the evaluation of English listening and speaking teaching mostly uses summative assessment, which mainly pays attention to learners' achievement indicators, thus ignoring the tracking and evaluation of the learning process. This evaluation method cannot

comprehensively reflect learners' academic situation, progress progress and improvement speeds. It also has no capability to quickly find out learning difficulties, hence it becomes hard to give support which is based on evidence for changing teaching methods.

At the same time, the mechanism of evaluation feedback is still not enough perfect. Teachers generally give feedback just after assessments finish, therefore leading to postponed answers which have no enough particularity and relation. This makes it difficult for learners to promptly find out their learning gaps and hence adjust their study strategies. In addition, the appraisal system depends mostly on evaluations given by teachers, and learners have only little involvement by way of self-evaluation or peer reviews. This kind of single-origin method cannot effectively bring about the activation of learners' initiative and the ability of self-reflection.

3.4 Insufficient technological application hinders the implementation of personalized teaching

Though a number of teachers have made attempts to employ multimedia technology in English listening and speaking teaching, the utilization of such technologies still stays at a primary phase, and is mainly restricted to simple video playing and audio explanations. Up till now, the deep integration between multimodal technologies and teaching content as well as instructional processes has still not been achieved. The absence of valid technical assistance hinders the accurate judgment of learners' personal requirements, individualized provision of teaching materials, and personalized improvement of teaching procedures.

At the same time, the technical ability that teachers have is still not enough. Majority of educational workers do not have enough understanding and practical abilities regarding multimodal technologies and artificial intelligence, which therefore hinders the abilities that they have for carrying out personalized teaching by means of these tools. In addition, some schools have not the necessary technical infrastructure and platforms—including speech recognition systems and multimodal interaction platforms—this therefore prevents effective use of multimodal learning models, hence makes personalized teaching hard to carry out.

4 Framework Construction for Implementing Multimodal Learning Models in Personalized Teaching of English Listening and Speaking Courses

4.1 Construction Principles

4.1.1 Learner-centered Principle

The establishment of the model always follows the learner-centered principle, it fully respects individual differences and pays attention to the personalized learning demands. Through bringing in learners' language level, learning abilities, hobbies and speed modes as core design standards, the model guarantees that customized learning experiences and teaching services which effectively arouse learners' go-aheadism and participation.

4.1.2 Multimodal Fusion Principle

The model structure puts emphasis on the combination of multi-mode information, which includes text, speech, images, videos, hand movements, and facial expressions. Through utilizing the mutual promotion effects on each sensory channel, it has built immersive learning environments for English listening and speaking which promote learners'

understanding and utilization of language knowledge, and at the same time enhance communication ability. Furthermore, this design puts the deep integration of multimodal technologies with instructional content and teaching processes in the first place, therefore to guarantee practical application possibility and measurable effect efficiency.

4.1.3 Principles of Precision and Intelligence

This model uses technologies including artificial intelligence, big data, speech recognition, to with accuracy collect and with intelligence analyze the multi-mode information of learners, thus make accurate learner portraits to confirm the individual demands and learning difficulties of learners. It at the same time lets directional content sending, intelligent teaching flow improvement, and accurate assessment feedback, hence therefore promoting the connected degree and effect efficiency of individual-centered education.

4.1.4 Principles of Operability and Scalability

The model arrangement puts practicality in the first place, therefore it guarantees that its core component parts, working flow procedures, and realization approaches are direct and simple to comprehend, hence helping the smooth taking-up by front-line education workers and learners. In addition, the model puts emphasis on extensibility through the inclusion of stipulations for technical promotions and function expansions. This method allows the timely optimization and promotion of model abilities in accordance with the progress of educational technology and changing teaching demands, therefore promoting its adaptability and long-term survival ability.

4.2 Overall Model Framework

This model adopts a closed-loop design which is “data gathering–intelligent analysis–adaptive instruction–effect assessment–iteration optimization”. It puts together multimodal information combination, individual need digging, self-adaption teaching decision, and dynamic feedback adjustment into a single frame, therefore guaranteeing systematic property, standard property, and expansion ability in the process of implementation.

4.3 Core Module Design

4.3.1 Multimodal Information Acquisition Module

The multi-mode information gathering module acts as the base for model running, mainly takes charge of collecting various mode data from learners in the processes of English listening and speaking study. This offers fundamental data assistance for constructing accurate learner portraits and carrying out individualized teaching adjustments. The collection content of this module mainly includes the following kinds:

(1) Speech modality information: It is collected that oral expression audio from learners by using speech recognition devices (e.g., microphones, headphones), which includes the data on pronunciation accuracy, fluency, intonation, and speech rate. At the same time, the audio materials which come from listening practice tasks are collected, in order to evaluate the speed of learners' listening responses and the accuracy rate of learners' listening.

(2) Text mode information: Gather text data which contains learners' listening manuscripts, spoken expression articles, learning notes, and assignment finished situation to analyze their language basic level, vocabulary amount, and grammar ability.

(3) Image modality message: Through catching learners’ facial expressions and body motions by cameras, the system carries out analysis on their learning conditions (e.g.,

concentration degree, tiredness level, interest degree) as well as mood changes, hence it assesses learning motivation and learning results.

(4) Behavioral mode information: Gather learners' study behavior data, including study time length, study advancement, study times, practice numbers, and error distribution situations, to carry out analysis on their study rhythm and study habits.

This module uses a method which combines together real-time and offline data gathering. It catches multi-mode information from learners when classroom teaching and on-line practice is going on in real time, meanwhile it gathers related data such as homework and examinations out of line to guarantee overall and prompt data obtaining. In addition, the data encryption technology has been carried out for the protection of learners' individual information and educational data, thus guaranteeing the security of data.

Multi-modal data offers a comprehensive information base for the accurate learner portrait drawing and personalized teaching decision making, thus laying a firm data foundation for the refined teaching management and intelligent intervening.

4.3.2 Learner Profiling Module

The learner character module acts as the core part for carrying out individualized teaching through this model. It mainly carries out intelligent analysis and processing on multi-modal information which is collected by the data acquisition module, constructs comprehensive and accurate learner portraits to find out individual learning demands and difficulties. The key function parts of this module include the following several aspects:

(1) Multimodal Information Pre-handling: The gathered multimodal data has passed through cleaning, noise cutting, and standardization procedures to remove invalid and interference data, therefore guaranteeing data accuracy and usability. As an example, speech data are handled by noise reduction process for eliminating background noise; The text data is undergone word cutting and mistake fixing for the standardization of formatting; Through the methods of facial recognition and action recognition, the analysis of image data is carried out, for the purpose of extracting key features.

(2) Extraction of Learner Features: Through utilization of technologies including natural language processing, speech recognition and computer vision, core characteristics of learners are extracted from multimodal data that has been preprocessed. These include language basic characteristics (vocabulary quantity, grammar mastery degree, pronunciation correctness), learning ability indexes (listening understanding, language expression abilities, self-guided learning ability), learning habit parameters (interest modes, learning ways, learning speed), and learning condition data (attention degree, participation degree, tiredness indexes).

(3) Learner Accurate Sorting: According to extracted learner characteristics, we use clustering algorithms to carry out accurate sorting, and put learners who have similar features and demands into the same classification. At the same time, for every learner, a personalized information file that has records is made! study advancement, difficult domains, and personal likings, hence forming a! accomplished! Study participant's characteristic file.

