# Development of Computer-Aided English Listening System based on BS Architecture

Jiya Ding

School of Foreign Languages, Ningxia Normal University, Guyuan, Ningxia 756000 China, 82001022@nxnu.edu.cn

Corresponding Author: Jiya Ding. E-mail: 82001022@nxnu.edu.cn

**Abstract.** In this paper, we study and analyze the computer-aided English listening system, and develop and apply the system using BS architecture. The system is deployed in a way that the front-end and back-end are separated and decoupled from each other. The back-end is programmed with Java language, designed with Model View Controller (MVC) pattern, developed with SSM framework, using Web-Service to provide an interface to realize the connection between modules, and Elasticsearch search engine to provide search services for the system. In the design and development phase, the system was developed based on the functional model constructed in the previous phase, using the English audio-visual tutorials as the basis for learning content and other English listening learning materials as supplementary resources to develop the adaptive learning listening system. Based on adaptive learning theory, constructivism theory, and second language acquisition theory as the main learning ideas, this study uses adaptive testing and learning traces to dynamically record learners' learning process, and uses algorithms to analyze data and recommend learning paths to design and develop an adaptive learning English listening learning system, aiming to assist in teaching English listening courses at university and provide learners with personalized learning. We aim to provide an ideal environment for personalized learning.

**Keywords:** BS architecture; computer-assisted; English listening; system development

**DOI:** https://doi.org/10.14733/cadaps.2022.S1.93-104

## 1 INTRODUCTION

With the rapid development of the Internet and information technology, education is not limited by space and time, especially information technology featuring multimedia, and gradually developing network technology is gradually applied to the field of education. Under the roaring reform of information technology and classroom integration, online education has become an important form to assist teachers' teaching and learners' independent learning. In the context of the modern Internet+ era, online education is not only not limited by physical space in terms of teaching methods, but also provides learners with personalized learning methods. Online education can establish a comprehensive evaluation system based on big data and realize accurate management in scientific decision-making. With the support of big data, it can reveal the national learning rules, realize the visual monitoring of the whole process and provide intelligent support services [1]. Under the demand of the development of the times, how to realize personalized learning support through the network environment has become a challenge for online education. The development of computer technology provides technical support for online education, even though the technology is becoming more mature, without core teaching ideas for teaching and learning, online education provides personalized learning for learners is just an empty talk. Both constructivism and adaptive learning theory have similar learning concepts, that is, the learning process should be a process of divergent thinking, in which learners should construct meaning along different paths and according to their original knowledge base following individual paths [2]. For this reason, online education should provide learners with personalized and diverse learning paths and learning support in the learning process, so that learners can learn according to their learning styles and original cognitive structures.

The popularity of the Internet allows every learner to access resources online for learning, especially for college students who already have a certain level of information literacy, learning through intelligent systems is a way of learning with immediacy and convenience. Learning is a process of a complex interaction between students and their surroundings, and to explore this process, it is necessary to choose the right entry point as well as appropriate research methods, tools, and techniques [9]. In the learning of intelligent learning systems, the learner is in a fully autonomous learning environment with appropriate learning path recommendations that enable the learner to find the entry point for learning and to better personalize the learning. Adaptive path learning is an intelligent form of learning and a learning style that allows learners to personalize their learning. Based on adaptive learning theory, constructivism theory, and second language acquisition theory as the main learning ideas, this study uses data collection and analysis technology to record learner data, construct learner student models, and design and develop an adaptive learning path for college English listening learning system [3]. The platform is designed with the learner-centered idea, and learners can take a test in this online platform to get a preliminary understanding of their pre-cognitive structure and basic level and learn independently in the system according to their learning needs.

At the same time, the system will recommend the optimal learning path based on the test results and the characteristics of the learner's learning traces and learning style, guide the learner to learn the corresponding content, and provide test questions of the corresponding difficulty at the end of learning and recommend learning resources of deeper difficulty. During the whole learning process of learners, the background will use the stored learner data for secondary development and utilization to build the learning style of learners and fully explore the learning needs of learners. The web platform provides timely feedback and diagnostic analysis of learning results, which enables learners to understand their real-time learning situation and make corresponding adjustments accordingly; teachers give timely guidance and adjust teaching according to learners' feedback on learning results; the web platform can better "tailor teaching" to learning results. The web platform can better "tailor" the learning results and recommend the most optimal adaptive path for learning content. Therefore, the e-learning system designed and developed in this study can provide learners with personalized learning and assistance.

