Computer Fraud and Security
ISSN (online): 1873-7056

Analysis of English Oral Speech Information Based on Intelligent Algorithms

Xiangyu Guo
College of Foreign Languages, Zhengzhou University of Technology, Zhengzhou, Henan, China

Abstract.

In order to improve the intelligent evaluation effect of spoken English, this paper analyzes the traditional spoken English test algorithm, and proposes an improved spoken English scoring algorithm based on the needs of intelligent English evaluation.Moreover, this paper proposes a framework of multi-index fusion pronunciation quality evaluation technology for reading questions, and constructs a functional module for English oral proficiency evaluation based on the needs of speech feature recognition. At the same time, this paper uses deep learning-based speech recognition technology to automatically recognize the tester's pronunciation, and uses the dual-threshold endpoint detection method of short-term energy and short-term average zero-crossing rate to divide the pronunciation of the speech sentence into syllables.In addition, this paper recognizes speech features by inputting the candidates' voice, and compares the recognition features with the standard database to score it. Finally, this paper analyzes the system performance by means of experimental research. The research results show that the system constructed in this paper has a certain effect.

Keywords: Speech analysis; spoken English; proficiency test; speech recognition

1. INTRODUCTION

The oral English test is one of the most representative tests for language learning and testing applications, and it is a veritable large-scale test. Due to the large number of learners, traditional manual teaching and testing methods can no longer meet the current teaching needs. At the same time, domestic second language learning has gradually expanded from English to many other languages, such as Japanese, French, and German.In the field of education in China, from elementary school to middle school to university, English is one of the focus of students' learning. Moreover, the arrangement of course learning fully reflects the importance of the education department. At the same time, it also reflects the urgent need of people to learn foreign languages in today's society, so as to improve the level of international communication among the people.The ultimate goal of learning a foreign language is to go hand in hand with comprehensive language ability. Therefore, "listening", "speaking", "reading" and "writing" should be used as both the content of learning and the means of learning. At the same time, in addition to the traditional written test, the important summative assessment of the college entrance examination should also include oral and listening tests to comprehensively examine the comprehensive skills of students' language use [1].In the current upsurge of "Internet +" and artificial intelligence, computer technology is more and more widely used in the field of education. Especially, in the grading of large-scale examinations, the scoring work is heavy, so it is imperative to use the computer to automatically score.The so-called computer automatic scoring refers to the scoring of the speech or text of the candidates that have been entered by the computer system. Although the research in this field in China has become increasingly mature, there is still a certain gap compared with related foreign research.At present, the computer-aided evaluation system [2] has become one of the research hotspots of education combined with artificial intelligence. Moreover, using computers to replace traditional teacher scoring will be a major change in the education sector.

Compared with manual scoring, the advantages of computer automatic scoring are as follows. First of all, the computer can repeat the operation without feeling tired, and can avoid the scoring error caused by the fatigue factor of the scorer. Moreover, it can continuously perform scoring work for 24 hours, which is more efficient.In addition, automatic scoring uses multiple computers to score according to exactly the same scoring standard, and pre-processing the voice signal before the scoring procedure, so that the evaluation result is not easily affected by the recording volume and the noise of the surrounding environment, and it is relatively objective and stable.However, manual scoring is highly subjective and emotions are prone to ups and downs. The review process may be affected by personal preference or the scoring environment. From this perspective, automatic scoring is more accurate than manual scoring [3].In addition, from the perspective of users of the automatic scoring system, for students, computer automatic scoring is more objective and efficient. For teachers, it saves time,

reduces labor intensity, and at the same time helps improve work efficiency. For the education management department, the computerized automatic scoring system can provide a large amount of background statistical data, which can be analyzed from different levels, such as the difficulty, discrimination, consistency, and reliability of the test questions, as well as the distribution of the overall knowledge level of the candidates.All in all, the oral automatic scoring system has many advantages such as objectivity, immediacy, speed and economy.

This paper combines the intelligent speech analysis technology to build the oral English proficiency test system, and verifies the system performance through system simulation research, so as to further promote the development of oral English proficiency test system.

2.RELATED WORK

With the rapid development of computer technology, people began to try to combine natural language processing and automatic speech recognition technology to develop automatic speech scoring system [4]. In the field of oral automatic scoring, Ordinate automatic scoring system and Speech RaterTM automatic scoring system are the two most representative oral automatic scoring systems [5].Orient is developed by Pearson for versant oral English test, and Speech RaterTM is developed by ETS for TOEFL oral test. The system scoring types constructed in literature [7] include listening vocabulary, repetition accuracy, pronunciation, reading fluency and repetition fluency.The advantages of the system constructed in [8] are easy to measure, high reliability, and high accuracy of score prediction. However, the system does not consider various components of communicative competence, such as "social communication skills", "cognitive function" and "world knowledge" [8].

