

## Research Article

# Multimedia Network English Reading Teaching Model Based on Speech Recognition Confidence Learning Algorithm

Chuanju Wang 

*College of Humanities, Shandong Management University, Jinan 250000, Shandong, China*

Correspondence should be addressed to Chuanju Wang; 14438120120302@sdmu.edu.cn

Received 25 August 2021; Revised 11 October 2021; Accepted 18 October 2021; Published 8 November 2021

Academic Editor: Sang-Bing Tsai

Copyright © 2021 Chuanju Wang. 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.

With deepening internationalization, English has become an increasingly important communication tool. Because traditional English teaching has short teacher-student interaction time, lack of oral English training environment, and single learning method, the oral English teaching is not ideal, and the students' "speaking" confidence is insufficient. Aimed at addressing the exposed problems of traditional English reading teaching, this paper proposes a multimedia-based English reading teaching mode. On this basis, establish a voice recognition phoneme network grid to detect the recognition results. Secondly, the lattice is used to generate the confusion network mesh, and the acoustic posterior probability is calculated. Then, the feature vector is input into the SVM classifier for confidence mark, and finally the feature is extracted by principal component analysis. The research shows that multimedia network teaching can teach more vividly, increasing the initiative of students. At the same time, it is shown that the speech recognition confidence learning algorithm can improve the language learning system.

## 1. Introduction

The phenomenon of "dumb foreign language" is not uncommon. At present, the current situation of oral teaching also shows that traditional oral English teaching has certain limitations in terms of educational methods and teaching methods, and there are many drawbacks [1–4]:

First, the traditional junior middle school oral English teaching is not intuitive in creating real situations, especially lacking in three-dimensionality and dynamics

Second, individualized teaching and autonomous learning cannot be effectively implemented without fully considering the individual differences of students

Third, there are too many classes, many aspects of group cooperative teaching are not well organized, and students are not in practice

Fourth, both the classroom and the after-school oral practice lack a language environment and cannot achieve remote oral communication

Many experts and English teachers have been discussing effective methods of oral English teaching. To analyze the characteristics of oral teaching, it is found that the tasks and activities of oral teaching are inseparable from the context, and the creation of real and virtual scenes needs to be accomplished by means of information technology. It is an urgent problem to solve in oral English teaching to enable learners to have the social adaptation ability and sustainable development in their future jobs. Learning spoken English through multimedia network has become a better way of learning [5, 6].

Multimedia network learning of spoken English has been realized by using language and phonetic processing technology. Real-time assessment of learners' spoken English improves quality.

As one of the most efficient and convenient ways of communication between human beings, information exchange has shown an explosive growth. As an important way of information exchange, phonetics has naturally become the focus of modern civilized society research. As early as the

1920s, the voice recognition machine “Radio Rex” appeared, but speech recognition technology research really originated in the 1950s. The 1950s witnessed the climax of machine translation, during which the famous Bell lab first introduced a system for identifying ten English isolated words, thus ushering in a new era [7]. Later, other research institutes, such as MIT, have implemented vowel and consonant recognition machines. In the late 1950s, the first computer speech recognition system was introduced.

The significance of confidence level research is illustrated as follows [8–10]:

Firstly, in the isolated word speech recognition system, it can reduce the system error operation and improve the reliability of the system. Therefore, the corresponding confidence level is obtained in the recognition process, which can reduce the misoperation of the system by threshold. Confidence level is one of the important indicators describing the uncertainty of the position of line elements and surface elements in GIS. The confidence level represents the degree of confidence in the interval estimation, and the span of the confidence interval is a positive function of the confidence level; that is, the greater the degree of confidence required is, the wider the confidence interval is bound to be obtained, which correspondingly reduces the accuracy of the estimation. Its threshold is set within this interval, and the higher the accuracy of the estimation, the smaller the false alarm rate.

Secondly, improving the detection efficiency of keyword spotting system: at present, the main mechanism of keyword detection system is to first give a large number of candidates, reduce the system missing report, then use confidence measure method, identify which of these candidates are credible, and reduce system false positives. Keyword detection system hopes to reduce false alarm rate, and its performance is related to the choice of confidence.

Third, in the unsupervised speaker adaptation system, the wrong recognition part will affect the adaptation effect. Therefore, if the confidence algorithm is introduced, the relatively reliable data can be picked out and the unreliable data can be removed; then, the adaptive effect can be improved to a certain extent.

