Research Article

A Recognition Method Based on Speech Feature Parameters-English Teaching Practice

Lili Zhu,[1] **Xiujing Yan** (iD)**,**[1] **and Jing Wang**[2]

[1]*School of Foreign Languages, Mudanjiang Medical University, Mudanjiang, Heilongjiang 157011, China*
[2]*School of Basic Medical Sciences, Mudanjiang Medical University, Mudanjiang, Heilongjiang 157011, China*

Correspondence should be addressed to Xiujing Yan; yanxiujing@mdjmu.edu.cn

In order to improve the effect of English teaching practice, this paper constructs an intelligent English phonetic teaching system combined with the method of phonetic feature parameter recognition. Moreover, this paper simulates the self-mixing interference signal containing noise by establishing a simulation, analyzes the size of the noise and its various possibilities, and selects the EEMD method as the English speech denoising algorithm. In addition, with the support of an intelligent denoising algorithm, this paper implements an English intelligent teaching system based on the recognition algorithm of English speech feature parameters. Finally, this paper evaluates the teaching effect of the intelligent English speech feature recognition algorithm proposed in this paper and the intelligent teaching system of this paper by means of simulation teaching. The research shows that the English teaching system based on the intelligent speech feature recognition algorithm proposed in this paper has a good effect.

1. Introduction

In English learning, language input and language output are closely related. Input is the premise and necessary preparation for output, and output is the ultimate goal of input and can also stimulate input. The two complement each other and jointly improve students' comprehensive language ability. Although the Internet provides a large number of online learning resources, the individual differences in English proficiency of college students are quite large. Moreover, students are not very motivated to actively obtain language input, and their ability to learn online resources independently is not good. Therefore, college students in the Internet age still have the problem of insufficient effective language input.

The teaching hours and teaching capacity of English classrooms in colleges and universities are limited. Public English courses are mostly taught in large classes. The number of students in each class ranges from fifty to sixty to hundreds of students. One class is 90 minutes, and the average output of each student is the language output. There are very few opportunities, and the students' language output in the classroom is far less than the language input, relying on the online learning platform to realize real-time online and offline interaction between teachers and students, expand teaching space and time, increase students' language output opportunities on online platforms and offline classrooms, and help solve the dilemma of traditional English classrooms. The content and form requirements of language output are not random. It should form an organically linked whole with the content and form of language input. It is not only necessary for teachers to regulate the quantity and quality of language input. Make the content of language input materials meet the interests and needs of college students, and ensure that the difficulty level of language input materials is in line with students' learning ability and learning level. It is even more necessary for teachers to properly supervise the online and offline language input process of students in classrooms and online platforms, promote students' online and offline autonomous learning, and achieve effective language output.

Oral expression and written writings are the main forms of language output and the main way to test the validity of language input. When designing the teaching content of

speaking, audio-visual, reading, and writing, attention should be paid to the relevance of thematic cohesion between modules to ensure that, under the same learning theme, students can learn vocabulary and sentence patterns layer by layer and practice repeatedly to gradually improve [1]. Oral content is mainly arranged in the course, and under the guidance of teachers, students train their oral expression or speech ability. In the case of students with low English proficiency, we can set up oral homework after class and encourage students to assist each other offline [2]. In the writing module, short videos, microlectures, or courseware materials for preclass preview can be provided for students to learn by themselves. If students have a poor English foundation, they should not set too many preclass writing tasks that are too difficult. We can focus on the input of writing knowledge, strengthen the teacher's intensive teaching and guidance in the middle of the class, encourage students to write on the spot in the classroom, and complete effective language output [3].

In English teaching, listening teaching runs through the whole teaching process, and the application conditions of intelligent speech synthesis technology in English listening teaching are analyzed from two aspects of teaching activities and lesson types in the textbook.

This paper combines the method of speech feature parameter recognition to construct an intelligent English pronunciation teaching system and applies it in English teaching practice to improve the effect of modern English teaching.

2. Related Work

The artificial neural network essentially imitates the human neuron structure, sets some judgments between the input and output, whether it is 1 or 0, and outputs data after a series of judgments and processing. After being able to solve some simple linear problems, it falls into the stagnation of development, and the XOR problem is difficult to solve [4]. The study of backpropagation has arisen. Backpropagation helps the model to learn the weight parameters independently, which solves the training problem of artificial neural networks. With the continuous development of Internet technology, artificial neural networks accumulate more and more data, and the learning speed is getting faster and faster. It has become a research hotspot in the world [4].

The deep learning method is developed on the model of artificial neural network, and the learning speed is faster. Literature [5] designed an 8-layer convolutional neural network for speech recognition. The deep learning model has strong learning ability and good calculation results, and the training difficulty is also better than that of artificial neural network, but because it is an emerging technology, there is still less research at present, and continuous exploration and attempts are required.

Reference [6] divides the nonstationary signal into several signals with small time intervals by adding a fixed time window to the signal, and the signal can be regarded as stationary within a sufficiently narrow time window, and then the Fourier transform is performed to obtain the signal. The time-frequency correspondence of the Gabor transform

is proposed. On the basis of the Gabor transform, the signal is decomposed by using window functions of different sizes to obtain the time-frequency analysis method of the Short-Time Fourier Transform (STFT) [7]. STFT relies heavily on the selection of the size and shape of the time window, and the selection of the time window is unique, which also leads to the singularity of the resolution of the STFT, and the size of the window function cannot be adjusted according to the frequency transformation. Reference [8] proposed the Wigner-Ville Distribution (WVD), which performs Fourier transform on the signal by using the instantaneous autocorrelation function. It is also an effective method to deal with nonstationary signals because it does not use a window function. The signal is decomposed so that its time-frequency focus is high, but there are serious cross-interference terms. Reference [9] proposes a time-frequency analysis method of NLMS transform (Wavelet Transform, WT) by matching the NLMS basis function with the signal. The parent NLMS selected by the NLMS transform can be scaled and shifted, and the shape of the NLMS basis can be changed according to the frequency characteristics of the signal. Therefore, the time-frequency analysis results have a certain "zoom." The NLMS transform essentially considers the signal in the time window to be approximately stationary, and sometimes the local signal time-frequency resolution deteriorates, which limits the application of the NLMS transform in seismic signals [10]. Reference [11] proposed S transform (ST), whose time window can be adaptively adjusted with frequency, which improves the defect of fixed time-frequency resolution in STFT and better meets the time-varying characteristics of seismic signals, but the fixedness of the selected NLMS basis function leads to certain limitations in its application. Reference [12] proposed the generalized S transform (GST), the window function of the generalized S transform can be flexibly adjusted with the frequency transformation, and the time-frequency analysis results with high time-frequency focus can be obtained. At the same time, the seismic signal can be flexibly analyzed according to actual needs.