The system in the literature [9] is a scoring system that provides test preparation practice for the TOEFL i BT test. Its purpose is to allow candidates to check whether their oral skills have reached the level that can participate in the TOEFL i BT. The scoring principle of the system in the literature [10] is to construct a multiple regression scoring model by extracting 29 meaningful features from the speech input. After a series of pruning, these characteristics can be used as representatives of English communication ability on the technical level, such as fluency, vocabulary, grammar and pronunciation.Typical characteristics include the average silent length of words, the average silent duration, the average number of words per second, and so on. In terms of reliability, the reliability of the system constructed in the literature [11] reached 0.60-0.70, and the complete agreement rate and adjacent agreement rate between human score and machine score ranged from 95% to 99%.The results of the literature [12] showed that the correlation coefficient between Speech Rater and manual scoring was only 0.57, while the research in literature [13] showed that the correlation coefficient between the system and manual scoring was as high as 0.97.The correlation between the automatic scoring system of the literature [14] and manual scoring can be accepted in the case of mock exams. However, compared with TOEFL-i BT (TPO) test score correlation (0.74), there is still a big gap. From the perspective of the correlation value, the quality of the attraction score constructed in the literature [15] does not seem to be satisfactory. This may be related to the selection of the sample. The reason is that the scores used by the researchers are mostly concentrated in the 2 to 3 bins, which lack difference.

Nowadays, many scholars have carried out the research and practice of various oral test methods to improve the teaching of oral English, and the specific test methods include direct oral test, semi-direct recording oral test and computer-assisted oral test [16].Some researchers believe that the traditional direct oral test, which is a face-to-face test with the examinee and the examiner, is the most close to real-life situation of oral communication. At the same time, most domestic colleges and universities still use this traditional method to evaluate students' oral communication skills. Because this method is similar in form to real communication, the examiner can also directly participate in the examinee's real communication activities, examine the content of the students' language expression, and observe the examinee's facial expressions and body language. All in all, it has the characteristics of "high surface validity, great flexibility, and strong pertinence" [17].However, due to factors such as human resources and the different criteria of the examiner, the score reliability of the direct oral test is usually low [18]. In the semi-direct recording oral test, because candidates do not need to face the psychological pressure from the examiner, the performance of their oral English will not be affected by the examiner's language level or emotions, and will pay more attention to the correctness of the speech output throughout the test. The semi-direct oral test is to score the answer recordings, so the test process and the scoring process can be carried out separately, so that more language samples can be collected, and it has the characteristics of high scoring reliability and

strong operability. However, the disadvantages of this form are also more prominent because of the lack of interaction in real oral communication [19].Literature [20] proposes that the computer-based oral English test should be able to detect and evaluate the pronunciation, grammar and content of the test taker like a traditional oral test, and should not be limited to the assessment of "follow-the-reading" imitation ability. The literature [21] compared the advantages and disadvantages of direct oral test and computer oral test, and showed that the latter not only saves time and effort, but also is easy to operate. At the same time, it also has many advantages such as easy storage of test corpus and relatively objective and fair scoring. The conclusions drawn are based on three college oral English computer test experiments, so it provides effective suggestions for colleges and universities to choose the form of oral English test.

3.MULTI-INDEX FUSION OF ENGLISH PRONUNCIATION QUALITY EVALUATION

The object-oriented research work of this paper is the pronunciation quality evaluation of reading aloud questions. In addition to speech emotion, it also includes pronunciation intonation, rhythm, intonation and speech speed. The framework flow chart of the multi-index fusion evaluation technology of pronunciation quality of reading aloud questions proposed in this paper is shown in Figure 1.