Fourthly, the application of slightly supervised acoustic model training: the biggest difficulty in some speech recognition systems is that the number of speech annotations is relatively small and the training data is insufficient. In practice, the commonly used method is to use the existing general automatic annotation corpus method to annotate the number of voices and use annotations for more reliable training. Application of confidence is similar to that in unsupervised adaptive applications, but the amount of data is larger.

Fifthly, merge the results recognized by different speech recognition systems, although the recognition rate is

similar, but the recognition results are often quite different, which shows that these systems have a certain complementarity. If these recognition results can be integrated into a new result according to the confidence level, the recognition rate of the system may be improved.

At present, the main application directions of speech recognition confidence are error detection and error correction, outlier detection, keyword validation, unsupervised and semisupervised speech training and adaptive technology, speech recognition multisearch technology, corpus error corpus selection, and so on [11–13].

Aimed at addressing the problems of oral English training environment and single learning methods in traditional oral English teaching, this paper proposes a multimedia web-based English reading teaching model. The main contributions of this paper are as follows:

- (1) Summarized problems such as the lack of time for teacher-student interaction, lack of an oral English training environment, and a single learning style
- (2) A teaching mode of English reading based on multimedia network is designed
- (3) The important human-machine communication is applied in the teaching mode, and the evaluation results are given

## 2. Proposed Method

*2.1. Introduction of Multimedia Network Technology.* The English word for multimedia is “multimedia,” which consists of two parts: “multi” and “media.” It is generally understood as a synthesis of multiple media. Multimedia technology is a kind of information technology that, through computer for comprehensive processing and control, can support a series of interactive operations [14].

Compared with traditional media, multimedia has the following five basic characteristics [15, 16]:

- (1) Integration: multimedia has the characteristic of collecting text, sound, picture, and many kinds of media. It is composed of different types of media, including text, sound, graphics, images, videos, and animation.
- (2) Independence: multimedia integration shows that the various single media that make up multimedia are also independent of each other, that is, one kind of media is processed without affecting other media. This feature makes it possible to synthesize different media at different application levels through computers.
- (3) Big data volume: the large amount of information of multimedia data is incomparable to any traditional data.
- (4) Real-time processing: multimedia data is required to be processed in real-time in many cases, such as

audio and video data, where the data are closely related to time information.

- (5) Interactivity of operations: that is to say, it can process, modify, and recombine all kinds of multimedia information at any time by real-time interactive operation mode.

## 2.2. The Construction of Multimedia Network English Reading Teaching Mode

2.2.1. *Principles of Constructing Multimedia Network English Reading Teaching Mode.* With its unique advantages, multimedia network teaching gradually replaces some traditional foreign language teaching methods. Although multimedia network oral English teaching has many advantages, it is not appropriate in oral English teaching; it is likely to achieve the opposite effect. Certain principles should be followed according to different teaching contents, different teaching purposes, and different teaching objects. These principles can be summarized as follows [17–19]:

First, implement the principle of auxiliary teaching and highlight the key points of teaching. However, multimedia can only play a supplementary role. It cannot be overwhelming.

Second, the principle of people-oriented is to regard the student as a living and complete person. Attach importance to everyone's learning process, so that students take the initiative to participate, are willing to explore, be diligent, and use their own knowledge to acquire and construct more knowledge. The transformation from knowledge base to people-oriented approach is the need of the development of the times. Knowledge economy is different from that of agricultural and industrial societies. It relies on high-quality people with innovative spirit and practical ability. Such a person should be a comprehensive, coordinated, free, and fully developed person with personality development [20].

Third, the principle of cooperative learning should be applied, with the multiple roles of instructors, partners, and promoters of students' learning activities. Students are no longer passive recipients, but active participants. The students have established healthy competition and cooperative interpersonal relationship. To advocate teaching democracy, we must create a democratic, relaxed, and harmonious psychological atmosphere. Teachers should fully understand and trust each student; respect their thoughts, emotions, and independent personality; create conditions for students to study independently; provide opportunities for success; and guide them to make a reasonable evaluation of themselves and others [21].

Fourth, pay attention to stimulating students' interests and satisfying the principles of individualized learning.

