

Retraction

Retracted: Study on Automatic Evaluation Method of Spoken English Based on Multimodal Discourse Analysis Theory

Security and Communication Networks

Received 26 December 2023; Accepted 26 December 2023; Published 29 December 2023

Copyright © 2023 Security and Communication Networks. 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.

This article has been retracted by Hindawi, as publisher, following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of systematic manipulation of the publication and peer-review process. We cannot, therefore, vouch for the reliability or integrity of this article.

Please note that this notice is intended solely to alert readers that the peer-review process of this article has been compromised.

Wiley and Hindawi regret that the usual quality checks did not identify these issues before publication and have since put additional measures in place to safeguard research integrity.

We wish to credit our Research Integrity and Research Publishing teams and anonymous and named external researchers and research integrity experts for contributing to this investigation.

The corresponding author, as the representative of all authors, has been given the opportunity to register their agreement or disagreement to this retraction. We have kept a record of any response received.

References

- [1] J. Ren, "Study on Automatic Evaluation Method of Spoken English Based on Multimodal Discourse Analysis Theory," *Security and Communication Networks*, vol. 2021, Article ID 1486575, 8 pages, 2021.

Research Article

Study on Automatic Evaluation Method of Spoken English Based on Multimodal Discourse Analysis Theory

Junlong Ren 

College of Preschool and Primary Education China West Normal University Sichuan, Nanchong 637000, China

Correspondence should be addressed to Junlong Ren; ren1988@st.btbu.edu.cn

Received 13 October 2021; Revised 26 October 2021; Accepted 18 November 2021; Published 27 December 2021

Academic Editor: Jian Su

Copyright © 2021 Junlong Ren. 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.

Aiming at the low confidence of traditional spoken English automatic evaluation methods, this study designs an automatic evaluation method of spoken English based on multimodal discourse analysis theory. This evaluation method uses sound sensors to collect spoken English pronunciation signals, decomposes the spoken English speech signals by multilayer wavelet feature scale transform, and carries out adaptive filter detection and spectrum analysis on spoken English speech signals according to the results of feature decomposition. Based on multimodal discourse analysis theory, this evaluation method can extract the automatic evaluation features of spoken English and automatically recognize the speech quality according to the results. The experimental results show that, compared with the control group, the designed evaluation method has obvious advantages in confidence evaluation and can solve the problem of low confidence of traditional oral automatic evaluation methods.

1. Introduction

Modern signal processing and automatic pattern recognition technology are used to distinguish the quality of oral English pronunciation. Combined with signal detection and speech signal feature extraction methods, oral English automatic evaluation is carried out to improve the objectivity and accuracy of oral English pronunciation quality evaluation. The research on the automatic evaluation method of oral English pronunciation quality is based on speech signal detection and feature extraction, using intelligent signal processing technology, combined with time-frequency feature analysis and spectral analysis of oral English pronunciation signals, so as to improve the automation and intelligence level of oral English pronunciation quality evaluation. The research on the optimization design method of the automatic evaluation system of oral English pronunciation quality has a good application value in the design of oral English teaching. The related research on the automatic evaluation system method of oral English pronunciation quality has attracted great attention. In the automatic evaluation system of oral English pronunciation quality, the oral English pronunciation signal is affected by the

disturbance and distortion of oral pronunciation channel, which leads to the poor accuracy of oral English pronunciation quality evaluation. The change of oral features produces speech attenuation and distortion, which leads to the decline of the accurate detection performance of the automatic evaluation system of oral English pronunciation quality. It is necessary to optimize the design of signal processing algorithm. Among the traditional methods, the research on oral English automatic evaluation methods mainly includes multiresolution feature detection method, wavelet analysis method, scale decomposition method, time-frequency analysis method, and fractional Fourier feature extraction method. Combined with artificial intelligence control and feature extraction, the performance of oral English pronunciation quality automatic evaluation system is improved, and some research results are obtained. Among them, the automatic evaluation model of oral English pronunciation quality based on neural network extracts the distributed features of the pronunciation signals of the automatic evaluation system of oral English pronunciation quality and combines the wavelet analysis method to carry out the blind separation processing of oral English pronunciation signals, so as to improve the ability of automatic

evaluation of oral English pronunciation quality. However, this method has high computational complexity and poor real-time performance. The time-frequency feature decomposition method is used for speech recognition of the automatic evaluation system of oral English pronunciation quality. The feature decomposition and correlation dimension feature registration of English pronunciation signals are combined with the wavelet multilayer reconstruction method for speech recognition and quality evaluation. The reliability of this method in pronunciation quality evaluation is not good.