## 2 CURRENT STATUS OF RESEARCH

In the context of the Internet + era, the digital learning environment composed of multimedia information technology provides a diverse platform for online education, and the concept of online learning has changed dramatically with the impact of the times, which also poses a huge challenge for the development of online education at the same time. The goal of software evolution research is to use the history of a software system to analyze its current situation and predict its future development [4].

Zhang et al. [5] divided the listening comprehension process into three interrelated and overlapping stages: perception, parsing, and application. Perception is the process of primitive encoding of the received information; parsing is when the learner breaks down the speech in short-term memory into words and phrases based on syntactic structure and semantic cues and constructs meaningful text-based mental representations based on linguistic structures, rules, and semantic principles, which are stored in long-term memory; and application is when the learner uses existing schemas and linguistic knowledge to connect with mental representations in the brain and correctly understand the speech information. Das et al. [6] viewed listening comprehension as a dynamic information processing process based on information processing theory. That is the process by which instructional material passes from sensory memory to working memory and finally to permanent memory. The learner retains the heard speech material in a raw, unanalyzed form for a short period through sensory memory; the task-relevant listening information is stored in working memory through selective attention.

Qin et al. [7] interpreted the listening comprehension process as "top-down" and "bottom-up" according to the information processing model. The "bottom-up" process involves the learner first identifying the phonological layer through syllables and phonemic positions, then extracting the words from semantic memory based on the identified phonological information, organizing the words into phrase structures until they form parts of speech, and finally linking them to the meaning of sentences in the preceding text to form higher-level units. The "top-down" approach emphasizes the recognition of phrases, vocabulary, and phonology in sentences through the overall discourse, and is a process of moving from higher to lower levels. Kurokawa et al. [8] found that the strategies used more often by learners with good listening skills than those with poor listening skills included attention and self-monitoring strategies. Sun et al. [9] found that high-level learners were more aware of problems in the listening process. Li et al. [10] noted that metacognitive strategy training was most effective for learners with low levels of listening comprehension and least effective for learners with low levels of listening comprehension due to phonological problems.

Whether changes in learning media and learning styles can make a qualitative difference to learners' learning outcomes remains a challenge for educators today. In the online environment, it is crucial to create a learning system that can be adapted to the learners' own learning needs. In college English listening learning, the need for adaptive pathways is even more important because of the differences in non-native learning and learners' cognitive levels. The development and design of adaptive learning paths using new technologies and corresponding pedagogical theories will allow for better-personalized learning for learners.

## 3 BS ARCHITECTURE ENGLISH LISTENING SYSTEM DESIGN

### 3.1 BS Architecture System Design

The Web browser is the main application on the client-side. After the client-side was unified by this model, the core part of how the system functions were implemented was centralized on the server. After installing Oracle, MYSQL, SQL Server, and other databases on the server, a browser, such as Internet Explorer or Netscape Navigator, is installed on the client, and the browser can interact with the database through the Web Server. The biggest advantage of the B/S structure is that it can be operated anywhere without installing any special software. The client only needs to install a computer with Internet access, so there is no installation and maintenance. B/S structure is used increasingly, and it has been changed from a demand state to a state that drives the development of AJAX technology.

The system architecture is designed to bridge the gap between technology and business needs, through the online education and training school OA system requirements analysis, the decision to use flexible enough to handle changes in hardware and software and business needs, weighing the overall impact of design decisions and quality attributes of the B/S architecture. Under the B/S architecture system, the user software work interface is realized through browser-assisted, both transactional logic relies on the user front-end to achieve, and scripted transactional logic relies on the server back-end main functions to achieve, which is conducive to the balance of the computer load and reduce response time. Under the B/S architecture, the administrative access rights can be set comprehensively, and the function of stabilizing the server database can also be realized. The use of contextual interfaces defined around business boundaries ensures a more stable system, and changes to business processes can be better reflected. Build tools that support B/S automation to reduce the significant labor costs of later iterative development and reduce the failure rate. Hide the internal implementation details, especially the services that expose REST-API, and have the same high compatibility for lower versions when updating.

