

Retraction

Retracted: Design of Automatic Scoring System for Oral English Test Based on Sequence Matching and Big Data Analysis

Discrete Dynamics in Nature and Society

Received 3 October 2023; Accepted 3 October 2023; Published 4 October 2023

Copyright © 2023 Discrete Dynamics in Nature and Society. 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This article has been retracted by Hindawi following an investigation undertaken by the publisher [1]. This investigation has uncovered evidence of one or more of the following indicators of systematic manipulation of the publication process:

- (1) Discrepancies in scope
- (2) Discrepancies in the description of the research reported
- (3) Discrepancies between the availability of data and the research described
- (4) Inappropriate citations
- (5) Incoherent, meaningless and/or irrelevant content included in the article
- (6) Peer-review manipulation

The presence of these indicators undermines our confidence in the integrity of the article's content and we cannot, therefore, vouch for its reliability. Please note that this notice is intended solely to alert readers that the content of this article is unreliable. We have not investigated whether authors were aware of or involved in the systematic manipulation of the publication process.

Wiley and Hindawi regrets 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 own 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] P. Li, H. Zhang, and S. Tsai, "Design of Automatic Scoring System for Oral English Test Based on Sequence Matching and Big Data Analysis," *Discrete Dynamics in Nature and Society*, vol. 2021, Article ID 3018285, 10 pages, 2021.

Research Article

Design of Automatic Scoring System for Oral English Test Based on Sequence Matching and Big Data Analysis

Ping Li,¹ Hua Zhang ,¹ and Sang-Bing Tsai ²

¹Foreign Language School, Weifang University, Weifang, Shandong 262100, China

²Regional Green Economy Development Research Center, School of Business, WUYI University, Nanping, China

Correspondence should be addressed to Hua Zhang; zhanghua198316@126.com and Sang-Bing Tsai; sangbing@hotmail.com

Received 10 September 2021; Revised 6 October 2021; Accepted 12 October 2021; Published 25 October 2021

Academic Editor: Gengxin Sun

Copyright © 2021 Ping Li et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

With the application of an automatic scoring system to all kinds of oral English tests at all levels, the efficiency of test implementation has been greatly improved. The traditional speech signal processing method only focuses on the extraction of scoring features, which could not ensure the accuracy of the scoring algorithm. Aiming at the reliability of the automatic scoring system, based on the principle of sequence matching, this paper adopts the spoken speech feature extraction method to extract the features of spoken English test pronunciation and establishes a dynamic optimized spoken English pronunciation signal model based on sequence matching, which could maintain good dynamic selection and clustering ability in a strong interference environment. According to the comprehensive experiment, the automatic scoring result of the system is much higher than that of the traditional method, which greatly improves the recognition ability of oral pronunciation, solves the difference between the automatic scoring of the system and the manual scoring, and promotes the computer automatic scoring system to replace or partially replace the manual marking.

1. Introduction

With the popularization of computers and networks and the improvement of related technical performance, the requirements for listening, speaking, reading, and writing skills in English were getting higher and higher [1]. However, the manual reviewing machine to assist oral test recording still required a huge labor cost. The computer-assisted oral test (computer-assisted oral test for short) had been gradually applied to various oral tests at all levels, greatly improving the efficiency of test administration [2]. In order to improve the objectivity of the computer-based test, it was necessary to design an automatic scoring system for the computer-based English test, combined with the intelligent scoring system for the computer-based English test, to perform speech recognition and semantic feature recognition for the output of the computer-based English test. According to the results of speech recognition, the automatic evaluation of oral English tests based on a computer was realized. Research on the optimization design method of the

automatic scoring system for spoken English testing was of great significance in improving the automatic scoring level of the computer-based oral English test and promoting the construction of the intelligent level of the computer-based English oral test [3]. The research on related system design methods had attracted great attention.

In the traditional methods, the automatic scoring system design of oral English computer test mainly included the automatic scoring method of oral English computer test based on spectrum analysis and the automatic scoring method of oral English computer test based on wavelet analysis. The correlation statistical feature analysis method was used to design the automatic scoring system of oral English computer test, so as to improve the automatic scoring ability of oral English computer test [4]. The frequency characteristics and phase characteristics of the pronunciation feature sequence of the oral English test were scattered, resulting in poor accuracy and low stability of the system. Therefore, it was necessary to optimize the signal processing part of the automatic scoring system of oral

English test; combined with the statistical analysis method of pronunciation sequence, this work improved the information processing and pronunciation sequence analysis ability of the automatic scoring system of oral English test [5].