Aiming at the disadvantages of this method, this paper proposes an automatic oral English evaluation method based on multimodality. The ability of spoken English is reflected in the auditory modal symbols that can be recognized by human auditory system through human vocal organs, that is, spoken language symbols. Therefore, auditory modal symbols are the most important modal symbols in college oral English classroom. However, the auditory modal symbols in modern oral English classroom include not only the auditory modal symbols issued by people, but also some other auditory modal symbols. Based on this, we can capture the multimodal symbols of oral English and extract the automatic evaluation features of oral English, so as to explore the multimodal symbols in college oral English teaching at the microlevel, so as to enlarge the details of oral English teaching, promote the long-term development of college oral English teaching, strive to solve the problems existing in the traditional oral English automatic evaluation methods, and fundamentally improve the confidence of oral English automatic evaluation.

2. Multimodal Technology

Multimodal, that is, multimodal biometrics, refers to the integration or fusion of two or more biometric technologies, making use of the unique advantages of its multiple biometric technology, combined with data fusion technology to make the authentication and identification process more accurate and safe [1, 2]. The main difference from the traditional single biometric method is that multimodal biometric technology can collect different biometrics (such as fingerprint, finger vein, face, and iris image) through an independent or integrated collector of multiple acquisition methods and identify and authenticate by analyzing and judging the eigenvalues of multiple biometric methods [1, 3]. Multimodal discourse refers to the use of hearing, vision, touch, and other senses to communicate through language, image, sound, action, and other means and symbolic resources. The traditional discourse analysis theory only analyzes the meaning of discourse from the perspective of language. However, as other symbol systems other than language also have an impact on meaning construction, this view has been unanimously recognized by scholars. In addition to language education symbols, image, sound, animation, smell, human expressions, and body movements are incorporated into the elements of meaning expression.

In the process of evolution, life gradually obtains five different perception channels: vision, hearing, smell, taste,

and touch. The acquisition of five perception channels leads to the following five communication modes: visual mode, auditory mode, tactile mode, olfactory mode, and taste mode. Multimodal research rose in the 1990s, then developed rapidly, and attracted more and more attention of domestic scholars. Stein (2000) first proposed multimodal teaching. He believes that all communicative activities in the classroom are multimodal. Due to the application of modern science and technology in all aspects of life, modern classrooms have let students and teachers get rid of the simple era of "one textbook, one chalk, and one blackboard." The use of computers, projectors, and PPT courseware has greatly enriched the classroom, and it is also the material condition for students to receive multimodal information. Scholars have different views on the definition of multimodality. For example, Kress and Van Leeuwen (2001) believe that "multimodality is the use of several symbolic modes, or the comprehensive use of several symbolic modes to strengthen the expression of the same meaning, or perform supplementary functions, or carry out hierarchical sorting." They have studied the relationship between modality and media for many times (1996, 2001). They believe that modality refers to the symbolic resources that synchronously realize the categories of discourse and communication, and modality can be realized through more than one production medium. They believe that media refers to the material resources used in the production of symbolic products and events, including the use of tools and materials, which provides a theoretical basis and research method for multimodal discourse analysis theory.

Multimodal discourse analysis theory was first proposed by O'Toole, Kress, and van Leeuwen. Its theoretical starting point is that language is a social symbol. It extends the function of language as a social symbol to other symbols except language and regards various symbols including language as independent and mutually friendly symbol resources. While analyzing language features, it emphasizes the role of visual, auditory, and behavioral symbolic modes such as image, color, sound, and action in discourse. At present, how the various modes interact and how to affect the construction of overall meaning is an important topic of multimodal discourse analysis theory [4, 5]. Among them, the study of grammatical rules of various modes restricts the further development of this field. Generally speaking, the academic community generally believes that the reasonable coordinated multimodal interaction will strengthen the meaning construction, while the unreasonable multimodal interaction will reduce the overall meaning. The application of reasonable and coordinated multimodal technology to the evaluation of oral English can maximize the meaning that the speaker wants to express and achieve the best effect through different expression methods of hearing, vision, and other senses. At the same time, we can also mobilize the senses of oral English speakers through the application of multimodal technology, mobilize the enthusiasm of learning oral English through rich emotional expression, and emphasize the cultivation of learners' diversified learning ability.

3. An Automatic Evaluation Method of Spoken English Based on Multimodality

In this paper, the design process of multimodal oral English automatic evaluation method is shown in Figure 1.

As shown in Figure 1, the detailed research content of the four-step main process in the figure is described below.

