Hindawi
Scientific Programming
Volume 2022, Article ID 8995398, 11 pages
https://doi.org/10.1155/2022/8995398

Research Article

An Improved BP Deep Neural Network Multimedia Used in Oral English Training

Lihua Huang

School of Foreign Languages, Anqing Normal University, Anqing 246133, China

Correspondence should be addressed to Lihua Huang; huanglh@aqnu.edu.cn

Received 13 December 2021; Revised 27 December 2021; Accepted 5 January 2022; Published 31 January 2022

Academic Editor: Tongguang Ni

In order to improve the effect of spoken English training, this paper combines multimedia information technology to reform the teaching of spoken English training, and integrates BP neural network English into spoken English training. Moreover, this paper combines the actual needs of spoken English training and the teaching framework of the multimedia system to construct the data set, clean up the data set, and implement the word vector representation of students and professionals. In addition, this paper constructs the entire system framework of the spoken English resource recommendation algorithm based on the graph convolutional neural network, and combines the BP deep neural network algorithm to construct the spoken English training system. Finally, this paper designs an experiment to evaluate the effect of this system. The experimental research results show that the multimedia based on the BP deep neural network proposed in this paper has a good effect in the application research of spoken English training, and can effectively promote the effect of spoken English training of students.

1. Introduction

Oral expression ability, as an important English language skill, promotes the formation of students' English pragmatic competence, and is also an important manifestation of the function of English communication tools. Therefore, it has attracted great attention from teachers and students [1]. In particular, in the current college English classroom, teachers increase the intensity of the students' oral expression training, and hope to seize the golden period of language ability formation at the university stage to lay the foundation for the development of students' oral expression ability [2]. However, many teachers find that students lack enthusiasm for oral training in English classrooms, and their participation is generally low, which leads to unsatisfactory oral training effects, and the students' oral level is not improved. In view of this situation, teachers need to take a variety of measures to explore the reasons for the students' problems in spoken English training, design oral training strategies suitable for students, and improve students' spoken English level [3].

With the continuous development of the times, in order to cultivate more useful talents for the society, we also deeply realize that the teaching activities of teachers must closely follow the pace of the development of the times, so as to provide more effective guidance for students. English teaching can be divided into four teaching sections: listening, speaking, reading and writing. With the continuous development of the times, we have to admit that the ability of students to "speak" is becoming more and more important. However, as far as the current college spoken English teaching is concerned, there are still some problems that affect the further improvement of students' oral ability. Therefore, this paper makes a simple analysis on it.

For university students, there is also a factor that affects the improvement of students' oral English in the learning of oral English, that is, the students' oral foundation is relatively weak. Spoken language learning, especially for college students, does not start from scratch. Therefore, students will be affected by their existing oral language skills in the process of oral learning. Students generally do not dare to speak, inaccurate pronunciation and other phenomena. Students can use English for flexible communication requires a long-term practice process, and the first step is for students to have the courage to “speak” and boldly communicate in English. Only when students have the courage to speak, can they truly go out of spoken English learning. first step. However, college students generally do not dare to “speak”. Students are afraid of inaccurate pronunciation and fear of saying wrong. This kind of fear of students has seriously affected the improvement of students’ oral English ability. Oral English learning is different from the learning of other English knowledge sections. Only in a certain environment can students truly achieve the purpose of oral practice, and only in a certain environment can students’ oral ability be truly improved. As far as the current oral learning of university students is concerned, there is no specific environment for students to practice a lot of oral English. Students’ oral practice is only limited to classroom conversations or some English activities organized by the school. In fact, this is far from enough for students’ oral practice.

For improving the current situation of college students’ spoken English, it is crucial to pay great attention to spoken English. Teachers should pay attention to spoken English teaching, and students should pay attention to spoken English learning. First of all, teachers need to pay attention to spoken English teaching. This requires teachers to change the traditional English teaching concept and take the new teaching concept as the guide. Moreover, teachers should not only pay attention to teaching students in other parts of English, but also pay attention to oral teaching to students, and be able to introduce effective spoken English teaching methods into classroom teaching to improve the effectiveness of spoken English classroom teaching. In addition, teachers should make students understand the importance of spoken English teaching, and spend part of their English learning time on spoken English learning.

This paper combines the BP deep neural network algorithm to construct an spoken English training system to improve the intelligent effect of spoken English training, and evaluates the performance of the system through experimental research.

2. Related work

Literature [4] believes that programmed language is “a part of automatic or semi-automatic memory reserve, even through the screening mechanism of dividing these words, they are still automatically stored”. Literature [5] believes that spoken language is a series of words that people habitually use together. Literature [6] proposed that spoken language is a way for the sharing of words in natural texts to reach a statistically significant number: from the perspective of language programming, Wray’s term “programming language” has a far-reaching impact. Formula language refers to: it is a combination of continuous or discontinuous words or other elements, is orderly and prefabricated: it is stored and retrieved as a whole in the process of language use, and is not affected by the generation and analysis of grammar Influence. Literature [7] proves that word chunks are cognitively processed as a single vocabulary unit. Moreover, word blocks are learning materials suitable for human cognition.

