

Retraction

Retracted: The Modular Design of an English Pronunciation Level Evaluation System Based on Machine Learning

Security and Communication Networks

Received 5 December 2023; Accepted 5 December 2023; Published 6 December 2023

Copyright © 2023 Security and Communication Networks. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] X. Han and T. Huwan, "The Modular Design of an English Pronunciation Level Evaluation System Based on Machine Learning," *Security and Communication Networks*, vol. 2022, Article ID 6804131, 11 pages, 2022.

Research Article

The Modular Design of an English Pronunciation Level Evaluation System Based on Machine Learning

Xue Han¹ and Trip Huwan² 

¹Henan Institute of Economics and Trade, Zhengzhou, Henan 450007, China

²Research Center of Environmental Science and Engineering, Bishkek, Kyrgyzstan

Correspondence should be addressed to Trip Huwan; dr.tribhuwan@mail.cu.edu.kg

Received 30 March 2022; Revised 14 May 2022; Accepted 20 May 2022; Published 7 June 2022

Academic Editor: Mohammad Ayoub Khan

Copyright © 2022 Xue Han and Trip Huwan. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

To improve the accuracy of English pronunciation level evaluation, we study the modularization of the English pronunciation level evaluation system unfolding based on machine learning. The S3C2440A chip is used as the main processor of the system, and the spoken English recordings are sent to the evaluation module through the speech upload module. In the evaluation module, the pronunciation signal is filtered by the multilayer wavelet feature scale transformation method, and the intonation, speed, pitch, rhythm, and emotion features are decomposed and extracted. The test results show that the misjudgment rate of different mispronunciations is less than 1% when the system is used to evaluate the English pronunciation level, which proves that it has high evaluation accuracy. In-depth study of speech recognition related theories, speech scoring, and pronunciation correction algorithms are discussed, and an assisted learning system based on AP scoring method and pronunciation resonance peak comparison technology is proposed for the problem of inaccurate pronunciation scoring and lack of effective feedback of speech recognition technology applied to oral learning. The English pronunciation training system has achieved the expected pronunciation following of English phonetic symbols and words, real-time pronunciation. The English pronunciation training system has achieved all basic functions such as pronunciation following, real-time pronunciation evaluation, and pronunciation correction of English phonemes and works as expected. After testing, the system has achieved high accuracy in pronunciation scoring, and the similarity with experts' scoring is over 90% for vowel and word pronunciation; the efficiency of pronunciation correction reaches 80%, which can improve learners' pronunciation level to a certain extent.

1. Introduction

Computers have been widely used in language evaluation for language learning as well as speech recognition, and speech recognition technology is an important manifestation in determining the level of language learning. A large amount of language signal data in the language learning process, the complexity of speech variation, and the high dimensionality of speech feature parameters lead to more difficult recognition of speech features [1]. The computational volume of speech evaluation and recognition is too large, which requires higher hardware resources as well as software resources to realize the high-speed processing of massive speech signals [2]. Machine learning algorithms are efficient algorithms that are widely used in many fields, and the

methods such as BP neural networks and support vector machines in machine learning algorithms have high data processing speed as well as accuracy [3]. Build a learning environment that is friendly to students and meets actual teaching needs within the platform. Under the guidance of the above theories and principles, in the speech diagnosis module, the feature extraction and recognition functions of speech are realized on the MATLAB platform. The current English pronunciation level evaluation system mainly has the following defects: most computer-aided language learning systems pay too much attention to grammar as well as word learning, ignore oral learning pronunciation evaluation, and have too few oral evaluation indexes; spoken English has a high subjective consciousness and has low stability with slow speed as the English oral test evaluation

index [4]. To address the above evaluation system defects, English speech is selected as the system research object, and the traditional English pronunciation level evaluation system is implemented to improve the system by using intonation, rhythm, pitch, and speed problems as the system evaluation indexes and using support vector machine method to achieve multi-index evaluation. An efficient feature extraction method is used to extract English spoken signal features, which makes the evaluation results more accurate and objective.

The studied system has a high application value to avoid the shortcomings of the assessment quality degradation due to the influence of vocal channel perturbations and distortions in the process of English-speaking level evaluation. The effect of multivariate feedback on English composition revision is based on an automatic evaluation system [5]. The multivariate feedback method is used to improve the English evaluation performance, but the system has high computational complexity and poor real-time performance. The English ubiquitous learning ecosystem based on big data analytics makes full use of the high computing level of big data analytics technology, but the system is less reliable. To study the modular design of the English pronunciation level evaluation system based on machine learning, the S3C2440A is selected as the core control chip, and the support vector machine method is used to achieve the final evaluation of the English-speaking level [6]. Through the gap between them, the quality of the pronunciation used can be judged, so as to better provide the user with pronunciation guidance, which is also the consideration of the platform design.

In recent years, with the maturity of speech recognition technology, research on computer-aided pronunciation learning based on speech recognition technology has been carried out at home and abroad. Speech recognition technology can make the pronunciation learning software have the function of pronunciation feedback to help learners correct the wrong pronunciation in time so that learners can avoid forming wrong pronunciation habits due to multiple repetitions. Although the current research is diverse and there is no unified design scheme, and there is a great distance from the large-scale popular application, the research in this field has shown great practical value as an innovative research topic. Therefore, this research on computer-aided pronunciation learning using speech recognition technology has great theoretical research value and application prospects. With the rapid development of mobile Internet technology today, mobile communication extends the network to everyone's pocket and is truly ubiquitous, providing a stage for the application of smartphones. Smartphones not only have traditional cell phone functions such as voice calls but also have more powerful data processing capabilities, more brilliant graphical user interfaces, and more powerful software loading capabilities. The most important point is that the price of smartphones and ordinary cell phones is compared to little difference, so the popularity of smartphones is an inevitable trend. The software-based on smartphone platform has great convenience and practicality compared with the traditional computer software.