3.1. Analyzing Spoken English Pronunciation Signals. In order to realize the automatic evaluation of spoken English, the signal analysis method of spoken English pronunciation is used for evaluation and signal feature extraction. The extracted oral English pronunciation quality features are adaptively matched to realize oral English pronunciation signal recognition and association information mining. The multiwavelet decomposition method is used to decompose the features of spoken English pronunciation signals. According to the position of spoken English pronunciation, vowel category, and spectral characteristics, the spoken English pronunciation is distinguished to realize the automatic evaluation and optimal recognition of spoken English pronunciation [6–8]. According to the above principle analysis, the signal detection analysis is carried out, and the sound sensor is used to collect the spoken English pronunciation signal. Let the original input characteristic sequence of spoken English pronunciation signal be $x = [x(0), \dots, x(n-1)]$, where $x(n)$ is a discrete spoken English pronunciation signal of finite length, $0 \leq n \leq N-1$; then the discrete Fourier transform (DWT) of spoken English pronunciation characteristic sequence of x is defined as follows:

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

In the formula, $0 \leq k \leq N-1$ represents the length of spoken English pronunciation signal. After the signal $x(n)$ is transformed by discrete orthogonal wavelet transform, $X = \text{DFT}\{x\}$ is used to represent the DFT of finite time series x of oral English pronunciation characteristics, that is,

$$X = [x(0), \dots, x(N-1)]. \quad (2)$$

Through formula (2), the finite length oral English pronunciation signal is reconstructed, and the oral English pronunciation signal E_j with different resolutions $j=0, 1, \dots, M$ is reconstructed. For the integer N_0, N_1 intelligent speech input signal v_0 . The speech control and multimedia control technology are used to analyze the speech characteristics combined with the expert system analysis method. The input and output relationship Z^N of the automatic evaluation system of oral English pronunciation quality is as follows:

$$Z^N = g \cdot X^N + W^N. \quad (3)$$

In formula (3), g represents the source, voice transmission channel, and coding channel for automatic evaluation of oral English pronunciation quality. The adaptive symbol location method is used for automatic evaluation of oral English pronunciation quality. WN is the additive noise

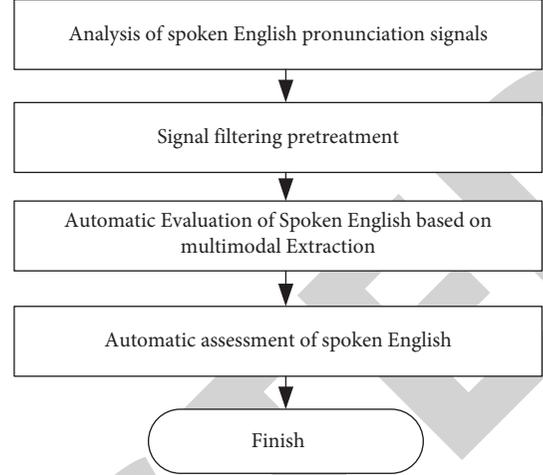


FIGURE 1: Automatic oral English evaluation process based on multimodality.

of oral English pronunciation signal on the transmission channel. N is a linear equalization cancellation code combined with wavelet entropy feature extraction for speech quality recognition. Obtain the satisfaction relationship of the linear relationship of speech feature extraction:

$$H(P_e^E) + P_e^E \log |S^N| \geq H(S^N | S_E^N) \geq NR - N\epsilon. \quad (4)$$

In formula (4), P_e^E is the sampling symbol rate for automatic detection of oral English pronunciation quality, and SN represents the acquisition data bit sequence of oral English pronunciation signal with length N [9, 10]. According to the analysis results of spoken English pronunciation signals, the automatic evaluation and analysis of English pronunciation quality is carried out.

3.2. Signal Filtering Preprocessing. The sound sensor is used to collect oral English pronunciation signals. The multilayer wavelet feature scale transform method is used for feature decomposition and filtering of spoken English pronunciation signals [11, 12]. The wavelet feature decomposition of spoken English pronunciation signal is carried out by using modulated pulse. Make the scale coefficient of signal decomposition $N_0^{(j)}$, frame the oral English pronunciation signal of the intelligent speech input system, Zn (N is the number of frames), the length of the input signal of the j -th filter bank, and $N_0^{(j)}$ and $N_0^{(j)}$ represent the pronunciation length of oral English pronunciation. Introduce pulse modulation variables R^N [13, 14]. The signal component phase rotation technology is used for linear coding, and the rotation moment of inertia of the output speech signal is obtained as follows:

$$\text{angle}(gX^N) = (\text{angle}(X^N) + \varphi g) \bmod (2\pi). \quad (5)$$

This paper reconstitutes the spoken English pronunciation signal and uses the multilayer wavelet feature scale transform method to denoise the spoken English pronunciation signal. The maximum likelihood detection of $\varphi g, R^N$, and WN parameters is carried out through arithmetic

coding, and the output positive correlation characteristic quantity of speech signal is obtained as follows:

$$gX^N = |g|R^N. \quad (6)$$

Simultaneous formula (6) gives

$$Z^N = |g|R^N + W^N. \quad (7)$$

Therefore, the adaptive filter detection and spectral analysis of oral English pronunciation signal are carried out according to the feature decomposition results, and the wavelet entropy feature of oral English pronunciation signal is extracted to improve the automatic detection ability of oral English pronunciation quality.