(4) Individualized Demand Analysis: Through utilizing learner portraits, we carry out deep analysis of individual learning demands, including content inclination (custom listening and speaking materials according to learners' basic knowledge), speed flexibility (self-adjustable learning time arrangements), teaching methods (multi-mode interaction and self-guided learning), and guidance demands (pointed support for difficult themes). This therefore offers solid proof for the making of individualized teaching methods.

The feature drawing module changes dispersed multi-mode data into structured and explainable learner features, therefore it supports exact locating of learning defects and scientific making of individual teaching schemes.

4.3.3 Personalized Teaching Adaptation Module

This module, according to learners' learning situation files, realizes the custom-made adaptation of teaching content, progress and methods, hence it formulates exclusive personal teaching plans for different students.

(1) Personalized Teaching Content Adaptation

It carries out matching for differentiated listening and speaking materials according to the language level that learners have. Beginners who study this subject obtain simple daily conversation contents to consolidate their foundational knowledge, while high-level learners get hard news and speaking materials. It also offers pointed resources for examination-focused learners and suggests interest-connected materials to enhance learning enthusiasm.

(2) Personalized Teaching Pace Adaptation

This system, according to the degree learners can accept, carries out dynamic adjustment on the rhythm of learning. The students can by themselves freely control their own learning speed, in order to review difficult knowledge and skip the content that they have already mastered. This system, moreover, can automatically carry out optimization on the whole progress, hence it can guarantee the learning that is stable and high in efficiency.

(3) Personalized Teaching Method Adaptation

It has adopted teaching methods that have targets for different study styles. For visual type learners, image and video materials are provided by teachers, for auditory type learners, the focus is put on audio and dialog practices, and for interactive type learners, they take part in scenario simulation and role-play tasks.

(4) Personalized Learning Resource Recommendation

Through combining with the multimodal resource storehouse, it pushes suitable listening materials, speaking manuscripts and practice resources according to need, thus enabling students to carry out high-efficiency self-governing learning at any time and any place.

4.3.4 Multimodal Interactive Teaching Module

The multi-mode mutual-action teaching module is mainly designed for making immersive English listening and speaking study surroundings, which lets multi-mode interaction among learners and the teaching system therefore promote learner involvement and language ability level. Its core function components include the below aspects:

(1) Immersive Situation Imitation: By utilizing technical means such as virtual reality (VR) and augmented reality (AR), this method builds real English listening and speaking situations—including daily communication scenes, work conversations, and cross-cultural contacts—soaking learners in true language environments to promote practice trueness and effect. For example, VR technique imitates environments such as airfields, eating places and meeting rooms, it lets learners practice oral talking in these places and promote language using abilities.

(2) Multimodal Interaction Characteristics: It enables many kinds of interaction ways between learners and the teaching system, which include voice talks, gesture handle, and facial expression feedback. For example, learners are able to practice oral communication by means of voice interactions, in which the system gives real-time analysis of pronunciation and personalized feedback. Hand movements make learners able to pick learning materials and change learning speed, and face feeling identifying lets the system watch over learning conditions and actively change teaching content and speed.

(3) Cooperative Group Study: Through building group study situations which let members conduct talks and role-play exercises via audio and video forms, it facilitates multi-model interactions between learners, hence promoting oral speech abilities and cooperative study abilities. For example, learners can be cut into groups to practice oral talking on given topics,

with members giving each other feedback and direction to together raise listening and speaking ability.

(4) Real-time Teaching Feedback: When learners carry out listening and speaking practice activities, the system, in a continuous manner, gathers multi-modal data for the analysis of study advancement and the provision of pointed direction. For example, in the practice of spoken language, it can find out wrong pronunciations and give modification suggestions that have standard pronunciation examples. Regarding the listening training, this system carries out analysis on the root reasons that lead to errors, and then recommends the related strengthening study materials.

4.3.5 Learning Outcomes Assessment Module

The Learning Effectiveness Evaluation Module is mainly on the side of carrying out comprehensive and scientific evaluations on learners' learning processes and results, offering proof for model optimization and teaching plan modifications. Through the combination of formative evaluation and summative evaluation, this teaching module has realized the diversification and precision of assessment methods. Its core function abilities include the below aspects:

(1) Process-oriented appraisal: It carries out continuous monitoring on the learning processes of learners through collecting data which includes study time length, progress recording, practice times number, error distribution situations, pronunciation correctness, and spoken language smoothness. This makes possible the dynamic measurement of learning condition and progression indexes. For example, the analysis of oral training data assesses promotes in pronunciation accuracy and smoothness, while listening training data analysis gauges increases in listening understanding ability.

(2) Formative Assessment: Carry out periodic overall appraisals of learners' listening and speaking abilities through multimodal assessment approaches like oral examinations, listening examinations, and situation-based simulated operations, to overall appraise their capability level. For an example, situation-based simulated activities let learners do practice of oral expression and listening understanding in true environments, therefore they make evaluation of their language using abilities.

(3) Multi-party participant evaluation frame: It is necessary to establish a comprehensive evaluation system that has integrated "teacher assessment, learner self-evaluation, and peer mutual assessment", therefore it can fully arouse learners' initiative and the ability of self-reflection. Through the model, teachers are able to carry out analysis on learners' academic data so as to give targeted feedback and recommendations; Learners who study this may use the platform to look over their own progress reports, for the purpose of carrying out self-assessment and reflection; The peer appraisals can enable the cooperative study and the mutual promotion among the students.

(4) Evaluation Feedback and Usage: Timely feedback of evaluation outcomes to learners and teachers is given to provide pointed learning suggestions for learners and help them to make adjustments to their study strategies; Teaching effect feedback is given to teachers to help them optimize teaching schemes and model parameters, hence promoting teaching quality.

The evaluation module completes the change from summative assessment to process-oriented and multiple-type evaluation, thus promoting the objectiveness, timeliness and comprehensiveness of teaching evaluation.

4.3.6 Model Optimization Module

The model optimization module mainly puts emphasis on continuously doing refinement to the model's parameters and functions on the basis of learning effect assessment outcomes, feedback from learners, and changing teaching demands, hence enhancing the model's adaptive ability and actual effect. Its core function parts include the below aspects:

(1) Data-driven optimization method: Through analyzing the learning data and evaluation results of learners, we identify problems in the operation process of the model, such as not precise enough adaptation of teaching content, not good enough multimodal interaction, and delayed evaluation feedback. We then carry out the targeted optimization of model parameters, therefore to promote the model’s performance level.

(2) Feedback-pushed optimization: Gather feedback from learners and teachers, in order to find problems and demands when the model is being put into use, improve model functions and working procedures, and promote user experience. For example, adjust teaching content difficulty degrees and recommendation arithmetic methods according to learner feedback, therefore optimize evaluation systems and teaching management functions in accordance with instructor feedback.

(3) Technique Promote and Perfect: Along with the progress of artificial intelligence, big data, and multi-mode technologies, therefore, it is a necessary thing to in time use new techniques and methods to promote model function abilities, promote intelligent abilities, and raise adaptive abilities. For example, putting more advanced speech recognition technologies into practice can raise the accuracy of pronunciation recognition, hence at the same time the usage of sentiment analysis methods allows more accurate evaluation of the academic achievements and emotional changes of learners.