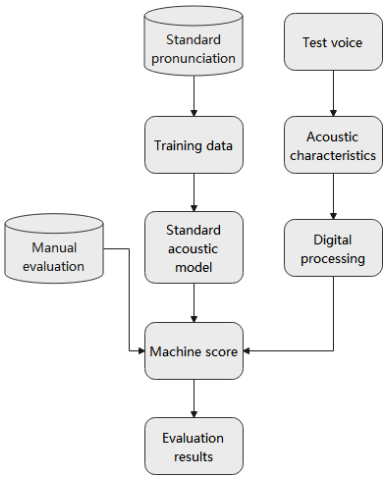

Figure 1 Flow chart of machine evaluation for reading aloud questions

Intonation evaluation is an evaluation of the tester's pronunciation content, which examines whether the tester has read the content of a given topic completely and accurately, and whether the pronunciation content is clear and audible.We assume that the pronunciation sequence of the tester's speech is $O=\{o_1,o_2,\cdots,o_n\}$, and the text sequence of a given title is $W=\{w_1,w_2,\cdots,w_m\}$. The intonation evaluation measures the degree of matching between the pronunciation sequence and the text sequence.The computer evaluation process of pronunciation accuracy includes two steps. One is to record the standard pronunciation sequence by the sound recorder, and the other is to calculate the similarity between the tester's pronunciation sequence and the standard pronunciation sequence. MFCC is a hearing mechanism based on the human ear, which simulates the response of the human ear to the voice signal, and can reflect the human perception of the voice. It is an important voice signal characteristic parameter. Moreover, it is also the most common feature parameter of speech signal used in current automatic speech recognition technology.In addition, the MFCC parameters can better digitize the pronunciation sequence, so the intonation evaluation also adopts the calculation method based on the MFCC parameters[22].

This article uses deep learning-based speech recognition technology to automatically recognize the tester's pronunciation, determine whether the content is a given topic sentence, and then calculate the similarity between the test speech and the standard speech (see the following formula). We assume that X is the test voice data, and Y is the standard voice data.

$$\mathrm{cov} = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \quad (1)$$

By fusing the results of the above two steps, the tester's pronunciation intonation is evaluated. The evaluation process is shown in Figure 2:

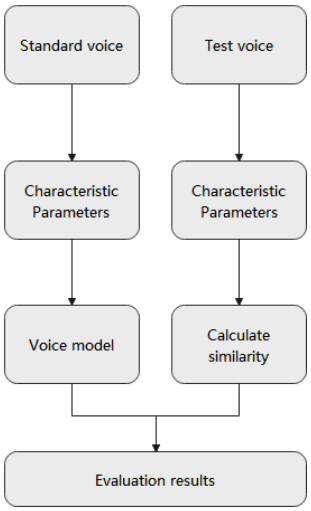

Figure 2 The evaluation process of pronunciation intonation

The final evaluation result Score is mapped from the recognition result *result* of the DBN speech model and the correlation coefficient *cov* of the standard speech and the test speech:

$$Score_1 = f(result, cov) \quad (2)$$

English is a typical stress-timed language. When native English speakers read English sentences, the sentence stress will appear rhythmically. The better the pronunciation, the more isochronous the interval between sentence stresses. Chinese is a syllable timed language. When reading Chinese sentences aloud, there is a phenomenon of isochronism between syllables and syllables. It is precisely because of the rhythm difference between English and Chinese that it is difficult for Chinese English learners to grasp the stress and rhythm when learning English.Therefore, rhythm evaluation is an important indicator of English pronunciation quality evaluation, and the Pairwise Variability Index (PVI) is a method that can effectively evaluate the rhythm level of English learners. This paper uses an improved PVI algorithm to compare the difference in stress distribution between the test speech and the standard speech as an evaluation measure of the rhythm level[23].

An important feature of stress in a speech signal is loudness, so when extracting accent features, the energy intensity in the original speech signal is mainly considered. The method in this paper first preprocesses and divides the speech signal into frames, and normalizes the test speech to the same duration as the standard speech, and extracts the energy intensity of the speech signal for each frame of speech data:

$$E_n = \sum_{m=-\infty}^{\infty}\left[s(n)\omega(n-m)^2\right] \quad (3)$$

According to the threshold, the stress unit in the sentence is divided into:

The stress threshold is:

$$T_y = \left(\max(sig\_in) + \min(sig\_in)\right)/2.5 \quad (4)$$

The non-stress threshold is:

$$T_n = \left(\max(sig\_in) + \min(sig\_in)\right)/10 \quad (5)$$

Through the PVI algorithm, the difference in the distribution of k accent durations between the test speech s_1 and the standard speech s_2 is calculated. At the same time, the tail t of s_1 and s_2 is considered. The calculation result is the basis for the rhythm evaluation of the test speech:

$$Score_2 = dPVI = 100 \times \sum_{k=1}^{m-1} |s_{1k} - s_{2k}| + |(s_{1t} - s_{2t})| / len \quad (6)$$

The rhythm evaluation process is shown in Figure 3:

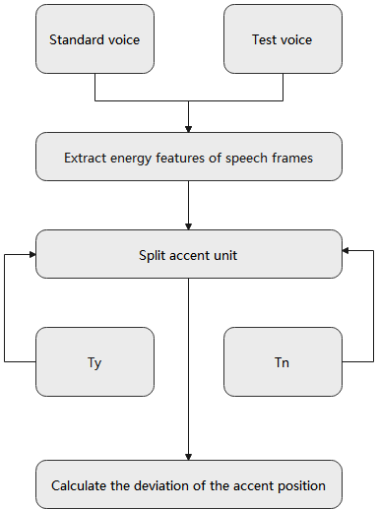

Figure 3 The rhythm evaluation process

Intonation refers to the tone of speech, which is the preparation and change of the tone of voice in a sentence. No language in the world is spoken in a single tone. Take English as an example. English has five basic intonations: rising $(\quad)$, falling $(\quad)$, rising-falling $(\wedge)$, falling-rising $(\vee)$, and flat $(\rightarrow)$. These tones will match different English sentence patterns, and intonation is a phonetic feature closely related to semantics.