Literature [8] pointed out that there are a large number of lexical sequences ylJ (sequences) in English, and a large number of prefabricated sentences of this kind exist, which explains to us why the language of second language learners can be “as fluent as native speakers.” And the answer to “choose like native speakers”. Literature [9] pointed out that many of the native language users’ knowledge are spoken languages that are not required and impossible to analyze outside of grammar. He explained that communicative competence is actually the ability to master a large number of assembled structures, formulaic formulas and a set of rules, and be able to make necessary adjustments according to the context. Literature [10] believes that language is not composed of traditional grammar and vocabulary, but composed of multi-word prefabricated spoken language. Literature [11] pointed out that the degree of language fluency is determined by the amount of programmed language stored in the brain’s memory, rather than by the knowledge of grammar. Literature [12] pointed out: learning spoken language is more important than learning grammar, language knowledge is spoken knowledge to a considerable extent, and grammar is second. Literature [13] believes that spoken language is the basis of second language learning, and the teaching of common phrases (stock phraSe) in language teaching should be as important as vocabulary and grammar teaching.

Literature [14] studied the differences between college students and middle school students in the use of spoken English, and pointed out the wrong ways of using spoken English by students. Chinese English learners seldom use multiple words and idioms in oral English, and there are differences in the frequency of use of phrase structure and sentence structure. There is a significant difference between the oral language use of domestic speakers and English speakers. There is no significant difference between college students and middle school students, while the oral language use of college students is significantly higher than that of middle school students. Spoken language errors in oral communication include mother tongue transfer errors, acquisition errors, and redundancy Errors, mixed use errors, etc.

Literature [15] believes that: in the aspect of language learning, oral input can reasonably configure the cognitive resources of information processing; in the aspect of oral language, prefabricated oral language can help to communicate meaning first and coherently; in the aspect of reading, oral language is effective Prefabrication can process low-end information more quickly, thereby speeding up reading; in terms of writing, spoken language has a good effect on preserving short-term memory, and can vividly express one’s thoughts in words. Literature [16] believes that the richer the storage of spoken language, the more proficient in calling the spoken language, which has a positive impact on all aspects of language learning, listening, speaking, reading and writing. Literature [17] pointed out: oral language also has its own limitations. In language learning, oral input cannot be one-sidedly emphasized, while grammar and other indispensable aspects of language learning are ignored. It is necessary to ensure that learners can stay in language learning for a long time. In order to make progress, only by correctly handling the direct relationship between oral input and grammar, and making up for each other's shortcomings, can the various abilities of the language develop in a co-ordinated manner. Literature [18] did research on whether the use of spoken language affects spoken English and writing. It is believed that the use of spoken language can affect students' English spoken and written scores, and the impact is greater. Compared with grammar knowledge, spoken language is in the structure of English knowledge. Obviously occupies a more important position. The oral English level of students largely depends on their ability to call the oral English in their minds. Students who can call as many oral English as possible more accurately tend to have higher oral English. The literature [19] found that the higher the student's writing score, the more proficient and richer the oral English.

3. Construction of spoken English resource recommendation model based on BP deep neural network

This article analogizes students as users, majors as projects, and students' class relationships as users' social networks, and establishes a professional recommendation model based on students' social interactions. The characteristics of students in social networks will be spread, which affects students' professional choices. Therefore, the influence propagation layer is introduced into the social recommendation network, and then a better loss function is selected to improve the accuracy of the recommendation result. The spoken language resource recommendation model proposed in this paper is divided into three parts. The first part is student modeling, which is divided into two models. The first model is to understand students' preference for majors through the interaction of the student-professional diagram. The second model extracts student characteristics from social networks, and integrates the characteristics of student friends into each student node in the social graph to model students. Combining the information of student-professional space and social space, the potential characteristics of students are intuitively acquired. The second part is professional modeling. In the interaction of the student-professional diagram, the evaluation of students of the same major is aggregated to reflect the potential characteristics of the major. The third part is the integration of the two models of students and spoken English. Through the full convolutional layer learning model parameters, the matching degree calculation between the student and the recommended spoken English is performed.

$U = \{u_1, u_2, \ldots, u_n\}$ and $V = \{v_1, v_2, \ldots, v_m\}$ are the sets of students and spoken English respectively, where n is the number of students and m is the number of spoken English. We assume that $R \in R^{n \times m}$ is the student-English speaking scoring matrix, which also known as the student-English speaking graph. If candidate u_i scores spoken English v_j, then r_{ij} is non-zero, otherwise we use 0 to indicate the score of spoken English v_j by candidate u_i, that is, $r_{ij} = 0.\Gamma = \{\langle u_i, v_j \rangle r_{ij} = 0\}$ is the set corresponding to the spoken English that student u_i has not evaluated. $T \in R^{n \times n}$ represents the social graph of students, where $T_{ij} = 1$ means that u_i and u_j are closely related.

The following is a detailed introduction to the spoken language resource recommendation algorithm based on the BP deep neural network.

The purpose of building a student model is to extract the potential characteristics of students. The student model is divided into two parts: the student-English speaking picture and the social network. The two parts are described separately below.

