

Research Article

Self-Study System Assessment of Spoken English considering the Speech Scientific Computing Knowledge Assessment Algorithm

Hui Tao¹ and Zhe Zhang ²

¹Inner Mongolia Vocational and Technical College of Communications, Inner Mongolia, China

²Henan Industry and Trade Vocational College, Henan, China

Correspondence should be addressed to Zhe Zhang; zhangzhe@hngm.edu.cn

Received 1 March 2022; Accepted 25 April 2022; Published 26 May 2022

Academic Editor: Chia-Huei Wu

Copyright © 2022 Hui Tao and Zhe Zhang. 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.

For the purpose of effectively improving the assessment accuracy of spoken English self-study quality of English learners, a speech scientific computing knowledge assessment algorithm is introduced into the assessment of the self-study system of spoken English. Through the combination of spoken English speech assessment and accuracy detection, a DSP-based self-study assessment system for spoken English is designed. This system is mainly divided into two parts: voice signal processing and hardware circuit design. The features of spoken English self-study speech are extracted based on the multilayer wavelet feature conversion method, and the English pronunciation is detected and analyzed by adaptive filtering based on the characteristic values obtained. The corresponding wavelet entropy features of spoken English self-study speech are automatically assessed and assessed for the self-study quality. Finally, a practical case is analyzed, and the results indicate that the system proposed in this paper has high accuracy and excellent stability in assessing the quality of self-study in spoken English. Hence, it is of practical value in self-study assessment of spoken English.

1. Introduction

Through the application of the modern signal processing and automatic recognition technology, the spoken English quality of self-study can be assessed effectively. At the same time, combined with signal detection and English phonetic characteristic advance methods, the quality in the self-study of spoken English can be effectively assessed and detected, so that the objectivity and accuracy in the assessment of self-study quality for spoken English can be effectively improved accordingly [1, 2]. Traditionally, the studies of self-study quality assessment methods for spoken English mainly focus on the detection of voice signals and the extraction of features by using intelligent signal processing methods. With the continuous progress of science and technology and the rapid economic growth, the performance of mobile terminals at home and abroad has been substantially improved, and various operators are also increasing the development of mobile phone built-in software constantly. Various speech recognition software has gradually

been applied extensively. However, the general voice assessment feature of mobile phones is a mobile phone function that provides users with interactions with other people. Self-study assessment of spoken English is a learning method that can help learners to tap their own potential more effectively and give full play to their potential. At the same time, it can also promote learners to show themselves fully and freely so as to learn English knowledge better. With the continuous development of big data, the whole process of self-study assessment of spoken English has gradually turned into a learning process of learners' personal hobbies, which can help teachers find out the self-expression capabilities of students and supervise English learners to engage in adaptive English learning more effectively. The research subject is changed from macroscopic groups to microscopic individuals. At the same time, it is also helpful to provide English educators with personalized teaching guidance that is highly needed based on the practical situation of the students [3, 4]. Moreover, self-study of spoken English is also an effective way to achieve effective self-study of spoken

English by individual learners. With the individual characteristics of learners as the basis, the methods that are most suitable for the self-study of spoken English by English individuals are explored. As learners are different individuals, it is necessary to develop assessment methods for the self-study of spoken English based on various learning situations.

In this paper, a self-study assessment system for spoken English speech is designed to collect the spoken information of English learners through the application of acoustic sensors. Subsequently, the use of a multilayer wavelet feature is combined to transform the short-haired English self-study speech for feature decomposition to acquire the wavelet entropy feature in the self-study of the spoken language, and the wavelet entropy feature is used as the basis for assessing the quality of English self-study. Finally, the results of the experimental test suggest that the system designed in this paper can be used to assess the quality of self-study of spoken English accurately.

2. Method for Assessing Scientific Computing Knowledge of Speech

Since the demand of English learners for the development of full, autonomous, and personalized learning continues to grow, teachers can help meet their personalized learning needs and develop reasonable and scientific learning plans for them. Hence, the self-study system assessment method of spoken English based on the speech scientific computing knowledge assessment algorithm can help teachers teach students in accordance with their aptitude and solve the problems encountered by learners in their individualized learning process effectively [5, 6]. Based on the respective situation of learners, the content and tasks of personalized English learning are planned, real-time feedback is provided, and personalized assessment is carried out in time. The English knowledge field mainly refers to the set of knowledge units of all experience summaries and theoretical methods in the field of English professional. However, in the computer field, it is possible to make a certain type of design for the knowledge contained in a specific field. Based on this method, self-study of spoken language and structured self-study of spoken English can be carried out through the existing computer storage, system organization, and management of multiple aspects such as easy operation and other characteristic knowledge groups. In this way, personalized learning resources can be provided for the self-study of spoken English. The knowledge model structure in the English learning field requires that the English knowledge system should have a sound structural relationship so that accurate choices can be made when resource paths are recommended for the self-study of spoken English.

3. Establishment of the Self-Study Model of Spoken English

Since English learners are the main participants and experiencers in the self-study of spoken English, they are also the main entities in the acquisition of learner resources. Therefore, the personalized design of English needs to comply

with the individual demands of English learners. For the purpose of better clarifying the attribution of learners in the self-study system of spoken English, it is necessary to establish the model based on the practical cases and learners. The self-study mode of spoken English mainly involves the self-study of spoken English to establish the user modules or acquire the real-time data through third-party agent software and develop the self-study plans for spoken English based on the data collected. Based on the LIP (Learner Information Package) standard of the IMS, a model for English learners is established in this paper from four perspectives as follows: the basic characteristics of oral English self-study, the learning characteristics, the knowledge level, and the English learning records:

LearnerModel = (BaseInformation, LearningStyle, CognitiveLevel, AccessRecords).

In the above expression, BaseInformation stands for the basic data information of English students, such as name, age, registered birthplace, education, and family background. The LearningStyle stands for the learning characteristics of English learners. The model is established based on the learning characteristics of Felder, which contains domain definitions and can mainly be divided into “intuitive-feeling,” “voice-visual,” “passive-active,” “globalization-serial type,” and so on. These types can be specially set in the self-study system of spoken English or complete the acquisition of data from the Felder learning style scale. The CognitiveLevel stands for the current level of knowledge acquired by English learners. It can mainly be divided into three levels: elementary, intermediate, and high. The primary feature of these three levels is that the system will carry out independent assessment with reference to the unit test scores of the participants [7]. The AccessRecords stands for the visit records of all English learners during their study period, which mainly contains basic data information such as the contact information of the visitors, visit time, and specific address of the visit.