3.3. Feature Extraction of Spoken English Automatic Evaluation Based on Multimodality. On the basis of signal filtering preprocessing, the sound sensor is used to collect oral English pronunciation signals, extract oral English automatic evaluation features, and optimize the algorithm for oral English pronunciation quality automatic evaluation. This paper presents an automatic evaluation method of pronunciation quality based on spoken English speech signal detection and dynamic synchronous recognition [15, 16]. Based on multimodal discourse analysis theory, the features of oral English automatic evaluation are extracted. Among them, the spoken language symbols include three parts: meta spoken language symbols, E_1 language symbols, and electronic audio spoken language symbols played through multimedia. Among them, the first two parts are the oral language symbols issued by people in the natural state, while the third part is the oral language symbols issued by people in the natural state transformed by electronic technology, which is unnatural [17]. Therefore, the choice of sound size, tone, audio, and accent of these meta spoken languages should serve their meaning construction. The purpose of meaning construction of E_1 spoken language symbols is secondary (not meaning transmission in the real sense), and its main purpose includes organizing language symbols to express the established meaning and imitating and expressing the authentic English language. Therefore, these spoken language symbols pay more attention to the tone, mood, and other factors of the language. The electronic audio oral language played by multimedia appears as an auxiliary means in the classroom [18, 19]. Because of their unnaturalness, these language symbols cannot be changed once they are selected for use. Because it is the special function of the exercise template, the tone, audio, and accent of the oral language of the selected object should be accurately selected.

The main purpose of musical symbols is to provide teaching scenes or create a relaxed atmosphere for foreign language teaching. This requires that not any type of music can be used at will. Music for the purpose of providing scenes shall be properly played according to the needs of the scene. For music aimed at creating an atmosphere, relaxed and soothing music is appropriate. Once the music is selected, it cannot be changed, but its volume should be

controlled so that it does not affect other modal symbols in the classroom [20, 21]. Other sound modal symbols are divided into two cases: noise from outside the classroom and sound effects specially made to simulate a scene or attract students' attention. The noise outside the classroom interferes with the classroom. It is unpredictable and irresistible. For example, the whistle issued by roads and vehicles around the classroom belongs to this kind. Such sound symbols are not under the control of teachers [22]. It cannot be adjusted, but it has an impact on teachers' selection of other modal symbols. If this kind of noise is too serious, the mode originally intended to use oral sound symbols for meaning transmission has to be adjusted to the written language symbol mode of visual mode [23]. The purpose of specially creating sound effects to simulate a scene or attract students' attention is to provide teaching scenes for teaching or remind students' attention. The use of these sound symbols should be adjusted according to the needs of the scene as a feature of oral English automatic evaluation.

3.4. Automatic Evaluation of Spoken English. After extracting the features of oral English automatic evaluation based on multimodality, the functional component analysis and development environment analysis of oral English automatic evaluation method are carried out. Taking STC12C5A60S2 chip as the core control chip, combined with the embedded program loading design method, the hardware integration design of the system is carried out [24]. The network of the automatic evaluation system for oral English pronunciation quality includes three basic entity objects: target, observation node, sensing node and perceived field of view [25]. Information collection and data processing analysis of oral English pronunciation quality are carried out in the central information processing unit. The sensor node of the spoken English pronunciation quality automatic evaluation system has the functions of raw data acquisition, local information processing, and remote data transmission. Taking FPGA as the programmable logic chip as the core, the design of oral English pronunciation quality automatic evaluation system can realize the rapid perception and collection of oral English pronunciation information [26, 27]. In the ZigBee dynamic network structure design of the automatic evaluation system of oral English pronunciation quality, the network adapter encapsulates the bottom layer of the wireless sensor network (wireless sensor network infrastructure, wireless sensor operating system) of the automatic evaluation system of oral English pronunciation quality [28]. The bus transmission module and integrated control module of oral English automatic evaluation method are designed by using ZigBee network and IPv6 network [29]. The SCSI bus transmission control of oral English automatic evaluation method is realized by booting from external Serial EEPROM. The external differential SCSI bus interface can be connected to HP E1562 F or other SCSI bus hard disks. In the upper computer communication module, the dynamic test data of the automatic evaluation system of oral English pronunciation quality are sent to two sets of SCSI bus interface