Because a student's score on a spoken English can tell the students' preference for the spoken English, the score of the spoken English and the characteristics of each spoken English can help students model. The student spoken English graph model in this paper is based on this to integrate spoken English features and spoken English evaluation to extract potential characteristics of students.

$$x_{ia} = g_v\left(\left[q_a \oplus e_r\right]\right). \tag{1}$$

Among them, x_{ia} is the multi-layer perceptron (MLP) of students and spoken English, x_{ia} is the interactive representation of each spoken English and score, q_a is the feature vector of spoken English, e_r is the feature quantity of five evaluation levels, and $\oplus$ represents the connection of two vectors. The calculation of formula (1) combines the scores corresponding to each spoken English and each spoken English to perform feature extraction. Since there are many spoken English that each student participates in the evaluation, it is necessary to aggregate all the spoken English that the students participate in the evaluation, and we introduce an aggregation function.

$$h_i^I = \sigma\left(W \cdot \text{aggretions}\left(\left\{x_{ia}, \forall a \in C(i)\right\}\right) + b\right). \tag{2}$$

Among them, $C(i)$ is the spoken English evaluated by student $u(i)$, W and b are the weights and biases of the neural network, and σ is the Relu activation function. The most common aggregation function is to directly aggregate the multiple spoken English evaluated by each student, that is, formula (3).

$$h_i^l = \sigma\left(W \cdot \left\{\sum_{a \in C(i)} \alpha_i x_{ia}\right\} + b\right), \tag{3}$$

$$\alpha_i = \frac{1}{|C(i)|}. \tag{4}$$

However, because the weight of each spoken English is different for students, the average aggregation cannot be directly calculated. In order to make up for the lack of average aggregation, the attention mechanism is introduced.

$$h_i^l = \sigma\left(W \cdot \left\{\sum_{acc(i)} \alpha_{ia} x_{ta}\right\} + b\right), \tag{5}$$

α_{ia} is the weight of the attention mechanism.

$$\alpha_{ia}^{*} = w_2^T \cdot \sigma\left(W_1 \cdot \left[x_{ia} \oplus p_i\right] + b_1\right) + b_2, \tag{6}$$

$$\alpha_{ia} = \frac{\exp\left(\alpha_{ia}^{*}\right)}{\sum_{a\in C(i)} \exp\left(\alpha_{ia}^{*}\right)}. \tag{7}$$

Among them, p^i is the feature vector of student u_i, and the interaction of spoken English and scoring is connected with the feature vector of the student. w_2^T, b_1, b_2 and W_1 are weights and biases, and the Softmax function is use to normalize the above attention weights to get the final attention weights. It can be understood as the contribution of interaction to the latent factors of the student-English speaking space. After the calculation of formula (1)-(7), we extracted the potential features of each student in the structure of the student-English speaking graph. The potential characteristics of students will also be reflected in the social relationships of students. Therefore, we introduce the student social network model to extract potential features.

This paper compares students as users, and oral English as projects, and compares students' class relationships to users' social networks, and establishes a model of oral English recommendation based on students' social interactions. In view of the fact that the social relationship of students will affect the students' choice of spoken English, the influence communication layer is introduced into the social recommendation network. The following is an improved social network model, as shown in Figure 1.

The characteristics of each student are divided into potential characteristics and obvious characteristics. Obvious characteristics can be reflected by students' evaluation of spoken English, and latent characteristics are reflected by students' social relationships. Therefore, we set up latent features, perform free embedding on the latent features of students, and continuously update the latent features through social relationship fusion.

$$\text{embedding} = nn \cdot \text{Embedding}(n, d). \tag{8}$$

The input of the fusion layer is the apparent feature vector and latent feature vector of each student. The two feature vectors are fused, and the fully connected layer is used for feature extraction. At this time, the output h_i^0 is the entire feature of the student.

$$h_i^0 = g\left(W^0 \times \left[p_i, m_i\right]\right). \tag{9}$$

Among them, W^0 is the weight matrix, $g(.)$ is a fully connected layer, p_i is the potential feature of the student, and m_i is the obvious feature.

The influence propagation layer is shown in Figure 2. Among them, the green node represents the current node u_i, the gray node is the first-order neighbor node of the current node, and the red node is the current node is the neighbor node of the first-order neighbor node.

Due to the change of time, the influence in social networks is spreading. For the current node u_i, h_i^0 is the input of the fusion layer, and h_i^1 is the feature obtained by averaging the first-order neighbor nodes (all gray nodes) of the current node u_i. At this point, the current node has the characteristics of the four neighbor nodes on the blue background. h_i^2 is the first-order neighbor node of the first-order neighbor (the result of the fusion of the red node, and it can be seen that the node u_i at this time has the characteristics of all the nodes in the green background) is also the feature of average fusion. Therefore, if h is the student's feature vector after impact diffusion, $S(i)$ is the set of neighborhood nodes of student u_i, and h_b^{k-1} is the neighborhood node feature of student u_i after $k-1$ social networking. Through the feature fusion of the neighboring nodes after $k-1$ times of social network influence, the characteristics of student u_i are obtained as:

$$h_i^k = \text{pool}\left(h_b^{k-1} b \in S_i\right), \tag{10}$$

$$\text{pool}\left(h_b^{k-1}\right) = \sum_{b\in S(i)} \frac{1}{|S(i)|} h_b^{k-1}. \tag{11}$$

The social network and the student-English speaking graph provide the characteristics of the student from different angles. Therefore, in order to extract all the potential characteristics of the student, it is necessary to combine the student-English speaking graph model and the social network model.

$$c_1 = \left[h_i^l \oplus h_i^k\right], \tag{12}$$

$$c_2 = \sigma\left(W_2 \cdot c \frac{n!}{r!(n-r)!} \frac{n!}{r!(n-r)!_1} + b_2\right), \tag{13}$$

$$h_i = \sigma\left(W_l \cdot c_{l-1} + b_l\right). \tag{14}$$

Among them, the output h_i is a combination of the potential features of the student-English speaking graph model and the social network model, and the student features are extracted through the full convolutional layer.