4. HMM across the State

In the process of the self-study of spoken English, there are multiple causes for the errors of spoken English learners. Some learners of spoken English fail to master the learning rules of English knowledge points very well, which lead to mistakes in their application process. In addition, the tension during the assessment process, the lack of focus, and many other unfavorable factors can result in a number of errors made by learners of spoken English [8, 9]. Compared with the self-study content of spoken English based on the basic knowledge of English, this learning mode is mainly developed based on the error or misunderstanding model of spoken English learners, which can effectively identify the errors or misunderstandings of spoken English learners by self-study. Through the analysis of the root cause for the errors or misunderstandings, they can be dealt with in time based on the error correction method to consolidate English knowledge, complete the internalization, and improve the learning efficiency of spoken English learners quickly. In the self-study of spoken English, the action data

of spoken English learners are minded based on the learning records to identify errors or misunderstandings quickly in the process of self-study of spoken English. The self-study model of spoken English is mainly used to address the errors or misunderstandings of spoken English learners. The system will identify the corresponding correction methods in the error database so that the spoken English learners can find out the content of the error and identify the type of error quickly. In addition, the system will send timely feedback on the corrective measures to the learners of spoken English. The error database can mainly be divided into the enumeration type and the generation type. The system will effectively determine the possible errors that may occur in the learners of spoken English by analyzing the experience of the designer and the experts and find out the cause of the errors based on the enumeration method. For the generation type of errors, a reference will be provided for the learners of spoken English based on the inventory system, which can independently collect and analyze the errors made by the learners of spoken English in the learning process (Figure 1). In addition, we have found that it is quite rare that all three states of a phoneme take up only one frame in each frame of the system, which will result in the state of only one corresponding frame in the two sound scientific computing knowledge assessment systems. If this state is skipped on the token transmission path, the HMM structure can still accurately describe the phoneme in the speech scientific computing knowledge assessment system, and such an HMM structure is shown in Figure 2.

If the conversion path of the HMM structure is changed, the computation of the token transmission will become more complicated. In any HMM with n states, it is required that the transmission probability distribution of the state transition matrix should be

$$\left. \begin{array}{l} a_{ij} \neq 0, \quad i = j \text{ and } i = j - 1 \\ a_{ij} = 0, \quad \text{others} \end{array} \right\}. \quad (1)$$

Thus, there are $2n + 1$ transmission paths for each factor. However, there are 7 transmission paths in the HMM in the three states. After the cross state conversion, there will be $3n + 1$ transmission paths for each voice factor. In addition, with regard to the 3 states, there are also 10 transmission paths. Although the increase in the number of conversion paths will increase the time complexity of the computation, the increase in the proportion is restrictive. Hence, the scientific computing knowledge assessment of English speech still has the advantages in terms of frame asynchronous system operation. For the purpose of obtaining more valuable phoneme information, it can ensure that there is an observation value in each English phoneme. Figure 3 shows the HMM structure across $n - 1$ states.

5. Equivalent Frame Shift of the Speech Scientific Computing Knowledge Assessment System

The level of English knowledge not only represents the level that the learners of spoken English have mastered the basic

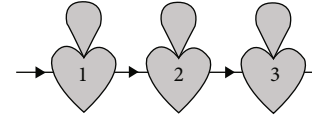


FIGURE 1: HMM state by state.

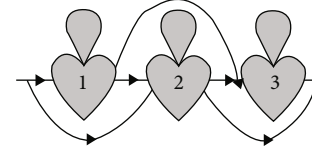


FIGURE 2: HMM across a single state.

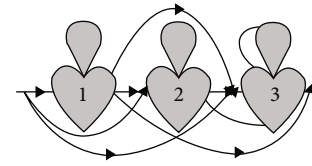


FIGURE 3: HMM across $n - 1$ states.

knowledge in the field at present but also is the most essential spoken self-study feature of the learners of spoken English and an intuitive manifestation of the self-study effect of spoken English [10]. Based on this learning mode, the cognitive level of the spoken English self-study knowledge points of each spoken English learner is analyzed according to the target classification of the learning content of spoken English learners, and the learning of English is adjusted in real time accordingly. The content difficulty factor and the learning sequence are used to establish a personalized learning method. At present, with the continuous deepening of the intelligence of self-study of spoken English, the capacity of the database in the system has gradually increased, and the difficulty of data processing has gradually increased. The system applies the main technology of the English and English learner model to measure the level of English knowledge mastered by oral learners accurately. For example, the MAEVIF and SoNITS systems adopt the semantic network technology to establish the knowledge level model of spoken English learners, respectively. It is proposed that the main design rules of the user model can be taken as the basis to establish a self-study system for the knowledge level and learning style of spoken English.

6. Noise Reduction Filter Processing of Spoken English Self-Study Speech Analysis

6.1. Speech Analysis in Self-Study of Spoken English. At present, with cloud computing, computer technology, big data analysis, and hypermedia technology as the basis for driving the continuous development of modern information technology, virtual teaching services and fast teaching services can be provided for English learners. Data and information technology services such as universally connected, intelligent, massive

data-scale data information mining and image-friendly learning interface can not only improve the environment for the self-study of spoken English effectively but also be a game changer to the previous self-study views of spoken English and the learning-style relationship between teachers and students and apply the practical English learning to the era of information development, serving as a primary factor that can contribute to the implementation of the “structural reform” of English education. The correct evaluation of the self-study system evaluation of spoken English, first of all, is necessary to establish an information sample mode that restricts the self-study system evaluation of spoken English. Combined with a speech scientific computing knowledge recognition algorithm, statistical evaluation of self-learning ability of spoken English is carried out. The evaluation constraint index parameters of the self-learning system of spoken English are a set of nonlinear time series. Constructing a high-dimensional feature distribution space represents the distribution model of spoken English evaluation parameters. Its main index parameters restrict the self-learning ability of spoken English, teacher level, investment in educational facilities, and policy relevance. Construct differential equations, and construct an information flow model that expresses the constraints of the self-learning ability of spoken English.

$$x_n = x(t_0 + n\Delta t) = h[z(t_0 + n\Delta t)] + \omega_n, \quad (2)$$

where $h(\cdot)$ is the multiple value function of self-study course assessment and evaluation of spoken English. ω_n is the evaluation error measurement function. In the high-dimensional feature distribution space, the solution vector evaluated in the self-study content of spoken English is calculated by the speech scientific computing knowledge recognition algorithm, and the feature training subset $S_i (i = 1, 2, \dots, L)$ of self-study evaluation of spoken English is obtained, which meets the following conditions.

$$\begin{aligned} \Sigma &= \text{diag}(\delta_1, \delta_2, \dots, \delta_r), \quad \delta_i = \sqrt{\lambda_i}, \forall i \neq j, \\ \bigcup_{i=1}^L S_i &= V - v_s. \end{aligned} \quad (3)$$

$x_{n+1} = \mu x_n (1 - x_n)$ indicates that the self-study evaluation index of spoken English adopts the conjugate solution of the statistical information model, which can reach the decomposition condition $U = \{u(t) | u(t) \in X, \|u\| \leq d, t \in I\}$ of the initial value, where $(I_i)_{i \in N} = \{x_1, x_2, \dots, x_m\}$. For multiple variable groups, the characteristic distribution sequence $x(n)$ corresponding to the self-study evaluation statistics of spoken English can be used to construct a self-study model of spoken English based on the measured value of the self-study level of spoken English in the previous period.

$$\begin{aligned} c_{1x}(\tau) &= E\{x(n)\} = 0, \\ c_{2x}(\tau) &= E\{x(n)x(n+\tau)\} = r(\tau), \\ c_{kx}(\tau_1, \tau_2, \dots, \tau_{k-1}) &\equiv 0, \quad k \geq 3, \end{aligned} \quad (4)$$

When $Q = 2$, the English classroom assessment teacher

strength level and educational resource distribution level meet the $(2 + 1)$ -dimensional continuous letter writing condition. In other words, self-study courses in spoken English should be assessed and evaluated.

$$\psi_x(\omega) = \ln \Phi_x(\omega) = -\frac{1}{2} \omega^2 \sigma^2. \quad (5)$$

Based on the exclusive analytical evaluation data information flow model for self-study of spoken English, it provides an accurate data input basis for the evaluation of self-study of spoken English and constructs a set of scalar sampling sequence components.