2. Current Status of Research

Computer-assisted language learning (CALL) refers to the use of computers for language teaching and learning in general. In recent years, with the continuous progress of computer technology, the application of computer-assisted language learning has become increasingly widespread [7]. Traditional language learning software has been gradually eliminated due to its single function and poor interaction with users, and intelligent language learning systems with multifunctional functions such as "listening," "speaking," and "reading" have been gradually eliminated. The FLUENCY pronunciation training system uses automatic speech recognition technology based on the SPHINX engine [8]. The system can point out syllable and rhyme errors, but only to the extent that it provides feedback on how to correct them [9]. The system scores the learner's pronunciation on a word-by-word phoneme basis and gives feedback on rhyme only in terms of pronunciation duration. The current state of research in this area has shown a great deal of practical value, but it has not yet moved into the large-scale application [10]. The current research results have used a variety of complex recognition methods, and there is no unified implementation scheme, and they are mainly based on large experimental systems on computer platforms, with few portable and simple terminal software programs suitable for most English learners. Therefore, the English pronunciation training system based on the Android platform in this topic has strong significance in both theoretical research and practical application [11].

The dynamic time regularization (DTW) algorithm was used for the first time in the analysis of speech signals to solve the problem of the unequal length of speech signals [12]. This greatly improved the efficiency and accuracy of speech recognition [13]. The techniques for extracting feature parameters gained significant progress, the linear predictive analysis started to be gradually applied, and vector quantization techniques opened a new chapter in speech recognition technology [10]. The execution of instructions with distributed storage and parallel collaborative processing has led to some key advances in model refinement, parameter extraction and optimization, and statistical adaptive techniques, which is another breakthrough in the development of speech recognition and is of milestone significance. Among them, the application of artificial neural networks is significant for the processing of speech signals [14]. Artificial neural networks are composed of individual neurons, and this structure makes the network system have the characteristics of distributed parallel processing, which makes the artificial neural network have a strong self-adaptive and self-learning ability [15]. However, due to the difficulty of training artificial neural networks, they were later gradually replaced by Gaussian mixture models. The English pronunciation level evaluation system includes a user management module, a voice upload module, an evaluation module, a data visualization module, and a database module. The rapid development of information technology based on computer technology, network technology, and modern communication technology has greatly

changed the traditional teaching methods and classroom capacity, providing a vivid and intuitive simulation teaching environment for language learning using multimedia technologies such as graphics, sound, and video [16]. Multimedia computer systems can provide language skills training for each student and provide rapid feedback to help students correct wrong language habits and expert correct language skills on time [17]. Their powerful graphic, audio, and visual processing functions and friendly interface interaction capabilities enable learners to integrate themselves into the learning situation and actively participate in independent learning.

The integration of information technology and English courses can create an authentic English learning environment for learners, and the application of multimedia technology effectively stimulates learners' interest in learning, thus forming a lasting motivation to learn. The application of new technologies enables them to learn independently on their terms, and this way of learning gives full play to the learners' subjective initiative. It can be said that English learning in the information technology environment can cultivate high-quality talents of the 21st century who meet the requirements of the information age, and such a learning style will receive increased attention from researchers [18]. However, in terms of research on the application of information technology to independent English learning, most researchers have focused their attention on higher education and adult learning, and most of them have concentrated on the application of information technology in English teaching or teaching single content such as vocabulary and grammar sentences, while few of them have involved the application of multimedia technology and speech matching technology in English phonetic learning. Therefore, the research on the application of English phonetic assisted learning platforms based on multimedia technology and speech recognition technology in English phonetic learning has some significance to promote independent learning of English in college.

3. Modular Design Analysis of English Pronunciation Level Evaluation System with Machine Learning

3.1. Machine Learning Algorithm Design for English Pronunciation Level Evaluation. The level of learners' pronunciation is important information that the intelligent oral pronunciation training system gives feedback to learners. Only accurate and reliable evaluation of learners' pronunciation can motivate learners to practice and improve their pronunciation level continuously. There are two ways to measure the pronunciation level: one is based on the standard speech reference template; the other is by using the training HMM reference corpus [19]. The method of training HMM reference corpus is more complicated, which requires a large amount of speech data in the training phase and repeated calculations to obtain the parameter model and is suitable for complex speech recognition systems such as continuous utterance recognition. In contrast, the method

based on standard speech reference templates is less computationally intensive, does not require additional training, and is more reliable for speech sounds such as monosyllables and small words, which is especially suitable for scoring systems with embedded peripherals or small speech learning machines. In this paper, we focus on scoring methods based on standard speech reference templates.

The spoken English pronunciation signal is filtered by a multilayer wavelet feature scale transformation method. The collected English spoken pronunciation signal is implemented as wavelet feature decomposition by modulated pulses [12] until the scale coefficients of the English spoken pronunciation signal decomposition. Therefore, the inspiration for the platform is that multiple forms of representation can be used for the same knowledge point, such as text, video, etc., so that users can choose the way they like to process information and then form a corresponding model after personal processing and finishing.

$$K_0^{(j)} = \alpha^j K, K_1^{(j)} \approx \alpha^{j+1} \beta K, \quad (1)$$

where K denotes the number of frames and $\alpha\beta$ are scale parameters. $K_0^{(j)}$ and βK , $K_1^{(j)}$ denote the length of the spoken English pronunciation and the length of the input signal of filter bank j , respectively, and the received English spoken signal is divided into frames. The signal component phase rotation technique is selected to implement linear coding using the pulse modulation variable GN to obtain the rotational moment of inertia of the output spoken English pronunciation signal as follows.