Intonation is expressed through changes in pitch, and pitch is determined by the level of frequency. The two are in direct proportion: the frequency is high, the pitch is "high", and vice versa, the pitch is "low". Therefore, for the evaluation of intonation, it is first necessary to extract the fundamental frequency parameters in the speech signal, and then compare the fundamental frequency characteristics of the test speech and the standard speech[24].

This article uses Auto Correlation Function (ACF) to extract the pitch of each frame of data in English sentences, as shown in the following formula:

$$acf(\tau) = \sum_{i=0}^{n-1-\tau} s(i)s(i+\tau) \quad (7)$$

T is the time delay in the unit of sampling point, n is the data length of a speech frame, and $s(i)$ is the speech windowing function processing. By setting the pitch threshold, the abnormal speech frame is excluded, and then the median filter is used to smooth the pitch value of the entire sentence, and finally the intonation curve of the speech sentence is obtained.The test voice and the reference voice are processed by the above algorithm respectively, and the intonation sequence graph data s_1 and s_2 of the two voices are obtained respectively. The dynamic time warping algorithm (DTW) is used to calculate the similarity between the two as a measure of intonation evaluation.

$$Score_3 = dtw(s_1, s_2) \quad (8)$$

The process of intonation evaluation is shown in Figure 4:

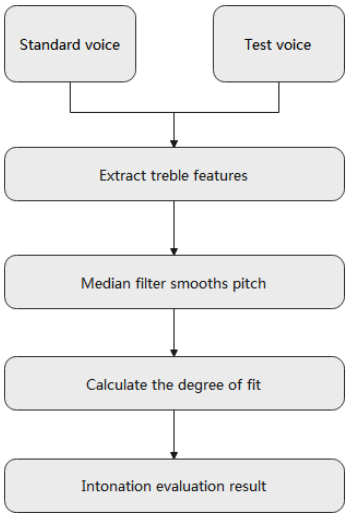

Figure 4 Intonation evaluation process

Speaking speed is the speed of the speaker. In the pronunciation evaluation of English reading aloud, the tester is required to imitate the speed of speaking according to the reference speech. Therefore, it is necessary to measure the speaking rate ratio of the test voice and the reference voice.Speaking speed can be reflected by the total length of the effective speech frame of the sentence. This paper first uses the dual-threshold endpoint detection method of short-term energy and short-term average zero-crossing rate to classify the pronunciation syllables of the speech sentence, and then sums the duration of each syllable to obtain the effective pronunciation duration *len* of the sentence.

The test speech duration is len_1, and the standard speech duration is len_2. The ratio of the durations is compared as a measure of speech rate evaluation:

$$Score_4 = \begin{cases} \dfrac{len_1}{len_2}, (len_1 \le len_2) \\ \dfrac{len_2}{len_1}, (len_1 \ge len_2) \end{cases} \quad (9)$$

The speech rate evaluation process is shown in Figure 5:

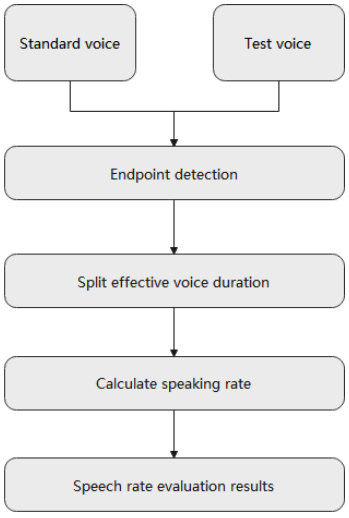

Figure 5 Speech rate evaluation process

The developed speech emotion pronunciation quality evaluation method uses the confidence probability output value of SVM corresponding to emotion classification as the metric value $Score_5$ of speech emotion evaluation.

The traditional multi-index fusion method mainly adopts the method of multiple linear regression, but the traditional method does not have the evaluation of each sub-index, so there is a lack of effective tools. The initial measurement value of the index is rated and mapped, and the linearity of each index cannot be accurately obtained. Polynomial weight; Second, the rating scores in this article are grade points, including 4 grades. The evaluation results of each indicator in this article are continuous variables. If the linear regression method is used to predict the scores, the result is a continuous variable value, which is inconsistent with the manual rating method . Based on the shortcomings of the above multiple linear regression methods, this article will adopt the method of machine classification and treat the machine rating problem as a classification problem.