controllers [30, 31]. The system includes information acquisition module, ad module, control module, host computer communication module, voice conversion module, receiving and sending module, bus control module, and man-machine interface module [32]. The information acquisition module of the automatic evaluation system of oral English pronunciation quality realizes the functions of voice signal acquisition and host computer communication [33, 34]. The data forwarding control of oral English pronunciation information is designed under ARM network control protocol. The spectral features of speech pronunciation signals are threshold, and the time-frequency analysis method is used to locate the signal source of each pronunciation feature in the k -interval, so as to obtain the pronunciation feature spectrum of the automatic evaluation system of oral English pronunciation quality. In the data transmission time of one frame, the output quantization information of the pronunciation feature vector $x(n) \in \mathbb{R}^N$ of each frame is described as $I(X_N, Z_N)$. The discrete Fourier transform (DWT) of tone feature sequence of spoken English pronunciation quality automatic evaluation system is defined as X_k . $0 \leq k \leq N - 1$ represents the length of speech signal. Extract the correlation feature of the pronunciation speech signal of the automatic evaluation system of oral English pronunciation quality. The empirical mode decomposition method is used for quantitative equalization of English pronunciation, and the domain model of English pronunciation signal is decomposed by wavelet transform. According to the wavelet entropy feature extraction results of spoken English pronunciation, the automatic recognition of pronunciation quality is C_s ; it is as follows:

$$C_s = \max [I(X_N; Z_N) - I(X_k; Z_N)]. \quad (8)$$

According to formula (8), the automatic recognition of spoken English is realized.

4. Experiment

This paper designs an automatic oral English evaluation method based on multimodality. In order to test the application effect of this method, the following experiments are carried out.

4.1. Experimental Preparation. In order to test the application performance of this method in oral English speech signal detection and automatic pronunciation quality evaluation, a system test is carried out. The experiment is designed by MATLAB simulation software. The system running MATLAB software is windows 10 system, equipped with i7 processor and 256g memory, which can fully meet the experimental requirements. The oral English pronunciation signals of 100 students in a school were collected as test samples, the sample test signal of spoken English speech signal adopts linear frequency modulation signal. The sampling time width of spoken English speech recognition is 1.2 s and the relative bandwidth is 0.4 dB. The acquisition frequency of oral pronunciation signals for different vocal cords is 1024 kHz, and the baseband signal frequency is

2 kHz–10 kHz. According to the above simulation environment and parameter settings, oral English speech signal detection and automatic evaluation of pronunciation quality are carried out, and the original signal acquisition results are obtained, as shown in Figure 2.

4.2. Experimental Index. The evaluation indexes set in the experiment are as follows:

- (1) Resolution of spoken English signal: the higher the resolution of oral English signal, the better the effect of oral English evaluation. On the contrary, the lower the resolution of oral English signal, the worse the effect of oral English evaluation.
- (2) Oral English confidence: the higher the confidence of oral English, the higher the credibility of oral English evaluation. On the contrary, the worse the confidence of oral English, the worse the credibility of oral English evaluation.
- (3) Oral English mastery rate: the higher the mastery rate of oral English, the better the teaching effect of oral English. On the contrary, the lower the mastery rate of oral English, the worse the teaching effect of oral English.
- (4) Time for oral English evaluation: the longer the time of oral English evaluation, the higher the efficiency of oral English evaluation. On the contrary, the shorter the time of oral English evaluation, the worse the efficiency of oral English evaluation.

4.3. Analysis of Experimental Results

4.3.1. Experimental Results and Analysis. According to the signal acquisition results in Figure 2, the automatic oral English evaluation method based on multimodal discourse analysis theory proposed in this paper is compared with the traditional oral English evaluation method, and the oral English pronunciation signal is decomposed. According to the results of feature decomposition, adaptive filter detection and spectrum analysis are carried out to realize signal detection and recognition. The test results are shown in Figure 3.

Through the analysis of Figure 3, it can be seen that the resolution of this method to detect and evaluate oral English signals is high, up to 1, while the traditional oral English evaluation method can only reach 0.75, indicating that this method has better effect in oral English pronunciation detection than the traditional method. In order to test the confidence of different automatic evaluation methods of oral English pronunciation quality, two different methods are still used to test the confidence. Table 1 shows the confidence test results of different methods.

It can be seen from Table 1 that, after four parallel tests, the confidence of the evaluation method designed in this paper is basically maintained above 88, up to 89.54. The confidence of traditional evaluation methods is always lower than 50, and the highest is 49.93. It can be concluded that the automatic oral English evaluation method based on

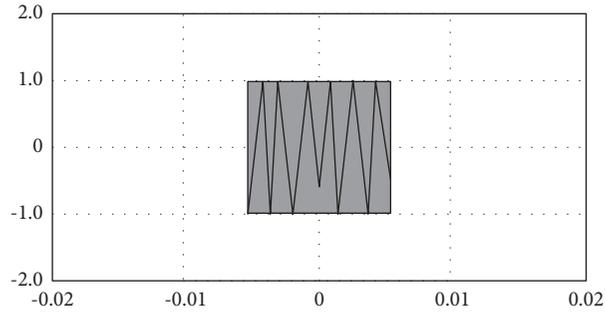


FIGURE 2: Signal acquisition results.