$$\text{angle}(gX^K) = (\text{angle}(gX^K) - \varphi_g) \bmod(4\pi). \quad (2)$$

The speech signal is generated by the airflow excitation of the vocal tract. The difference between the three types of speech sounds, clear, turbid, and burst, lies mainly in the way of articulatory excitation. Clear tones are generated by air turbulence (like noise) in the constricted area of the vocal tract, turbid tones are generated by quasiperiodic pulse signal excitation at the sound gate, and burst tones are formed by the sudden release of accumulated air pressure at a closed point of the vocal tract. As the most natural way of human-machine information interaction, the basic idea of speech recognition technology takes speech as the research object and converts the input speech signal into the corresponding text command through the process of machine recognition and understanding, to realize the control of speech to the machine. In the development of speech recognition technology, although different researchers have proposed many different solutions, the basic principles are similar. In terms of processing speech signals, any speech recognition system can use Figure 1 to represent its approximate recognition principle, and the most important modules of the speech recognition system are speech feature extraction and speech pattern matching.

The first step of speech recognition is speech signal preprocessing, which is the prerequisite and foundation of speech recognition and a very critical step for feature extraction of speech signals [20]. Unvoiced sounds are excited by air turbulence (like noise) in the constricted area of the

used in this project. The endpoint detection technique based on short-time energy and short-time zero rates is a time-domain analysis, which is relatively simple and computationally small, and ensures a certain degree of reliability. Figure 2 shows the endpoint detection diagram of a speech segment. The combination of the two thresholds, short-time energy and short-time zero rates, will determine the start and endpoints of the speech segment and prepare the signal for further processing later.

Mel frequency represents the perception of frequency by the general human ear. Studies have shown that the sound level perceived by the human ear does not correspond to the actual frequency level of the sound. The human auditory system is a special nonlinear system with different auditory sensitivities for signals of different frequencies. However, the basic principles are similar. In the processing of speech signals, any speech recognition system can express its general recognition principle. The most important modules of speech recognition systems are speech feature extraction and speech pattern matching. At low frequencies, the human ear is more sensitive, while at high frequencies, the human ear is increasingly coarse. Below 1000 Hz, the perceptual sensitivity and frequency are roughly linear, and above 1000 Hz, the relationship is approximately logarithmic. By transforming the frequency domain to the Mel frequency domain, the extracted parameters can better reflect the auditory characteristics of the human ear. The delta band-pass filter in parameter extraction serves two main purposes: one is to smooth out the speech spectrum and reduce the harmonic effects, thus highlighting the resonant peaks of the initial speech. Therefore, the pitch or volume of a speech segment does not affect the extracted MFCC parameters; that is, the speech recognition system with MFCC as the feature parameter does not change due to the change of pitch or volume of the input speech, which is very suitable for English spoken pronunciation recognition.

The matching similarity between the test and reference template feature vectors can be expressed by the distance between the vectors, where a larger distance indicates a smaller matching similarity. The distance between the feature vectors $T(n)$ and $R(m)$ is usually expressed by the Euclidean distance as follows:

$$d[T(n), R(m)] = \sum_{i=1}^p (t_i + p_i)^2. \quad (6)$$

The dynamic time regularization algorithm is to find the time regularization function that maps the time axis n of the test template nonlinearly to the time axis m of the reference template, such that the total matching distance D between the test and reference templates is minimized, i.e.,

$$D = \max \sum_{n=1}^M d[T(n), R(m)]. \quad (7)$$

When the system performs speech recognition, the minimum matching distance D is obtained by matching the test template with the reference template. This minimum matching distance is used as a measure of the similarity of

pronunciation between the reference template and the test template, which can provide a more reliable and comprehensive response to the similarity of linguistic features.

$$D(i_n, i_m) = d(T(i_n), R(i_m)). \quad (8)$$

The speech recognition part is a key component of the intelligent English spoken pronunciation training system. The basic theory and algorithm of speech recognition and the processing flow of speech recognition such as pre-processing, feature extraction, and pattern matching are mainly introduced. Based on the analysis of the commonly used speech recognition technologies, the study selects a recognition scheme suitable for this system. According to the characteristics of limited data processing capacity of the platform and the requirements of real-time and reliability of the system, the algorithms of feature extraction using MFCC and pattern matching using DTW dynamic time regularization are improved and selected, and the specific parameter details are given. The research in this chapter lays the foundation for the research in the subsequent chapters and the implementation of the system.

Since the scoring parameter generation module and the pronunciation scoring module are located on the same mobile device, the operational parameters of pronunciation scoring are generated based on expert scoring training before pronunciation learning is performed, the generated scoring parameters reflect the characteristics of the current system hardware platform, and the scoring scores have high similarity with the expert experience scoring. In the process of processing voice signals, including voice recognition or voice comparison, whether the endpoint detection is accurate is related to whether the subsequent recognition results of voice recognition are credible. Correctly identifying the starting point and ending point of the voice signal can accurately obtain voice information with necessary conditions. Therefore, the AP-based method has strong adaptivity, high accuracy, and reliability and greatly improves the compatibility of the system.

When processing the speech signal, the system divides the whole speech signal into several frames, and each frame has a very short time (20–30 ms), which can be regarded as a short-time smooth signal. In each frame, it is determined that the mouth shape and tongue position are relatively immobile so that the resonance peaks of each frame reflect the position of the mouth shape and tongue position of the pronunciation. The resonance peaks vary from frame to frame, which reflects the change of mouth shape and tongue position over time. Therefore, the system uses the resonance peaks comparison graph to reflect the whole articulatory shape change on the platform.