In the corpus collection process of this article, the human scorer did not score the sentences by item, but adopted the overall scoring method. Although the overall scoring method pays attention to the integrity, in the scoring process, the scorer will still refer to the judgment of each indicator. The scorers who adopt the overall scoring method model pay different attention to the indicators. For example, the intonation-rhythm-oriented rater first pays attention to the intonation. If the tester thinks that the intonation is good, then he pays attention to the rhythm index, and then he cares about other indexes; if the intonation is incorrect, the rater will first evaluate the tester as a whole. Can only get grades below C. The scoring process described above (see Figure 6(a)) is similar to a top-down binary tree structure search, which can be abstracted as a human scorer to obtain an overall rating through an if-then model. The Decision Tree is aA classification method with a tree structure. In the classification problem, it represents the process of classifying instances based on features, which can be considered as a set of if-then rules. Figure 6(b) is a simple example. The decision tree uses loan applicants The basic information to determine whether to pass the application. The comparison of the two figures shows that the decision tree classification method is similar to the overall scoring method of the manual rater. Both use feature judgment to find the next path to continue matching rules.

Therefore, this paper adopts the method of decision tree to simulate the overall scoring mode of manual raters to construct a computer comprehensive evaluation model of multi-index pronunciation quality.

(a)Intonation decision tree (b)Emotional decision tree

Figure 6 Comparison of the overall manual scoring process and decision tree rules

In this paper, five scoring values $Score_i$ are used as the judgment index $\left[Score_1, Score_2, Score_3, Score_4, Score_5\right]$ of the decision tree, and the manual rating is used as the classification result to construct the decision tree.

This paper uses the ID3 algorithm to build a decision tree. ID3 selects features based on information gain criteria and uses a top-down greedy strategy to build a decision tree.The information gain $g\left(D, A\right)$ of the rating index $Score_i$ is obtained by subtracting the empirical entropy $H\left(D\right)$ of the rating data set D and the empirical conditional entropy $H\left(D|A\right)$ of the index $Score_i$ on the data set D.In this article, there are 4 kinds of rating results C, so $k \in \left[1,4\right]$ and j are the value interval n of $Score_i$.

$$H\left(D\right)=-\sum_{k=1}^{4}\frac{\left|C_k\right|}{D}\log_2\frac{\left|C_k\right|}{D} \quad (10)$$

$$H\left(D|Score_i\right)=\sum_{j=1}^{n}\frac{|D_j|}{D}H\left(D_j\right)=-\sum_{j=1}^{n}\frac{|D_j|}{D}\sum_{k=1}^{4}\frac{|D_{jk}|}{|D_i|}\log_2\frac{|D_{jk}|}{|D_i|} \quad (11)$$

$$g\left(D,A\right)=H\left(D\right)-H\left(D|A\right) \quad (12)$$

The algorithm process is as follows:

step1.if all rating indicators are processed, return; else go to step2;

step2. The index $Score_i$ with the largest information gain is calculated, and the index is used as the node for judgment; if index $Score_i$ can judge the rating separately, return; else go to step3step2;

step3. For each possible value interval v of the index $Score_i$, the following operations are performed:

i. The samples of all index $Score_i$ whose value is j are regarded as a subset D_j of D;

ii. The index set $Score_T = Score - Score_i$ is generated;

iii. The rating data set D_j and the index set $Score_T$ are used as input, and the ID3 algorithm is executed recursively;

It can be seen from the generation rules of the decision tree that the closer to the root node is, the index and interval with the larger information gain value. From the evaluation decision tree constructed in this article, it can be seen intuitively that intonation (node 1) is the most important indicator affecting comprehensive evaluation, which is consistent with our general experience: that is, the pronunciation content is complete and accurate, which is the basic requirement for pronunciation quality evaluation. . The second layer of nodes includes emotion (node 2) and rhythm indicators (node 3), indicating that these two indicators are also important classification criteria. As mentioned above, English rhythm patterns are different from Chinese rhythm patterns. Therefore, it is difficult for Chinese English learners to learn English rhythm patterns, and it is also an important reference for distinguishing the quality of pronunciation. The presence of emotional indicators in the second layer of nodes means that the evaluation results of speech emotional pronunciation quality developed in this article are very important for the prediction of the comprehensive evaluation results of the tester's pronunciation quality. On the one hand, it reflects that in the corpus of this article, the voice emotion index is an important indicator of pronunciation quality evaluation. On the other hand, it also proves that the voice emotion pronunciation quality evaluation method developed in this article is effective.