Multimedia assisted foreign language teaching should adapt to students of different intelligence types, enjoying students' body and mind, stimulating their interest in learning, and developing their personalities. The application of multimedia to oral English teaching in English class has the characteristics of visualization, diversity, novelty, intuition, richness, and interest. Another notable feature of multimedia and network education is compared with that of traditional education [22].

2.2.2. *The Ideal Process of Multimedia Network Reading Teaching.* Assuming that oral English teaching can be carried out in a computer room, the ideal process design of oral English teaching should include the following points [23]:

- (1) The website is mainly for the front desk. Through the course list on the homepage, students can select the courses offered this semester and browse the task list to get a preliminary understanding of the tasks to be completed in this class.
- (2) Teachers can focus on playing videos, import situations and roles, and then demonstrate the new words and sentence patterns.
- (3) Students can correct their pronunciation by using online speech synthesis software. Students use voice recognition software to follow the human-machine reading, human-machine dialog, and online completion of the lesson's oral practice tasks. When the accuracy of pronunciation reaches more than 80%, the task can be considered complete.
- (4) Teachers organize each group to have a group dialog. Speech intercom through the microphone does not affect other students. Teachers can join the discussion to understand the dynamics of students' communication.
- (5) Students log on to the website to complete their homework. Using the form of human-computer dialog, students complete the dialog recorded into audio files and uploaded to the site's own workspace, and teachers play the files online and check the completion of the situation, to explain the common problems.
- (6) In the unit test and final examination, the teacher issues oral dialog questions, and the students have a human-machine dialog. The students' completion is checked at the prescribed time. The speech recognition software automatically scores the questions. After the students complete the dialog task, they will submit the students' spoken dialog files to the teacher's computer automatically. The teacher will play the students' recording one by one to evaluate them.

Finally, there is another step, which is feedback. It is the basic concept of cybernetics, which refers to the process of returning the output of the system to the input and changing the input in some way, thereby affecting the function of the

system. In the teaching model, it is reflected in the learner or reader.

*2.2.3. Construction of Multimedia Network English Reading Teaching Mode.* If the classification is based on the characteristics of the input speech and the modeling method, the current recognition engine is mainly divided into engines such as isolated word recognition, continuous speech recognition, and connected word recognition. They can all use the feedback results of the students and help them correct their pronunciation. Multimedia network English reading teaching model constructed in this paper is shown in Figure 1.

*2.3. Confidence Recognition Algorithm for Speech Recognition.* Exploration technology is the technology for obtaining intelligence, and this intelligence may contain information that is very useful to explorers. After the two opposing systems understand each other's situation through this information, they can adjust their actions to achieve the desired goals. The main purpose of this article is to explore the learning algorithm of speech recognition confidence and make it more suitable for multimedia network English teaching mode. From the previous section, we can see how to correctly identify the learners' pronunciation and make judgments. The key to speech recognition is the distance between models, while the key to speech learning is to calculate the learner's pronunciation and the corresponding relationship with the speech. Based on the distortion of the standard pronunciation model, an accurate assessment can be made to improve the learning status of the learner, using the speech recognition belief learning algorithm to recognize and judge the learner's speech. Confidence is generally defined as a function, used to measure the matching degree between the model and the observed data in speech recognition. Its algorithms basically have a one-to-one correspondence effect, and all relevant algorithms are listed below. The algorithm is described as follows.

#### 2.3.1. Confidence Based on Posterior Probability.

where  $L$  is lattice and  $w_1^M$  is hypothetical word sequence. Direct computation of all  $p(x_{s_m}^{t_m}|w_m)p(w_m|w_1^M)$  in lattice is tedious.

Define forward probability:

$$\alpha_{s_m}(w_m) = p(x_{s_m}^{t_m}|w_m) \sum_{s_{m-1}} \alpha_{s_{m-1}}(w_{m-1})p(w_{m-1}|w_1^{m-1}). \quad (3)$$

Formula (1) can be rewritten as

Posterior probability is the most effective confidence feature currently proved to be used alone. Here, we use the obfuscation network as a storage format for identifying intermediate information to calculate the posterior probability of words [24]. The confidence level is generally expressed as a percentage, so the confidence interval on the confidence level of 0.95 can also be expressed as 95% confidence interval. The two ends of the confidence interval are called the confidence limits. For an estimate of a given situation, the higher the confidence level, the larger the corresponding confidence interval. The calculation of the confidence interval usually requires assumptions about the estimation process (so it belongs to parameter statistics), for example, assuming that the error of the estimation is normally distributed:

- (1) Calculate posterior probability for each arc of lattice
- (2) For each word in the optimal path, an alternative competitor is found and a confusion set is generated, that is, a confusion network [25, 26]
- (3) The posterior probability of words is calculated using the confusion network

Lattice's arcs represent words and nodes represent points of time. Each arc records alternative word information including start and end times, historical paths, acoustical models, and language model scores.