In this paper, we presented a design of an automatic scoring system for the oral English test based on sequence matching. The algorithm design of the automatic scoring system for the oral English test was carried out by using the speech signal processing method [6], the speech signal collection of oral English was carried out by using the time series analysis method, the collected oral English pronunciation sequence was mixed by using the decision feedback equalization adjustment method, and the statistical information feature of oral English pronunciation sequence was extracted. Decision feedback equalizer (DFE) is an equalization method commonly used in Rx in SerDes at present. Appropriate delay time and weight H1 and H2 can ensure the reduction or complete elimination of intersymbol interference of input data, so it can effectively improve RX reception performance. The correlation function feature matching method was used for the state evaluation and standard comparison of oral English pronunciation sequence, the spectral feature quantity of oral English pronunciation sequence was extracted, the extracted spectral feature quantity of oral English pronunciation sequence was matched with the standard pronunciation feature quantity, and the automatic test and scoring of oral English were realized according to the difference of comparison. The pronunciation sequence matching algorithm of oral English was loaded into the hardware module, and the hardware development and design of the system were carried out by using the B/S architecture system and DSP [7]. The hardware development and design of the automatic scoring system of the oral English computer test were realized, and an effective conclusion was drawn.

Based on the above analysis, the sections of this paper would be arranged as follows. Section 2 would analyze the research status of the existing oral scoring system and point out the advantages and disadvantages of various algorithms. In Section 3, based on the pronunciation sequence distribution signal, the spatial source modeling of the collected signal model was carried out by using the continuous digital speech recognition method, and the automatic scoring system model based on sequence matching was designed according to the characteristics of the signal model. Section 4 would analyze the man-machine consistency results, use the sequence matching dynamic optimization method to further improve the system, and get the test results of scoring accuracy.

2. Related Work

The research on automatic oral scoring had achieved remarkable results [8]. In the traditional methods, the design of the automatic scoring system for oral English computer test mainly included the automatic scoring method for oral English computer test based on spectrum analysis and the automatic scoring method for oral English computer test based on wavelet analysis [9]. The correlation statistical

feature analysis method was used to design the automatic scoring system for the oral English computer test, so as to improve the automatic scoring ability of the oral English computer test. Researchers tested the reliability of machine scoring results by comparing the correlation between machine evaluation and human evaluation results, absolute difference, average difference, absolute consistency rate, large difference proportion, serious error rate, etc. [10]. Adaboost-ELM algorithm was used to replace the support vector machine module in SVM-GSV automatic recognition system. Compared with the support vector machine algorithm, it had a faster training speed and similar classification accuracy [11–13]. The score of the MyET oral test system had strong stability and generally achieves a similar correlation with the manual score, but it was slightly lacking in discrimination. According to the feature decomposition results, the adaptive filter detection and spectral analysis of oral English pronunciation signals were carried out, and the wavelet entropy feature of the signal was extracted to improve the automatic detection ability of oral English pronunciation quality. However, the automation level of automatic scoring of oral English computer tests by this method was not high. The American Educational Testing Service (ETS) used a variety of indicators in the validity study of its developed oral automatic scoring system SpeechRater [14] for evaluating the online training task TPO (TOEFL Practice online) [15]. When SpeechRater was applied to evaluate TEFT (test of English for teaching) for English teachers, the Research Report of ETS adopts Pearson's correlation coefficient and kappa coefficient to measure the performance and quality of the system. In the validity study of the automatic scoring system developed by Orient Company for evaluating the telephone oral test Phonepass SET-10, the total scores of human evaluation and machine evaluation and the correlation coefficients of scores in each dimension were mainly reported [16]. There was no essential difference in the technical paths adopted by ordinate and SpeechRater scoring systems at the level of speech recognition. The automatic speech recognition programs of the two systems were responsible for processing the original speech files, that is, speech segmentation of vocabulary units and conversion of acoustic spectrum, so as to prepare for feature parameter extraction and score calculation [17]. The speech recognition programs of the two systems were established according to the hidden Markov model (HMM), which can be used to recognize the speech of nonnative speakers.