3.2. Modular Design of the Evaluation System. Timely feedback and error correction can quickly clear many obstacles on the way of learning English phonetics for students and help them find and solve problems, which should be the starting point and fundamental principle of the platform design. By providing an objective, efficient, and targeted learning aid platform, learners can gradually improve the

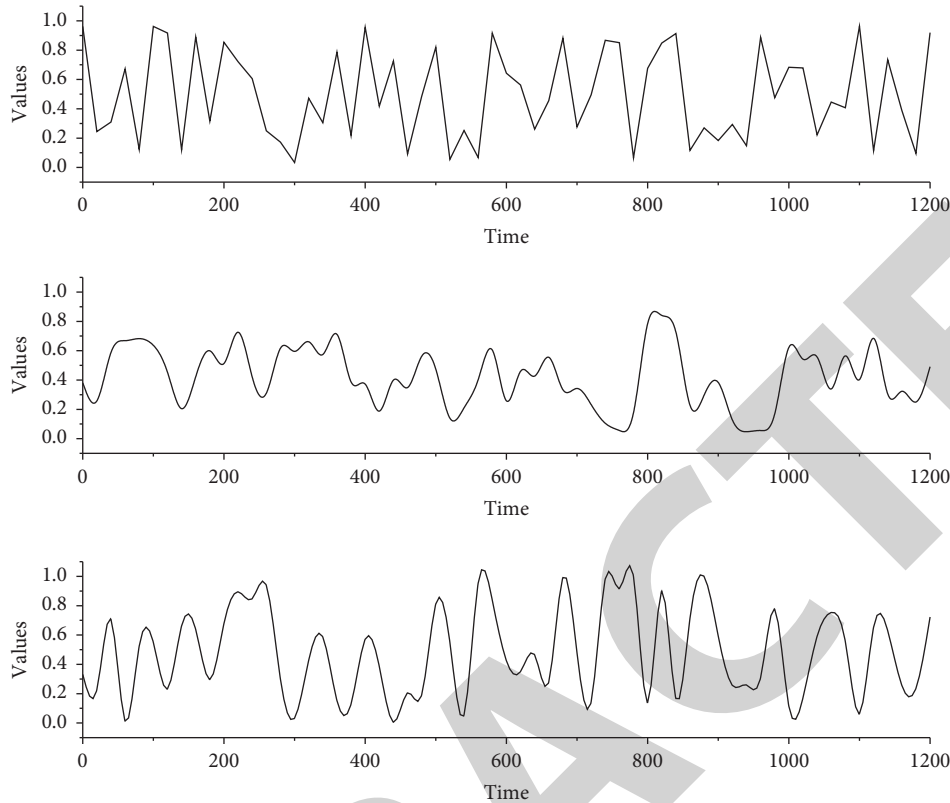


FIGURE 2: Endpoint detection of the speech signal.

accuracy of English pronunciation in a good English learning environment. Therefore, it is necessary to use behaviorism, constructivism, and other educational theories to theoretically guide the construction of the platform and to fully consider the characteristics of students as a group in the design and arrangement of teaching resources and the teaching of knowledge points, to build a learning environment that is friendly to students and meets the actual teaching needs within the platform. Under the guidance of the above theories and principles, in the speech diagnosis module, the feature extraction and recognition function of speech is implemented in the MATLAB platform and then packaged into a dynamic link library and published to the server through web server technology. The website of the platform and the App side of the platform perform the speech recognition operation through the open interface of web server and finally present the results according to the results returned from the server-side, while adding a series of multimedia methods such as graphics, video, and audio to build a supplementary learning platform. According to this idea, we design and implement the English speech-assisted learning platform.

The platform presents the knowledge of English phonetic symbols in various ways. At the same time, the platform is designed to establish a standard voice database, so that users can compare their recorded voices with the standard voices and judge the quality of pronunciation by the difference between them, to provide better pronunciation guidance to users. Using short-term energy, speech signal and noise signal can be easily distinguished in the case of high signal-to-noise ratio. However, in an environment with a low

signal-to-noise ratio, the short-term energy cannot clearly distinguish the speech signal from the noise signal. Meanwhile, on the website, the Asp.net language is mainly used for development, and the functional division is the same as the mobile side. Therefore, the framework of this paper is mainly based on the App side, supplemented by the website, with complementary advantages, to improve students' English phonetic pronunciation, then improve students' English-speaking skills, and finally achieve the purpose of promoting students' overall English proficiency. The core framework of the system is shown in Figure 3.

Based on the needs of English pronunciation level evaluation, the system is designed using C/B-S architecture. Learners can use cell phone clients and computers to browse through the web page; teachers and system administrators use the web page to realize data interaction. The ThinkPHP framework with MVC architecture is used to develop the web server of the system, and the web server is used to realize the system's speech signal preprocessing, English-speaking level evaluation, and evaluation-related data storage. The English pronunciation level evaluation system includes a user management module, speech upload module, evaluation module, data visualization module, and the database module [21]. At the same time, it can also guarantee a certain reliability. English learners use the speech upload module to realize spoken expressions and upload the completed expressions to the system. The evaluation module of the system preprocesses the English learners' spoken English recordings and extracts the intonation, speed, intonation, rhythm, and emotion features from the English spoken recordings after