Just like the student model, in order to learn the latent features of spoken English, the spoken English model aggregates the evaluations of students who select the same spoken English to reflect the latent characteristics of spoken English.

$$f_{jt} = g_u\left(\left[p_l \oplus e_r\right]\right). \tag{15}$$

Among them, f_{jt} is the potential feature of spoken English after connecting the feature vector of student t and the feature vector corresponding to the score, and extracted by the full convolution layer. Since there is more than one student participating in the evaluation of spoken English, it is necessary to aggregate all the students who evaluate the spoken English, and for this purpose, the student's aggregation function is designed.

$$z_j = \sigma\left(W \cdot \text{aggre}_{\text{users}}\left(\left\{f_{jt}, \forall t \in B(j)\right\}\right) + b\right). \tag{16}$$

Considering that each student has different meanings for extracting the potential features of spoken English, the fusion function needs to introduce an attention mechanism to give students who choose the same spoken English different weights.

$$\mu_{jt}^{*} = w_{2}^{T} \cdot \sigma\left(W_{1} \cdot \left[f_{jt} \oplus q_{j}\right] + b_{1}\right) + b_{2}, \tag{17}$$

$$\mu_{jt} = \frac{\exp\left(\mu_{jt}^{*}\right)}{\sum_{t \in B(j)} \exp\left(\mu_{jt}^{*}\right)}, \tag{18}$$

$$z_{j} = \sigma\left(W \cdot \left\{ \sum_{EB(j)} \mu_{jt} f_{jt} \right\} + b \right). \tag{19}$$

Among them, z_j is the potential feature of spoken English extracted by the aggregation function after introducing the attention mechanism.

The recommendation algorithm of spoken English is simply to match the characteristics of the students with the characteristics of the spoken English. Proved that the fully connected layer can fit arbitrary functions. Therefore, we choose to use multiple full convolutional layers to learn the similarity function between students and spoken English features.

$$g_{1} = \lfloor h_{i} \oplus z_{j} \rfloor. \tag{20}$$

Among them, h_i is the latent feature of the student, z_j is the latent feature of spoken English, and $\oplus$ means connecting the two vectors.

$$g_{2} = \sigma\left(W_{2} \cdot g_{1} + b_{2}\right), \tag{21}$$

$$g_{l-1} = \sigma\left(W_{l} \cdot g_{l-1} + b_{l}\right), \tag{22}$$

$$r_{ij}' = w^{T} \cdot g_{l-1}. \tag{23}$$

Among them, l represents the number of hidden layers, and r_{ij}' represents the matching degree of student u_i to spoken English v_j.

The above process is to perform feature extraction of multiple fully connected layers (MLP) on the result of the splicing. Finally, the obtained feature matrix is mapped to a one-dimensional vector, that is, the predicted value of each spoken english.

The loss function of BP deep neural network based on social recommendation is:

$$\text{Loss} = \frac{1}{2|O|} \sum_{i,j \in O} \left(r_{ij}' - r_{ij}\right)^{2}. \tag{24}$$

Among them, O is a set of tuples (i, j), where tuples represent student u_i and graded spoken English v_j. Therefore, this loss function is to calculate the loss for all scoring spoken English. For the spoken English without scoring, the loss is not considered, and the L2 loss function. When the predicted value is significantly different from the original value, the loss penalty is too large. When the gradient descent method is used to solve the problem, the gradient is large, which may cause the gradient to explode.

Therefore, we smoothed the L2 loss function and chose the Smooth Loss loss function. The function is as follows:

$$\text{Smooth Loss} = \begin{cases} \lambda\left(x - x'\right)^{2}, & \left|x - x'\right| < 1, \\ \left|x - x'\right| - \lambda, & x - x' < -1,\ x - x' > 1. \end{cases} \tag{25}$$

It can be seen that when $x - x'$ is less than 1, there is only one parameter λ difference from the original second order. However, when the absolute value of $x - x'$ is large, compared with the original L2 loss function, the Smooth loss function reduces the loss penalty and becomes a first-order loss. Therefore, using the Smooth Loss function can reduce the proportion of outliers and alleviate the model's deviation to the outliers compared to using the L2 loss function.

In order to prove the basis for the selection of the loss function, we conducted experiments on common loss functions and compared the commonly used loss functions in the recommendation system. For the MSE loss function, Exp exponential loss function, and Smooth loss function with different parameters, according to experiments, it is found that when the parameter in the Smooth loss function is 0.6, the MAE calculation results, RMSE calculation results, and accuracy are the best.

$$\text{Smooth Loss} = \begin{cases} \lambda\left(x - x'\right)^{2}, & \left|x - x'\right| < 1, \\ \left|x - x'\right| - \lambda, & x - x' < -1,\ x - x' > 1, \end{cases} \tag{26}$$

$$\text{MSE}(y, \hat{y}) = \frac{1}{n} \sum_{i=1}^{n} \left(y_{i} - \hat{y}_{i}\right)^{2}, \tag{27}$$

$$\text{loss}\left(y, y'\right) = \exp\left(-y y'\right). \tag{28}$$

4. Research on application of multimedia in spoken English training based on BP deep neural network

The voice part of the audio can be detected and extracted to enter the text-to-speech alignment process. For the sake of continuity, this paper first analyzes and discusses the forced alignment algorithm based on Viterbi decoding. On this basis, this paper proposes an improved fault-tolerant alignment algorithm based on extended matching network. Moreover, this paper presents a detection method for insertion, deletion, and replacement errors at the word and phrase level, and a larger-scale dynamic alignment algorithm for sentence level. Figure 3 shows the basic flow of text-to-sentence matching of a single sentence.