3. Design of Oral English Sequence Matching Model and Scoring System

3.1. Pronunciation Sequence Distribution Signal in Oral English Test. In order to realize the accurate detection and parameter estimation of the pronunciation sequence signal of the oral English test, it was necessary to first construct the pronunciation sequence signal model of the oral English test. Generally speaking, each note included the fundamental frequency and harmonic component. Combined with the dynamic acquisition method, the dynamic correlation

detection in oral English test was carried out to directly track the detection statistics of the fundamental frequency. The continuous digital speech recognition system has two parts: training and recognition. This training can be regarded as the process of modeling HMM. By reevaluating the parameters and adjusting various parameters of the model, a model with good robustness is obtained. Improving and optimizing the basic model can effectively improve the accuracy and obtain a better recognition rate. The recognition process can be considered as the process of using the existing HMM model base, data dictionary, and syntax control to form a recognition network and using the search algorithm to find the best match. Firstly, the speech signal waiting for recognition is sampled and then converted into an electrical signal. It was obtained that the measure where the human ear could still accurately perceive oral English was ν , the reliability of the oral test was evaluated based on the data-driven method, $r(t)$ was the characteristic quantity of dynamic pronunciation sequence received by the system, and then,

$$r(t) = \sqrt{s_0}f(s_0(t - \tau_0)) + n(t). \quad (1)$$

Among them, $s_0 = (c + \nu)/(c - \nu)$ was called the time scale factor, which was called scale for short, and represents the candidate pitch of each frame. The output spoken English pronunciation sequence was $f(t)$. c was the estimated spread delay of each frame signal, $\sqrt{s_0}$ was the normalization factor; τ_0 was the oral output delay in the higher frequency range, and $n(t)$ was the background interference.

The statistical characteristic quantity of the pronunciation sequence of the oral English test was detected under the interference background, as shown in Figure 1. Through the scale and time delay estimation of the pronunciation sequence signal of the oral English test, the spectral analysis method was used for wavelet scale decomposition [18], and the dynamic distribution characteristic quantity of the oral English test was obtained as follows (see Figure 1).

$$W(s, \tau) = \sqrt{s} \int r(t)f^*(s(t - \tau)). \quad (2)$$

For short-time Fourier transform (STFT), it has a certain resolution in the time domain and frequency domain, and the time-frequency resolution of STFT is the same in the global range. However, due to the restriction of the Heisenberg uncertainty principle (that is, the uncertainty principle in quantum mechanics), the area of each time-frequency window is fixed; that is, the time resolution is inversely proportional to the frequency resolution, so the two resolutions cannot be very high at the same time. Combining the time domain and frequency domain of the pronunciation sequence signal of the oral English test, the candidate pitch of the oral English test was estimated, and a spectral peak distribution function $P(t, f)$ of the amplitude spectrum was constructed. By continuously sliding the window on the time axis, each spectral peak could be adjusted with two parameters. The short-time Fourier transform of pronunciation sequence distribution $x(t)$ of oral English test was defined as

$$\text{STFT}_x(t, f) = \int_{\tau} [x(\tau)h^*(\tau - t)]e^{-j f \tau} d\tau. \quad (3)$$

In the above formula, τ was the window function of short-time Fourier transform, and f was the frequency domain decomposition feature of short-time Fourier transform and was the time of amplitude modulation of the original pronunciation sequence. The pronunciation sequence signal model of the oral English test constructed above was used for signal analysis.

In the actual process, the physical process of Korean speech signal generation is different from the above three models but is approximately equivalent. This also verifies that the Korean speech signal is a short-term stable signal and a signal that changes over time. In addition, the fricatives in voiced sounds have both unvoiced and voiced excitation sources at the same time and cannot be obtained by simply superimposing the two.