For the first step in the algorithm, we need to calculate the posterior probability of a segment of arc in lattice. Let  $w_m$  be an arc in lattice, corresponding to an observation value  $x_{s_m}^{t_m}$ .  $s_m$  and  $t_m$  are their time end points. A posterior probability for arcs  $w_m$  can be represented by

$$p(w_m | x_{s_m}^{t_m}) = \frac{p(x_{s_m}^{t_m}|w_m)p(w_m|w_{m-n+1}^{m-1})}{p(x_1^T)}, \quad (1)$$

where  $p(x_{s_m}^{t_m}|w_m)$  is the likelihood of acoustic models and  $p(w_m|w_{m-n+1}^{m-1})$  is the score of the language model.  $p(x_1^T)$  is the probability of observation, which can be obtained by calculating the sum of  $x_1^T$ 's hypothesis on all phoneme sequences:

$$p(x_1^T) = \sum_{w_1^M \in L} p(x_1^T|w_1^M)p(w_1^M) = \sum_{w_1^M \in L} \prod_{m=1}^M p(x_{s_m}^{t_m}|w_m)p(w_m|w_1^M), \quad (2)$$

$$p(w_m | x_{s_m}^{t_m}) = \frac{p(x_{s_m}^{t_m}|w_m)p(w_m|w_{m-n+1}^{m-1})}{\sum_{w_1^M \in L} \sum \alpha_{s_{M-1}}(w_{M-1})}. \quad (4)$$

Generally speaking, we do not require other conditions, because the amount of information is large; this not only cannot improve the performance of the speech recognition system, but also increases the complexity of the system. Therefore, the simpler the steps, the better. For the second step, in order to calculate the posterior probability of words, the lattice is aligned according to the optimal path. Each

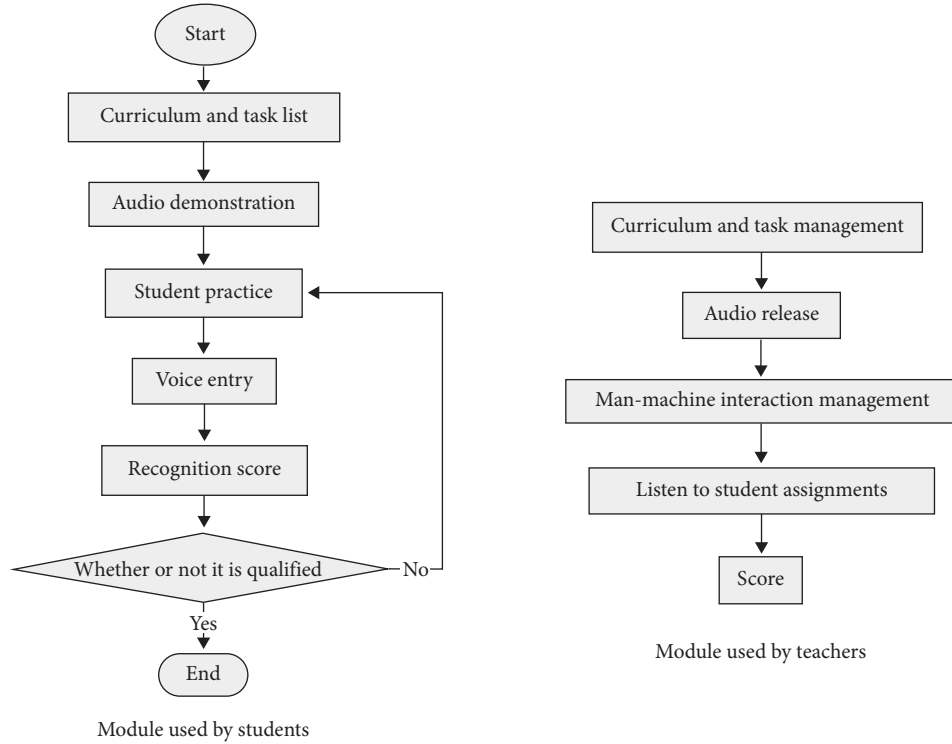


FIGURE 1: Multimedia network English reading teaching model.