4.ENGLISH SPEAKING PROFICIENCY TEST SYSTEM BASED ON INTELLIGENT SPEECH ANALYSIS

On the basis of the above research and analysis of the spoken English recognition algorithm, an English spoken proficiency test system based on intelligent speech analysis is constructed. The functional modules and system architecture of the system are discussed below.

The overall structure of the intelligent scoring system for spoken Englishproficiency test is shown in Figure 7.

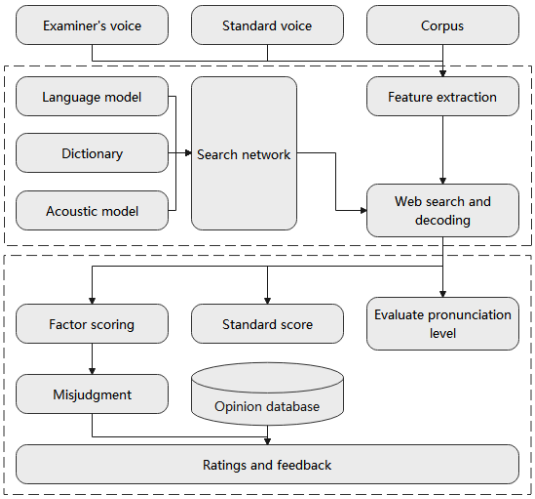

Figure 7 The overall structure of the intelligent scoring system

The related action flow from inputting speech to obtaining English spoken pronunciation evaluation is connected in series and the program flow of speech recognition module is shown in Figure 8.

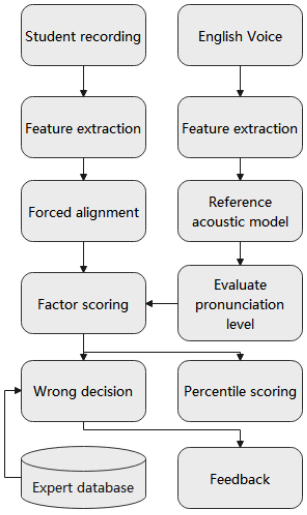

Figure 8 Program flowchart of speech recognition module

Through the above analysis, aspoken English proficiency test system based on intelligent speech analysis is constructed, and then the performance of the system needs to be verified through experimental research.The system constructed in this paper is mainly applied to the oral proficiency test. It recognizes the speech features of the examinee's speech, compares the recognition features with the standard database, and scores the speech on this basis.Therefore, the operation of the system includes two stages. The first stage is speech recognition, and the second stage is system scoring. The two stages are tested separately.First, this paper conducts a test of spoken English speech recognition. A total of 80 sets of data are collected to study the accuracy of speech recognition. The results are shown in Table 1 and Figure 9.

Table 1 Statistical table of the accuracy of speech recognition

Num	Spoken language recognition	Num	Spoken language recognition	Num	Spoken language recognition
1	90.2	28	79.8	55	90.0
2	90.2	29	87.7	56	80.6
3	79.3	30	87.8	57	83.6
4	87.5	31	92.1	58	88.2
5	81.7	32	83.9	59	84.7

6	81.0	33	83.8	60	79.7
7	87.6	34	82.3	61	90.5
8	88.5	35	87.0	62	79.6
9	80.4	36	81.6	63	87.8
10	89.9	37	81.2	64	89.9
11	83.8	38	85.5	65	85.1
12	81.3	39	84.1	66	84.3
13	91.0	40	92.8	67	81.8
14	87.0	41	89.4	68	85.9
15	87.4	42	82.7	69	92.2
16	86.5	43	79.5	70	79.7
17	90.2	44	87.4	71	87.9
18	80.9	45	90.1	72	80.8
19	91.0	46	82.3	73	80.1
20	87.6	47	85.2	74	83.6
21	92.4	48	81.4	75	90.2
22	90.0	49	88.6	76	88.5
23	90.0	50	80.0	77	90.7
24	91.5	51	89.9	78	84.4
25	83.7	52	89.0	79	79.7
26	81.9	53	91.0	80	82.0
27	80.9	54	80.5		

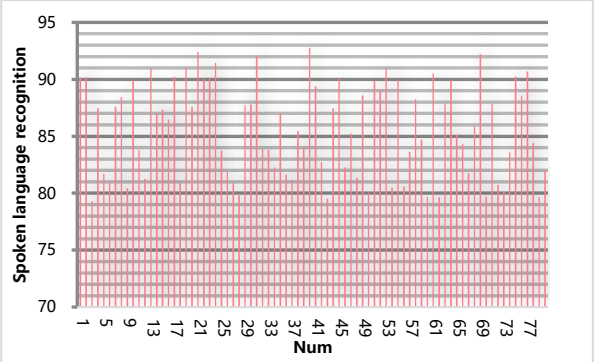

Figure 9 Statistical diagram of the accuracy of speech recognition

From the above experiments, it can be seen that the spoken English proficiency test system based on intelligent speech analysis constructed in this paper is effective. After that, this paper verifies the system's scoring effect on the spoken English level, and the results are shown in Table 2 and Figure 10.