Using the speech scientific computing knowledge recognition algorithm to evaluate the spoken English self-learning system and the evaluation of the big data information model, the control objective function for constructing the prediction and estimation of the ability of the spoken English self-learning system is

$$\begin{aligned} \max_{x_{a,b,d,p}} & \sum_{a \in A} \sum_{b \in B} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} V_p, \\ \text{s.t.} & \sum_{a \in A} \sum_{d \in D} \sum_{p \in P} x_{a,b,d,p} R_p^{bw} \leq K_b^{bw}(S), \quad b \in B. \end{aligned} \quad (6)$$

Therefore, the self-study evaluation of spoken English in the information-based learning environment has been specifically evaluated.

Constructing a constraint parameter index evaluation system for the self-study system evaluation of spoken English, using speech scientific computing knowledge recognition algorithms to evaluate the self-study system of spoken English based on big data information system evaluation, in order to improve the quantitative assessment ability of the English classroom scheduling level, we propose a method: a self-study assessment method for spoken English based on the speech scientific computing knowledge recognition algorithm. Speech scientific computing knowledge recognition is to seek the consistent estimation value of the resource constraint vector β of the English spoken self-study system, so that $\|Y - X\beta\|$ is minimized, where $\|\cdot\|$ is the F -norm in the European algebra norm, and the entropy of the constraint feature information of the English classroom scheduling ability is obtained. The feature extraction value is

$$P_{\text{loss}} = 1 - \frac{1 - p_0}{\rho} = \frac{p_0 + \rho - 1}{\rho} = \sum_{n=1}^N p_{k,n}. \quad (7)$$

Given that d_i is the perturbation feature vector of the self-study assessment of spoken English, the estimation formula of the English classroom scheduling ability is transformed into the least square solution:

$$z(t) = x(t) + iy(t) = a(t)e^{i\theta(t)} + n(t), \quad (8)$$

where $x(t)$ is the real part of the time series for evaluating the distribution of big data and $y(t)$ is the imaginary part

of the evaluation constraint index sequence of the self-learning system of spoken English.

7. Self-Study Quality Assessment Algorithm

7.1. Feature Extraction. The spoken English self-study speech is broken down based on the multilayer wavelet feature ratio transform method, and the output independence phase of the self-study quality is obtained. It can be observed that there is a one-to-one mapping relationship between the phase R^N and X^N :

$$\begin{aligned} p(R^N = r_i) &= p(X^N = x_i | x_i| = |r_i|, \text{angle}(x_i)) \\ &= (\text{angle}(r_i) - \varphi_g) \bmod (2\pi). \end{aligned} \quad (9)$$

When the phase distribution angle(X^N) of the self-study spoken English speech output by noise reduction is uniformly distributed on $[0, 2\pi)$, R^N is independent of φ_g . The phase information $I(R^N; \varphi_g | Z^N) = 0$ of the voice signal within the distribution range of the energy set $\{P_1, P_2, \dots, P_j\}$ can be obtained and introduced into the above equation to obtain the following:

$$H(X^N | Z^N) = H(R^N | Z^N) + H(\varphi_g | Z^N). \quad (10)$$

The feature quantities of wavelet entropy in the self-study English spoken speech are extracted as follows:

$$\begin{aligned} H(R^N) &= - \sum_{i=1}^M p(r_i) \log(p(r_i)) \\ &= - \sum_{i=1}^M p(x_i) \log(p(x_i)) \\ &= H(X^N). \end{aligned} \quad (11)$$

In the above equation, M stands for the number of elements in the symbol set. The self-adaptive filter coefficient of spoken English self-study speech is $N^{(1)} = N$, $N^{(j)} = N_0^{(j-1)}$, $2 \leq j \leq J$. Based on the signal analysis above, intelligent speech assessment is carried out to improve the capacity of assessing the quality of spoken English self-study [11].

7.2. Assessment of Self-Study Quality. Thresholding is carried out on the spectral features of self-study speech, and the signal source of each self-study feature in the k interval is located based on the time frequency analysis method to obtain the self-study feature spectrum of the English self-study quality assessment system.

$$\begin{aligned} H(X^N | Z^N) &= H(R^N, \varphi_g | Z^N) = H(R^N | Z^N) \\ &\quad + H(\varphi_g | Z^N) - I(R^N; \varphi_g | Z^N). \end{aligned} \quad (12)$$

In the frame number data transmission time, the output quantization information of the self-study feature vector for each frame can be obtained as follows:

$$\begin{aligned} I(X^N; Z^N) &= H(X^N) - H(X^N | Z^N) = H(X^N) - H(R^N | Z^N) \\ &\quad - H(\varphi_g | Z^N). \end{aligned} \quad (13)$$

The definition of the Discrete Fourier Transform (DWT) in the tone feature sequence of the self-study quality assessment system for spoken English is expressed as follows:

$$X(k) = \sum_{n=0}^{N-1} x(n) \exp\left(-j \frac{2\pi}{N} nk\right). \quad (14)$$

In the above equation, $0 \leq k \leq N - 1$ indicates the length of the voice signal, and the relevant feature quantities of the self-study voice signal in the self-study quality assessment system of spoken English are extracted. The quantification and balance processing are carried out on the self-study of speech based on the empirical mode decomposition method. The domain model for the self-study speech of English speech is broken down by wavelet transform, and the self-study quality is automatically assessed based on the result of wavelet entropy feature extraction from the self-study of spoken English [12].

$$\begin{aligned} C_S &= \max_{X^N} [I(X^N; Z_B^N) - I(X^N; Z_E^N)] \\ &= \max_{X^N} [H(R_E^N | Z_E^N) - H(R_B^N | Z_B^N) + H(\varphi_E | Z_E^N) - H(\varphi_B | Z_B^N)] \\ &= \max_{X^N} [H(W_E^N | Z_E^N) - H(W_B^N | Z_B^N)] + [H(\varphi_E | Z_E^N) - H(\varphi_B | Z_B^N)]. \end{aligned} \quad (15)$$

In the above equation, φ_B and φ_E , R_B^N and R_E^N , and W_B^N and W_E^N are multilayer wavelets of various interclass bit sequences. Based on the design of the algorithm described above, the automatic assessment of the self-study quality of spoken English can be implemented.