Consider limiting the impact of sacrificing availability and consistency on failed components in the event of a failure. Service call identification enables the tracking of the service call process and the use of service logs and data to observe the entire call chain. The school network, hardware environment, and data backup should be considered during the system implementation. Also, while ensuring the stability of the system, the security of the system should be considered to ensure the safe operation of the system. The business processes involved in the student users and back-end management users of the system need to be fully investigated to ensure strong usability of the system. Running the system should consider the system's information transfer promptly to ensure a high execution rate of the system, as shown in Figure 1.

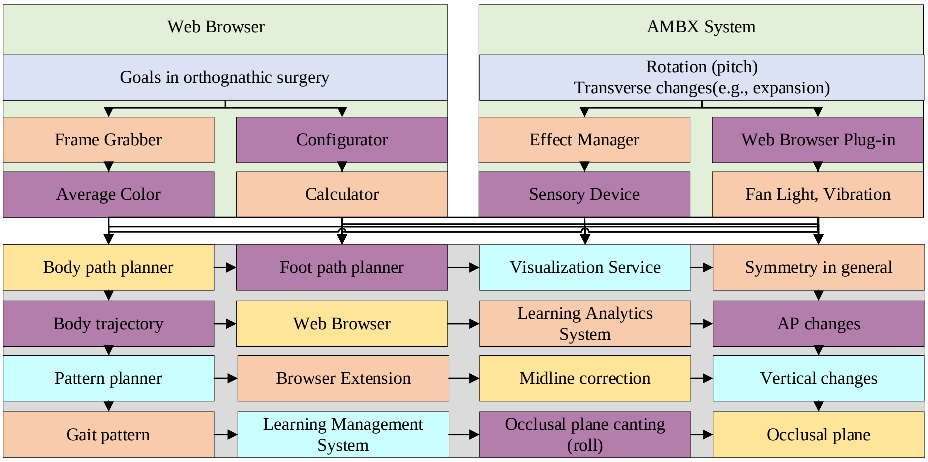


**Figure 1:** Overall system architecture.

The front-end of the system is responsible for sending requests and receiving returned data, and presenting the returned data in a visual form. The server side is responsible for receiving and processing requests, realizing the operation of the database, and returning the data to the front-end. The back-end of the system adopts a layered design based on the MVC model, dividing the back-end into view layer, control layer, and business model layer. The user sends the request to the front-end control through the operation view layer, and then calls the specific controller after the front-end controller's query resolution to hand the request to the business model layer for processing, and returns the processing result to the controller. After parsing, the controller returns to the view layer and finally displays it to the user. The layered design of the system can avoid the view layer directly accessing the database, thus reducing the coupling degree. The control layer is mainly responsible for interacting with the business logic layer and the view layer. The control layer receives the request from the front-end and calls the business model layer to process the request, and the business model layer returns the result of the request processing to the controller.

The knowledge model is an abstract representation of the learning content, mainly the learning tests, learning resources, and knowledge points of the course content, with the constituent elements within each knowledge and the interrelationships between the elements as the intrinsic form of presentation. The knowledge model is a prerequisite for adaptive path learning as well as learning resource recommendations. In the teaching process, knowledge is divided into teaching units and knowledge points, which are teaching contents including basic concepts and basic processes. The information that learners want to learn, like theories, exercises, learning topics, and multimedia resources, etc., are arranged in a certain order, and this arrangement constitutes the teaching units; and the knowledge model, which is the structured presentation of this knowledge to learners. The knowledge model is a structured presentation of this knowledge to the learner and enables dynamic assessment of the learner's learning. A good knowledge model helps learners to build a cognitive system of subject knowledge.

The adaptive learning system is learner-centered, so the student model is the core of the whole functional module and should be designed according to the learning needs of the learners, and the generation of a series of adaptive actions should be based on learner-related characteristics, and most of the current studies on learner modeling show individual characteristics such as the learning style, interest background, cognitive level, and learning behavior of learners, and learner-related characteristics are the specific content in the student model, and constructing the student model is to construct the learner characteristics, and the specific content of the student characteristics is shown in Table 1.

| Characteristics | Constituent elements |
|---|---|
| Basic learner information | Name, class, contact information |
| Learning style | Divergence, assimilation, convergence, mediation |
| Cognitive level | Knowledge base, cognitive ability (remembering, comprehension, application, analysis, synthesis) linear resource sequence, skip, topic, and cooperative discussion learning |
| Learning characteristics | Length of study, study resources, number of tests, number of studies, learning path and resources |
| Historical learning record | Constituent elements |