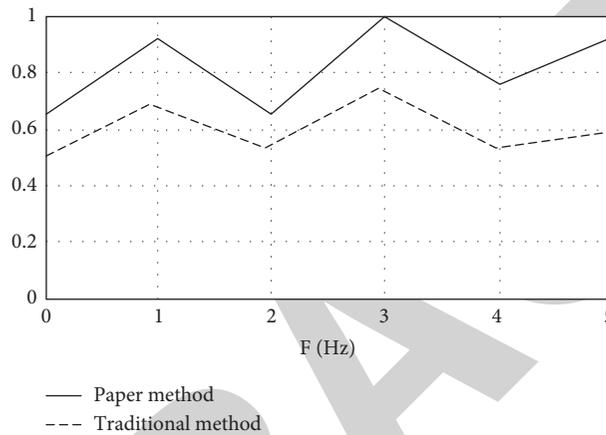


FIGURE 3: Evaluation results of different evaluation methods.

multimodal discourse analysis theory proposed in this paper is obviously better than the traditional methods in terms of evaluation confidence, and its evaluation results have a certain reliability.

To sum up, the automatic oral English evaluation method based on multimodal design has high accuracy for automatic oral English evaluation, and it is reasonable to apply it directly to practice.

4.3.2. Analysis of Different Students' Mastery of Reading Attributes. This section will analyze the students' mastery of various reading attributes in different groups and explore the cognitive characteristics of different groups in the process of reading. First of all, students are divided into three grades according to their grades. Based on the results of G-DINA model MLE, students are divided into two categories: "not mastered" and "mastered" by "0" and "1". For categorical data on such nonnormal distributions, a Chi-square test is used to analyze significant differences among variables.

The chi-square test found (Table 2 and Figure 4), comparing the mastering probabilities of the five attributes of the three level populations, that the five P values were lower than the critical values ($\alpha = 0.005$, $P < 0.001$), and it was found that the mastering rates of the five attributes of the three level populations were significantly different. In other words, for the 5 reading attributes, the mastering probability of A-level group is significantly higher than that of B-level group, and that of B-level group is significantly higher than

that of C level group. The three groups show completely different characteristics in the mastery of their attributes. The C level group performed poorly on the reading test as a whole, especially for the A2, A4, and A5 attributes with lower mastery probability ($P < 0.5$); the C level group had better mastery of the A1 and A3 attributes, indicating that the two attributes were easier to mastery than the others; the B-level group had lower mastery except for the attributes.

4.3.3. Time for Oral English Evaluation. In order to verify the efficiency of this method in the process of oral English evaluation, the traditional method and this method are used to verify the time of oral English evaluation, and the results are shown in Table 3.

By analyzing Table 3, it can be seen that, under different numbers, the oral English evaluation time of the traditional method and the method in this paper is different. When the experiment number is 1, the oral English evaluation time of the traditional method is 125 s, and the oral English evaluation time of this method is 12 s. When the experiment number is 3, the oral English evaluation time of the traditional method is 132 s, and the oral English evaluation time of this method is 15 s. When the experiment number is 5, the oral English evaluation time of the traditional method is 127 s, and the oral English evaluation time of this method is 16 s. The evaluation time of this method is significantly higher than that of the traditional method, which shows that the evaluation efficiency of this method is better.

TABLE 1: Comparison of evaluation confidence.

Number of experiments	The evaluation confidence of this method	Evaluation confidence of traditional methods
01	89.13	46.71
02	89.02	48.65
03	89.54	49.21
04	88.62	49.93

TABLE 2: Reading attribute mastery probability and difference test of three level groups.

	Sample size	Attributes				
		A1	A2	A3	A4	A5
Class A	418	0.9018	0.8534	0.9167	0.8747	0.8268
Class B	1383	0.7761	0.5761	0.7072	0.5180	0.4623
Class C	438	0.5143	0.4867	0.5449	0.4617	0.3024
Chi-square value		425.6	238.04	411.3	462.4	33.12
Degree of freedom		2	2	2	2	2

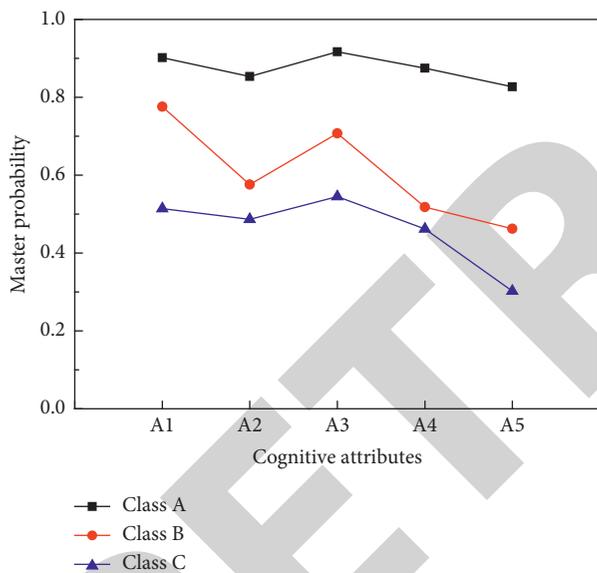


FIGURE 4: Reading attribute mastery probability of three horizontal groups.