7.3. Design of the Function for the Assessment of Spoken English Self-Study. The assessment function is the basis for implementing the self-study assessment function of spoken English. Its flow chart is shown in Figure 4. Firstly, the English speech input by the user is characteristically processed based on the Mel frequency spectrum coefficients. The result of feature processing is subjected to unit matching and English part-of-speech decoding. Finally, grammatical and meaning analysis is carried out and then output [13].

Mel frequency collapse coefficient is the essence of the spoken self-study assessment function. This coefficient has energy features. If it is applied to spoken self-study assessment, a nonlinear relationship with audio can be established. The English voice signal is $\omega(n)s(n)$, the frame sequence is n , the frame signal is $s(n)$, and the frame number of the window function $\omega(n)$ is N . Among them, the window function has three patterns as follows: rectangle, Hamming, and

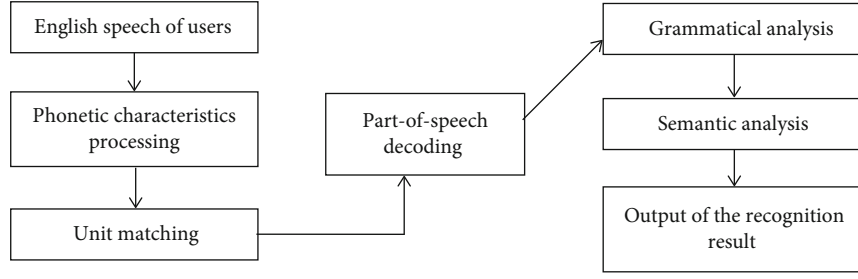


FIGURE 4: Flow chart of the self-study assessment function of spoken English.

Hanning, which are expressed specifically in equations (16), (17), and (18), respectively.

The rectangular pattern is expressed as follows:

$$\omega_R(n) = \begin{cases} 1, & 0 \leq n \leq N, \\ 0, & n > N \text{ or } n < 0. \end{cases} \quad (16)$$

The Hamming pattern is expressed as follows:

$$\omega_{H_m}(n) = \begin{cases} 0.54 - 0.46 \cos\left(\frac{2n\pi}{N-1}\right), & 0 \leq n \leq N, \\ 0, & n > N \text{ or } n < 0. \end{cases} \quad (17)$$

The Hanning pattern is expressed as follows:

$$\omega_{H_n}(n) = \begin{cases} 0.5 \left[1 - \cos\left(\frac{2n\pi}{N-1}\right) \right], & 0 \leq n \leq N, \\ 0, & n > N \text{ or } n < 0. \end{cases} \quad (18)$$

In the self-study assessment of spoken language, the Fourier transform is used to analyze the grammar and part-of-speech decoding to acquire the voice signal spectrum. The frame sequence after the Fourier transform is set to k , and the following equation can be obtained:

$$F(k) = \sum_{n=0}^{N-1} \omega(n)s(n)e^{-2n\pi k/N}. \quad (19)$$

The speech energy after feature processing is transformed into an equation through cosine transformation. The result of parsing the grammar is shown in

$$F^*(k) = \sum_{n=0}^{N-1} X_i \cos\left[\frac{n\pi(n+0.5)}{N}\right]. \quad (20)$$

Since the two changes described above can generate partially invalid data, the storage range of valid data is set to $1 \leq n \leq 12$, and the assessment result F is used as the equation through the normalization processing equation, as shown in

$$F = \arg \max Q(P, U) \times \prod Q(P_1, P_2, P_3). \quad (21)$$

English phonetic scoring is the core of the whole learning system. At present, most similar systems are assessed by measuring the similarity with the standard speech models. It is assumed that the i -th phoneme is associated with phoneme q ; the corresponding starting point of the time period is τ_i . The eigenvector observed at the time t is o_t , and the corresponding HMM state is s_t . Then, the likelihood score based on HMM can be obtained as follows:

$$M_i = \sum_{t=\tau_i}^{\tau_{i+1}-1} \lg[P(s_t|s_{t-1})P(o_t|s_t)]. \quad (22)$$

The English speech is scored based on HMM posterior probability as follows:

$$M_i = \sum_{t=\tau_i}^{\tau_{i+1}-1} \lg P(q|o_t) = \sum_{t=\tau_i}^{\tau_{i+1}-1} \lg \frac{P(o_t|q)P(q)}{P(o_t|q_i)P(q_i)}. \quad (23)$$

In the above equation, Q stands for the set of all phonemes.

7.4. Design of the Self-Study System for Spoken English. The operation process of the online English speech self-study assessment system designed in this paper is shown in Figure 5. The self-study of spoken English and the dynamics of the activity input system are mainly to facilitate changes in the assessment of the self-study system of spoken English, and it is necessary to perform measurement indirectly through the obvious teaching mode of the outside world. Based on the analysis of the teaching process between the self-study of spoken English and the students, the types and features of teaching activities are used to assess the health indicator system of the English self-study platform. The oral English self-study system assesses the preparation of the primary English self-study platform, English classroom education, and provision of special counseling. The types of student education activities include student learning on the self-study platform and mutual support among students in the self-study platform. The features of the self-study English activities are that the enthusiasm in the application of the self-study platform and the time used in the teaching process based on the self-study platform can be measured. The features of the learning modes of students can be explained by their enthusiasm for learning and the amount of time spent on learning. The teaching enthusiasm

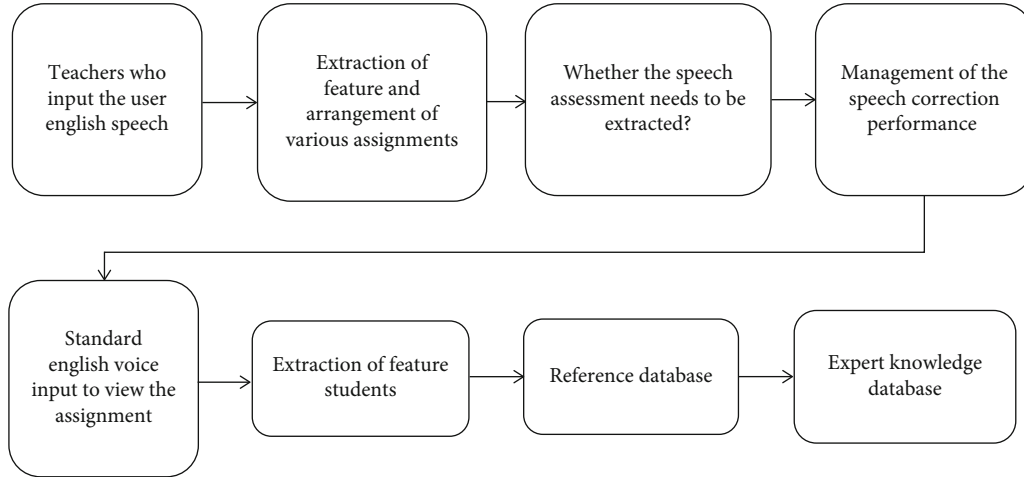


FIGURE 5: Process flow of the online English speech self-study assessment system.

of the English self-study platform is used to assess the degree of effort of the self-study platform for the self-study of spoken English, and it is also an active psychological activity that is presented in the self-study of spoken English as a teaching model. Based on the modern information, the motivation of the higher self-study platform assessment system is input. In addition to the two main teaching models of the self-study platform and students, the assessment basis of the assessment system is operated based on data informatization. In the past, English education required a self-study platform and a teaching model shared by the students. As the teaching model of the education system, modern informatization mainly makes use of the information-based teaching resources and self-study of spoken English.