word in the optimal path corresponds to an obfuscation set [27]. The method is as follows:

- (1) Firstly, we prune the path with lower probability value in lattice
- (2) Secondly, we compress the hypothetical paths of the same word within a certain time range into one path and add their posterior probabilities
- (3) Finally, we consider different words that occur in a certain time range as alternative competitors, aligned to the same segment of the obfuscation network and forming an obfuscation set

For step 3, the arcs with the same word in each confusion set are merged into an arc, and the probability is equal to the sum of the posterior probabilities of the merged arcs. Finally, the posterior probabilities of each confusion word are normalized [28].

2.3.2. *SVM Classifier*. The basic idea can be illustrated by the 2D case of Figure 2. In the graph, the two types of samples are represented by solid points and hollow points, respectively. We can get the following:  $(x_i, y_i)$ ,  $i = 1, 2, \dots, n$ ,  $x \in R^d$ ,  $y \in \{+1, -1\}$  satisfy

$$y_i[(w \cdot x_i) + b] - 1 \geq 0, \quad i = 1, 2, \dots, n. \quad (5)$$

In addition, the training sample points on  $H_1$  and  $H_2$  are called support vectors.

We can use the Lagrange optimization method: under constraint  $\sum_{i=1}^n y_i \alpha_i = 0$  and  $\alpha_i \geq 0, i = 1, 2, \dots, n$ ,

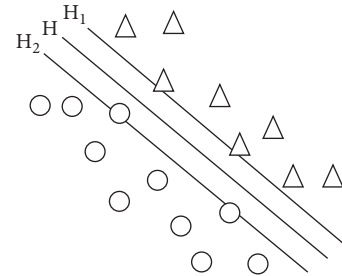


FIGURE 2: Schematic diagram of optimal classification plane.

$$Q(\alpha) = \sum_{i=1}^n \alpha_i - \frac{1}{2} \sum_{i,j=1}^n \alpha_i \alpha_j y_i y_j (x_i \cdot x_j). \quad (6)$$

It can be seen that formula (6), after solving the above problem, can be expressed as

$$f(x) = \text{sgn}\{(w \cdot x) + b\} = \text{sgn} \left\{ \sum_{i=1}^n \alpha_i^* y_i (x_i \cdot x) + b^* \right\}. \quad (7)$$

The summation in formula (7) is actually carried out only on support vectors.

The above discussion is linear, but many practical problems are not linearly separable:  $y_i[(w \cdot x_i) + b] - 1 \geq 0, i = 1, 2, \dots, n$ .

$$y_i[(w \cdot x_i) + b] - 1 + \xi_i \geq 0, \quad i = 1, 2, \dots, n. \quad (8)$$

If the target is changed,

$$(w, \xi) = \frac{1}{2}\|w\|^2 + C \left( \sum_{i=1}^n \xi_i \right), \quad (9)$$

where  $C > 0$  is a constant.

If a problem is not linearly separable in the space it defines, we may consider transferring the problem to a new space by constructing a new eigenvector. This space is generally larger than the original space dimension, but the nonlinear discriminant function in the original space can be realized by using the linear discriminant function.

**2.3.3. Feature Extraction.** The most commonly used feature extraction method is principal component analysis (PCA) [29, 30]. In order to get the effect of dimensionality reduction, PCA method is used to extract feature vectors. PCA is derived from the idea of linear transformation. It is mainly that linear transformation is equivalent to coordinate transformation. Using coordinate transformation, we can get new features with the same number from the original features, and the first few of these features may contain the main information of the original features.

### 3. Experiments

In the experimental simulation work, the computer hardware configuration is as follows:

- (1) Processor: Intel i5 2.50 GHz
- (2) Memory: 4 GB
- (3) Operating system: Windows 7 Ultimate 64-bit

### 4. Discussion

Firstly, this paper makes a preliminary investigation into the students' oral English learning by means of a questionnaire survey and explains through statistics and analysis of survey data.

200 questionnaires are analyzed by the software of online questionnaire service, among which the questionnaire about the reasons why oral proficiency cannot be improved is as shown in Table 1. It should be noted that the difference in pronunciation of women is smaller than that of men. Therefore, the issue of gender ratio must be considered in the process of systematic experiments. If you want to apply it to real life, you must fully consider real-life problems.