**Table 1**: Required learner characteristics.

In this learning system, the test model is a cyclic and bidirectional system. The cyclic characteristic lies in the fact that learners are tested at different difficulty levels before they start learning, and the initial test difficulty is chosen as a self-assessment by the learners, and the test model records the corresponding results according to the correctness and completion and automatically enters the learning page suitable for the learners' current cognitive level; the initial test is a trip evaluation; after entering a stage of learning, learners will be tested again to update their, learners will take another test to update their current cognitive level to enter a higher stage of learning. In the continuous learning-testing cycle, the test model can continuously record the cognitive changes of learners in stages, and this test is a summative evaluation. The two-way feature is that the test data stored in the test model and the learning results judged are the basis for the judgment of the adaptation model and the data source for the learning content recommendation, while the adaptation model continuously updates the learners' learning paths and makes learning resource recommendations that meet the learners' knowledge level at this stage based on the feedback from the test model. The learning path in the adaptation model marks the learning process of the learners, and this learning process dynamically grasps the changes of the learners' knowledge level, and the test model must match the corresponding test according to the feedback of such changes.

The purpose of accurate phonological discrimination listening instruction is to enable learners to correctly identify different speech sounds, phonological combinations, intonation, etc. by listening to them. This emphasizes the importance of giving a lot of phonological stimuli. Only when the stimuli reach a level where the learner can respond can the learner output successfully. During the implementation of the activity, the teacher needs to provide sufficient phonological stimuli for the learners to correctly perceive each speech sound and to facilitate the learners' speech recognition. This listening activity focuses on the perception of listening content, and the resources are designed to be short and unconnected, with contextual clues removed or kept to a minimum to induce learners to rely on their ears for listening. BS technology facilitates the design and development of resources. Teachers can design arbitrary and adequate listening resources according to learners' and instructional needs, without having to record themselves or find others to do so, or sift through a vast array of listening resources to provide learners with sufficient speech stimuli and create a good environment for speech input.

### 3.2 English Listening Experiment Design

This learning system is learner-centered, with self-directed learning as the key function and teacher background management as an aid to learner learning. The system must be logged in before use and is divided into two roles, student, and teacher. Different roles of users have been entered in the background database, so users can log in directly. After logging in, students enter the self-test question area and can enter the learning page only after the test is completed; after logging in, teachers enter the teacher management page directly to manage some learning resources and questions and can view the overall learning situation of students in this page. The system development implementation of this study will be detailed from the management pages of both roles. Before the first study, the learner enters the test module and makes a level estimation for himself according to his current learning situation. After the learner selects a difficulty level, he/she starts to enter the corresponding difficulty level to answer the test, and the system reviews the results and produces the results. If the learner chooses the intermediate difficulty, with a score out of 100, he will go to different pages depending on the different scores obtained from the test results. In the first case, if he gets only 30 points, he cannot enter the learning page, and the system will automatically enter the test questions of primary difficulty, and he can enter the learning page only if he reaches 60 points or more in the primary test questions; if he gets more than 60 points, he will enter the learning page of the intermediate learning content. After the learner logs in as a student and enters the learning system for the first time, he/she needs to take an initial adaptive test, which is to check the learner's current knowledge proficiency. When the learner selects the difficulty level of the test, it is a pre-judgment of their current learning proficiency. The test difficulty level contains three levels: easy, moderate, and hard, and the learner selects a difficulty level to enter. The test questions mainly contain objective questions (mainly multiple-choice questions) and subjective questions (listening content to answer questions), where the rules of the test questions are set: the system randomly draws a combination of questions from each difficulty level, and stipulates that the same question will not be drawn again if it is done correctly five times. After the test is completed, the system matches the learner's answers to the database and generates an analysis of the answers. The test results are the result of the learner's access to the learning pages of different difficulty levels and the results are stored in the database for the learner to compare and analyze later. The test results include test completion time, correct rate, number of correct answers, number of wrong answers, and result analysis.