This paper uses the matching network shown in Figure 4 to extend the search network of the traditional forced alignment algorithm, so as to achieve a fault-tolerant mechanism at the word and phrase level. SIL stands for silent model, OOV stands for garbage model.

This article further discusses how to provide a solution based on the idea of dynamic programming for the alignment of continuous corpus and incomplete matching corpus on a large scale (sentence level). The overall steps of the algorithm are shown in Figure 5.

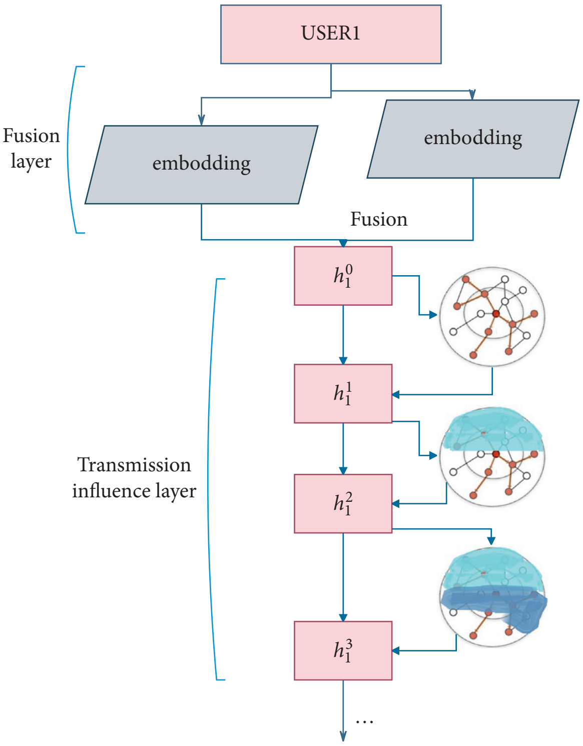

Figure 1: Improved social network model

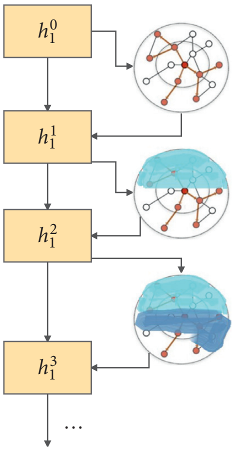

Figure 2: Influence propagation layer

Due to insertion, deletion, and replacement errors, not all models in the matching process can have corresponding nodes, especially the context-dependent HMM model (TRIPHONE) located at the word boundary. In the TRIPHONE model corresponding to syllable K as shown in Figure 6, since the word YOU is skipped during matching, the corresponding TRIPHONE model changes from <NG-K-Y> to <NG-K-V>. Therefore, in addition to comparing the corresponding nodes, it is also necessary to compare the word boundaries to prevent the expansion of the search subspace caused by some models in the second stage of matching because they cannot find the corresponding nodes.

The system front-end is based on the standard algorithm for feature extraction of the distributed speech recognition front-end of the European Telecommunications Standards Agency as shown in Figure 7. In order to ensure sufficient accuracy of the front-end features, a single-frame amplitude normalization technology is applied to make full use of the processing length of the processor to make the accuracy of fixed-point operations meet the requirements of the back-end recognition engine.

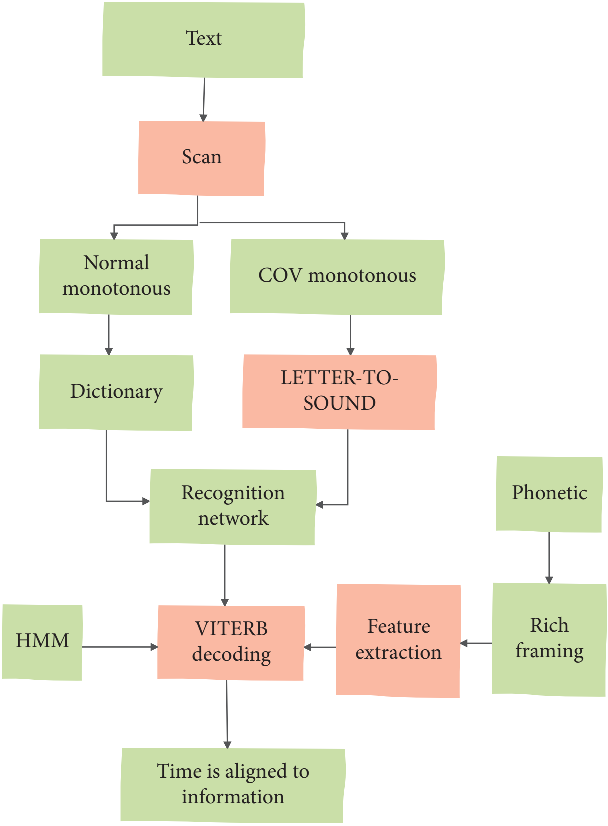

Figure 3: Flow chart of text-language matching.