TABLE 3: Time for oral English evaluation.

Number of experiments	Time for oral English evaluation (s)	
	Traditional method	Paper method
01	125	12
02	107	16
03	132	15
04	142	12
05	127	16
06	128	18

5. Conclusion

Using intelligent signal processing technology, combined with the time-frequency feature analysis and spectral analysis of oral English pronunciation signals, the

automation and intelligence level of oral English pronunciation quality evaluation are improved. This paper presents an automatic oral English evaluation method based on multimodal design and designs the system hardware combined with integrated DSP. The system design is divided into two parts: speech signal processing algorithm design and hardware circuit design. The sound sensor is used to collect oral English pronunciation signals. The feature decomposition of spoken English pronunciation signal is carried out by using multilayer wavelet feature scale transform. According to the results of feature decomposition, the adaptive filtering detection and spectral analysis of spoken English pronunciation signals are carried out. The wavelet entropy feature of spoken English pronunciation signal is extracted, and the pronunciation quality is automatically recognized according to the wavelet entropy feature extraction results of spoken English pronunciation. Taking STC12C5A60S2 chip as the core control chip, combined with the embedded program loading design method, the hardware integration design of the system is carried out. The research shows that the accuracy of automatic evaluation of oral English pronunciation quality using this system is high. It has good stability and good application value.

Data Availability

The raw data supporting the conclusions of this article will be made available by the author, without undue reservation.

Conflicts of Interest

The author declares that there are no conflicts of interest regarding this work.

References

- [1] J. Xu, Z. Xiong, J. Liu, and R. Wang, "A dynamic vector-formed information sharing algorithm based on two-state Chi square detection in an adaptive federated filter," *Journal of Navigation*, vol. 72, no. 1, pp. 101–120, 2019.
- [2] Y. Koda, J. Park, M. Bennis et al., "Communication-efficient multimodal split learning for mmWave received power prediction," *IEEE Communications Letters*, vol. 24, no. 6, pp. 1284–1288, 2020.
- [3] L. Jiang and H. Zhang, "Redundant measurement-based second order mutual difference adaptive Kalman filter," *Automatica*, vol. 100, pp. 100396–100402, 2019.
- [4] T. Jian, Y. He, H. Wang, and Y. Liu, "Persymmetric generalised adaptive matched filter for range-spread targets in homogeneous environment," *IET Radar, Sonar & Navigation*, vol. 13, no. 8, pp. 1234–1241, 2019.
- [5] L. Ma, T. A. Gulliver, A. Zhao, C. Ge, and X. Bi, "Underwater broadband source detection using an Acoustic Vector Sensor