The initial score of the test is rounded to zero and matched with the answers stored in the database after the questions are completed. Each set of test questions has a total score of 100 points, with 3 points for each multiple-choice question, 4 points for each listening question, and 10 points for each subjective question. The questions are randomly selected by the system, and the same question will not be drawn again if it is done correctly five times. If the score is higher than 60, the learner will enter the page of the corresponding difficulty level, and if the score is lower than 60, the learner will take another test at a lower level. At the end of the test, the database records the user's current test difficulty level according to the score, which is stored in the learner data table as the test difficulty factor.

After the learners complete the system's adaptive test, they enter the learning pages of different difficulty levels according to the test result levels. Although the learning content and learning resource types have different levels of difficulty, the framework of the learning page is the same for each learner, including three parts: knowledge learning, learning management, and learning recommendation. Among them, knowledge learning stores learning materials such as graphics, text, audio, and video; learning test in learning management is to test learners' stage learning results; learning recommendation is the recommendation of adaptive learning path, which is the core function of this learning system.

## 4 ANALYSIS OF RESULTS

### 4.1 System Performance Test Results Analysis

University students have relatively more time to study outside of class, and everyone is equipped with a smartphone. In terms of using the learning system, 140 students installed and used it smoothly. After using it for some time, according to the comprehensive situation of the data collected in the background and the interview survey, the frequency of using this learning system was divided into times and lengths: more than twelve times as often, five to twelve times was sometimes, one to four times was occasionally, and not once was never. The percentage of frequency of use is shown in Figure 3. 55% of the students use it frequently because mobile learning allows them to consolidate and supplement their knowledge anytime and anywhere, and it can meet their existing learning situation and can provide learning needs; 35% of the students use it sometimes, they have a general attitude toward the mobile learning system, and they think it has some effect on there and it is more convenient to use; 7.5% of the students 7.5% of the students seldom use the learning system, and it is difficult for them to say whether the system is of practical help to them; 2.5% of the students think that the system adds some burden to their learning, and they still prefer the traditional classroom teaching model.

Due to time constraints and some objective technical factors, the research results cannot objectively reflect students' learning effects from various indicators in a short period. It is concluded from the interviews with the learners that most of the learners are positive and supportive of this learning system, believing that this system is operable, has a simple interface, clear navigation, and relevant content, and most importantly, can be used for learning anytime and anywhere, where the learning module of extracurricular resources facilitates the direct provision of extracurricular listening resources and effectively enhances the learning of listening. The adaptive test and content recommendations allow them to grasp their current real knowledge level, and overall it is a practical help for them to learn English listening learning.

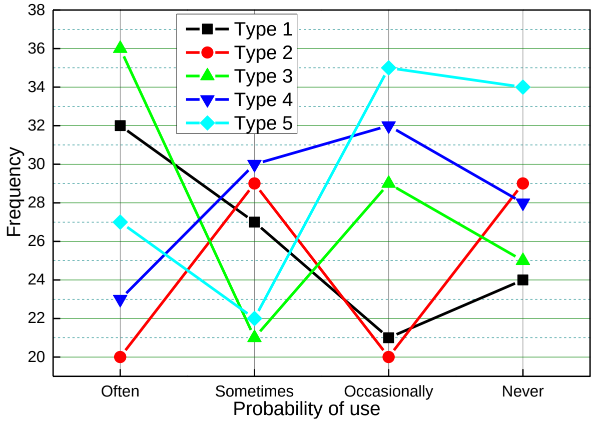


**Figure 2:** Percentage of the frequency of use of listening learning system.

The final oral test scores for the four experimental groups are shown in Figure 3.