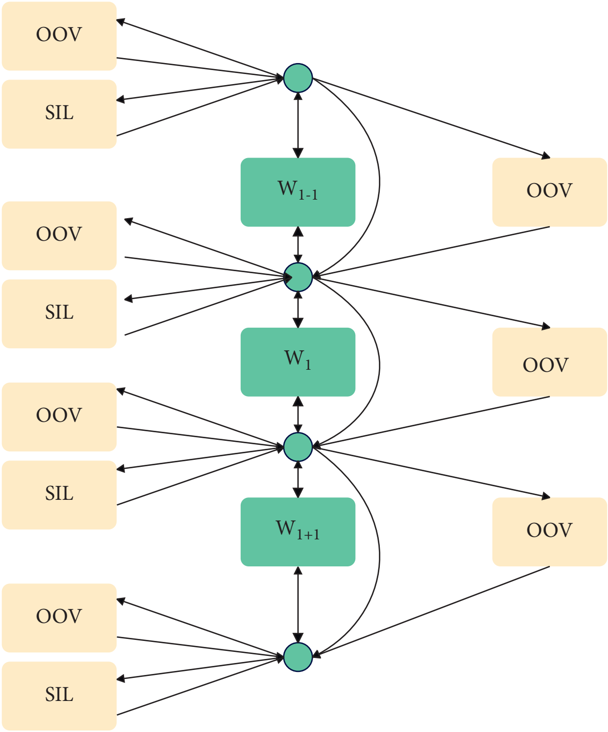

Figure 4: Extended network of fault-tolerant alignment algorithm.

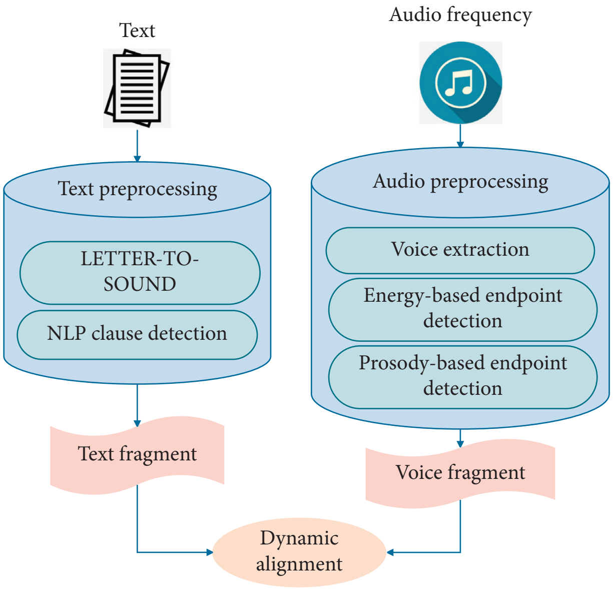

Figure 5: The flow of the alignment algorithm for a large number of continuous non-exact matching corpora.

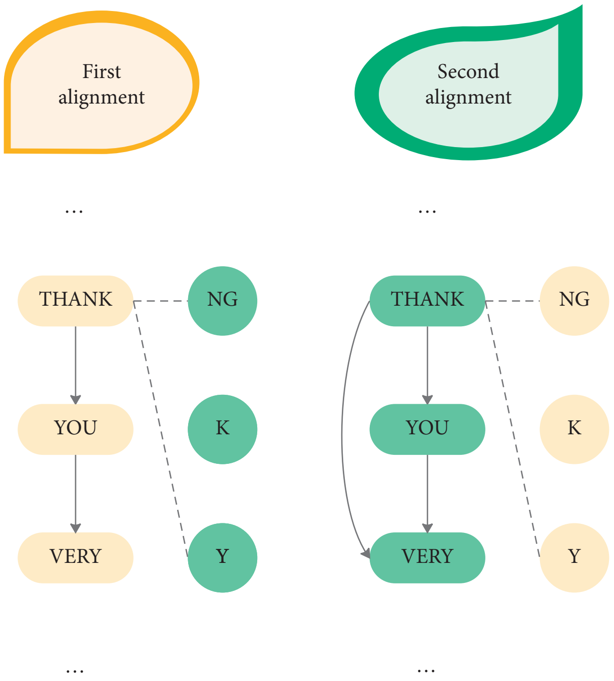

Figure 6: TRIPHONE changes in word boundaries caused by errors.

Figure 8 shows a schematic diagram of the voice processing flow of the system.

The modules of the speech recognition system include five parts: feature value extraction, phoneme recognition, phoneme association, pronunciation evaluation, and error detection. The speech recognition module is shown in Figure 9:

The external function realizes the visual window. The example sentence library view is played according to the specified example sentence or learning strategy. The information view displays sentences and phonetic symbols, displays monophony scores and corrections, displays sentence prosody scores and corrections, and plays corrections. The user management interface displays the user's transcripts, study files, and analyzes easy-to-mispronounce phonemes and common errors. The system can recognize English learners with a strong Chinese accent.