4.4 Model Technical Support

The putting into use of multimodal study models in individualized teaching for English listening and speaking classes depends on many advanced technologies, mainly including the following groups:

(1) Speech recognition technique: It is mainly utilized for gathering and analyzing speech data of learners to evaluate pronunciation correctness, smoothness, tone and speaking speed, therefore it enables real-time feedback and revision in the process of oral practice. For example, deep learning-based speech recognition algorithms have increased accuracy and anti-noise ability, hence permitting accurate recognition of spoken words among different accents and speaking rates.

(2) Natural Language Processing Technique: It is mainly used for analyzing the text data of learners—including listening transcripts and materials of spoken expression—to extract the language features such as vocabulary scale, grammar level, and logic consistency. This here allows accurate learner feature depiction and tailored teaching content sending. In addition, this thing makes the natural language mutual action between teaching systems and learners become easier, therefore it raises the fluency and the accuracy of communication.

(3) Computer vision technique: It is mainly utilized for gathering and analyzing learners' vision data, for example facial expressions and body motions, to recognize learning conditions and emotional changes, hence it provides basis for regulating teaching speed and carrying out personal teaching guidance. For example, the technology that recognizes facial expressions can judge whether learners keep their attention or feel tired, hence it allows people to make timely changes to teaching content and teaching pace.

(4) Big data technology: It is mainly used to store and analyze the multimodal learning data of learners, which includes learning behavior data and learning outcome data, so as to recognize the learning modes and personalized requirements of learners, hence thus providing

data support for model optimization and the adjustment of teaching plans. In addition, this allows accurate sending of teaching materials and precise assessment of study results.

(5) Virtual Reality (VR)/Augmented Reality (AR) technique: It is primarily used for making immersive English listening and speaking situations, which lets learners do practice of language abilities in true environments and improve language ability level. For example, VR technique can simulate cross-cultural communication situations, letting learners do oral dialogue practices in the simulated environments, hence hence increasing the learning authenticity and the degree of participation.

(6) Machine learning arithmetic methods: Mainly used for exact study worker classification, individual character need digging, and teaching content adjustment. Through the training and the analysis of learners' educational data, prediction models are constructed, which can accurately predict learning demands and results, hence promoting the intelligent abilities of the models.

4.5 Model Execution Flow

The operation work procedure of multi-mode study models in individualized instruction for English listening and speaking classes includes six core steps, which forms a closed-loop operation system:

First Step: The Collection of Multi-Modal Information. By means of the multimodal information gathering module, real-time obtaining of learners' multimodal data (including speech, text, images, and behavioral modes) in English listening and speaking learning processes is carried out, with the gathered data saved on a big data platform to give data support for later analysis and processing.

Step 2: The constructing work of learner accuracy profiles. By means of the learner accurate portrayal module, the collected multi-modal information has gone through preprocessing, feature extraction, and classification for constructing learner accurate portraits, hence it can recognize individual learning requirements and learning problems.

Third Step: Individualized Teaching Adjustment. By means of the individualization teaching adjustment module, the teaching content which fits each person is given out on the basis of the accurate basic situation of learners and the demands of each person, the flexible teaching speed is set up, and the suitable teaching ways are used for making the teaching plans which are customized.

Step 4: The Teaching of Multimodal Mutual Interaction Type. By means of the multi-mode mutual-action teaching module, immersion type listening and speaking study situations are built, thus to let multi-mode interaction happen between learners and the teaching system, and also among learners themselves. This makes convenient the practice of listening and speaking and the one-to-one guidance, and at the same time it can collect the study data of learners in real time.

The Fifth Step: Assessment of Learning Results. Through the module for evaluating study results, a mixture of formative and summative evaluations is utilized by people to comprehensively assess the study processes and results of learners, produce study reports, and supply pointed feedback and suggestions.

Sixth Step: The Model’s Optimization Work. By means of the model optimization module, the model’s parameters and functions are refined on the basis of learning effect assessment results and feedback that comes from learners and teaching staff. Teaching schemes are modified to guarantee continuous fitting to learners' individual demands and teaching needs, hence promoting teaching effect.

5 Implementation Path of Multimodal Learning Models in Personalized Teaching of English Listening and Speaking Courses

5.1 Preliminary Preparation: Establishing Technical Platforms and Resource Libraries

5.1.1 Establishing a Multimodal Learning Platform

The construction of a multimodal learning platform is the foundation which supports the realization of models. The platform has the necessity to integrate core functions which include multimodal information gathering, learner feature description, personalized teaching adjustment, multimodal interactive teaching, and learning result assessment, therefore to realize closed-loop model operation. The development of this platform must satisfy the hereunder requirements:

(1) All-round functions: The platform must have core characteristics which contain multi-mode information gathering, learner feature description, individual customized content sending, multi-mode mutual action, study evaluation, and model optimization, hence it satisfies all demands of individual customized teaching.

(2) Friendly-to-user operation: The design of this platform's interface must be simple and easy-to-perceive, with clear operation steps to help the use of frontline teachers and learners. Through the platform, teachers are able to obtain learners' learning data and make teaching plans become better, meanwhile, through this platform, learners can get learning resources which suit individuals, do practice of listening and speaking, and look over learning reports.

(3) Technology Stable Character: The platform must use a stable and dependable technical frame structure to guarantee the correctness of multi-mode information collection, data transmission safety, and system movement working stability, hence preventing problems such as slow operation or system breakdown.

(4) Many kinds of device compatibility: The platform needs to support different terminal devices which include computers, tablets, and smartphones, for realizing multi-device synchronization, hence it lets learners carry out study and practice at any time and any place, thus enhancing the convenience of study.

At the same time, schools ought to be furnished with necessary technical apparatuses such as speech recognition systems, cameras, and VR/AR facilities to offer hardware support for platform running; strengthen network basic establishment to guarantee unblocked connection and help real-time gathering and sending of multi-modal information.

5.1.2 Building a Multimodal Teaching Resource Library

The multi-mode teaching resource storage acts as the core support for personalized teaching adjustment, therefore it needs the integration of various kinds and difficulty degrees of English listening and speaking teaching resources, hence to satisfy the individual demands of different learners. The building of the resource storage bank must follow the below principles:

(1) Diversity Principle: The resource storage bank should include various-mode teaching materials that contain text, sound, pictures, videos, and situation simulations, for instance, listening materials, speaking scripts, situation videos, VR/AR environments, and practice question banks. This method increases the kinds of resources, in order to satisfy the different study hobbies that learners have.

(2) Distinction Principle: The resource storehouse should classify materials into three

layers—elementary, middle-level, and high-level—according to learners' language ability level. Every level includes listening and speaking study materials which have different difficulty degrees, in order to suit learners who are in different learning stages. For example, basic level resources mainly have daily communication talks and easy listening materials; Materials of intermediate level put their focus on dialogues that happen in workplaces and listening exercises which are related to news; while high-grade resources put stress on speeches and cross-culture communication materials.

(3) Principle of Practicability: The resources inside the repository must have close alignment with daily life and actual usages, including listening and speaking materials for many situations such as daily interchange, work places, study abroad, and language level tests, hence it guarantees learners can effectively put the obtained knowledge into use. In addition, the resources ought to be rapidly renewed for the inclusion of newly added content and current backgrounds, hence their timeliness can be promoted.

(4) Scalability Principle: The resource storage bank ought to reserve connecting ports for resource uploading and renewing, thus enabling teachers and learners to upload and share high-quality teaching resources on the basis of teaching and studying demands, therefore continuously enriching the storage bank's content.