In literature [13], the Hilbert transform is proposed, which is an analysis method suitable for spectral decomposition of seismic signals, which well excavates the physical significance of seismic signals and obtains three-instantaneous parameter attributes with geological significance. Reference [14] proposed the Hilbert-Huang Transform (HHT) on the basis of the Hilbert transform, which combined the Empirical Mode Decomposition (EMD) method with the Hilbert transform. EMD decomposes the signal into several subsignals, that is, Intrinsic Mode Function (IMF), and performs Hilbert transform on each M component to obtain the instantaneous properties of each subsignal. EMD decomposition has the advantages of completeness and self-adaptation but also has the defects of mode aliasing, endpoint effect, and insufficient sieving conditions. Reference [15] applied high-order spline interpolation to EMD to improve its calculation accuracy. Reference [16] improves EMD and proposes Ensemble Empirical Mode Decomposition (EEMD), which adds white noise to the original signal, which effectively solves the problem of modal aliasing, but

also causes residual signal reconstruction and noise. Reference [17] proposed the Complete Ensemble Empirical Mode Decomposition (CEEMD), which is also a noise-assisted method. It aims to add white noise to the first segment of the decomposition, which greatly reduces the influence of noise. From the above examples, although the decomposition method of EMD has achieved certain results, the overall calculation process is complicated and difficult to handle because there is no mathematical theoretical basis in the process. Reference [18] proposes an adaptive and nonrecursive signal decomposition method, which can locate the signal homeopathic spectrum more accurately because the modes decomposed in this method have their own center frequencies, and the modes are different from the modes. It is very compact between states, which can effectively eliminate the modal aliasing phenomenon commonly found in EMD, CEEMD, and other methods.

3. Denoising and Recognition of Speech Waveform Features

Signal denoising is an important step in the whole system process, and it is a basis for the following work. Only by obtaining relatively pure self-mixing interference signals after denoising can the following theories and experiments ensure the rigor and reduce unnecessary errors.

EMD is a form of decomposing the entire signal data into multiple IMFs and then superimposing them on each other. By analyzing the IMF subsignals, respectively, a more meaningful frequency and power can be obtained. The whole is the screening process, which is to simplify the complexity. In the whole process, the signal is iterated countless times, which can eliminate its previous error and eliminate the waveform superposition. Compared with the former, it is more symmetrical, and it is also the significance of decomposing into multiple IMF subsignals.

EEMD is an Ensemble Empirical Mode Decomposition, and the EEMD algorithm is a noise-assisted data analysis algorithm proposed on the basis of the EMD algorithm, which makes up for the insufficiency of EMD. The processing method based on EEMD is to assist the analysis by adding noise. In the EMD method, the obtained IMF is the required condition, that is, the distribution of the extreme points of the signal. If the distribution of extreme points is not uniform, modal aliasing will occur.

The principle of the EEMD algorithm is similar to that of EMD, but different from EMD is that modal aliasing will occur in EEMD, which is also unique to EEMD. EEMD is based on EMD by adding white noise signals with different degrees of uniformity, and the white noise will be evenly distributed in each signal. According to the characteristics of zero-mean noise, each time the noise is removed, it is superimposed and finally averaged, which can cancel each other out. Simply put, the principle of EEMD is to add white noise on the basis of EMD and then remove it. The specific steps of the EEMD algorithm are as follows:

(1) The algorithm introduces white Gaussian noise $\omega(t)$ with spectral mean value of 0 into the signal to be decomposed x(t) and normalizes it to obtain the signal $X(t)$, as shown in formula (1):

$$X(t) = x(t) + \omega(t). \tag{1}$$

(2) The algorithm performs EMD operation on the signal $X(t)$ and decomposes it to obtain n IMF components $C_j(t)$ and 1 participating component $r_n(t)$, as shown in formula (2):

$$X(t) = \sum_{j=1}^{n} C_j(t) + r_n(t). \tag{2}$$

(3) The algorithm repeats test steps 1 and 2 and continues to introduce white noise that satisfies the normal distribution into the signal to be processed. The total number of tests is N, which can be obtained as shown in formula (3):

$$X(t) = \sum_{j=1}^{n} C_{ij}(t) + r_{in}(t). \tag{3}$$

(4) According to the characteristics of the zero-mean value made above, it can be canceled as nothing, and the influence of the added white noise on the signal can be removed. Finally, the IMF components $C_j(t)$ and $r(t)$ of each order are obtained by decomposition, as shown in formula (4):

$$\begin{cases} C_j(t) = \dfrac{1}{N}\sum_{j=1}^{n} C_{ij}(t) r(t) = \dfrac{1}{N}\sum_{i=1}^{n} r_{in}(t), & i,j = (1,2,3,\ldots n). \end{cases} \tag{4}$$

(5) The final reconstructed signal is shown in formula (5):

$$X(t) = \sum_{j=1}^{n} C_j(t) + r_m(t). \tag{5}$$

In the above formulas, $C_{ij}(t)$ is the j-th IMF component obtained by decomposing the i-th white noise, r is the time, N is the number of white noise sequences added, $X(t)$ is the original signal, $r_n(t)$ is the residual term after decomposition, and n is the number of times of decomposition. In formula (5), $C_j(t)$ and $r_m(t)$ are the IMF components of each order and the final residual components obtained after EEMID decomposition, respectively. For the convenience of understanding, the flowchart of EEMD decomposition is drawn as shown in Figure 1.