This paper conducts training research on the model of this paper, and the results are shown in Figure 10 below. It can be seen from Figure 10 that the student's speech

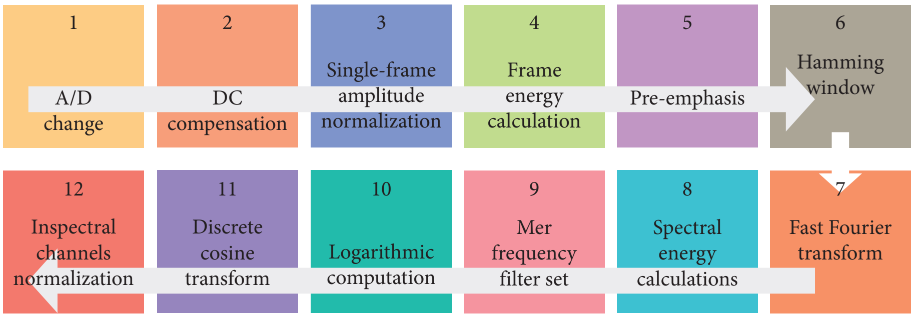

Figure 7: Front-end processing flow of text matching engine.

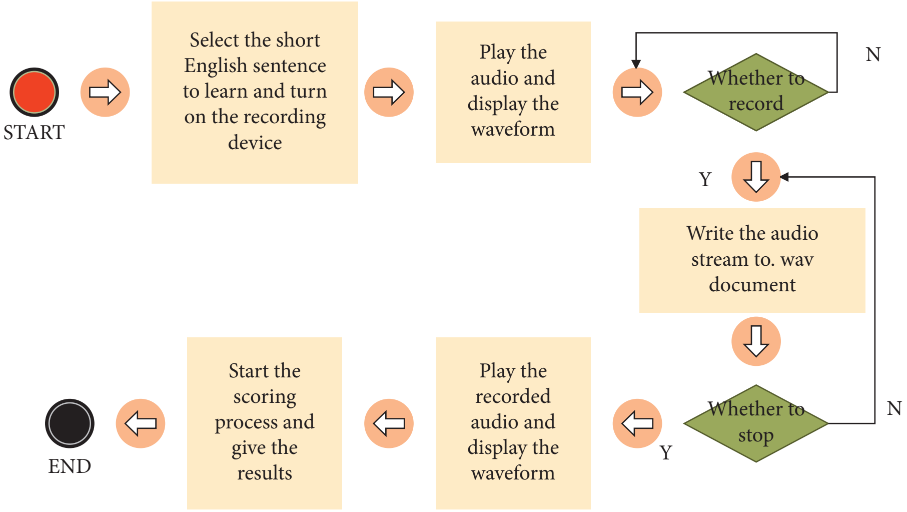
START
Select the short English sentence to learn and turn on the recording device
Play the audio and display the waveform
Whether to record
N
Y
Write the audio stream to. wav document
Whether to stop
N
Y
Play the recorded audio and display the waveform
Start the scoring process and give the results
END

Figure 8: Schematic diagram of system processing.

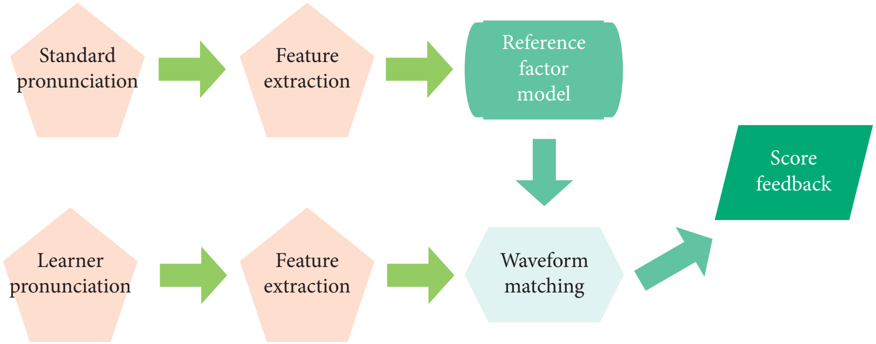

Figure 9: Schematic diagram of the speech recognition module.

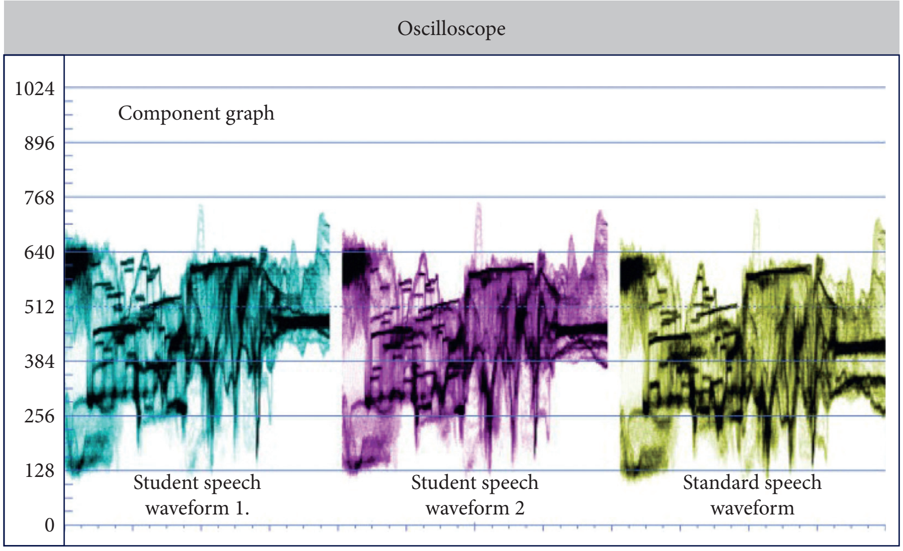

Figure 10: Comparison of speech waveforms in spoken English training.