Table 2 Statistical table of system scoring effect

Num	Rating effect	Num	Rating effect	Num	Rating effect
1	86.4	28	79.6	55	89.6
2	77.6	29	74.8	56	82.7
3	88.2	30	72.7	57	86.8
4	75.8	31	79.9	58	91.0
5	81.0	32	79.7	59	83.5
6	72.6	33	85.5	60	78.1
7	73.5	34	84.9	61	69.9
8	76.5	35	86.3	62	71.3
9	71.2	36	87.7	63	83.1

10	90.3	37	76.8	64	77.4
11	77.0	38	69.5	65	80.4
12	82.0	39	76.9	66	72.4
13	78.5	40	79.8	67	79.5
14	77.7	41	81.5	68	81.5
15	84.3	42	74.3	69	82.9
16	69.4	43	79.6	70	70.5
17	84.4	44	70.5	71	80.1
18	70.4	45	77.9	72	90.3
19	77.9	46	85.5	73	88.0
20	87.1	47	86.3	74	80.8
21	76.8	48	73.1	75	82.6
22	90.7	49	81.4	76	76.9
23	75.6	50	74.0	77	75.1
24	74.0	51	75.0	78	71.1
25	76.8	52	73.3	79	89.8
26	70.6	53	88.8	80	72.3
27	76.8	54	72.5		

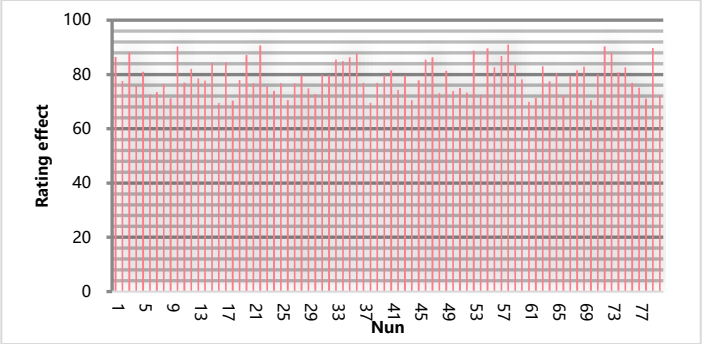

Figure 10 Statistical diagram of system scoring effect

From the experimental research, it can be seen that the spoken English proficiency test system based on intelligent speech analysis constructed in this paper is effective, and it can be used for auxiliary scoring in the spoken test.

5.CONCLUSION

The automatic scoring result of the English listening and speaking test is used as a part of the intelligent test. Whether it can judge the tester's oral level more accurately is related to whether it can replace the traditional manual scoring to maximize the efficiency of the computer. At the same time, it is also related to whether other application software in the field of automatic scoring can become an auxiliary tool to effectively improve the level of English learning.There are many factors that affect the system validity of the automatic scoring of spoken English, such as the scoring principle and the parameter settings of each analysis module. This article builds an English speaking proficiency test system based on intelligent speech analysis algorithms. Moreover, this paper uses the input of the examinee's speech to recognize the speech features, and compares the recognition features with the standard database to score the spoken language.Finally, this paper uses experimental research to test the performance of the English speaking proficiency test system. From the research results, we can see that the method proposed in this paper is feasible.

ACKNOWLEDGE:

The research in this paper was supported by Research and Application Project for Higher Education Teaching Reforms in Henan (NO.2021SJGLX292), by Teacher Education Curriculum Reform and Research Project of

Henan Province (NO.2024-JSJYZD-056) and also was supported by Teaching Reforms of Zhengzhou University of Technology (NO.ZGJG202403HAB).