Signal denoising has always been a key area that people generally pay attention to. There are many denoising methods with different denoising effects and different degrees of simplicity. Therefore, how to choose a better method for denoising is very important, and denoising needs evaluation indicators to evaluate. Nowadays, the more practical evaluation indicators are signal-to-noise ratio, mean square error, and correlation coefficient. The following will use these three evaluation indicators to select and
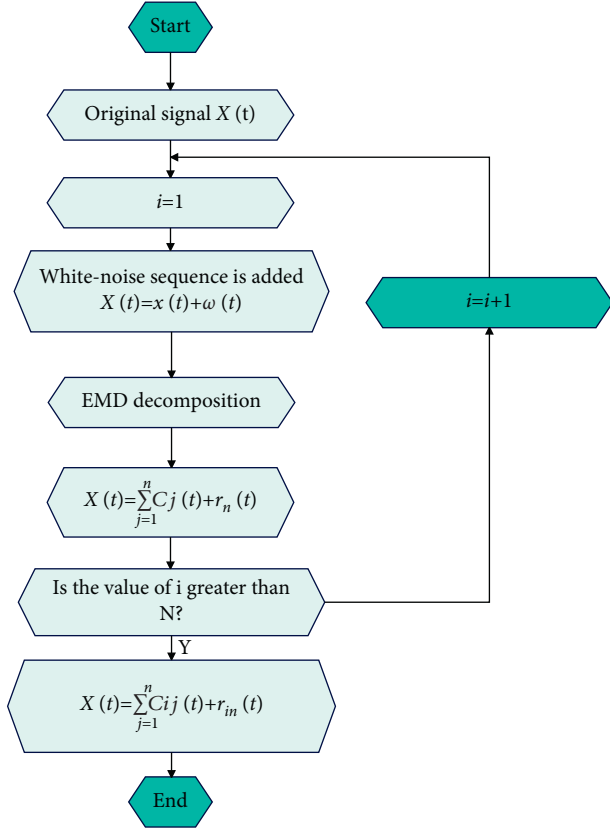
Figure 1: Flowchart of EEMD algorithm.

compare different denoising methods in order to select the optimal algorithm. First, the three evaluation indicators are introduced.

3.1. Signal-to-Noise Ratio (SNR).
The signal-to-noise ratio can be expressed as the ratio of two energies in the mathematical sense, and the expression is

$$\text{SNR} = 10 \times lo_{10}\left(\frac{\text{Power}_{\text{signal}}}{\text{Power}_{\text{noise}}}\right). \tag{6}$$

In the formula, $\text{Power}_{\text{signal}} = (1/n)\sum_{n}[f(i) - f(i)]^2 = \text{RSME}^2$ is the frequency of the original signal; $\text{Power}_{\text{signal}} = (1/n)\sum_{n}[f(i) - \widehat{f(i)}]^2 = \text{RSME}^2$ is the power of the noise. Among them, $\text{Power}_{\text{signal}} = (1/n)\sum_{n}f^2(n)$;

$$\text{Power}_{\text{signal}} = \frac{1}{n}\sum_{n}[f(i) - \widehat{f(i)}]^2 = \text{RSME}^2. \tag{7}$$

From the current research, the signal-to-noise ratio is often used as a measure of noise. It can not only reflect the size of the noise but also use it to evaluate the effect of denoising. In theory, the higher the signal-to-noise ratio, the better the denoising effect.

3.2. Root Mean Square Error (RMSE).
The root mean square error usually refers to the ratio of the square of the deviation of the reconstructed signal and the original signal to the

number of measurements. It can well express the precision of the measurement process and is also called the standard error. Its expression is

$$\text{RSME} = \sqrt{\sum_{n}[f(i) - \widehat{f(i)}]^2}. \tag{8}$$

In the formula, $f(i)$ is the original signal, $\widehat{f(i)}$ is the decomposed reconstructed signal at a certain scale, and n is the total length of the signal.

From the current research, the mean square error is often used as a measure to remove noise. In theory, the smaller the mean square error, the better the denoising effect.

3.3. Correlation Coefficient.
The correlation coefficient is a statistical indicator first designed by statistician Carl Pearson. It is a measure of the degree of linear correlation between variables. It is generally represented by the letter r, and its expression form is

$$r(X, Y) = \frac{\text{cov}(X, Y)}{\sqrt{\text{Var}[X]\text{Var}[Y]}}. \tag{9}$$

In the formula, $\text{cov}(X, Y)$ is the covariance of X and Y, $\text{Var}[X]$ is the variance of X, and $\text{Var}[y]$ is the variance of Y.

The correlation coefficient is often used as a strong indicator of how well the noise is removed. In theory, the larger the correlation coefficient, the better the denoising effect, and the smaller the correlation coefficient, the worse the denoising effect.

We add varying degrees of noise to the pure, undisturbed simulated signal. Moreover, we add three different levels of interference with SNR of 20 dB, 25 dB, and 30 dB to the signals of $f_t = 20$ Hz, $f_s = 2000$ Hz, $a_1 = 7\pi$, $a_2 = 120\pi$, $n = 2000$, and $C = 3$, respectively, as shown in Figure 2.

It can be seen from Figure 2 that, with the increase of the added SNR, the noise becomes smaller and smaller. It can simulate different degrees of interference in real experiments, which is closer to reality and can conduct more comprehensive simulation and analysis so as to better determine the effective feasibility of the method. EEMD generates each IMF by decomposition, as shown in Figure 3.