From Table 1, we can see that only 25% of the students can not improve their spoken English due to their own reasons, such as lack of time, interest, and self-confidence, while 75% of them cannot do for external reasons. This shows that most of the students are willing to learn spoken English but lack conditions. The student-centered teaching method can well promote students' interest in learning. When the oral English level improves, the student's self-confidence problem can be solved. To illustrate the reasons for not achieving improvement, Table 1 is drawn into a pie chart as shown in Figure 3.

This paper also investigates whether students are willing to use multimedia network to learn spoken English. The results are shown in Table 2. 81% of the students are willing to use multimedia network for learning spoken language and training; only 5% are unwilling.

In this paper, a recognition system with a recognition error rate of 13.98% is selected as the baseline system, and the word posterior probability, word position, language model score, difference of posterior probability, number of contenders, and score of frame acoustic model are used. This paper uses speech recognition and confidence evaluation subsystems to form a complete system, so that the existing recognition effect cannot be changed, and the recognition effect is guaranteed on the basis of the recognition system structure, while reducing the cost. Ten dimensional features, such as sentence length, time length, and word length, are input into SVM classifier to determine confidence, and then PCA is used to extract features. In this paper, the classification error rate (CER) is used as the basis for evaluating the classification of confidence features. The formula can be expressed as

$$\text{CER} = \frac{\text{FA} + \text{FR}}{N} * 100\%. \quad (10)$$

Among them, FA is false recognition, but it is classified as reliable number; FR is correct recognition, but it is classified as unreliable number; and  $N$  is the total number of recognition results.

Table 3 shows that the classification error rate of 10-dimensional features is 10.16%, which is 3.82% lower than the baseline system. The comparison results show that the 10-dimensional features selected in this paper have a good prediction effect. On this basis, PCA is used to process the selected 10-dimensional features, and the classification error rate is 9.05%, which is 1.11% lower than that before processing. This shows that PCA can effectively extract the main information of the original features, remove noise interference, and improve the classification effect. Figure 4 is a histogram of the classification error rate of the speech recognition confidence.

In order to design an English learning system, first of all, an alternative hypothesis model library must be constructed. Second, choose an effective step to adjust the threshold of knowledge. Third, effectively accumulate and synthesize the decoded information to calculate the confidence of the entire sentence, and evaluate the accuracy of the entire sentence on this basis. The robustness and effectiveness of the system are simulated. Conducting robust design experiments involves the use of many of the quality engineering tools mentioned earlier. The success of this work depends on correctly selecting team members, focusing on optimizing and testing appropriate performance, and following methodological guidelines.

Robustness means that a qualified speech learning system should ensure that its FA and FR reach the minimum, and the size of both depends on the rejection threshold. The FA, FR, and threshold of the system designed in this paper are shown in Table 4. From the table, we can see that if the

TABLE 1: A questionnaire about the reasons why oral English is not promoted.

The reasons for not improving oral English level	Percentage
There is no time to practice speaking	5
No confidence to speak	10
Language environment is not good; there is no person communicating in English	40
Lack of interest	10
The teacher does not have the right way of teaching	15
Schools only pay attention to the education mode of grammar and examination results	20



FIGURE 3: The reasons for not achieving improvement.

TABLE 2: The willingness of students to use multimedia network to learn spoken English.

Are you willing to use multimedia network to learn oral English?	Percentage
No	5
Yes	81
Does not matter	14

TABLE 3: Classification results with different characteristics.

Characteristics	CER (%)
Recognition system	13.98
10-dimensional characteristics selected in this paper	10.16
After PCA processing	9.05

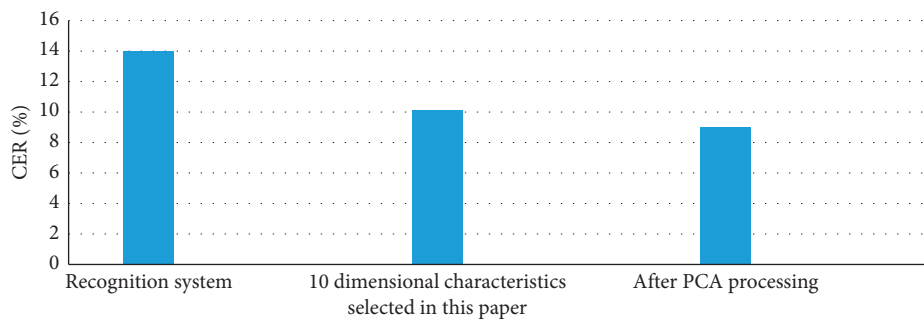


FIGURE 4: Different feature classification effect diagram.