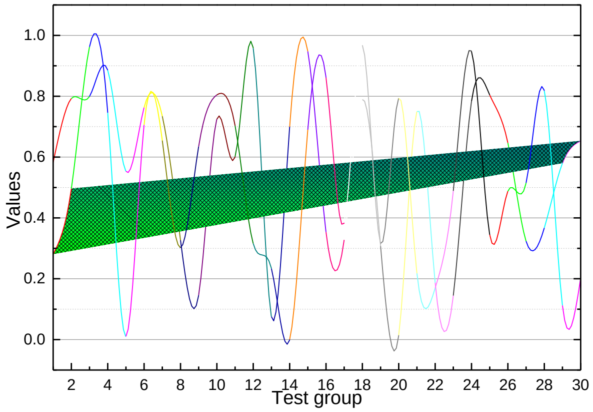


**Figure 3:** Descriptive statistics of the test scores of the experimental group.

The mean score of the oral test group of group 1 was 99.10 higher than the mean score of the oral test group of group 2, 290.30, and the mean score of the oral test group of group 3, 86.50, was higher than the mean score of the oral test group of group 4, 80.80. The web resource-assisted experimental group included group 1 and group 3, and the non-web resource-assisted experimental group included group 2 and group 4, i.e., the oral test score of the web resource-assisted experimental group was higher than the mean score of the non-web resource-assisted experimental group. The mean score of the web resource-assisted experimental group was higher than that of the non-web resource-assisted experimental group. The average score of the oral test group of group 1 is 99.10 higher than the average score of the oral test group of group 3 is 86.50, and the average score of the oral test group of group 2 is 90.30 higher than the average score of the oral test group of group 4 is 80.80. The mobile technology-assisted experimental group includes group 1 and group 2, and the non-mobile technology-assisted experimental group includes group 3 and group 4, i.e., the average score of the oral test group of the mobile technology-assisted experimental group is higher than the average score of the non-mobile technology-assisted experimental group. The mean score of the mobile technology-assisted experimental group was higher than the mean score of the non-mobile technology-assisted experimental group.

We used randomly selected students, formed classes, and conducted a total of 10 tests, with 60 users randomly selected for each test, after which we sequentially calculated the error between the user's choice of the exercise score set and the predicted value and used Pearson similarity to calculate the minimum root mean square error, and the results are shown in Figure 4.

From the above test results, we can see that the result is less than 0.1, which means that the recommendation algorithm is within 0.1 of the user's real rating error in the prediction of the recommended exercises, and the algorithm can recommend the exercise papers to the user more accurately according to the user's preference. According to the load test and stress test conducted by LoadRunner, the maximum throughput QPS of the online education and training school system can reach 358 under high concurrency, and the maximum response time is 1.2 s, which can fully satisfy the normal use of users.

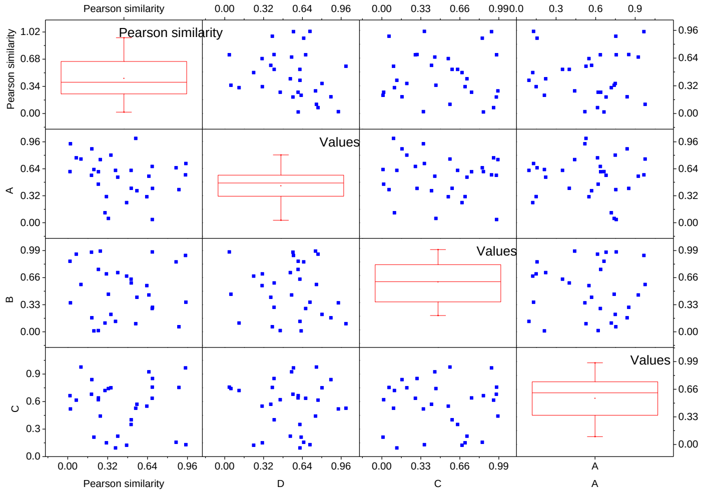


**Figure 4:** Minimum root mean square error results.

## 4.2 Analysis of Test Results

The analysis of the listening scores tested whether there was a significant difference between the listening scores of the two classes. The scores were selected as the final scores at the third grade. Independent samples t-tests were conducted by SPSS on the pre-test scores of the two classes, and the results were obtained, as shown in Figure 5.

As can be seen in the results of Figure 6, the value of the F statistic is 0.967, which is greater than 0.05, and it can be concluded that there is no significant difference in the variance of the two classes. Therefore, it is necessary to determine the value of the column assuming equal variance, because the two-sided value is 0.371, which is greater than 0.05, yielding the result of the independent samples T-test: there is no significant difference between the listening scores of the two classes, which can be judged as a basic agreement between the two classes in terms of the basis of English listening scores.

This study focuses on examining learners' listening discrimination, acquisition, and comprehension, and listening communication skills. The learners are currently in the third-grade level. Therefore, the author and the instructor designed and tested the test according to the requirements of the compulsory education curriculum standards and the focus of listening instruction in the second semester of Grade 3. The SPSS independent samples t-test was conducted on the post-test scores of the two classes, and the results were obtained as shown in Figure 6.