The self-study assessment index for spoken English is taken as a platform for the self-study of English and the individual application of teaching mode by students, and the input energy from teaching is rationally allocated as an international teaching system. A certain teaching model can be established to ensure smooth energy flow and high efficiency. The self-study platform assessment system can run the sequence in a healthy manner. In the modern information environment, the self-study platform assessment system includes the self-study platform and the interaction between students and teaching resources, detailed self-study platform learning, and classroom learning, which can be used to explore the learning motivation of students to ensure that the input system energy complies with the characteristics of all aspects in these mutual exchanges; the measurement is suitable for the organizational structure of English internationalization assessment and the construction of teaching assessment indicators. The self-study platform and students are required to meet the requirements of each other, and the mutual adaptability is very high. Based on their respective teaching activities, they can supervise each other's work. In this way, the channels of energy transfer between each other have become more fluent, and the number of energy conversions is increased as well. On the whole, fitness should comply with the requirements in four aspects, that is, purpose fitness, content fitness, method

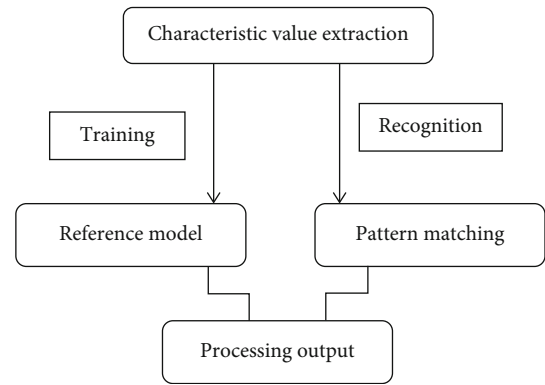


FIGURE 6: System flow chart.

fitness, and attitude fitness, which in turn play a substantial role in the assessment system of the English self-study platform. Various activity goals, teaching content, and learning attitudes are all involved to comply with the teaching level of both parties.

7.5. Speech Training Module. In the design of the system put forward in this paper, professors and professional teachers (from different regions) from 100 foreign language schools in China were organized to record the language resources. The standard sampling mode is adopted, and each person records 50 minutes of English speech. A total of 1850 words and 2500 dialogues are recorded. In addition, the words are assessed and marked to generate language resources. The sampling method is described as follows.

In the speech training module, the tone mark resources include 28 auxiliary tone marks and 20 original tone marks (8 two vowels and 12 vowels). In the expert knowledge database and reference database, the probability of three words appearing at the same time is calculated when the phonetic symbols are analyzed, and equation (24) can be obtained:

$$Q(P_1, P_2, P_3) = Q(P_1)Q(P_2|P_1)Q(P_3|P_2). \quad (24)$$

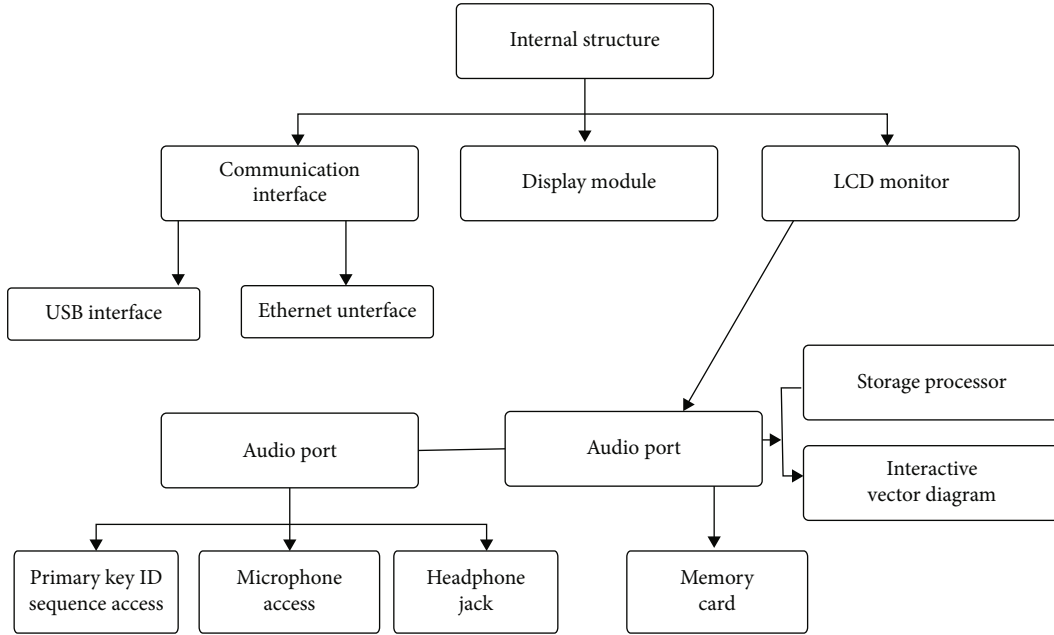


FIGURE 7: Chip structure.

For the probability of a single word, $Q(P_1)$, $Q(P_3 | P_2)$, and $Q(P_2 | P_1)$ are the probability of two words that appear at the same time. The above equation has determined the working intensity of the expert knowledge database and the reference database.

In the second stage of the decoding process, the Markov model is applied. There are sound models, language models, and dictionary databases. An acoustic model is obtained through the training of a large number of standard pronunciation samples based on the Baum-Welch algorithm. The language model is a probability model for the vocabulary or multiple phrases obtained by counting words in a corpus. The dictionary provides the range of words that can be assessed and the phoneme mapping relationship corresponding to the words. These two models are the basis of the whole HMM probability data, and their quality can directly affect the accuracy of the assessment.

Figure 6 shows the process of the speech assessment system, in which the training process is the characteristic value for a large number of standard pronunciations, and the calculation is repeated based on the Baum-Welch algorithm. The result thus obtained is a reference model. The assessment process is the voice characteristic value input by the user and is decoded based on the Viterbi algorithm.