The establishment of a resource storage bank can utilize a "self-development + integration of existing resources" method. On one hand, the school arranges teachers to by themselves make multimodal teaching resources, for example situation-based videos and oral practice scripts. On the other hand, current high-quality resources are got together, including excellent English listening and speaking textbooks from home and abroad sources, as well as materials from network study platforms, hence realizing that resource integration is optimized.

5.2 Mid-term Implementation: Phased advancement of personalized teaching

5.2.1 Phase 1: Learner Precision Diagnosis and Profile Construction (1-2 weeks)

The core purpose of this stage is to finish accurate learner judgment, build learner feature files, and find out individual learning requirements. The concrete execution steps are as what follows:

(1) Multi-modal information gathering: By a multi-modal study platform, first-phase multi-modal information of learners is collected, including voice and character materials from fundamental tests (listening and speaking), character messages from study interest inquiry tables, as well as vision and behavior materials from class study activities, hence to comprehensively know the learners' first-phase situation.

(2) Foundation Test and Requirement Investigation: Arrange learners to accept basic English listening and speaking examinations, which cover listening understanding and spoken expression to assess their starting language ability level. By the methods of questionnaire investigation and conversation talk, we collect the individual-related information of learners' interesting points, study speeds and education goals.

(3) Build learner figure: Through the learner figure module, the collected multi-modal information and test data are analyzed by us to extract core characteristics of learners, therefore we establish accurate learner figures. For every learner, personalized individual profiles are constructed, which clearly ascertain their language base foundation, study difficulties, study preferences, and study requirements.

(4) The Verification and Regulation of Profile: Provide the learner profile which has been constructed with feedback to both learners and instructors, collect feedback viewpoints, and carry out adjustment and optimization on the profile, hence to guarantee its accuracy and

practical applicability.

5.2.2 Phase II: Development and Implementation of Personalized Teaching Plans (Weeks 3-16)

The core goal of this stage is to work out personalized teaching schemes on the basis of accurate learner portraits and carry out individual teaching through multi-modal interactive teaching modules. The concrete execution steps are as what follows:

(1) Establishment of Customized Teaching Schemes: According to the accurate personal files of learners, and in accordance with the goals of English listening and speaking courses, teachers are the persons who make individual teaching schemes for every student, which clearly state teaching content, progress, means and study assignments. For example, learners who have worse basic conditions get concentrated training on elementary pronunciation and easy conversations, while those who have better basic conditions participate in complicated situation speaking training and listening understanding tasks.

(2) Personalized teaching content transmission: By a personalized teaching adaptation module, custom-made multimodal teaching resources—inclusive of listening materials, speaking manuscripts, and scene videos—are sent in real time according to learners' teaching plans, hence permitting self-directed study at the speed they themselves have.

(3) Implementation of Multimodal Interactive Teaching

This research has employed multi-modal mutual action modules for offline class teaching and online individual practice. Inside classroom, VR and AR scene simulation, role playing and group talking are utilized by teachers to construct immersion-type listening and speaking surroundings. Teachers carry out real-time monitoring on students' learning conditions, and thus provide guidance that has specific targets. On the internet, learning persons complete self-guidance trainings, amend speech sounds through speech identification technique, and carry out situation-grounded exercises. This system carries out automatic recording of all learning data, in order to support subsequent personalized adjustment of teaching work.

(4) Real-time Teach and Question Answer: By a multi-mode study platform, teachers are able to watch learners' advancement and participation in real time, thus providing personal support that is matched to each person's difficulties. For example, learners who have difficulty on pronunciation can obtain directional guidance through audio showings and video teaching classes, while those who need to promote listening comprehension abilities can get practice materials and listening skills to promote their capacities. This platform also permits online question sending, which lets teachers in time deal with questions by means of written messages, sound messages, or video interactions, hence ensuring non-stopped study experiences.

5.2.3 Phase III: Learning Outcomes Tracking and Teaching Adjustment (Weeks 17-18)

The core goal of this stage is to follow the study results of learners, quickly adjust teaching methods according to assessment outcomes and study data, and therefore guarantee the effect of individualized teaching. The concrete executing steps are as below:

(1) Study data tracking and analysis: By a multi-modal learning platform, real-time gathering of learners' study data is carried out, including study progress, practice times, mistake distribution, pronunciation correctness, and spoken fluency. This method allows for the dynamic carrying out of analysis on learners’ study situation and study results, thus finding out current problems and insufficient points that exist inside their study processes.

(2) Carrying out stage-by-stage evaluation work: Organize learners to accept regular examinations through combining formative evaluation and summative evaluation, so as to comprehensively assess the promotion of learners' listening and speaking abilities. At the

same time, we carry out questionnaire investigations and talk with persons to get learners' satisfying degree on individual teaching schemes and collect feedback opinions, therefore this gives foundation for teaching adjustment work.

(3) The Optimization of Teaching Schemes: On the basis of learning data analysis and periodic examination outcomes, personal teaching schemes are adjusted to deal with learners' special needs. As an example, students who learn more slowly may get advantages from a decreased teaching speed with added foundational practicing, whereas those who have problems with pronunciation problems need more strengthened training courses. To learners who have low study engagement, specially-made study materials and adjusted teaching ways are used to raise study motivation and study passion.

5.3 Post-implementation Optimization: Model Iteration and Teaching Refinement

5.3.1 Model Optimization Iteration

After finishing teaching execution, the multimodal study model carries out repeated optimization via analysis of learners' information, evaluation outcomes, and feedback from both teachers and students. Crucial promotion points contain improving personal adaptation calculation methods to promote content transmission correctness, promoting multi-mode mutual action functions for more smooth participation and stronger user experience, and ameliorating learning assessment systems to guarantee more scientific and pointed evaluation works. In addition, the most front multi-modal technologies which include the advanced emotion identification and voice synthesis are got together promote the lifting of model ability and raise the level of intelligent performance.

Furthermore, a durative mechanism for model optimization ought to be set up, which includes regular gathering of feedback from learners and teachers, tracking of tendencies in educational technology development, and continuous model promotion to guarantee consistency with individualized English listening and speaking teaching demands, hence thus promoting teaching effect.

5.3.2 Summary and Promotion of Teaching Experience

This research summarizes the execution experience of multi-mode learning models in individualized English listening and speaking teaching, finds out current difficulties and hence puts forward measures to build reproducible and extendable teaching methods. By means of campus communication activities, teaching discussion meetings and academic published articles, we share teaching experience points and study results to provide reference materials for other schools and education workers who carry out self-designed English language study methods.

At the same time, we ought to reinforce the training for teachers, thus promoting their proficiency in the application of multimodal technologies and the teaching methods with personalization. This will permit more educational workers to masterfully use multimodal learning modes in English listening and speaking teaching, hence promoting the wide spread application of multimodal technologies in this domain, thus pushing forward the digital and personalized transformation of English language education.

6 Experimental Validation of Multimodal Learning Models in Personalized Teaching of English Listening and Speaking Courses

6.1 Experimental Design

6.1.1 Objective of the Experiment

The main purpose that this experiment has is to prove that multimodal learning models have good effect in personalized English listening and speaking teaching. Key evaluation standards include: evaluating whether the model correctly judges learners' individual needs, for the purpose of promoting the relevance of teaching; evaluating its ability to make big promotion of students' language ability in listening and speaking; and carrying out examination on its latent capacity to promote learners' engagement and satisfaction degrees. These results hence can supply experiment-based proof for the model's wider utilization.