3.2. An Analysis of the Characteristics of Spoken English Pronunciation Signal Model. For the collected signal model, the continuous digital speech recognition method was used for spatial source modeling, and a continuous digital cutting method with variable time window length was used for adaptive adjustment of the source of the automatic scoring system for the oral computer test. The spatial source distribution of the output speech information of the automatic scoring system for oral computer test was obtained as follows:

$$\begin{aligned} \Lambda_E &= \text{diag}[\rho_E(i)], \\ \Lambda_F &= \text{diag}[\rho_F(i)], \\ \Lambda_G &= \text{diag}[\rho_G(i)]. \end{aligned} \quad (4)$$

In the above, Λ_E was the pronunciation continuous number cutting point of Λ 's oral English computer test, and $\{\rho_G(i)\}_{1 \leq i \leq l}$ and $\{\rho_F(i)\}_{1 \leq i \leq l}$ were nonzero eigenvalues of matrices F and G , respectively. Because E , F , and G have correlation coupling, the more connections between modules, the stronger their coupling. At the same time, it shows that their independence is worse, and they have the same eigenvector as A . The feature detection was carried out according to the multiparameter constrained evolution method, and the joint detection method was used to obtain the pronunciation signal frequency discrimination output of the oral English computer test. The feature detection was carried out according to the multiparameter constrained evolution method. The joint detection method was adopted to obtain the pronunciation signal frequency discrimination output of oral English computer test as follows:

$$\hat{\nu} = a_{0k} + \delta a_k = \frac{(2F) |X_{\bar{a}_{0k}}(\bar{u}_{0k})|}{(2N + 1 |A_{\bar{a}_{0k}}|)}. \quad (5)$$

The high-order hidden Markov model based on piecewise linear processing was used for spectrum analysis [19]. Combined with the corresponding relationship of spectrum features, the semantic correlation analysis was carried out on

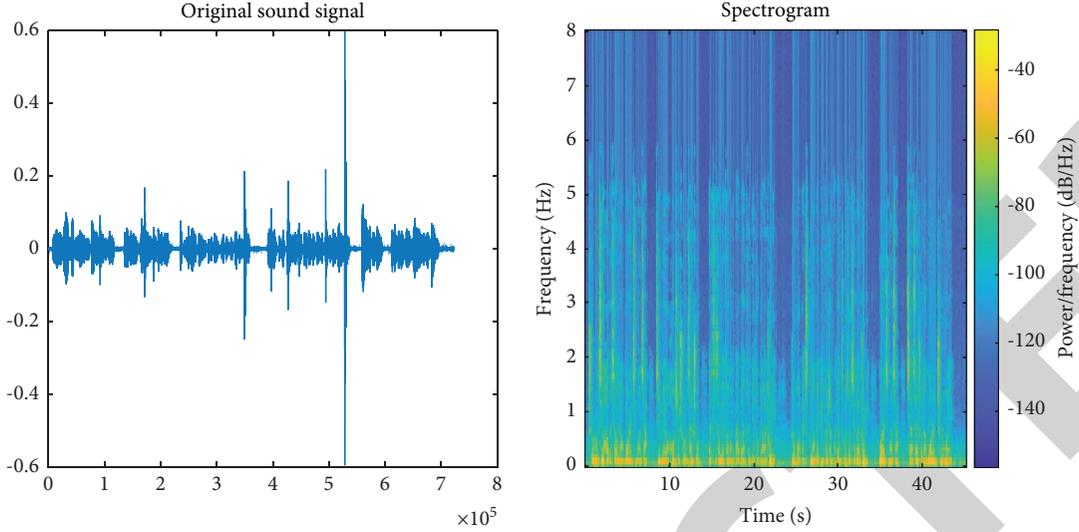


FIGURE 1: Original signal acquisition for automatic scoring of oral English test.

the pronunciation signals of oral English computer test, and the output spectrum features were obtained:

$$S(\tau, \lambda) = \int_{\tau} R_{CC_{\dots}}(\tau, \alpha) e^{-j2\pi\lambda\alpha}. \quad (6)$$