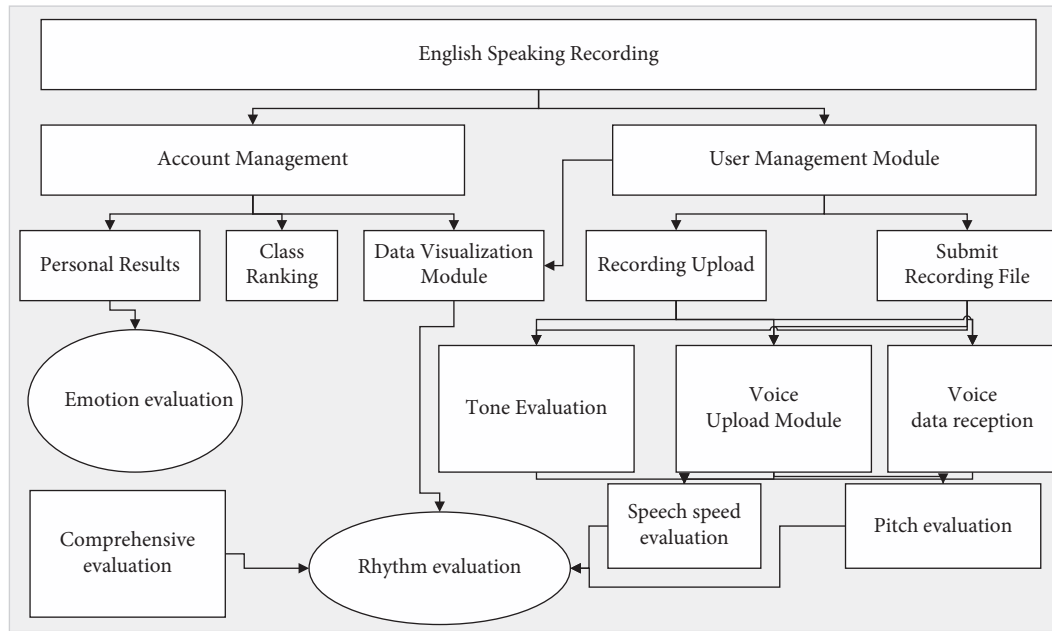


FIGURE 3: Overall system structure diagram.

the preprocessing is completed and then implements a comprehensive evaluation using the extracted features through the support vector machine in the machine learning algorithm. The evaluation results are displayed to the user using the data visualization module, and the evaluation results are saved in the system database for subsequent data analysis and recall.

The system's data visualization module presents the English pronunciation level evaluation results to the user in various forms such as pie charts, line graphs, radar charts, and bar graphs. The English-speaking level evaluation results include the evaluation results of each index so that learners can visualize the English pronunciation level evaluation results through the evaluation results. The evaluation results can help teachers to clarify the problems of different learners' English pronunciation and provide theoretical support for English-speaking teaching.

As a supplementary teaching software to facilitate learners' English phonetic learning, all knowledge transfer can only be done through virtual networks, which cannot provide the same teaching methods as real classroom situations and therefore can only be presented through multimedia teaching resources provided by the platform and cannot be ensured to apply to every student and must take into account students' cognitive development, in which multimedia learning cognitive theory can provide a viable solution for the platform in a theoretical way [22]. According to the two-way channel hypothesis of multimedia cognitivism learning theory, the two information processing channels, visual representation and auditory representation, can be used to convert knowledge by switching the way of representation, so the inspiration for the platform is that multiple forms of representation can be used for the same knowledge point, such as text, video, etc., so that users can choose their preferred way of processing information. And

after personal processing for each information processing channel, it is not unlimited to processing any amount of information, so the knowledge arrangement and the allocation of teaching resources in the platform should be simple and clear, and the information conveyed should be accurate and progressive at the same time; learners can cognitively process the learned knowledge according to their situation and then add it to their knowledge system. The platform should take this into account the overall design according to this feature and form a three-dimensional and multilevel knowledge structure, as shown in Figure 4.

There are obvious differences among college students in terms of the purpose and motivation of English learning. Since many colleges and universities directly link the evaluation and graduation with certificates, many students learn English only to get certificates for examinations, just to cope with the examinations, lacking deep-seated motivation, such as learning English for personal interests, and even if they do, they only account for a small part. Below 1000 Hz, there is a roughly linear relationship between perceptual sensitivity and frequency and an approximate logarithmic increase above 1000 Hz. By transforming the frequency domain into the Mel frequency domain, the extracted parameters can better reflect the auditory characteristics of the human ear. Of course, some students learn English to study abroad or to pave the way for future work, which is out of consideration for their future development, but this is also forced by the current situation, it is necessary to learn, and in general, it may become more disgusted with learning English.

The S3C2440A chip is the main control and speech processing chip of the English-speaking level evaluation system, and the software module of the system is realized by this chip, which can realize the LCD, user control function, and evaluation control of the system. The audio bus-based

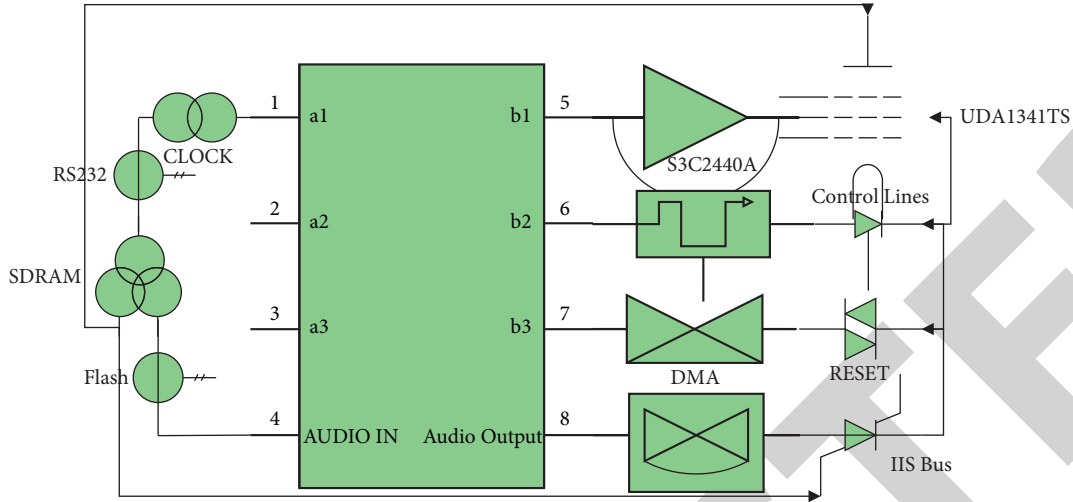


FIGURE 4: System hardware structure diagram.