6.1.2 Experimental Subjects

This research has recruited 200 grade one students who do not take English as their major from a higher education school as study objects, which are randomly separated into one experiment group and one comparison group, each of which has 100 students. Statistics-based analysis did not find obvious differences between the two groups on English ability, study capability, or study enthusiasm ($P>0.05$), therefore it guarantees experiment fairness and scientific validity. The experiment group used a multi-mode learning model for individual English listening and speaking teaching, while the comparison group employed traditional teaching ways for English language gaining.

6.1.3 Experimental Variables

In this experiment, the independent variable is the teaching mode, therefore the experimental group uses a multimodal study mode to carry out individualized teaching, thus the control group uses traditional teaching methods. The dependent variables have contained learners' English listening and speaking ability level, learning activeness, as well as satisfaction degree. Control variables have included teaching length, course content, and teacher qualifications therefore to ensure unchanging teaching conditions among all groups and thus reduce disturbance from unrelated factors.

6.1.4 Experimental Tools and Materials

(1) Experiment apparatus: Multi-mode study platform (which contains core modules like multi-mode information gathering, learner feature description, and individual teaching adjustment), English listening and speaking examination papers, study activity questionnaire, study content degree questionnaire, and talk frame.

(2) Experiment materials: The experiment group used resources from the multi-modal teaching resource storehouse, including listening materials, spoken language scripts, situation videos, and VR environments, which were personally made according to learners' individual requirements. The group which does not receive experiment treatment uses traditional textbooks and matched sound materials, and it follows a unified teaching arrangement.

6.1.5 Experimental Procedure

The experiment time length was 18 weeks, which is consistent with the personal teaching implementation cycle that was talked about before. The concrete operation flow is as what follows:

(1) Pre-test preparation work (First Week): Carry out fundamental English listening and speaking capability examinations for two groups of students in order to evaluate their starting language abilities; Send out learning active degree questionnaires and learning satisfactory degree questionnaires to assess their starting learning situation and satisfactory degrees; Carry out training work on the multimodal learning platform for the experimental group, therefore, to ensure that they can carry out proficient operation work on the platform.

(2) Experiment Putting Into Practice (Weeks 2-17): The experiment group used a many-mode learning model for personalized teaching, following an organized flow of "accurate diagnosis-character building-plan making-mutual teaching-result assessment-model improvement." The control group used traditional teaching ways, carried out teaching by standard content, speed, and methods without using multimodal study platforms or resources.

(3) Post-experiment Evaluation and Questionnaire (Week 18): Carry out final English listening and speaking examinations for the two groups of students, thus to assess their promotion in language ability. Again distribute learning initiative and satisfaction questionnaires for the comparison of the changes of students' learning engagement and satisfaction degrees. From each of the two groups, namely the experimental group and the control group, we select twenty students and twenty teachers to carry out interviews, for the purpose of obtaining deep opinions on teaching effect and finding out the existing difficulties.

6.2 Experimental Results and Analysis

6.2.1 Analysis of English Listening and Speaking Test Results

Before the experiment carries out, statistical analysis on the test scores of English listening and speaking ability between the two groups has not found obvious differences (P>0.05), thus this shows that the starting English listening and speaking levels of students in these two groups are the same. After the experiment finished, we did statistical analysis on the final test scores of the two groups, and the results are shown in the following table:

Table 1: Comparison Table of English Listening and Speaking Test Scores Between Experimental Group and Control Group (Full Score: 100)

group	number of **people**	Divide equally before the experiment	After the experiment, divide equally.	Average score increase percentage	standard error	P price
experimental group	100	62.35	78.62	16.27	5.32	<0.05
control group	100	62.41	68.75	6.34	6.15	<0.05

The test outcomes prove that both the experiment group and comparison group had increases in English listening and speaking examination marks. However, the experiment group obtained a obviously larger promotion (16.27 points) when compared with the control group (6.34 points), with a P-value <0.05 that shows a statistics meaningful difference on the promotion of performance. These research results indicate that the multimodal study model effectively promotes learners 'English listening and speaking abilities, while its individualized

teaching method better satisfies learners' study requirements and promotes teaching effects.

Further deep analysis on test marks in every sub-index (listening and speaking) discovered that the experiment group got average mark rises of 15.89 points for listening and 16.65 points for speaking, therefore the control group obtained average mark increases of 6.12 points for listening and 6.56 points for speaking. These outcomes prove that the multi-modal learning model greatly promotes learners' listening and speaking abilities, hence especially obvious promotions are in spoken language level. The imitation of multi-mode mutual action situations and instant pronunciation feedback have obviously enhanced learners' fluency and correctness in spoken expression.

6.2.2 Analysis of Learning Initiative Questionnaire Results

In the time period before the experiment and the time period after the experiment, questionnaire investigations were carried out for evaluating learning activeness among students in two groups. The question paper used a 5-point Likert scale, to carry out evaluation on four dimensions: study motivation, self-directed study frequency, study objectives clarity, and study engagement, therefore its total score is 100 points. Higher score values have indicated that the learning initiative of people is stronger. The statistical outcomes have been put in the table that follows:

Table 2: Comparison of Learning Initiative Questionnaire Scores Between Experimental Group and Control Group (Full Score: 100)

group	number of people	Divide equally before the experiment	After the experiment, divide equally.	Average score increase percentage	P price
experimental group	100	60.23	79.56	19.33	<0.05
control group	100	60.35	65.42	5.07	<0.05

The questionnaire results give demonstration that the experiment group has a significantly bigger promotion in learning initiative (19.33 points contrast 5.07 points in the control group, $P<0.05$), hence it indicates that the multimodal learning model can effectively promote the degree that learners are engaged in learning. This point can therefore be ascribed to that the model provides personalized content and flexible speed arrangement, hence it lets learners by themselves choose learning materials and time arrangements according to each person’s own interests. In addition, the immersive multi-modularity mutual action environment promotes the study pleasure and joining degree, therefore it stimulates the learners' study motivation and the initiative study behaviors.

6.2.3 Analysis of Learning Satisfaction Questionnaire Results

When the experiment had been finished, a questionnaire investigation was carried out for learning satisfaction among the students of two groups. The investigation paper used a 5-point Likert scale to assess four aspects: teaching content, teaching ways, teaching results, and study experience, with a whole score of 100 points. Higher numerical marks have pointed to a higher degree of the learning satisfaction. The outcome of statistics is put in the following table:

Table 3: Comparison of learning satisfaction questionnaire scores between the experimental group and control group (out of 100 points)

group	number of people	average	Excellent rate (≥80 points)	Good rate (70-79 points)	Pass rate (60-69 points)	Failure rate (<60 points)
experimental group	100	80.12	68%	22%	8%	2%
control group	100	66.78	23%	35%	32%	10%

The questionnaire outcomes make manifest that the experimental group obtained significantly higher learning satisfaction marks (80.12 vs. 66.78) and significantly better academic achievement indicators (68% excellence ratio vs. 23% in the control group, having a 2% failure ratio vs. 10%). These result points show that the individual teaching mode which rests on multi-mode learning has effectively solved learners' personal demands, hence promoting both the learning experience and the satisfaction degree. Through personal content provision, immersive mutual action environments and real-time response guidance, interviews have discovered that most students in the experimental group said they have obtained promoted study connection, measurable advancement, and raised interest in English listening and speaking abilities.