TABLE 4: Robustness table of the system designed in this paper.

Threshold	FA (%)	FR (%)
No threshold	3.5	0
-1.134	1.09	0.55
-0.768	0.99	0.67
-0.659	0.89	0.69
-0.113	0.77	2.1

TABLE 5: Differences between English teacher evaluation and systematic evaluation.

Gender	Differential pronunciation number	Indifference pronouncing number	Difference rate (%)
Male	13	387	3.36
Female	9	391	2.3

threshold is set lower and lower, the FR is lower and lower, but the FA is higher and higher. As the threshold is set higher and higher, FA is getting lower and lower, but FR is getting higher and higher.

Comparing the systematic evaluation with the subjective scores of English teachers, we obtain two differences. The experiment is carried out 50 times, and the average difference is shown in Table 5.

As can be seen from Table 5, the difference rate for males and females is 3.36% and 2.3%, respectively. This shows that the system recognizes and judges pronunciation, and the difference between the result and that of professional English teachers does not exceed 3.36%. After the overall grasp of the experimental data in this article, we found that these data are not unrelated; on the contrary, there is a mutual echoing relationship between them. We can learn from this that the system can identify students' pronunciation very efficiently and get the results. The evaluation is in real time.

## 5. Conclusions

This paper is aimed at addressing the unsatisfactory situation of spoken English teaching in China. When the simulation verifies the merits and demerits of the teaching mode designed in this paper, the necessity of the design of multimedia network oral English teaching is illustrated through questionnaire survey. The shortcomings of this article are as follows: (1) The research object is not suitable for actual needs, the research direction is not quite researchable, and the quality of research needs to be strengthened. (2) The research tools are relatively old and cannot keep up with the times. Therefore, in the following research, we should focus on the improvement of these two shortcomings and strive to enable this system to have a broader application.

## Data Availability

This article does not cover data research. No data were used to support this study.

## Conflicts of Interest

The author declares no conflicts of interest.

## Acknowledgments

This work was supported by Shandong Management University Qihang Project, Data-Driven Teaching Research and Practice of Business English Based on Online Learning (no. QH2020R13).

## References

- [1] Y. E. Yang and L. Xu, "An analysis on senior high school oral English teaching present situation & problems & countermeasures," *Journal of Jiamusi Vocational Institute*, 2017.
- [2] N. Gan and B. Y. Shao, "The present situation and improvement strategy of oral health of students in colleges and universities," *Education Teaching Forum*, 2018.
- [3] S. Nie, "On the reform of oral English teaching in higher vocational college based on spoken English contest," *Journal of Sichuan Vocational & Technical College*, 2016.
- [4] H. Chen and L. I. Jingna, "Reflections on the current situation of phonological assessment: A survey among raters of standardized oral tests in China," *Foreign Languages & Their Teaching*, 2017.
- [5] V. Vukotic, "Deep neural architectures for automatic representation learning from multimedia multimodal data," 2017.
- [6] G. Xue and L. Lingling, "A comparative study on cooperative learning in multimedia and network environment used by English majors between China mainland and taiwan," *Advances in Language and Literary Studies*, vol. 9, no. 1, Article ID 127, 2018.
- [7] L. J. Zhang, "Study on the transformation of scientific and technological achievement of Bell laboratory," *Journal of Guangxi University for Nationalities*, vol. 17, no. 2, pp. 20–24, 2011.
- [8] C. Baulig and F. Krummenauer, "Statistical significance is not equivalent to clinical relevance: interpretation of significance level, P value, and confidence interval," *Implantologie*, vol. 25, no. 2, pp. 127–132, 2017.
- [9] X. D. Ren, Y. K. Zhang, D. W. Jiang, T. Zhang, and G. F. Sun, "A model for reliability and confidence level in fatigue statistical calculation," *Theoretical and Applied Fracture Mechanics*, vol. 59, no. 1, pp. 29–33, 2012.
- [10] X. S. Liu, "Comparing sample size requirements for significance tests and confidence intervals," *Counseling Outcome Research and Evaluation*, vol. 4, no. 1, pp. 3–12, 2013.