UDA1341Ts audio CODEC from Philips and the S3C2440A chip are used as the main hardware parts of the system, and both the S3C2440A microprocessor and the UDA1341Ts codec have 115 audio coding and decoding interfaces, which can be directly connected to enhance the simplicity of the system. The system uses DMA as the information sending and receiving method to achieve real-time recording of spoken English as well as real-time playback. UDA1341Ts codecs can amplify, filter, and A/D convert English pronunciation speech signal; the digital signal is realized by S3C2440A microprocessor; speech signal is passed to FIFO buffer to DMA buffer by DMA controller, which provides convenience for microprocessor processing. SDRAM and Flash are used for system data and application operation and BIOS storage.

4. Analysis of Results

4.1. Algorithm Evaluation Results. To test the practical validity of the modular design of the English pronunciation level evaluation system based on machine learning, the following experiment was designed. The experimental environment is as follows: a student majoring in business English in a foreign language school is selected as the experimental object, and MATLAB software is used to program this system. The linear FM signal of 8 segments of the student's spoken English pronunciation is collected, and the time width and relative bandwidth of the collected speech samples were 1.5 s and 0.5 dB, respectively, and the collection frequency and baseband signal frequency of the spoken English pronunciation signal with different vocal folds were 1024 kHz and 3~9 kHz, respectively. To avoid overfitting problems of the experimental results, the range of the penalty parameter was set to 0.1~0.5. On this basis, the classification performance of the support vector machine used for different radial basis parameters was calculated and the statistical results are shown in Figure 5. Machine learning algorithms are high-efficiency algorithms that are widely used in many fields. Methods such as BP neural networks and support vector machines in machine learning

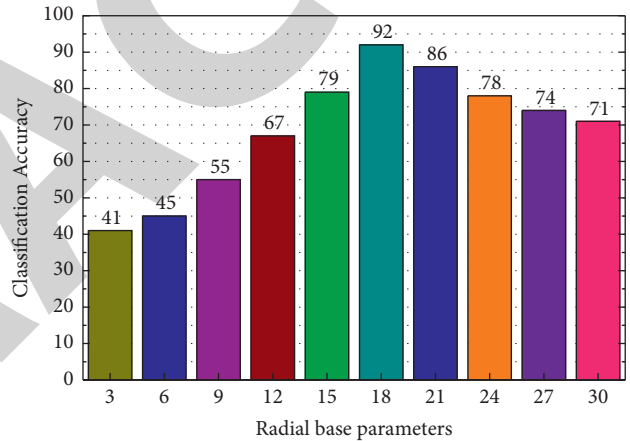


FIGURE 5: Support vector machine classification performance with different radial basis parameters.

algorithms have high data processing speed and accuracy. The analysis of Figure 5 shows that the classification performance of the support vector machine is optimal when the radial basis parameter is 18. Therefore, by setting the radial basis parameter to 18, the classification performance of the support vector machine used in the system with different penalty parameters is counted, and the statistical results are shown in Figure 5.

The results of the Pearson correlation coefficients of the three systems are shown in Figure 6. The results of the Pearson correlation coefficients of the three systems are shown in Figure 6.

The system can effectively evaluate five indicators of English pronunciation: intonation, speed, intonation, rhythm, and emotion, and use the evaluation results of each indicator to make the final evaluation of English pronunciation level. Spoken English has a high subjective consciousness, and the slow speed is used as the evaluation index of the oral English test, which has low stability. The system can evaluate learners' English pronunciation levels from different directions and has high evaluation effectiveness. To

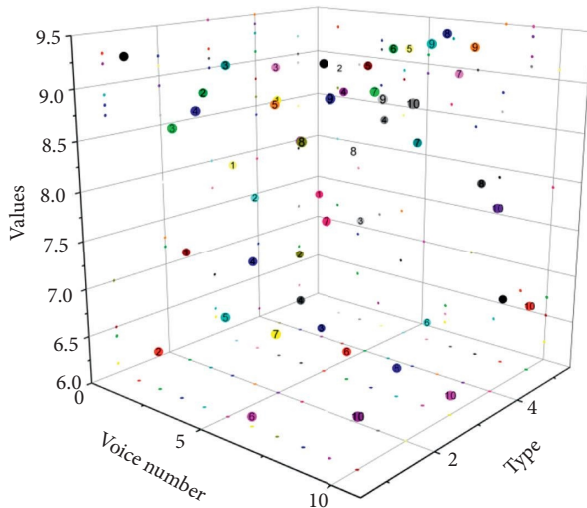


FIGURE 6: English pronunciation level evaluation results.

further verify the effectiveness of the modular design of the English pronunciation evaluation system based on machine learning, the exact agreement rate, adjacent agreement rate, and Pearson correlation coefficient are selected as important indicators for the performance evaluation. To demonstrate the evaluation performance of the system more intuitively in this paper, the multiple feedback system and the big data analysis system are selected as comparisons. The English pronunciation level evaluation experiment is to examine the degree of agreement between different system evaluations and manual evaluations for the same English speech samples. The test process is to judge the confidence level of manual evaluation and to evaluate the degree of agreement between system evaluation results and manual evaluation results when the manual evaluation results are reliable.