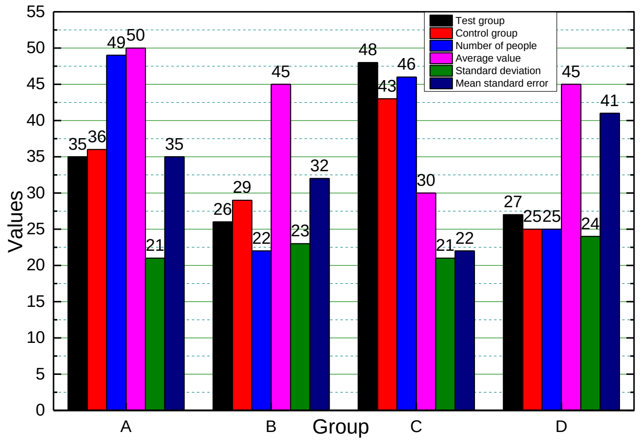


**Figure 5:** Listening score pre-test.

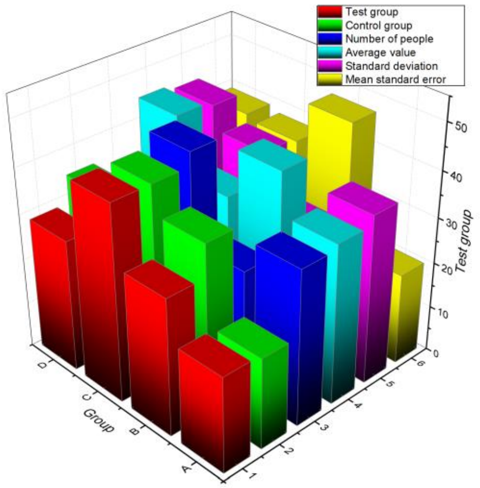


**Figure 6:** Listening post-test scores.

As can be seen in Figure 7, the sig value of the F statistic is 0.582, which is greater than 0.05, and it can be concluded that there is no significant difference in the variance of the two classes. Therefore, it is necessary to determine the value of the column assuming equal variance, since the two-sided value is 0.046, which is less than 0.05, yielding the result of the independent samples t-test: there is a significant difference between the listening scores of the two classes. In the group statistics, the mean value of the experimental group was 36.62 and the mean value of the control group was 34.67, the experimental group was higher than the control group, therefore, it can be determined that the listening teaching activities can improve the learners' academic performance.

The system also has a statistical analysis module, and the platform provides a large amount of data that can be analyzed and adjusted by the lecturer so that the data can be fully, comprehensively, and effectively used. Since the plug-in of highchairs is used in the foreground of the teaching statistics analysis module, the data display interface effect is beautiful and friendly, for example, Figure 7 shows the statistics of all the students' qualifications of the whole school or this teacher and the statistics of the number of online learners of this course in each month (day).

The satisfaction survey on the implementation of the activities shows that the learners are highly satisfied with the fine-processing listening teaching activities and the situational communicative listening teaching activities, and are motivated and like to participate in the listening teaching activities, which improve the learners' interest in learning and self-confidence. Although the fine-processing listening instructional activities improved learners' listening discrimination ability, the improvement of learners' motivation and interest was still lacking. However, in general, technology-based listening activities in teaching can promote teachers' teaching and learners' listening skills and abilities.