The simulation of this paper is carried out on the operating platform of Matlab, and EEMD is used to denoise the self-mixing interference signal. In the simulation, we set $f_t = 20$ Hz, $f_s = 2000$ Hz, $a_1 = 7\pi$, $a_2 = 120\pi$, $n = 2000$, $C = 3$, and $\alpha = 3$ and add Gaussian white noise with a signal-to-noise ratio (NSR) of 20 dB.

It can be clearly seen from the first 8 IMFs in the figure that the high-frequency signals are mainly concentrated in the first 4 IMFs, and the signals decomposed later are relatively pure signals. What we mainly remove is the high-frequency signal. It can be seen from the first four IMF diagrams that the noise mainly comes from the transition point, and the interference generated by the transition point is the mainstream interference. The key point to filter out is the noise of the transition point so that the transition point can be pinpointed when processing the signal, so the first 4

A self-mixed interference signal added with 20 dB noise

A self-mixed interference signal added with 25 dB noise

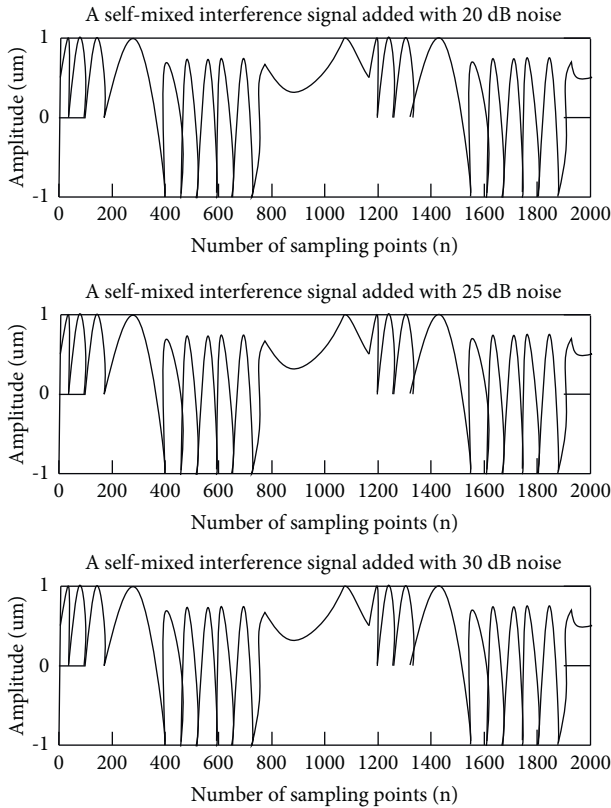A self-mixed interference signal added with 30 dB noise

Figure 2: Self-mixing interference signal with different noise added.

signals are filtered out, and all the remaining IMF signals are synthesized, as shown in Figure 4.

From the above figures, it can be clearly seen that most of the noise can be basically removed, and residue is the signal that removes the residual high-frequency noise. It can be seen that the noise is negligible, a pure interference signal is obtained, and the transition point is easy to see, which lays the foundation for the subsequent phase unwrapping and displacement reconstruction.

In order to observe the simulation effect more intuitively, we comprehensively compare the influence of each signal-to-noise ratio on the signal. The simulated signal parameters are $f_t = 20$ Hz, $f_s = 2000$ Hz, $a_1 = 7\pi$, $a_2 = 120\pi$, $n = 2000$, $C = 3$, and $\alpha = 3$. Table 1 shows the data of each denoising evaluation index when $C = 3$.

4. English Teaching Based on Speech Feature Parameter Recognition

When setting up teaching units in English courses, the topics of chapters can be modularized, and each teaching unit adopts a combination mode of life themes, workplace, or post themes. In addition, the commonly used teaching content can also be selected according to the module combination of vocabulary, audio-visual, speaking, reading, writing, grammar, and translation (interpretation and translation). Each learning module is configured with teaching content according to the combination of before class, one class and one after class, online learning + online tutoring, and offline teaching + offline activities. According to the teaching content, design various activities of online teaching and classroom teaching. At the same time, it is necessary to optimize the online learning mode and teacher-student interaction mode of the course, broaden the path of language task output, and design corresponding homework, tasks, and their supporting evaluation methods. The overall framework of the modular hybrid teaching model is shown in Figure 5.