- with an adaptive passive matched filter,” *Applied Acoustics*, vol. 148, pp. 162–174, 2019.
- [6] Ma, Lin, Gulliver et al., “Underwater broadband source detection using an acoustic vector sensor with an adaptive passive matched filter,” *Applied acoustics*, vol. 148, pp. 162–174, 2019.
- [7] Y. Zhang, “The development of an evaluation model to assess the effect of online English teaching based on fuzzy mathematics,” *International Journal of Emerging Technologies in Learning (ijET)*, vol. 16, no. 12, p. 186, 2021.
- [8] W. Ma, “Study on the evaluation method of students’ English classroom performance based on big data analysis,” *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 31, no. 1, p. 1, 2021.
- [9] N. K. Ibrahim, H. Hammed, and A. A. Zaidan, “Multi-criteria evaluation and benchmarking for young learners’ English language mobile applications in terms of LSRW skills,” *IEEE Access*, vol. 99, p. 1, 2019.
- [10] B. Hu, “The evaluation method of English teaching efficiency based on language recognition technology,” *International Journal of Continuing Engineering Education and Life Long Learning*, vol. 30, no. 4, p. 445, 2020.
- [11] M. Piotrowska, A. Czyżewski, T. Ciszewski, G. Korvel, A. Kurowski, and B. Kostek, “Evaluation of aspiration problems in L2 English pronunciation employing machine learning,” *Journal of the Acoustical Society of America*, vol. 150, no. 1, pp. 120–132, 2021.
- [12] G. Grami, “An evaluation of online and automated English writing assistants: collocations and idioms checkers[],” *International Journal of Emerging Technologies in Learning (ijET)*, vol. 15, no. 4, p. 218, 2020.
- [13] S. B. Ahmed, S. Naz, and M. I. Razzak, “A novel dataset for English-Arabic scene text recognition (EASTR)-42K and its evaluation using invariant feature extraction on detected extremal regions,” *IEEE Access*, vol. 7, p. 1, 2019.
- [14] S. B. Ahmed, S. Naz, M. I. Razzak, and R. Yusof, “A novel dataset for English-Arabic scene text recognition (EASTR)-42K and its evaluation using invariant feature extraction on detected extremal regions,” *IEEE Access*, vol. 7, 2019.
- [15] H. Lee, W. Park, and S.-C. Lee, “English proficiency and vowel insertion after word-final stops in English,” *Journal of the Acoustical Society of America*, vol. 146, no. 4, Article ID 2842, 2019.
- [16] A. Sd, B. Lf, and B. Mlc, “Computer-assisted assessment of phonetic fluency in a second language: a longitudinal study of Japanese learners of French–ScienceDirect,” *Speech Communication*, vol. 125, pp. 69–79, 2020.
- [17] M. C. Alfred, D. M. Neyens, and A. K. Gramopadhye, “Learning in simulated environments: an assessment of 4-week retention outcomes,” *Applied Ergonomics*, vol. 74, pp. 107–117, 2019.
- [18] Y. R. Oh, K. Park, H. B. Jeon, and J. Park, “Automatic proficiency assessment of Korean speech read aloud by non-natives using bidirectional LSTM-based speech recognition,” *ETRI Journal*, vol. 42, no. 10, 2020.
- [19] F. Jiang, Y. Chiba, T. Nose, and A. Ito, “Automatic assessment of English proficiency for Japanese learners without reference sentences based on deep neural network acoustic models–ScienceDirect,” *Speech Communication*, vol. 116, pp. 86–97, 2020.
- [20] J. Wang, “Speech recognition of oral English teaching based on deep belief network,” *International Journal of Emerging Technologies in Learning (ijET)*, vol. 15, no. 10, p. 100, 2020.
- [21] D. Korzekwa and B. Kostek, “Deep learning model for automated assessment of lexical stress of non-native English speakers,” *Journal of the Acoustical Society of America*, vol. 146, no. 4, pp. 2956–2957, 2019.
- [22] A. R. Johnson, A. F. Doval, M. D. Granoff et al., “A cm assessment of English and Spanish carpal tunnel syndrome materials,” *Journal of Surgical Research*, vol. 238, pp. 64–71, 2019.
- [23] A. R. Johnson, A. F. Doval, M. D. Granoff et al., “Comparative multimetric assessment of English and Spanish carpal tunnel syndrome materials,” *Journal of Surgical Research*, vol. 238, pp. 64–67, 2019.
- [24] C. Chen and Q. Ling, “Adaptive convolution for object detection,” *IEEE Transactions on Multimedia*, vol. 21, no. 12, pp. 3205–3217, 2019.
- [25] C. G. Clopper and E. Dossey, “Phonetic convergence to sa,” *Journal of the Acoustical Society of America*, vol. 147, no. 1, pp. 671–683, 2020.
- [26] J. Elmer, B. L. Jones, V. I. Zadorozhny et al., “A novel methodological framework for multimodality, trajectory model-based prognostication,” *Resuscitation*, vol. 137, pp. 197–204, 2019.
- [27] Y. Chu, “On English translation of Chinese original picture books from the perspective of multimodality,” *Open Access Library Journal*, vol. 7, no. 3, pp. 1–19, 2020.
- [28] W. Xiaohui, “Theory and practice of multimodality English language material development and design–review on “English language teaching materials: theory and practice”,” *Modern Linguistics*, vol. 9, no. 1, pp. 227–230, 2021.
- [29] B. Khoshnevisan, “Spilling the beans on understanding English idioms using multimodality: an idiom acquisition technique for Iranian language learners,” *International Journal of Language, Translation and Intercultural Communication*, vol. 8, p. 128, 2019.
- [30] L. Liu, “Research on IT English flipped classroom teaching model based on SPOC,” *Scientific Programming*, vol. 2021, Article ID 7273981, 9 pages, 2021.
- [31] Y. Wang and P. Zhao, “A probe into spoken English recognition in English education based on computer-aided comprehensive analysis,” *International Journal of Emerging Technologies in Learning (ijET)*, vol. 15, no. 3, p. 223, 2020.
- [32] X. Qin, “Evaluation of English intercultural communication ability based on machine learning and fuzzy mathematics,” *Journal of Intelligent and Fuzzy Systems*, vol. 40, no. 6, pp. 1–13, 2020.
- [33] P. R. Goundar, “Bridging the gap for English language testing study in Fiji: proposing an evaluation of the writing proficiency level of pre-first year and post-first year undergraduate students,” *Journal of Language Teaching and Research*, vol. 11, no. 3, p. 351, 2020.
- [34] T. N. Huynh and V. N. Long, “Applying corpus linguistics to English textbook evaluation: a case in Viet Nam,” *VNU Journal of Foreign Studies*, vol. 36, no. 5, 2020.