REFERENCES

[1]. Rhodes, Richard. Aging effects on voice features used in forensic speaker comparison[J], international journal of speech language & the law, 2017, 24(2):177-199.
[2]. Ngoc Q. K. Duong, HienThanh Duong. A Review of Audio Features and Statistical Models Exploited for Voice Pattern Design[J], computer science, 2015, 03(2):36-39.
[3]. Sarria-Paja M , Senoussaoui M , Falk T H . The effects of whispered speech on state-of-the-art voice based biometrics systems[J], Canadian Conference on Electrical and Computer Engineering, 2015, 2015(1):1254-1259.
[4]. Leeman A , Mixdorff H , O'Reilly M , et al. Speaker-individuality in Fujisaki model f0 features: Implications for forensic voice comparison[J], International Journal of Speech Language and the Law, 2015, 21(2):343-370.
[5]. Hill A K , Rodrigo A. Cárdenas, Wheatley J R , et al. Are there vocal cues to human developmental stability? Relationships between facial fluctuating asymmetry and voice attractiveness[J], Evolution & Human Behavior, 2017, 38(2):249-258.
[6]. MarcinWoźniak, DawidPołap. Voice recognition through the use of Gabor transform and heuristic algorithm[J], Nephron Clinical Practice, 2017, 63(2):159-164.
[7]. Haderlein T , Michael Döllinger, VáclavMatoušek, et al. Objective voice and speech analysis of persons with chronic hoarseness by prosodic analysis of speech samples[J], LogopedicsPhoniatricsVocology, 2015, 41(3):106-116.
[8]. Liu S M , Chen J H . A multi-label classification based approach for sentiment classification[J], Expert Systems with Application, 2015, 42(3):1083-1093.
[9]. Shang L , Zhou Z , Liu X . Particle swarm optimization-based feature selection in sentiment classification[J], Soft Computing, 2016, 20(10):3821-3834.
[10]. Alejandro MoreoFerna´ndez, Esuli A , Sebastiani F . Distributional correspondence indexing for cross-lingual and cross-domain sentiment classification[J], Journal of Artificial Intelligence Research, 2016, 55:131-163.
[11]. Harer S , Kadam S . Sentiment Classification and Feature based Summarization of Movie Reviews in Mobile Environment[J], International Journal of Computer Applications, 2014, 100(1):30-35.
[12]. Zhang Z , Wang Z , Gan C , et al. A double auction scheme of resource allocation with social ties and sentiment classification for Device-to-Device communications[J], Computer networks, 2019, 155(MAY 22):62-71.
[13]. Li Y , Wang J , Wang S , et al. Local dense mixed region cutting + global rebalancing: a method for imbalanced text sentiment classification[J], International journal of machine learning and cybernetics, 2019, 10(7):1805-1820.
[14]. Phu V N , Dat N D , Ngoc Tran V T , et al. Fuzzy C-means for english sentiment classification in a distributed system[J], Applied Intelligence, 2017, 46(3):717-738.
[15]. Giacoumidis E, Matin A, Wei J, et al. Blind nonlinearity equalization by machine-learning-based clustering for single-and multichannel coherent optical OFDM[J]. Journal of Lightwave Technology, 2018, 36(3): 721-727.
[16]. Tsoi K K F, Chan N B, Yiu K K L, et al. Machine learning clustering for blood pressure variability applied to Systolic Blood Pressure Intervention Trial (SPRINT) and the Hong Kong Community Cohort[J]. Hypertension, 2020, 76(2): 569-576.
[17]. Li H, Kafka O L, Gao J, et al. Clustering discretization methods for generation of material performance databases in machine learning and design optimization[J]. Computational Mechanics, 2019, 64(2): 281-305.
[18]. Cheng L, Kovachki N B, Welborn M, et al. Regression clustering for improved accuracy and training costs with molecular-orbital-based machine learning[J]. Journal of Chemical Theory and Computation, 2019, 15(12): 6668-6677.

[19]. Mydhili S K, Periyanayagi S, Baskar S, et al. Machine learning based multi scale parallel K-means++ clustering for cloud assisted internet of things[J]. Peer-to-Peer Networking and Applications, 2020, 13(6): 2023-2035.
[20]. Sun X, Young J, Liu J H, et al. Prediction of pork loin quality using online computer vision system and artificial intelligence model[J], Meat science, 2018, 140(7): 72-77.
[21]. Nourani V, Baghanam A H, Adamowski J, et al. Applications of hybrid wavelet–artificial intelligence models in hydrology: a review[J], Journal of Hydrology, 2014, 514(9): 358-377.
[22]. Bui X N, Nguyen H, Choi Y, et al. prediction of slope failure in open-pit mines using a novel hybrid artificial intelligence model based on decision tree and evolution algorithm[J], Scientific reports, 2020, 10(1): 1-17.
[23]. Yaseen Z M, El-Shafie A, Jaafar O, et al. Artificial intelligence based models for stream-flow forecasting: 2000–2015[J], Journal of Hydrology, 2015, 530(7): 829-844.
[24]. Laird J E, Lebiere C, Rosenbloom P S. A standard model of the mind: Toward a common computational framework across artificial intelligence, cognitive science, neuroscience, and robotics[J], Ai Magazine, 2017, 38(4): 13-26.