There are four teaching layers in Figure 6, which are cognitive construction layer, embodied learning layer, situation interaction layer, and innovation generation layer. The cognitive construction layer and the embodied learning layer are the process layer of acquiring knowledge and skills, the situational interaction layer is the situational layer of problem solving, and the innovation generation layer is the practical layer of knowledge and ability. In the process layer, teachers are students' facilitators, supporters, and classroom observers. Students actively acquire tools and materials through "physical skills," actively seek help and support from human resources, and complete the construction of self-cognition through repeated iterations of action and thinking. At the situational level, teachers use examples familiar to students to introduce teaching, creating a learning atmosphere and motivation for students. Moreover, they set up problem situations to make students live in the teaching content, let students understand the truth that learning is life and life is learning, exert students' associative ability and innovative thinking, and stimulate students' strong desire to solve problems. In the practice layer, students use imagination and open thinking in brainstorming to conceive a variety of implementation plans and possibilities, open up new ideas for problem solving through hands-on practice, and creatively organize experience and knowledge to solve new problems. The implementation of teaching links, the growth and development of students, and the improvement of ability and literacy all run through the four teaching layers in the model. Each of the four teaching layers plays a different role. Among them, the cognitive construction layer is a collection of experience and creation, and experience is transferred and applied in innovation. The embodied learning layer gives full play to students' initiative and encourages students to use their hands and brains and experience learning through "doing." The situational interaction layer realizes the transfer of the situation, which can incite the emotion of the students, arouse the students' resonance, strengthen the interaction of the classroom, and make the classroom teaching start, proceed, and end under the sense of substitution of the situation. The innovation generation layer is the practice field for students to solve problems and an important activity layer for cultivating students' innovative thinking and creativity. Each layer is independent of each other but not isolated and disconnected, and multiple teaching activities may be carried out in one teaching layer. For example, in the cognitive construction layer, there will be learning activities such as experience reproduction, comprehension of new knowledge, and design and creation. At the same time, the same teaching
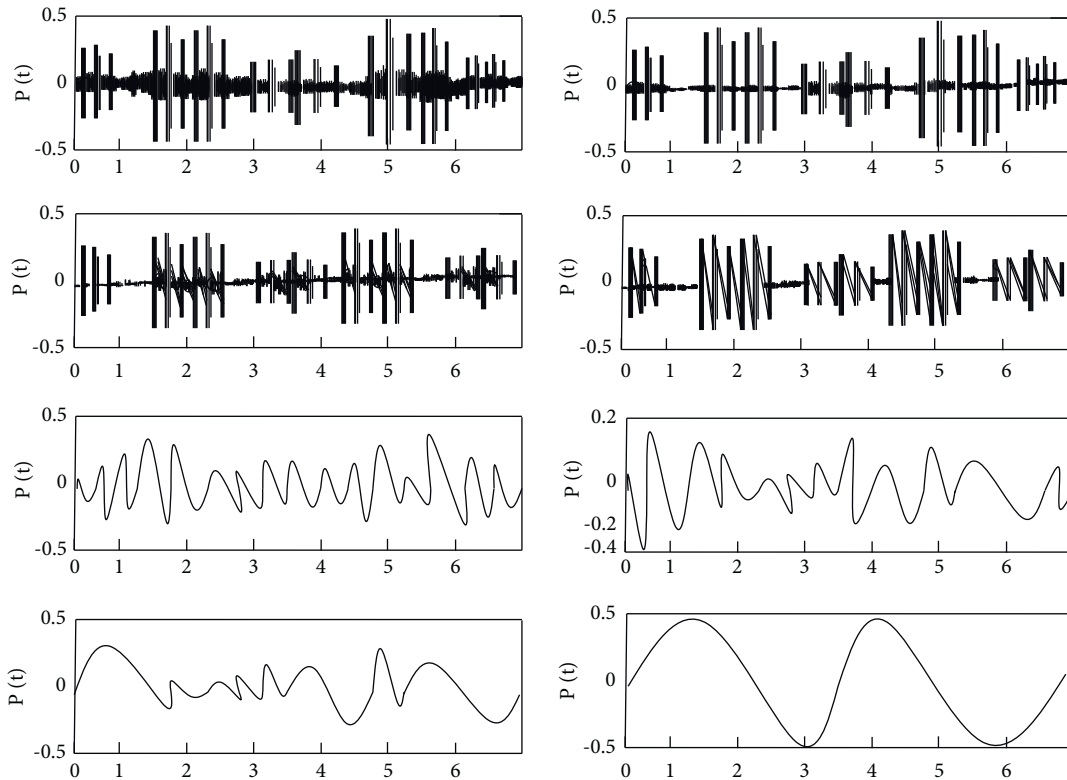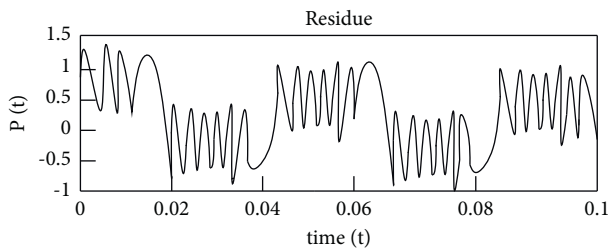
Figure 3: Decomposed different IMF diagrams.

Figure 4: Self-mixing interference signal after filtering.

task will be completed in more than one teaching layer, such as students' inquiry activities occurring in the cognitive construction layer, the embodied learning layer, and the innovation generation layer. The four teaching layers are an important part of the migration teaching model, which echoes the top layer of the model, conforms to the theoretical concept of this study, supports the implementation of the three paths in the middle layer of the model, and is a teaching field that nurtures teaching, development, and ability literacy.

Speech recognition technology, broadly speaking, refers to semantic recognition and voiceprint recognition. In a narrow sense, it refers to the understanding and recognition of speech semantics, also known as Automatic Speech Recognition (ASR). As the most natural way of human-computer information interaction, the basic idea of speech recognition technology takes speech as the research object and converts the input speech signal into corresponding text commands through the process of machine recognition and understanding so as to realize the control of the machine by speech. In the development of speech recognition technology, although different researchers have proposed many different solutions, the basic principles are the same. In the processing of speech signals, any speech recognition system can use Figure 7 to represent its general recognition principle. The most important modules of speech recognition system are speech feature extraction and speech pattern matching.

In the view of behaviorist learning theory, human learning is a process of stimulus-response-reinforcement. In the learning process, there is only stimulus-response, and the lack of reinforcement will greatly reduce the learning effect. However, if we blindly provide stimuli and ignore the learner's original knowledge level, learning needs, and interest in learning, it will be difficult for the stimulus to stimulate the learner's response; that is, it is difficult for learning to occur. On the other hand, because people's information processing methods are different, if only a single learning stimulus is provided and the differences in the learners' individual information processing methods are ignored, it will lead to difficulties for learners to process new knowledge using their original cognitive experience. The manifestation is that it is difficult to learn new knowledge. At the same time, in English pronunciation learning software, many learning systems lack an overall score for the learner's phonetic pronunciation, nor can they give timely and accurate feedback on the learner's pronunciation. Moreover, it does not provide corrective guidance for learners'

TABLE 1: The recognition effect of the English teaching system based on the intelligent speech feature recognition algorithm on spoken English features.