6.2.4 Interview Result Analysis

When experiment was finished, we did face-to-face talks with students and teachers that came from the experimental group and also the control group. The results of these talks are shown as below:

(1) Student interview results

Students of the experimental group said that the personalized teaching of the multimodal platform conforms to their personal learning level and needs, hence it breaks the restrictions of traditional unified teaching. Immersive VR/AR environments very much promote study enthusiasm, while real-time voice pronunciation correction helps them correct mistakes in time and raise study efficiency. Through comparison, the students in the control group hold that the traditional teaching method possesses single and identical content. It has no lively practice situations, cannot adapt to individual study differences, hence causes low study motivation and hence slow spoken English promotion.

(2) Teacher interview results: Teachers of the experimental group stated that the multimodal learning model allows them to accurately find learners' individual needs, hence assisting "customized teaching" and thus lowering teaching work burden. Through the platform, learners' data is real-time monitored by them, therefore they can timely find out learning difficulties, hence provide pointed guidance, thus promote teaching effect. On the opposite side, teachers who are in the control group have pointed out that traditional teaching methods have difficulty in adapting to individual differences that exist among learners, cannot accurately evaluate learning conditions and requirements, provide feedback that is late, and produce relatively limited education results.

6.3 Experimental Conclusions

Through this experimental verification, the following conclusions were drawn:

(1) The multimodal study model can accurately hold the individual demands of learners, build accurate learner feature files, and realize the custom adaptation of teaching content, speed, and methods. Therefore, it greatly promotes the relevance and effect of English

listening and speaking teaching.

(2) The multi-mode study model can effectively promote the promotion of learners' English listening and speaking abilities, especially in the aspect of oral expression. By means of the immersion type scene simulation and the real time pronunciation feedback, it has obvious enhancement to learners' fluency and correctness in spoken language.

(3) The multimodal study models can promote the study initiative and satisfaction of the learners. Through providing individual study experiences and attractive mutual action scenes, they arouse learners' interest and internal drive, thus helping the completion of the change from passive study to active study.

(4) The multi-modal learning model gives firm technical support to frontline English teachers, therefore enabling them to break through the restrictions of traditional teaching methods, hence realize teaching that is personalized for each individual learner, and thus promote both teaching efficiency and teaching quality.

7 Optimization Strategies for Applying Multimodal Learning Models in Personalized Teaching of English Listening and Speaking Courses

7.1 Optimize model functionality to enhance intelligent capabilities

7.1.1 Improve multimodal information acquisition and analysis capabilities

Carry out optimization on the multimodal information obtaining module to expand the scope of data collection, incorporate multi-dimensional measurement indicators such as learners' emotional conditions and attention degrees to promote the comprehensiveness and accuracy of data. Make refinement on preprocessing algorithms for the purpose of decreasing noise disturbance and raising data availability. Strengthen learner feature extraction and profile building algorithms to purify learner profiles, thus enabling more accurate confirmation of individual learning demands and teaching difficulties.

7.1.2 Optimization of Personalized Adaptation Algorithm

Through the combination of experiment outcomes and learner's feedback information, we carry out optimization on the algorithm of personalization teaching adaptation modules, for the enhancement of the accuracy of content transmission. This system, on the basis of the progress that learners make and the results they get from learning, dynamically makes adjustments to the difficulty of teaching content and the frequency of content delivery. We carry out refinement on the teaching rhythm adaptation feature to let more flexible pacing adjustments be made, hence better letting it match with various kinds of learning habits. In addition, we promote teaching method adaptive abilities through automatically recommending suited methods on the basis of learners' fondness, hence hence promoting the whole learning experiences.

7.1.3 Upgrade multimodal interaction functionality

Promote the upgrading of the multimodal interactive teaching module through optimizing the VR/AR scene simulation ability, and add more real-life situations with practical uses to increase the degree of authenticity and participation. Through the promotion of speech recognition accuracy and noise resistant ability, we optimize the voice interaction functions, which may support many kinds of accents and speaking rates. We introduce extra interaction

ways which include gesture control and facial expression, for the purpose of diversifying interaction patterns and guaranteeing smooth running, therefore a more immersion study environment can be constructed.

7.2 Improve teaching resource development and enhance resource compatibility

7.2.1 Diversified Multimodal Teaching Resource Types

Enlarge the building scope of the multimodal teaching resource storehouse through making resource types various, which includes VR/AR scene resources, mutual practice materials, and real-time live-broadcast teaching resources, in order to satisfy learners' different fondnesses and demands. According to the present development tendency and the actual demands of learners, we will import professional resources including working place English, cross-culture communication and examination preparation materials, thus to promote the practical property and time effect of these resources.

7.2.2 Optimization of Resource Stratification and Classification

According to learners' language ability and study goals, we further improve resource layering through dividing materials into many levels: basic, high-level, middle-level, and special breaking-through levels. Every level holds resources that have different difficulty degrees and categories to guarantee accurate matching with learners’ ability situations. A standardized resource label marking system is put into practice to give precise metadata tags, thus enabling learners to with efficiency find and get customized learning materials, hence at the same time optimizing the efficiency of resource utilization.

7.2.3 Establish a long-term mechanism for resource renewal

Establish the sustainable mechanism for the renewal of multimodal teaching resources through the regular collection of resource demands from learners and educators, therefore guaranteeing timely renewals and supplements. Let professional teachers and technical specialists carry out independent development of high-quality multi-modal teaching materials, and at the same time carry out the optimization integration through the integration of excellent domestic and international education resources. Advocate that learners should upload and share their learning materials, cultivate a cooperative resource construction mode of "teacher-leading development + learner-pushing sharing" to continuously enrich the resource storage bank.

7.3 Enhancing Teacher Competencies to Promote Effective Model Application

7.3.1 Conduct multimodal technology training

Make systematic cultivation projects for teachers that place key attention on multimodal technical methods, artificial intelligence usage, and comprehensive study platforms, for the enhancement of the technical ability that educators have. This lets teachers, therefore, utilize multimodal study models to carry out personalized teaching, data-based study analysis, and optimized teaching methods. We ought to organize regular teaching seminars and professional exchange activities, so as to make the knowledge sharing of multimodal teaching methods become smooth, therefore we can cultivate collaborative learning and mutual professional promotion among educational workers.

7.3.2 Enhancing Teachers' Personalized Teaching Competence

Strengthen the training of personal teaching doctrines for education workers to cultivate a student-centered method, fully understand the importance of individual distinctions, and heighten their consciousness of customized teaching. Carry out professional training on personalization-oriented teaching methods and skills to let teachers have necessary abilities which include learner requirement evaluation, making of individualized teaching plans and teaching flow improvement, hence promoting their ability for carrying out personalized teaching work.

7.3.3 Establishing a Teacher Incentive Mechanism

Set up and perfect teacher encouragement systems to urge education workers to actively use multi-modal learning modes for individual teaching, therefore acknowledgement and prizes are given to excellent teachers on multi-modal teaching practices. Put the carrying out of multimodal teaching into the assessment of teachers’ work performance to raise teaching motivation and initiative, therefore promote the wide use of multimodal learning models in English listening and speaking teaching.