4.2. Evaluation of System Performance Results. At the same time, to get more feedback on the platform, so that the platform can continue to improve and help more learners, the platform participated in the competition as a project type and won the first prize in the competition after layers of selection. However, as a frontier research topic, research in this field has shown great practical value. Therefore, the research on computer-assisted pronunciation learning using speech recognition technology in this subject has great theoretical research value and application prospect. This is a great recognition of the platform, and at the same time for the series of suggestions people made to the platform in the competition, the platform will continuously improve in the subsequent design and development of the platform. For the actual teaching effect of the platform, the author also conducted a corresponding teaching experiment. Twenty university students were selected for the experiment, and the experiment lasted for 2 weeks, during which each person learned 30 minutes through the platform every day. After the experiment, the data were collected by issuing questionnaires and interviews, among which the designed questionnaire had twelve questions, which were mainly divided into four dimensions, and the results are shown in Figure 7.

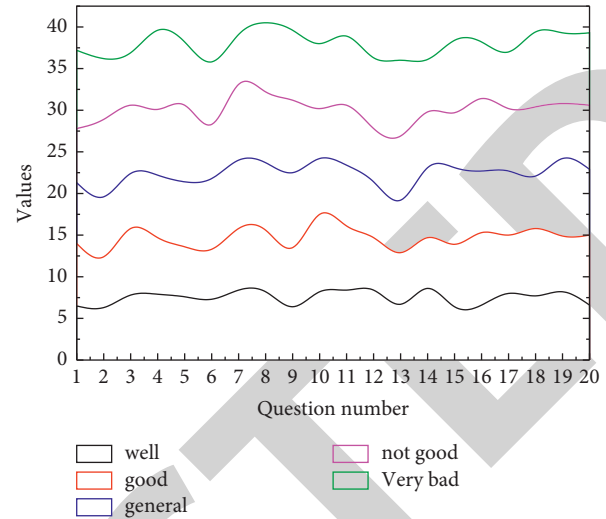


FIGURE 7: Statistical results.

From the statistical results above, the testers all considered their pronunciation level to be on the lower side of average; 80% of the students thought that the platform was helpful to their pronunciation learning, and 85% of them could improve their pronunciation level after using the platform for learning, while 80% of the test subjects indicated that the platform could help them well in their English-speaking learning. In particular, the phonetic diagnostic module gives immediate feedback through the system recording and comparing with the sample voice library and corrects the correct audio after the error, which is a very interactive way to help learners learn English pronunciation. The influence of multiple feedback is based on automatic evaluation system on English composition revision. The multifeedback method is used to improve the English evaluation performance, but the system has high computational complexity and poor real-time performance. In addition, the platform's interface, operation, and functions are satisfactory. Finally, most of the suggestions for the platform are to increase the learning resources of the platform and to increase the error book, which is also the place to continue to improve the platform in the future.

In addition to the basic test pages, the general testing principles and methods are more critical to link testing. This is also the biggest difference between software testing and web testing. The most obvious and intuitive impression of a website comes from the links, and to significantly improve user satisfaction with the website, it is necessary to avoid possible link errors and remove more invalid links from the website. Since links and web pages are the main content that makes up a web page, and links are a method used to guide customers and switch pages, they need a lot of attention. The content shown in Figure 8 is a graph of the results of the efficiency and load situation of the test system made using load impact software. The results of this test provide a realistic and reliable overall assessment of the system performance. The principle of the software used is to simulate business transactions and user logins, gradually increasing

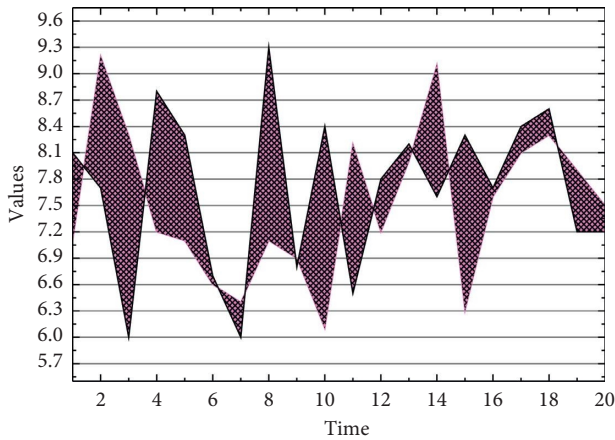


FIGURE 8: Load testing with the load impact testing tool.

the number of threads of operations and keeping track of the results.

Through the above performance test data, the performance indexes of the intelligent speech teaching aids system constructed in this paper are good, with low response time, and can meet the requirements of the system performance in the case of medium-scale user use. The final evaluation of spoken English proficiency is realized by using the support vector machine method. It is verified by experiments that the system has high evaluation performance, can provide English learners with accurate and objective evaluation results, and can quickly help learners to identify defects in English pronunciation.

5. Conclusion

The system implementation in this paper applies the idea of modularity and describes the process of implementing the functional modules of the intelligent speech teaching aids system. The system testing was carried out in a way oriented to the use of English intelligent speech teaching aids, mainly on the interface, interaction, business logic, and teaching workflow of the system, etc. The system was tested to run stably and meet the expected design goals. In this study, the S3C2440A chip is selected as the system microprocessor, the multilayer wavelet feature scale change method is selected to implement noise reduction processing for English spoken pronunciation signals, the wavelet full feature quantity is extracted based on the feature decomposition results of the modulated pulse, and the support vector machine in the machine learning algorithm is used to achieve English spoken pronunciation level evaluation. One is the method based on the standard speech reference template; the other is the method using the training HMM reference corpus. The method of training the HMM reference corpus is relatively complicated. It needs to provide a large amount of speech data in the training phase, and the parameter model can be obtained through repeated calculation, which is suitable for complex speech recognition systems such as continuous sentence recognition. The modularized English pronunciation level evaluation system is verified to have a high

evaluation performance through simulation tests, and the evaluation results are found to be highly accurate and reliable, with high application value. Although this study has achieved certain results, the performance of the English pronunciation level evaluation system in terms of timeliness and other applications still needs to be improved. Therefore, in the next research phase, the English pronunciation level evaluation system will be further optimized from the perspective of improving the evaluation timeliness.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