| Num	Speech recognition	Num	Speech recognition	Num	Speech recognition
1	93.86	19	94.96	37	92.01
2	94.69	20	95.95	38	92.78
3	90.34	21	92.28	39	96.53
4	94.60	22	93.82	40	89.42
5	95.25	23	94.53	41	93.98
6	89.96	24	93.94	42	89.65
7	91.98	25	96.06	43	96.58
8	94.75	26	93.59	44	91.86
9	89.02	27	94.00	45	94.05
10	89.26	28	96.20	46	89.01
11	92.20	29	89.79	47	89.25
12	95.90	30	95.86	48	93.09
13	90.66	31	89.37	49	92.00
14	89.19	32	96.92	50	92.35
15	93.19	33	89.28	51	96.90
16	90.36	34	91.91	52	94.39
17	91.00	35	93.09	53	95.87
18	96.60	36	96.61	54	92.33
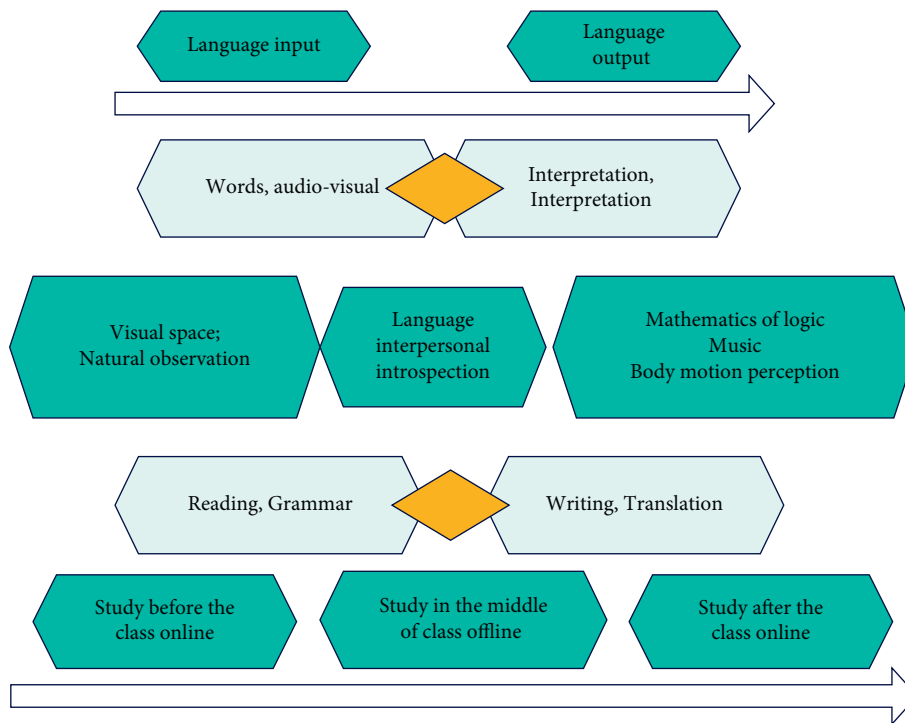
FIGURE 5: Framework diagram of the mixed mode of "language + multiple intelligences."

mispronunciation, and it is difficult for learners to find and correct their own pronunciation errors due to their limited level.

The integration of information technology and English courses is not simply to present the learning content on the computer, but the following points need to be considered. One is to use information technology and multimedia technology to systematically integrate the knowledge of English phonetic transcription. The second is to present knowledge to best promote learners to absorb. The third is to strengthen and consolidate the learning of the learners. In view of the problems existing in the above-mentioned English phonetic pronunciation learning software, behaviorist learning theory, multimedia cognitivist learning theory, and related theories of information technology and curriculum integration should be taken into account when designing the platform. In the implementation of the system, by constructing an English phonetic transcription learning
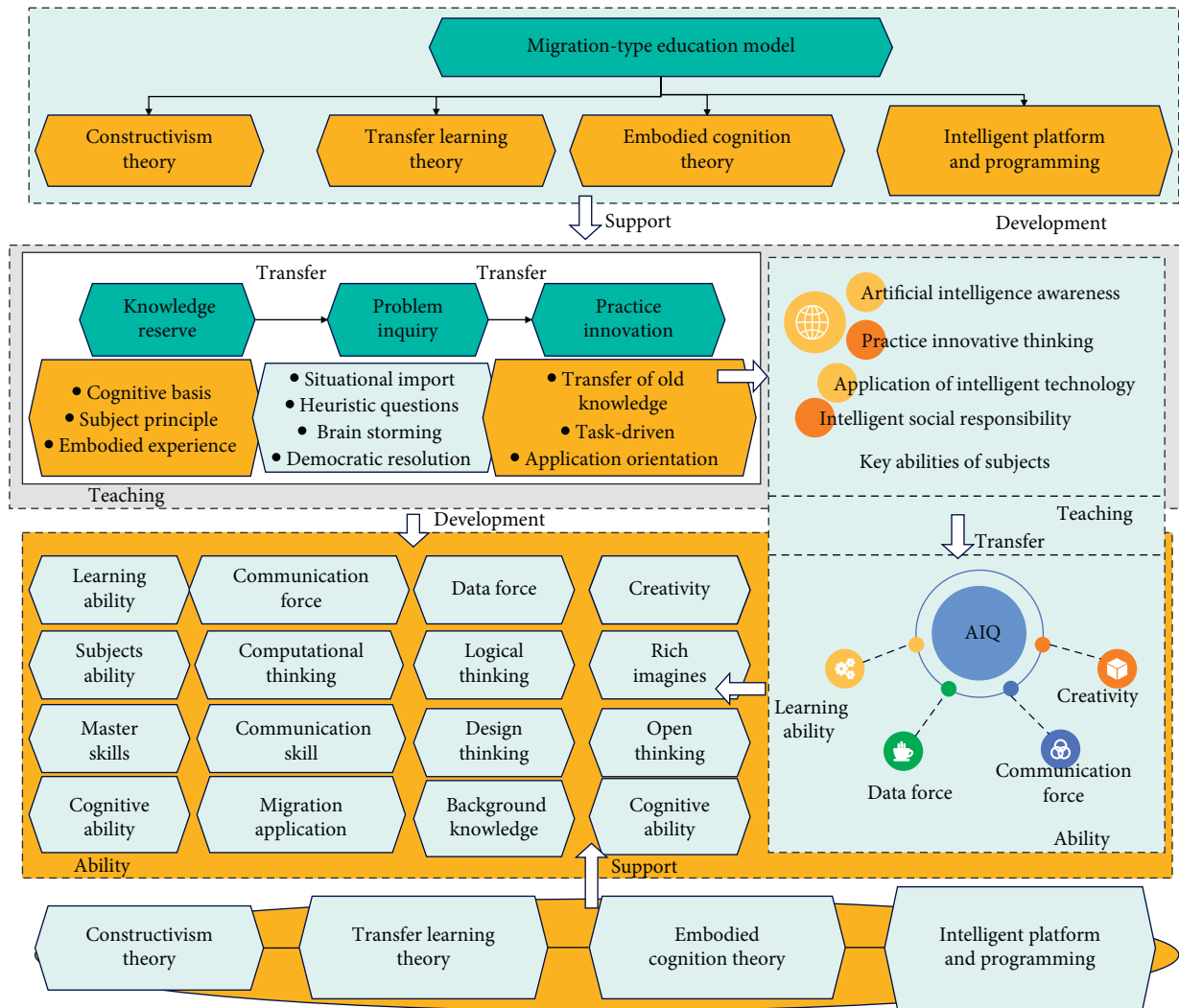
Figure 6: English teaching mode diagram.