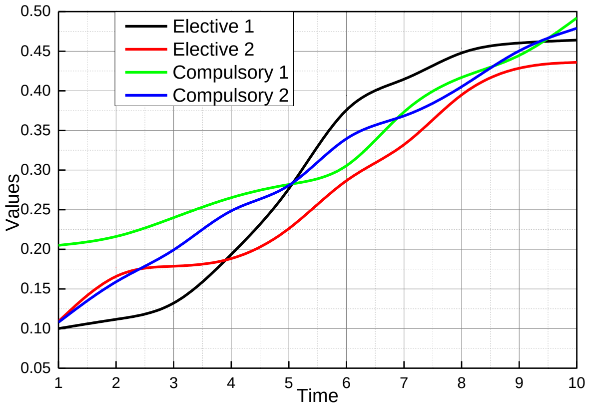


**Figure 7:** Statistics on the number of learners per month.

## 5 CONCLUSION

In this paper, we design and implement an interactive teaching platform based on BS architecture, BS architecture, SSH framework, and other technologies. From the problems in listening teaching and the requirements of English listening teaching by education curriculum standards, it is found that using the functions of TTS technology can effectively improve and optimize English listening teaching. This improvement and optimization are mainly reflected in three aspects: BS technology can provide learners with a standard speech environment; BS technology can provide learners with rich listening teaching materials; BS technology can support the implementation of multiple teaching modes. In turn, it can meet the teaching needs of teachers and the learning needs of learners, and promote the realization of listening teaching objectives. Based on the needs of each activity, the listening activities were designed with the support of BS technology and theoretical support from behaviorist learning theory, information processing theory, and constructivist theory, respectively. Different types of activities require different aspects in the design and implementation process, for example, the precision-variant listening activities need to provide learners with a large amount of standard speech input to stimulate learners' speech perception and discrimination ability; the fine-processing listening activities need to meet the needs of learners in the information processing process, providing learners with strategies for processing and speech materials at different speech speeds; the contextual communicative listening activities are designed to provide learners with sufficient learning scaffolds to help learners learn independently and cooperatively, and to accomplish the goal of communicative situations.

*Jiya Ding*, https://doi.org/0000-0002-1094-5986

**ACKNOWLEDGEMENTS**

The study was supported by Construction of First-class Disciplines in Ningxia Colleges and Universities (Pedagogy) (Grant No. NXYLXK2017B11).

**REFERENCES**

[1] Agarwal, C.; Chakraborty, P.: A review of tools and techniques for computer aided pronunciation training (CAPT) in English, Education and Information Technologies, 24(6), 2019, 3731-3743. https://doi.org/10.1007/s10639-019-09955-7

[2] Shadiev, R.; Wu, T.-T.; Sun, A.; Huang, Y.-M.: Applications of speech-to-text recognition and computer-aided translation for facilitating cross-cultural learning through a learning activity: issues and their solutions, Educational Technology Research and Development, 66(1), 2018, 191-214. https://doi.org/10.1007/s11423-017-9556-8

[3] Mao, L.: Application of Browser/Server Architecture in College English Online Learning System Design, International Journal of Emerging Technologies in Learning (iJET), 13(3), 2018, 129-140. https://www.learntechlib.org/p/182433/

[4] Watanabe, S.; Hori, T.; Kim, S.; Hershey, J.-R.; Hayashi, T.: Hybrid CTC/attention architecture for end-to-end speech recognition, IEEE Journal of Selected Topics in Signal Processing, 11(8), 2017, 1240-1253. https://doi.org/10.1109/JSTSP.2017.2763455

[5] Zhang, L.; Zhao, Z.; Ma, C.; Shan, L.; Sun, H.; Jiang, L.; Gao, C.: End-to-end automatic pronunciation error detection based on improved hybrid ctc/attention architecture, Sensors, 20(7), 2020, 1809. https://doi.org/10.3390/s20071809

[6] Das, A.; Li, J.; Ye, G.; Zhao, R.; Gong, Y.: Advancing acoustic-to-word CTC model with attention and mixed-units, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(12), 2019, 1880-1892. https://doi.org/10.1109/TASLP.2019.2933325

[7] Qin, C.-X.; Zhang, W.-L.; Qu, D.: A new joint CTC-attention-based speech recognition model with multi-level multi-head attention, EURASIP Journal on Audio, Speech, and Music Processing, 2019(1), 2019, 1-12. https://doi.org/10.1186/s13636-019-0161-0

[8] Kurokawa, T.; Kai, A.; Kondo, H.: Effects of End-to-end ASR and Score Fusion Model Learning for Improved Query-by-example Spoken Term Detection, In 2020 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC),2020, 654-661. IEEE. https://ieeexplore.ieee.org/abstract/document/9306333

[9] Sun, S.; Guo, P.; Xie, L.; Hwang, M. Y.: Adversarial regularization for attention based end-to-end robust speech recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 27(11), 2019, 1826-1838. https://doi.org/10.1109/TASLP.2019.2933146

[10] Li, R.; Wang, X.; Mallidi, S.-H.; Watanabe, S.; Hori, T.; Hermansky, H.: Multi-stream end-to-end speech recognition, IEEE/ACM Transactions on Audio, Speech, and Language Processing, 28, 2019, 646-655. https://doi.org/10.1109/TASLP.2019.2959721