References

- [1] K. Lopez-de-Ipina, N. Barroso, P. M. Calvo et al., "Multilingual audio information management system based on semantic knowledge in complex environments," *Neural Computing & Applications*, vol. 32, no. 24, pp. 17869–17886, 2020.
- [2] H. Zhao, Z. Liu, X. Yao, and Q. Yang, "A machine learning-based sentiment analysis of online product reviews with a novel term weighting and feature selection approach," *Information Processing & Management*, vol. 58, no. 5, Article ID 102656, 2021.
- [3] Y. Mei, D.-p. Ye, S.-z. Jiang, and J.-r. Liu, "A particular character speech synthesis system based on deep learning," *IETE Technical Review*, vol. 38, no. 1, pp. 184–194, 2021.
- [4] C. Agarwal and P. Chakraborty, "A review of tools and techniques for computer aided pronunciation training (CAPT) in English," *Education and Information Technologies*, vol. 24, no. 6, pp. 3731–3743, 2019.
- [5] Y. Yang and Y. Yue, "English speech sound improvement system based on deep learning from signal processing to semantic recognition," *International Journal of Speech Technology*, vol. 23, no. 3, pp. 505–515, 2020.
- [6] R. Satapathy, E. Cambria, A. Nanetti, and A. Hussain, "A review of shorthand systems: From brachygraphy to microtext and beyond," *Cognitive Computation*, vol. 12, no. 4, pp. 778–792, 2020.
- [7] A. Hair, K. J. Ballard, C. Markoulli et al., "A longitudinal evaluation of tablet-based child speech therapy with Apraxia World," *ACM Transactions on Accessible Computing*, vol. 14, no. 1, pp. 1–26, 2021.
- [8] M. Arora and V. Kansal, "Character level embedding with deep convolutional neural network for text normalization of unstructured data for Twitter sentiment analysis," *Social Network Analysis and Mining*, vol. 9, no. 1, pp. 1–14, 2019.
- [9] Y. Wang and P. Zhao, "A probe into spoken English recognition in English education based on computer-aided comprehensive analysis," *International Journal of Emerging Technologies in Learning (IJET)*, vol. 15, no. 3, pp. 223–233, 2020.
- [10] A. Holzinger, M. Plass, M. Kickmeier-Rust et al., "Interactive machine learning: Experimental evidence for the human in the algorithmic loop," *Applied Intelligence*, vol. 49, no. 7, pp. 2401–2414, 2019.

- [11] S. Wu, J. Li, T. Wang, and Z. Chen, "Quality evaluation system of online courses based on review data," *International Journal of Design, Analysis and Tools for Integrated Circuits and Systems*, vol. 8, no. 1, pp. 19–23, 2019.
- [12] S. Dong, "Intelligent English teaching prediction system based on SVM and heterogeneous multimodal target recognition," *Journal of Intelligent and Fuzzy Systems*, vol. 38, no. 6, pp. 7145–7154, 2020.
- [13] E. Y. Oh and D. Song, "Developmental research on an interactive application for language speaking practice using speech recognition technology," *Educational Technology Research & Development*, vol. 69, no. 2, pp. 861–884, 2021.
- [14] M. Shahin, U. Zafar, and B. Ahmed, "The automatic detection of speech disorders in children: Challenges, opportunities, and preliminary results," *IEEE Journal of Selected Topics in Signal Processing*, vol. 14, no. 2, pp. 400–412, 2020.
- [15] P. Li, M. Wang, and J. Wang, "Named entity translation method based on machine translation lexicon," *Neural Computing & Applications*, vol. 33, no. 9, pp. 3977–3985, 2021.
- [16] A. Vitui and T. H. P. Chen, "MLASP: Machine learning assisted capacity planning," *Empirical Software Engineering*, vol. 26, no. 5, pp. 1–27, 2021.
- [17] T. Jaiswal, "Image captioning through cognitive IOT and machine-learning approaches," *Turkish Journal of Computer and Mathematics Education (TURCOMAT)*, vol. 12, no. 9, pp. 333–351, 2021.
- [18] A. S. Alrumayh, S. M. Lehman, and C. C. Tan, "Emerging mobile apps: Challenges and open problems," *CCF Transactions on Pervasive Computing and Interaction*, vol. 3, no. 1, pp. 57–75, 2021.
- [19] B. Syiem and L. J. Singh, "Exploring end-to-end framework towards Khasi speech recognition system," *International Journal of Speech Technology*, vol. 24, no. 2, pp. 419–424, 2021.
- [20] W. C. Latorza and C. M. Ambayon, "Explicit phonetic instruction and pronunciation skills of grade 10 high school students," *Journal of Critical Studies in Language and Literature*, vol. 1, no. 4, pp. 47–54, 2020.
- [21] U. Tariq and A. Aldaej, "Advancing an in-memory computing for a multi-accent real-time voice frequency recognition modeling: A comprehensive study of models & mechanism," *Multimedia Tools and Applications*, vol. 79, no. 37, pp. 27705–27720, 2020.
- [22] L. Hu, "Visual analysis of English teaching model based on scientific programming," *Mobile Information Systems*, vol. 2022, Article ID 8592802, 12 pages, 2022.