environment, learners can choose the learning content independently. The system presents learners with various forms of learning stimuli, which facilitates learners to link old and new knowledge through various information processing channels and form their own knowledge structure. At the same time, the system uses template matching technology in speech recognition to compare the difference between the tester's pronunciation and the standard pronunciation and strengthens the learning results by testing the learner's pronunciation and giving pronunciation feedback. This process of stimulation, response, connection, and

reinforcement is more in line with the law of human learning, so it can better promote the occurrence of learning. As mentioned in the review, the English speech learning assistance platform studied in this paper is shown in Figure 8.

On the basis of the above research, the intelligent English speech feature recognition algorithm proposed in this paper and the intelligent teaching system of this paper are evaluated by means of simulation teaching, and the results shown in Table 1 and Table 2 are obtained.

From the above research, it can be seen that the English teaching system based on the intelligent speech feature
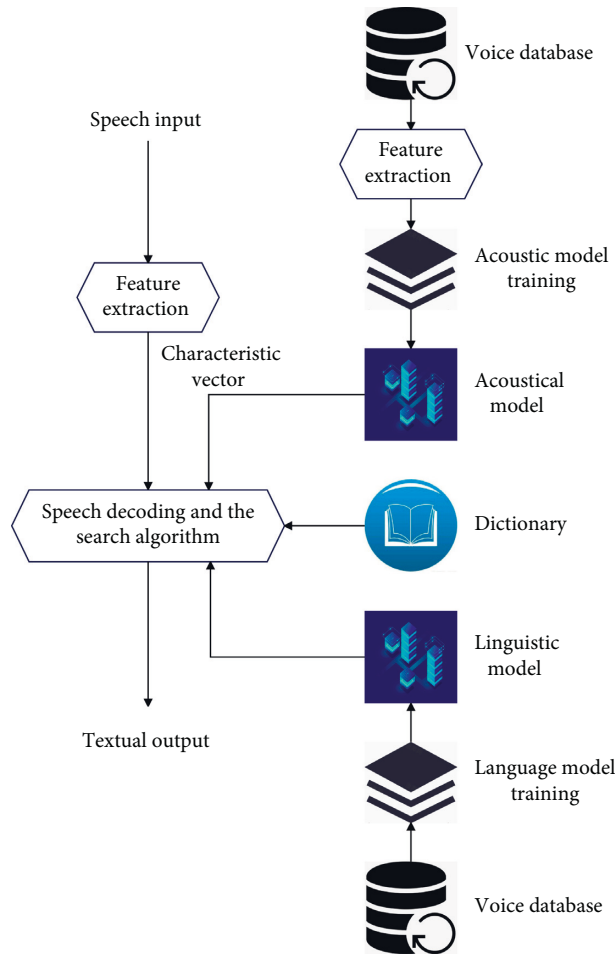
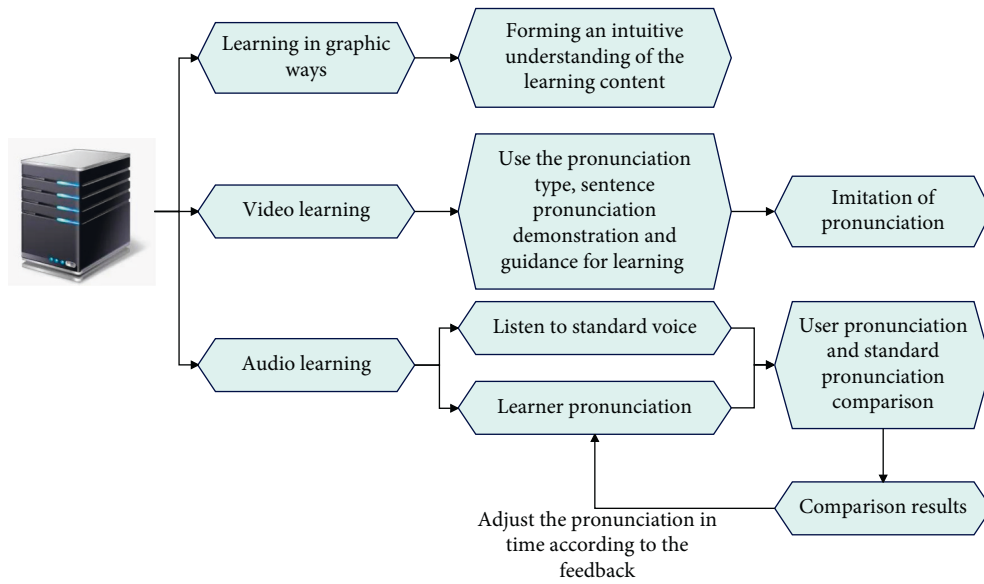FIGURE 7: Principle block diagram of English speech recognition system.