7.6. Design of the Assessment Module. The system designed in this paper can implement functions such as voice correction and database assessment and perform a large number of computation tasks in the operation of the module. Hence, it allows the processing chip to have a consistent computing power. The DSP signal processing chip has a very strong digital signal processing capacity, a low cost, and a small size, which makes it more suitable for users of tablet computers and mobile phones. In addition, the chip has high-speed processing features and online interactive functions. The

resilience of the English course assessment system refers to the capacity of the teaching assessment system to maintain its capacity to create tasks and function normally when it is subject to any threat from the outside world. For the purpose of maintaining the normal operation of teaching assessment, the assessment system can play an overall adjustment role and enhance the protection elements to reduce the risk. This requires the English self-study platform to have an awareness of self-reflection in the teaching process and can consciously correct the existing risk factors in a timely manner. The leading influencing factors in the assessment system are the fatigue of the English self-study platform in the teaching process, the lack of motivation of students, the incompatibility that is present in the information environment, and the teaching organization structure subject to threat from the outside world. They are derived from the professional fatigue of the self-study platform in the interaction between the learning of students and the teaching activities of the self-study platform, the purpose, teaching content, teaching attitudes, and teaching methods that fail to meet the practical demands from different perspectives. It is not applicable to the modern information teaching environment, and it is necessary to overcome the existing problems constantly. This requires that the self-study platform and students should have the capability to make reflections, detect the existing factors of potential risks, and take the corresponding measures to overcome the existing difficulties. It should be a platform in which self-study and the educational capabilities of students can be effectively used, and risk factors can be detected in a timely manner. The excellent state of the learning motivation and organizational teaching structure is equivalent to the key protective factors in the English classroom assessment system, so as to promote the reflection of the English self-study platform and the education level of students effectively and ensure that students have better learning capabilities.

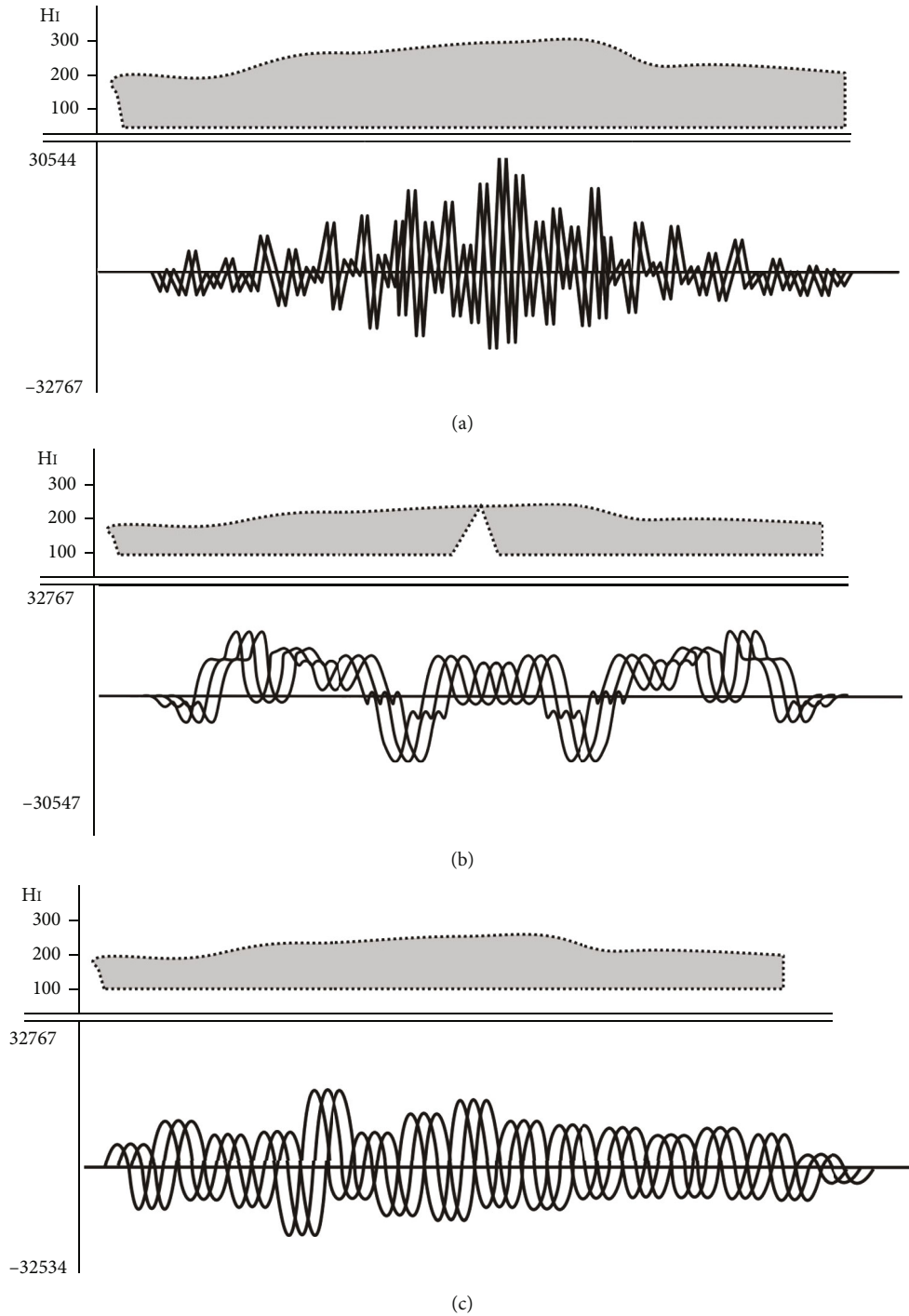


FIGURE 8: Prosodic correction.

The chip has a 250 KB memory processor, which can implement accompanying storage and on-demand functions. This feature allows the chip to buffer LCD data and sound data. The memory space of the system can be expanded with a memory card. Audio frame buffering can be implemented with interactive vector graphics. Voice interaction can be implemented with a 250 KB storage processor. In addition, it can support the multiprocessing mode. The powerful memory management function of the chip can implement the assignment of self-study assessment tasks in spoken English and migrate assessment

functions through the Ethernet interface. For the purpose of accomplishing powerful processing functions, the system can effectively improve the efficiency of self-study assessment of spoken English. The OMP AP 5912ZG chip is used as the main processor, and the tablet computer or mobile phone processor is used as the master processor (Figure 7).

7.7. *System Test Experiment Analysis.* For the purpose of testing the method put forward this paper, the application performance of the detection of spoken English voice signal

and the assessment of self-study quality are implemented, and the system test is carried out accordingly. In the experiment, the MATLAB simulation software is used to design the sample test signal of the spoken English voice signal. The linear frequency modulation signal is taken as the test signal. The spoken English speech assessment sample is 1.2 s wide, and the relative bandwidth is 0.4 dB. The acquisition frequency of self-study spoken language and self-study speech of different vocal cords is 1024 kHz, and the signal frequency of the baseband is 2 kHz~10 kHz.

If the audio segment is eliminated in the synthesized speech (for example, the blank part is inserted by the learner by mistake), it is necessary to connect the two sound units before and after the synthesis. Since the end points of the front and rear connection units have been attenuated, the envelope of the basic frequency and amplitude will present a jump change, and appropriate spline points should be selected to avoid the interruption and jump of the basic frequency curve accordingly.