- [11] L. Berke, "Displaying confidence from imperfect automatic speech recognition for captioning," *ACM SIGACCESS-Accessibility and Computing*, no. 117, pp. 14–18, 2017.
- [12] C. H. Lee and K. Bang, "Speech recognition system and speech recognition method," *Journal of the Acoustical Society of America*, vol. 126, no. 3, Article ID 1648, 2016.
- [13] T. Stonehocker, K. Mohajer, and B. Mont-Reynaud, "System and method for performing dual mode speech recognition," US Patent, 2017.
- [14] T. Zhao and H. Shi, *Application of Multimedia Network Technology in College Russian Teaching*, Jiangsu Science & Technology Information, Zhenjiang, China, 2016.
- [15] C. W. D. Silva and N. Correia, "Analysis of interactive multimedia features in scientific publication platforms," in *Proceedings of the International Conference on Theory and Practice of Digital Libraries*, Thessaloniki, Greece, September 2017.
- [16] D. Blank, A. Henrich, and S. Kufer, "Using summaries to search and visualize distributed resources addressing spatial and multimedia features," *Datenbank-Spektrum*, vol. 16, no. 1, pp. 1–10, 2016.
- [17] P. A. Lei, *The Principles of Designing Multimedia Teaching Courseware in College English Teaching*, Academics, London, UK, 2010.
- [18] L. I. Hou-Ye, L. I. Hao-Yu, and H. University, "Principles of eco-teaching in college English classes in multimedia network environment," *Journal of Harbin University*, 2014.
- [19] H. B. Huang, "Principles of multimedia-based college English teaching—a learning style perspective," *Journal of Xuzhou Normal University (Natural Science Edition)*, 2012.
- [20] N. N. Hurrah, S. A. Parah, N. A. Loan, J. A. Sheikh, M. Elhoseny, and K. Muhammad, "Dual watermarking framework for privacy protection and content authentication of multimedia," *Future Generation Computer Systems*, vol. 94, pp. 654–673, 2019.
- [21] Y. Zhang, Q. He, Y. Xiang et al., "Low-cost and confidentiality-preserving data acquisition for internet of multimedia things," *IEEE Internet of Things Journal*, vol. 5, no. 5, pp. 3442–3451, 2017.
- [22] S. B. Tsai, W. Wu, S. Ma, C. H. Wu, and B. Zhou, "Benchmarking, knowledge inertia, and knowledge performance in different network structures," *Enterprise Information Systems*, vol. 14, 2019.
- [23] N. Feng, "Empirical research on the teaching of college oral English with the view of input and output theory," *Journal of Shanxi Youth Vocational College*, 2017.
- [24] B. Fanous and H. N. Stefansson, "Parallel implementation of maximum a posteriori probability decoder," US Patent, 2016.
- [25] S. Ghannay, Y. Estève, and N. Camelin, "Enriching confusion networks for post-processing," in *Proceedings of the Statistical Language and Speech Processing*, pp. 119–130, Le Mans, France, October 2017.
- [26] G. Jagfeld and N. T. Vu, "Encoding word confusion networks with recurrent neural networks for dialog state tracking," 2017, <https://arxiv.org/abs/1707.05853>.
- [27] L. Liu and C. Cao, "A seed-based method for generating Chinese confusion sets," *ACM Transactions on Asian and Low-Resource Language Information Processing (TALLIP)*, vol. 16, no. 1, Article ID 5, 2016.
- [28] P. Singh and R. Agrawal, "A customer centric best connected channel model for heterogeneous and iot networks," *Journal of Organizational and End User Computing*, vol. 30, no. 4, pp. 32–50, 2018.
- [29] A. C. Eckert-Gallup, C. J. Sallaberry, A. R. Dallman, and V. S. Neary, "Application of principal component analysis (PCA) and improved joint probability distributions to the inverse first-order reliability method (I-FORM) for predicting extreme sea states," *Ocean Engineering*, vol. 112, pp. 307–319, 2016.
- [30] K. W. Tan and C. B. Weng, "Evaluation of water quality and benthic macroinvertebrates fauna relationship using principal component analysis (PCA): a case study of cameron highlands Malaysia," *Environmental Management and Sustainable Development*, vol. 5, no. 1, Article ID 187, 2016.