FIGURE 8: English speech learning assistance platform.

TABLE 2: Evaluation of the English teaching effect of the English teaching system based on the intelligent speech feature recognition algorithm.

Num	Teaching English	Num	Teaching English	Num	Teaching English
1	82.71	19	83.38	37	85.31
2	92.44	20	82.40	38	83.74
3	89.75	21	85.83	39	82.54
4	89.95	22	92.05	40	82.41
5	84.45	23	84.62	41	88.85
6	85.28	24	85.80	42	89.95
7	86.39	25	87.98	43	92.67
8	90.15	26	86.55	44	90.04
9	83.58	27	92.38	45	90.46
10	92.84	28	91.76	46	88.36
11	83.25	29	87.44	47	84.36
12	90.57	30	90.97	48	92.56
13	89.49	31	91.81	49	87.40
14	87.02	32	84.75	50	88.97
15	85.83	33	91.57	51	83.14
16	87.12	34	83.21	52	90.89
17	90.70	35	92.35	53	85.25
18	84.24	36	87.58	54	85.01

recognition algorithm proposed in this paper can not only effectively improve the effect of spoken English feature recognition but also effectively improve the effect of English teaching.

5. Conclusion

The application of intelligent voice technology in education and teaching is never to replace the role of teachers but to help teachers carry out teaching, make up for teachers' deficiencies, improve teaching efficiency and effect, and achieve the effect of "reducing burden and increasing efficiency." The application conditions of technology in education and teaching also need to be analyzed. Under the application conditions, teachers can play their original role, and technology can also reflect the functional advantages of technology itself. In English listening teaching links or teaching scenarios, intelligent speech technology can give full play to its technical advantages and better assist teachers in English listening teaching, so that the role of teachers and the functions of intelligent speech synthesis technology can complement each other. In this paper, an intelligent English pronunciation teaching system is constructed by combining the method of speech feature parameter recognition. The experimental study shows that the English teaching system based on the intelligent speech feature recognition algorithm proposed in this paper can not only effectively improve the recognition effect of spoken English but also effectively improve the effect of English teaching.

Data Availability

The labeled dataset used to support the findings of this study is available from the corresponding author upon request.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

References

[1] A. S. Fatimah, S. Santiana, and Y. Saputra, "Digital comic: an innovation of using toondoo as media technology for teaching English short story," *English Review: Journal of English Education*, vol. 7, no. 2, pp. 101–108, 2019.

[2] B. Ayçiçek and T. Yanpar Yelken, "The effect of flipped classroom model on students' classroom engagement in teaching English," *International Journal of Instruction*, vol. 11, no. 2, pp. 385–398, 2018.

[3] N. Guzachchova, "Zoom technology as an effective tool for distance learning in teaching English to medical students," *Bulletin of Science and Practice*, vol. 6, no. 5, pp. 457–460, 2020.

[4] M. S. Hadi, "The use of song in teaching English for junior high school student," *English Language in Focus (ELIF)*, vol. 1, no. 2, pp. 107–112, 2019.

[5] A. Mahboob, "Beyond global Englishes: teaching English as a dynamic language," *RELC Journal*, vol. 49, no. 1, pp. 36–57, 2018.

[6] H. Sundari, "Classroom interaction in teaching English as foreign language at lower secondary schools in Indonesia," *Advances in Language and Literary Studies*, vol. 8, no. 6, pp. 147–154, 2017.

[7] A. Gupta, "Principles and practices of teaching English language learners," *International Education Studies*, vol. 12, no. 7, pp. 49–57, 2019.

[8] B. S. M. Abdelshaheed, "Using flipped learning model in teaching English language among female English majors in majmaah university," *English Language Teaching*, vol. 10, no. 11, pp. 96–110, 2017.

[9] T. Ara Ashraf, "Teaching English as a foreign language in Saudi Arabia: struggles and strategies," *International Journal of English Language Education*, vol. 6, no. 1, pp. 133–154, 2018.

[10] O. Tarnopolsky, "Principled pragmatism, or well-grounded eclecticism: a new paradigm in teaching English as a foreign language at Ukrainian tertiary schools?" *Advanced Education*, vol. 5, no. 10, pp. 5–11, 2018.

[11] N. I. Sayakhan and D. H. Bradley, "A nursery rhymes as a vehicle for teaching English as a foreign language," *Journal of University of Raparin*, vol. 6, no. 1, pp. 44–55, 2019.

[12] D. A. W. Nurhayati, "Students' perspective on innovative teaching model using e in teaching English phonology: a virtual class development," *Dinamika Ilmu*, vol. 19, no. 1, pp. 13–35, 2019.

[13] S. Zou, "Designing and practice of a college English teaching platform based on artificial intelligence," *Journal of Computational and Theoretical Nanoscience*, vol. 14, no. 1, pp. 104–108, 2017.

[14] F. Kong, "Application of artificial intelligence in modern art teaching," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 13, pp. 238–251, 2020.

[15] M. Pantic, R. Zwitserloot, and R. J. Grootjans, "Teaching introductory artificial intelligence using a simple agent framework," *IEEE Transactions on Education*, vol. 48, no. 3, pp. 382–390, 2005.

[16] C. Yang, S. Huan, and Y. Yang, "A practical teaching mode for colleges supported by artificial intelligence," *International Journal of Emerging Technologies in Learning (iJET)*, vol. 15, no. 17, pp. 195–206, 2020.

[17] K. Kim and Y. Park, "A development and application of the teaching and learning model of artificial intelligence education for elementary students," *Journal of The Korean Association of Information Education*, vol. 21, no. 1, pp. 139–149, 2017.

[18] O. Zawacki-Richter, V. I. Marín, M. Bond, and F. Gouverneur, "Systematic review of research on artificial intelligence applications in higher education–where are the educators?" *International Journal of Educational Technology in Higher Education*, vol. 16, no. 1, pp. 1–27, 2019.