Figures 8(a)–8(c) show the basic frequency diagram of the waveform of the female voice, male voice, voice, and prosodic correction voice during reading aloud. The fundamental frequency is calculated based on the short-term amplitude difference (ADMF) method. In addition, the width envelope has yet to be corrected. The channel filter modulates the excitation signal to acquire the voice signal accordingly. Relative to the fundamental frequency, the spline described above may not be the channel spectrum. However, the fundamental frequency is continuous, and the channel spectrum is not continuous. The synthesized voice sounds unnatural at the seams, but it seems that they are from different language environments. However, the time to synthesize the sound is long, and the basic frequency and the other rhythm elements are a complete imitation of the reference sound and can maintain the timbre of the learning sound.

Based on the simulation environment and parameter settings described above, the detection of spoken English voice signals and the assessment of self-study quality are carried out. Figure 9 shows the results of the original signal collection.

The self-study spoken English speech collected in Figure 9 is a test sample, and the self-study spoken English speech is characterized based on the multilayer wavelet feature ratio transformation method. According to the results of feature decomposition, self-study spoken English speech is subjected to adaptive filtering detection and spectral analysis, and the detection result of signal detection and assessment is shown in Figure 10.

Through the analysis in Figure 10, it can be observed that the application of the method put forward in this paper can distinguish between the detection and assessment of spoken English self-study speech very well. Different methods are tested, and the accuracy of self-study quality of spoken English is assessed accordingly. The comparison results are shown in Figure 11.

Through the analysis in Figure 11, we have found that the self-study quality assessment of spoken English based on this system has high accuracy and excellent stability.

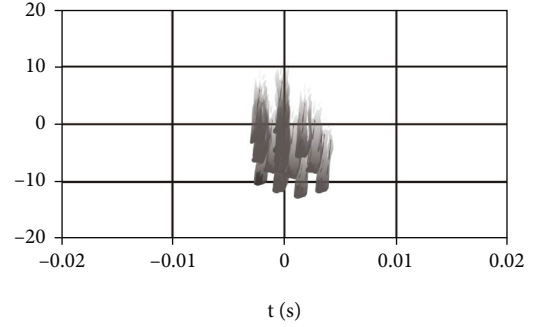


FIGURE 9: Results of signal acquisition.

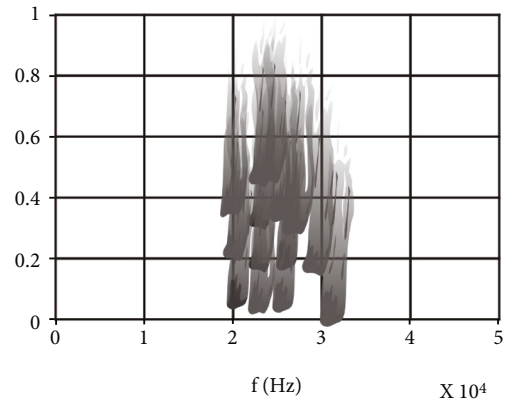


FIGURE 10: Test results in the quality assessment of self-study spoken English speech.

For the purpose of verifying the effectiveness of the system designed in this paper, the author extracted a sentence with 20 words, “Zhejiang FTZ is the only FTZ in China to focus on the development of the oil and gas industrial chain”, from the standard assessment database and assigned it to the training set. The 24-dimensional Hamming window is used as the voice signal window, the frequency is 30 kHz, the length is 20 ms, the frame shift is 80 points, and the frame is divided into 252 points. The parameters for the phonetic characteristics of the content in the training set are processed based on vectorization (52 codes). After processing, a 48×72 order matrix is formed. After training, a speech model can be output from the 29 English words. A mobile phone is used for testing, the processor is Kirin 990, and the memory is 8 G.

For the purpose of verifying the performance of the system, the HMM system, the endpoint detection system, and the Audry system, which are popular spoken English self-study assessment systems at present, are used to compare with the system designed in this paper. The overall control and result output of the system are processed by using MATLAB software. Based on the four systems, spoken self-study assessment and feature extraction are carried out, and a speech model is established. After training for 10 times, the average assessment rate and the assessment time of the 4 assessment systems are obtained, as shown in Table 1.

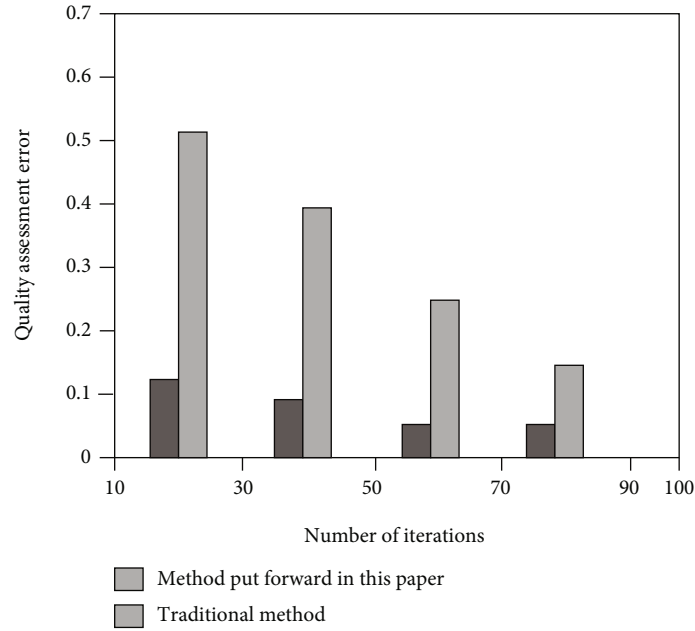


FIGURE 11: Comparison of the accuracy in the self-study quality assessment of spoken English.

TABLE 1: Comparison results of assessment systems.

Serial number of the speech model	Endpoint monitoring system		Audry system		HMM system		The system designed in this paper	
	Assessment rate (%)	Assessment time (ms)	Assessment rate (%)	Assessment time (ms)	Assessment rate (%)	Assessment time (ms)	Assessment rate (%)	Assessment time (ms)
1	98.253	24.569	98.776	15.698	92.125	13.698	99.818	12.635
2	98.126	20.119	99.123	16.002	91.003	12.589	99.268	12.964
3	97.663	22.369	98.695	15.289	92.535	18.216	98.968	13.563
4	96.595	19.848	97.664	15.669	94.216	17.625	98.901	14.536
5	97.212	22.147	98.013	16.987	89.215	12.336	99.053	13.684
6	93.214	28.693	98.096	17.002	89.336	13.528	99.525	16.279
7	95.113	27.629	96.215	16.592	91.025	14.256	98.773	17.061
8	94.126	28.139	95.125	15.365	92.216	12.693	98.693	12.113
9	98.001	22.169	97.662	15.841	93.114	13.547	98.525	11.986
10	97.629	25.551	98.218	16.035	95.216	14.006	99.147	13.516
11	95.365	24.336	99.003	15.256	90.963	13.694	99.981	12.589
12	96.333	27.219	98.675	14.969	90.003	13.546	98.927	14.369
13	97.125	29.329	98.762	16.259	91.326	12.589	98.796	12.654
14	96.123	28.647	96.251	16.576	88.779	17.216	99.669	12.156
15	93.655	27.664	95.796	15.295	94.252	12.697	99.785	12.954
16	94.558	25.125	98.256	16.286	93.251	12.254	99.367	13.816
17	95.216	23.958	98.698	15.321	94.105	13.664	98.864	16.021
18	97.001	24.569	99.126	17.069	91.189	14.654	99.125	13.624
19	97.921	20.325	99.331	16.696	92.336	15.329	98.138	14.941
20	96.129	23.693	98.276	15.643	90.017	14.369	98.714	12.694