7.4 Optimize the teaching implementation process to enhance teaching effectiveness

7.4.1 Refine the learner precision diagnosis process

We carry out optimization on the learner diagnosis process through the integration of multimodal data gathering and foundation examinations to obtain all-round understandings of learners' language ability, study requirements, and study liking habits. A dynamic learner characteristic updating mechanism has been established, which continuously optimizes learner portraits according to progress tracking and study results, therefore guaranteeing their accuracy and actual application possibility. This therefore provides the reliable data support for the teaching that adapts to personalized students.

7.4.2 Strengthen dynamic monitoring and adjustment of the teaching process

Strengthen dynamic supervision of the teaching process through employing multimodal learning platforms to record learners' progress, engagement condition, and learning results in real time, thus allowing timely discovery of learning difficulties. Set up a dynamic regulating mechanism for teaching schemes that allows timely changes to teaching content, progress and methods on the basis of learners' data and feedback, hence ensuring that teaching results are targeted and effective.

7.4.3 Improve the learning outcome evaluation and feedback mechanism

We shall further consummate the learning result evaluation system through increasing the weight of form evaluation, and through consummating evaluation indexes, thus to reach comprehensive and accurate evaluation on the learning processes and results of learners. The feedback mechanism shall have optimization carried out to guarantee timely and direction-specific reactions, thus giving concrete learning suggestions to learners for helping them adjust their study methods. In addition, we shall expand the scope of evaluation participation through strengthening self-assessment and peer evaluation, therefore fully stimulating the self-reflection abilities and learning initiative of learners.

8 Conclusion

This research carries out investigation on the putting into practice of multimodal learning models in the personalized teaching that is for English listening and speaking courses. Via the reviewing of related theories, the analyzing of present teaching practices, the constructing of implementation frames, the designing of teaching experiments, and the proposing of optimization strategies, below key conclusions are hence drawn:

(1) At present, English listening and speaking teaching meets problems including unified teaching modes, insufficient teaching resources, uniform assessment ways, and not enough combination with technology. These problems come from old teaching ideas, not enough technical support, teachers' ability has shortcomings, and resource building is slow. The utilization of multimodal study models can effectively solve these issues, hence giving firm support for individual-based English listening and speaking education.

(2) The combination of multi-mode study models with personalized English listening and speaking teaching has logical correctness. The information with multiple modes gives comprehensive support to the diagnosis of individual students' learning requirements, the teaching resources with multiple modes guarantee that content can be adjusted in a custom way, and the interaction with multiple modes helps the personalization of teaching procedures. This combined effect makes teaching have different differences, therefore it promotes the related degree and effect of English listening and speaking education.

(3) The multimodal study learning model frame that this research has made includes six core modules: multimodal data gathering, learner feature description, individual teaching adjustment, interactive multimodal instruction, study result evaluation, and model improvement. By using technologies which include speech recognition, natural language processing and computer vision, it has built a closed-loop working system that includes "data gathering-analysis-adaptation-instruction-evaluation-improvement", it therefore effectively realizes individualized teaching in English listening and speaking classes.

(4) Experiment examination and proof show that the multimode study model can exactly seize learners' individual demands, hence effectively promote their English listening and speaking ability, hence improve learning activeness and satisfaction. It gives operable teaching references for front-line English teachers, promoting the digital and personalized change of English listening and speaking teaching.

(5) For solving problems that exist in the using process of multimodal learning models, we put forward optimization methods, which include increasing model functions, promoting the construction of teaching resources, promoting teacher ability improvement, perfecting teaching practice steps, and reinforcing technical guarantee supports. These measures can further promote the model's adaptive ability and actual effect, therefore hence promoting its wide adoption and utilization.

Funding

This research got the support from 2025 Project of the 14th Five-Year Scheme for Educational Science in Shaanxi Province “Study on the Integration Way of Intelligent College English Teaching and Curriculum Ideology and Politics under the Visual Angle of Educational Ecology” (Project Number: SGH25Q589).

About the author

Dandan Yang was born in Xi'an，Shaanxi。P.R. China, in 1994. She obtained a Master's degree from Shaanxi Normal University in China. I am currently working at the School of Science and Engineering, Xi'an Kedagaoxin University. My main research direction is higher education in English and British and American literature. y15202952790@163.com

References

[1] Payala K V, Jeet A A, S A.Enhanced multimodal deep learning framework for emotion classification with Aquila optimizer based ensemble fusion[J].Array, 2026, 30100760-100760.DOI:10.1016/J.ARRAY.2026.100760.

[2] Meng X, Bachmann M, Yang F, et al.Porosity prediction in laser beam welding with a multimodal physics-informed machine learning framework[J].Advanced Engineering Informatics, 2026, 74(PA):104611-104611.DOI:10.1016/J.AEI.2026.104611.

[3] Liu H, Shi L, Shi Y, et al.A semi-supervised multimodal fusion framework with adversarial contrastive learning for Alzheimer's disease diagnosis[J].Engineering Applications of Artificial Intelligence, 2026, 174114537-114537. DOI:10.1016/J. ENGAPPAI.2026.114537.

[4] Man J, Yang C, Chen D, et al.Wearable multimodal sensing with deep online learning for real-time onboard crew behavior prediction[J].Engineering Applications of Artificial Intelligence, 2026, 174114582-114582.DOI:10.1016/J.ENGAPPAI.2026.114582.

[5] Lee J, Lee S, Lee J Y, et al.Predicting seismic floor response for nuclear power plant structures with time-series uncertainty propagation using attention-enhanced multimodal deep learning[J]. Reliability Engineering and System Safety, 2026, 272(P2): 112582-112582.DOI:10.1016/J.RESS.2026.112582.

[6] Batool A, Kim W Y, Byun C Y.MAVL-DRL: Multimodal foundation using multi-agent deep reinforcement learning for intelligent predictive maintenance of wind turbine energy systems[J].International Journal of Electrical Power and Energy Systems, 2026, 177111781-111781.DOI:10.1016/J.IJEPES.2026.111781.

[7] Zheng Y, Zhao C, Kong W.Spectral filtering and multi-view graph representation learning for multimodal recommendation[J]. Applied Soft Computing, 2026, 196115092-115092.DOI:10.1016/J.ASOC.2026.115092.

[8] Wang C, Zhang W, Chen G, et al.Divide-and-conquer: Prompt-based distribution learning for multimodal sentiment analysis[J].Information Fusion, 2026, 133104293-104293.DOI:10.1016/J.INFFUS.2026.104293.

[9] Arzu E G, Umar M, Khan A, et al.Adaptive multimodal emotion detection for mental health monitoring using deep learning[J].Information Sciences, 2026, 744123385-123385.DOI:10.1016/J.INS.2026.123385.

[10] LUAN Q.TRANSFORMER FOR INDIVIDUALIZED SPORTS MOTION ANALYSIS AND HEALTH-RELATED ACTIVITY RECOGNITION BASED ON MULTIMODAL

DEEP LEARNING METHOD[J].Journal of Mechanics in Medicine and Biology, 2026, (prepublish):DOI:10.1142/S0219519426400373.

[11] Li Z, Lu J, Cui J, et al.Functional customization of peptide linkers in fusion proteins through multimodal deep learning approach[J].Synthetic and Systems Biotechnology, 2026, 13362-375.DOI:10.1016/J.SYNBIO.2026.02.003.

[12] Xu Z, Chen X, Xu J, et al.Multi-sensor signals augmented multimodal MAML-1DCNN-RBEAM deep transfer learning algorithm for wind turbine bearing fault diagnosis[J].Nondestructive Testing and Evaluation, 2026, 41(4):1897-1925. DOI: 10.1080/10589759.2025.2487927.