Table 1: Test data of spoken English training.

Number	Training effect	Number	Training effect	Number	Training effect
1	79.95	13	79.23	25	80.87
2	86.17	14	91.71	26	89.86
3	86.92	15	89.46	27	88.14
4	86.96	16	79.26	28	84.91
5	81.40	17	89.62	29	88.07
6	88.60	18	83.92	30	86.97
7	86.94	19	83.53	31	93.84
8	90.22	20	81.77	32	80.81
9	84.05	21	89.08	33	88.86
10	81.89	22	83.90	34	80.04
11	87.04	23	87.65	35	84.74
12	91.89	24	90.66	36	82.48

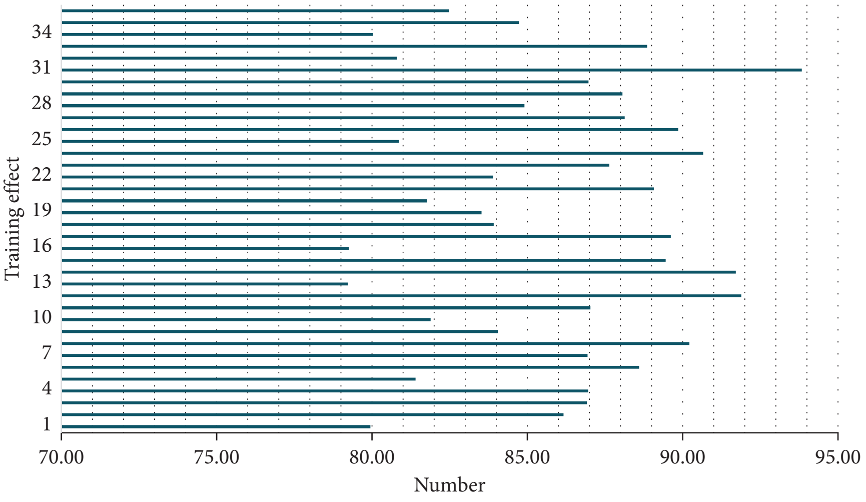

Figure 11: Evaluation of the training effect of spoken English.

waveform is similar to the standard speech waveform, which means that the student's spoken English h training is very effective and the student's spoken English is very standard.

On the basis of the above analysis, the effect of the spoken English training of this system is evaluated, and the results shown in Table 1 and Figure 11 below are obtained.

It can be seen from the above research that the multimedia based on BP deep neural network proposed in this paper has a good effect in the application research of spoken English training, and can effectively promote the effect of spoken English training of students.

5. Conclusion

In recent years, with the continuous advancement of new curriculum reforms, it is required to pay great attention to spoken English teaching. Although oral teaching has been improved to a certain extent compared with the previous teaching, it is still difficult to achieve the purpose of effectively training students' oral skills, which seriously affects the improvement of students' oral communication skills. The current intelligent spoken English learning system needs to provide functions such as recognition of the user's pronunciation, comparison with expert pronunciation, and error correction. The basis of all these functions is speech recognition. The accuracy of speech recognition and the robustness of the recognition algorithm will directly determine the overall performance of the learning system. This paper combines the BP deep neural network algorithm to construct the spoken English training system, improves the intelligent effect of spoken English training, and evaluates the performance of the system through experimental research. The experimental research results show that the multimedia based on the BP deep neural network proposed in this paper has a good effect in the application research of spoken English training, and can effectively promote the effect of spoken English training of students.

Data Availability

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

Conflicts of Interest

The author declares no competing interests.

Acknowledgments

This work was sponsored in part by the year of 2020 Quality Engineering Project of Anhui Province (2020jxtd155).

References

[1] 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.

[2] 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.

[3] 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.

[4] 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.

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

[6] N. Guzachchova, "Zoom technology as an effective tool for distance learning in teaching English to medical students," Бюллетень науки и Црактики, vol. 6, no. 5, pp. 457–460, 2020.

[7] 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.

[8] 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.

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

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

[11] A. B. Rinekso and A. B. Muslim, "Synchronous online discussion: teaching English in higher education amidst the covid-19 pandemic," *JEE*, vol. 5, no. 2, pp. 155–162, 2020.

[12] 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.

[13] 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.

[14] 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.

[15] A. S. N. Agung, "Current challenges in teaching English in least-developed region in Indonesia," *SOSHUM : Jurnal Sosial dan Humaniora*, vol. 9, no. 3, pp. 266–271, 2019.

[16] M. A. Saydaliyeva, E. B. Atamirzayeva, and F. X. Dadaboyeva, "Modern methods of teaching English in Namangan state university," *International Journal on Integrated Education*, vol. 3, no. 1, pp. 8-9, 2020.

[17] L. B. Kelly, "Preservice teachers' developing conceptions of teaching English learners," *Tesol Quarterly*, vol. 52, no. 1, pp. 110–136, 2018.

[18] A. Coşkun, "The application of lesson study in teaching English as a foreign language," *İnönü Üniversitesi Eğitim Fakültesi Dergisi*, vol. 18, no. 1, pp. 151–162, 2017.

[19] N. Sadat-Tehrani, "Teaching English stress: a case study," *TESOL Journal*, vol. 8, no. 4, pp. 943–968, 2017.