From the analysis above, it can be known that the assessment rate of the online English speech self-study assessment system designed in this paper is significantly higher than that of the HMM system, and the assessment rate of the end-

point detection system is slightly higher than that of the Audry system. In terms of the assessment time, the system is almost the same as the HMM system, and its assessment time is significantly shorter than that of the Audry system

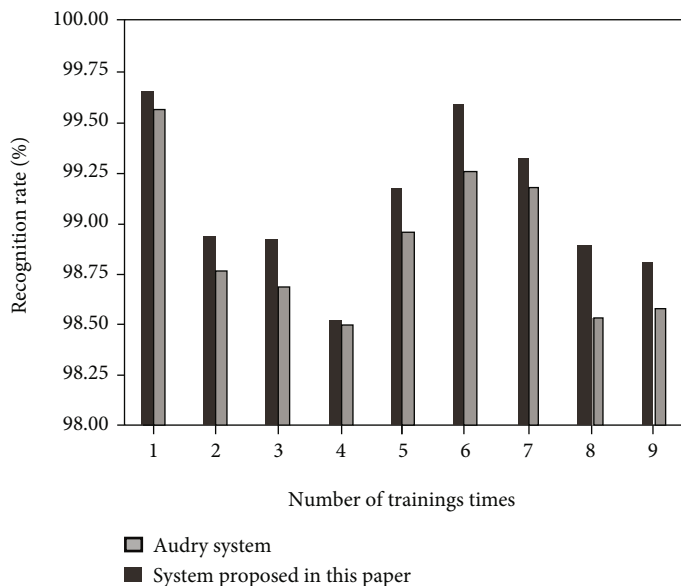


FIGURE 12: Comparison of the assessment rates between the Audry system and the system proposed in this paper.

and the endpoint detection system. For the purpose of enhancing the persuasiveness that the assessment rate of the system designed in this paper is higher than that of the Audry system, as shown in Figure 12, a curve is plotted for the self-study assessment rate of single training for the spoken language in the 10 training sessions based on the training groups described above. The results suggest that the assessment rate of the system designed in this paper is higher than that of the Audry system.

In summary, the online English speech self-study assessment system designed in this paper has certain advantages over the mainstream spoken English self-study assessment systems in terms of the recognition rate and the assessment time. Hence, the system is highly effective.

8. Conclusions

For the purpose of addressing the existing problems in the oral English self-study assessment system at present, an English self-study assessment system based on the speech scientific computing knowledge assessment algorithm is designed in this paper. In the system, a reference database is mainly used to compare with an expert knowledge database to identify errors. At the same time, assessment is made based on the phonetic characteristics of English learners. This method can effectively improve the assessment rate of English pronunciation and substantially reduce the assessment time. Finally, in the experimental section, three self-study assessments, that is, HMM system, endpoint detection system, and Audry system, are compared. The results indicate that the system designed in this paper has significant advantages over the other systems in the assessment rate and assessment time of spoken English and can effectively improve the accuracy and stability of self-study quality assessment of spoken English.

Data Availability

The data used to support the findings of this study are available from the corresponding author upon request.

Conflicts of Interest

The authors declare no conflicts of interest.

References

- [1] K. Yang, "The study of oral English with ESP in universities of engineering," *Education Modernization*, vol. 7, no. 8, pp. 128–135, 2017.
- [2] B. Yang, H. Tang, and L. Mou, "Research on higher English internationalization education model and evaluation index system based on multi-source information fusion," *Computational Intelligence and Neuroscience*, vol. 2021, 8 pages, 2021.
- [3] L. Mugglestone and J. R. Schwyter, "Dictating to the Mob: The History of the BBC Advisory Committee on Spoken English Oxford, U.K.: Oxford University Press. 2016. xiv +281 pp. Hb (9780198736738) 25.00.[J]," *Journal of Sociolinguistics*, vol. 21, no. 2, pp. 289–293, 2017.
- [4] L. Mugglestone and J. R. Schwyter, "Dictating to the Mob: The History of the BBC Advisory Committee on Spoken English. Oxford, U.K.: Oxford University Press. 2016. xiv + 281 pp. Hb (9780198736738) £25.00," *Journal of SocioLinguistics*, vol. 21, no. 2, pp. 289–293, 2017.
- [5] Y. Yang, "Study on the construction of multimodal interactive oral English teaching model," *Open Journal of Modern Linguistics*, vol. 8, no. 4, pp. 137–142, 2018.
- [6] M. C. Pennington and P. Rogerson-Revell, "English pronunciation teaching and research (contemporary perspectives)," *English Pronunciation Teaching and Research: Contemporary Perspectives*, pp. 1–55, 2019.
- [7] N. Nakamura, N. Kinoshita, and E. Yanagisawa, "How the learners grasp the special morae: improvement of pronunciation with the movement of the body[J]," *The journal of*

Japanese Language Education Methods, vol. 22, no. 3, pp. 72-73, 2015.

- [8] N. Cavus, "Development of an intelligent mobile application for teaching English pronunciation," *Procedia Computer Science*, vol. 102, pp. 365-369, 2016.
- [9] C. Liu, "Artificial intelligence interactive design system based on digital multimedia technology," *Advances in Multimedia*, vol. 12, 5 pages, 2021.
- [10] S. Tipmontree and A. Tasanameelarp, "Using role playing activities to improve Thai EFL students' oral English communication skills," *International Journal of Business and Society*, vol. 21, no. 3, pp. 1215-1225, 2021.
- [11] S. Yan, "The washback effect of web-based oral English assessment system on oral English teaching," *Journal of Jiamusi Vocational Institute*, vol. 21, no. 3, pp. 1215-1225, 2017.
- [12] H. Wen and Department, FL, "On the assessment mode for oral English curriculum in colleges and universities from the perspective of needs analysis," *Journal of Putian University*, vol. 34, no. 2, pp. 265-291, 2017.
- [13] Y.-J. Hung, "Group peer assessment of oral English performance in a Taiwanese elementary school," *Studies in Educational Evaluation*, vol. 59, no. 6, pp. 19-28, 2018.