[13] Bresalier S R, Eidens R M, Krammes L, et al.A MULTIMODAL STOOL RNA, FIT AND MACHINE LEARNING CONCEPT FOR DETECTION OF ADVANCED PRECANCEROUS LESIONS AND COLORECTAL CANCER.[J].Cancer prevention research (Philadelphia, Pa.), 2026, DOI:10.1158/1940-6207.CAPR-25-0425.

[14] Patel D, Dhavale V S, Mhetre B B.Personality in 3D: multimodal deep learning framework for big five trait prediction[J].Neural Computing and Applications, 2026, 38(7):212-212.DOI:10.1007/S00521-026-11979-3.

[15] Kamran J, Zedler M T, Schmitt M, et al.Comparative analysis of transfer learning architectures for multimodal microscopy based IBD histopathology[J].Discover Imaging, 2026, 3(1):2-2.DOI:10.1007/S44352-026-00024-7.

[16] Neethu V T, Kanaga M G E.Coyote and Badger Makeup Artist Optimization based Hybrid Deep Learning for Multimodal Sentiment Classification with Emotion Recognition[J].SN Computer Science, 2026, 7(4):302-302.DOI:10.1007/S42979-026-04895-9.

[17] Vu A T, Chuyen T M, Anh D T N, et al.Multichannel Learning Framework for Enhanced ECG Signal Classification Using Wavelet and MFCCs Features[J].Iranian Journal of Science and Technology, Transactions of Electrical Engineering, 2026, (prepublish):1-11.DOI:10.1007/S40998-026-01062-X.

[18] Qu J, Bi D, Liu Y.Adaptive multimodal fusion-driven reinforcement learning for robotic peg-in-hole assembly[J].Journal of Intelligent Manufacturing, 2026, (prepublish):1-19.DOI:10.1007/S10845-026-02831-5.

[19] Sha X, Wang S, Sun X, et al.Outlier-aware orthogonal latent space learning for multimodal alzheimer's disease diagnosis[J].The European Physical Journal Plus, 2026, 141(3):323-323.DOI:10.1140/EPJP/S13360-026-07572-1.

[20] Bao G, Zhang Q, Miao D, et al.Incomplete Multimodal Federated Learning via Masking and Contrasting Prototypes.[J].IEEE transactions on neural networks and learning systems, 2026, PPDOI:10.1109/TNNLS.2026.3658522.

[21] Kuo, Chun I.An Action Research Study on Implementing Flipped Classroom and Online Learning Resources in an English Listening and Speaking Course for University Military Service Program Students[J].New Explorations in Education and Teaching, 2026, 4(1):DOI:10.70711/NEET.V4I1.8520.

[22] Rong A.The New Connotation of Cross - Cultural Communication Competence and Innovation of Teaching Models in the AI Era - A Case Study of College English Listening and Speaking Courses[J].Journal of Higher Education Teaching, 2025, 2(3): DOI:10.62517/JHET.202515335.

[23] Zhao Y.A Study on Teaching Strategies for "Telling Chinese Stories Well" in Advanced English Listening and Speaking Courses Under the Guidance of POA Theory[J].Higher Education and Practice, 2025, 2(3):DOI:10.62381/H251317.

[24] Zhao Q.A Study on the Practical Application of Multimedia Technology in College English Listening and Speaking Classroom Teaching[J].Applied Mathematics and Nonlinear Sciences, 2025, 10(1):DOI:10.2478/AMNS-2025-0631.

[25] Kong N, Zhang M.The Practice of Ideological and Political Education in College English Listening and Speaking Courses within the Framework of Intelligent Learning[J].Curriculum and Teaching Methodology, 2024, 7(9):DOI:10.23977/CURTM.2024.070907.

[26] Jiaxue C.The Challenges and Countermeasures in Integrating Chinese Traditional Culture into English Listening and Speaking Course[J].Journal of Contemporary Educational Research, 2024, 8(9):303-309.DOI:10.26689/JCER.V8I9.8074.

[27] Hu X, Fan M, Li Z.The Captivating Wine: A Case Study of Situational Teaching Method in High School English Listening and Speaking Classes[J].The Educational Review, USA, 2024, 8(9):DOI:10.26855/ER.2024.09.007.

[28] Pan J.Inquiry into Teaching Listening and Speaking in High School English Classes Based on an Activity Perspective[J].Academic Journal of Management and Social Sciences, 2024, 8(2):114-117.DOI:10.54097/0WCM9Z22.

[29] Yaqin L.A Study on the Teaching Design of Blended College English Listening and Speaking Course Based on Unipus[J].International Journal of New Developments in Education, 2024, 6(7):DOI:10.25236/IJNDE.2024.060717.

[30] Li S.Application of OBE Educational Philosophy in English Listening and Speaking Courses in Vocational Colleges under the Context of Industry-Education Integration[J]. International Journal of New Developments in Education, 2024, 6(7):DOI:10.25236/IJNDE.2024.060738.

[31] Xinhan L, Caihong Z, Xueli Z, et al.Instructional Design of Front Vowels in College English Listening and Speaking Course Based on BOPPPS Teaching Model[J].Lecture Notes on Language and Literature, 2024, 7(1):DOI:10.23977/LANGL.2024.070104.

[32] Clark S J, Terrett M.Developing a second language listening and speaking assessment instrument for Use in student-led seminars in a chinese middle school english language acquisition class[J].International Journal of Chinese Education, 2024, 13(1): DOI:10.1177/2212585X241234335.

[33] LIU C.Practice and reflection on online and offline blended teaching of cross-school study in building and sharing news English listening and speaking courses[J].Region - Educational Research and Reviews, 2023, 5(5):DOI:10.32629/RERR.V5I5.1451.

[34] Wang L, Liu H, Chen G, et al.Research on Ideological Teaching Practice of College English Listening and Speaking Course under the Optimized Teaching Model[J]. International Journal of New Developments in Education, 2023, 5(23):DOI:10.25236/IJNDE.2023.052330.

[35] Li X.Research on the Application of English Movies in Junior High School English Listening and Speaking Classroom Teaching[J].International Journal of New Developments in Education, 2023, 5(14):DOI:10.25236/IJNDE.2023.051413.

[36] Wei L.A Case Study on the Multimode Teaching Design of College English Listening and Speaking Course Combined with Chinese Culture under the Guidance of the Production-Oriented Approach[J].Frontiers in Educational Research, 2023, 6(2): DOI:10.25236/FER.2023.060217.

[37] Chen B.Teaching Mode Reform of English Listening and Speaking Course under the Background of Smart Education[J].Curriculum and Teaching Methodology, 2022, 5(11):DOI:10.23977/CURTM.2022.051113.

[38] Danlu L.Optimization of Classroom Teaching Strategies for College English Listening and Speaking Based on Random Matrix Theory[J].Mathematical Problems in Engineering, 2022, 2022DOI:10.1155/2022/8563978.

[39] Song L.The problem-based learning mode for teaching English to college students[J].International Journal of Continuing Engineering Education and Life-Long Learning, 2022, 32(5):640-649.DOI:10.1504/IJCEELL.2022.10038810.

[40] Zhou L.Brief analysis of teaching design of English listening and speaking courses under e-learning environment[J].BioTechnology: An Indian Journal, 2014, 10(15):