We extracted the power spectral density characteristics of the pronunciation signal of the oral English computer test [20], transformed the speech signal Z into polynomial $s(t)$ for time delay expansion, and obtained the original group delay function as W . When calculating the group delay function, the envelope amplitude was obtained according to the convolution of channel response:

$$\tilde{s}(t - \tau) = \sum_{\tau} \tilde{s}(t - nT) \frac{\sin(m(\tau - nT))}{m(\tau - nT)}, \quad (7)$$

where $m = 2\pi W^2$ was the baseband bandwidth on the circumference of the unit circle, which was used for signal detection. Combined with the sequence dynamic selection method, this paper analyzed the speech correlation characteristics of oral English computer test, established the automatic feature matching model of oral English computer test, as shown in Figure 2, and then used the block matching and template matching methods to detect and recognize the speech correlation of oral English computer test, correlation-based template matching is actually another matching based on gray value, but its characteristic is to use a normalized cross-correlation matching (normalized cross-correlation, NCC) to measure the relationship between the template image and detection image. Different from the classical matching algorithm based on gray value, its speed is much faster. Compared with the matching algorithm based on shape template, its advantage is that it can retrieve some detection images with slight shape change, complex texture, or blurred focus; see Figure 2.

The main object of dynamic selection is complex blocks with complexity greater than the national value. In order to prevent irreversible block classification due to complexity

change at the extraction end, the complexity after processing must be greater than the original complexity without affecting the pixel sorting sequence.

3.3. Sequence Matching in Oral English Test. Based on the above feature analysis of oral English speech signal acquisition using the time series analysis method, the automatic scoring system of the oral English test was designed, and an automatic scoring method of oral English test based on sequence matching was proposed. In the multipitch estimation stage, the optimal value $R_{MDMMA}(k)$ of crosstalk judgment met

$$\text{abs}[|z(k)|^2 - R_{MDMMA}(k)] = \min_i \text{abs}[|z(k)|^2 - R_{MDMMA_i}]. \quad (8)$$

Let r_i and θ_i be the phase information of oral English pronunciation sequence, respectively, and the modulation signal [21] of oral English sequence was obtained as follows:

$$x(t) = As(t)r_i + n(t)\theta_i. \quad (9)$$

Combining significance and continuity constraints, the output frequency response representation of pronunciation sequence in oral English test under strong interference environment was presented:

$$C_{T'}(f) = \sum_{k=-K^2}^{K^2} c_k e^{-jfk^2T'}. \quad (10)$$

In the above, c_k was the amplitude equalization coefficient of the pronunciation sequence of the oral English test, N was the sampling length of the directly corrected distorted waveform, P was the melody track and spectrum, when $T' = (M/N)T$, the quotient function was R_a , the symbol width of the pitch candidate was T_a , $T_a = 1/R_a$, and the instantaneous frequency coefficient was a_n . The matching

analysis was carried out between the extracted spectral feature quantity of the oral English pronunciation sequence and the standard pronunciation feature quantity, which was described as follows:

$$\text{sgn}(z_R^2(k) - R_{\text{MDMMA}_R}) = \text{sgn}(z_R^2(k) - \mu_R(k)). \quad (11)$$

In the above, $\mu_R(k)$ was the modulation error of spoken English speech signal, and $z_R^2(k)$ was the statistical test quantity. In addition,

$$\text{sgn}(z_I^2(k) - R_{\text{MDMMA}_I}) = \text{sgn}(z_I^2(k) - \mu_I(k)). \quad (12)$$

$z_I^2(k)$ was the harmonic component of the perceived pitch, $\mu_I(k)$ was interference error, short-time Fourier transform was used for sequence matching of oral English test, and the output was

$$d(t) = a(t)c(t) = \sum_{n=0}^{\infty} d_n g_c(t - nT_c). \quad (13)$$

The extracted spectral feature quantity of the spoken English pronunciation sequence was matched with the standard pronunciation feature quantity, and the extended sequence was used to modulate the carrier to improve the accuracy of spoken English detection.

3.4. Design of Test Automatic Scoring System. In order to help the school to simplify the examination process and improve the marking efficiency, the system included three models: (1) speech recognition model, which was used to recognize the subjects' words; (2) standard pronunciation model, used to judge the accuracy of pronunciation; and (3) general score mapping model, which extracted scoring dimension features by collecting a large number of oral test data distinguished according to question types. The extracted main scoring dimensions and specific features are described in Figure 3. Experts were hired to score the oral test recording. Based on SVM (support vector machine) classifier and nonlinear regression mapping algorithm, a high-precision mapping model from dimensional features to manual scores (overall scores) and a mapping model from features to individual scores (such as pronunciation, fluency, etc.) could be realized. For closed oral test tasks such as reading aloud and the following reading, the system could score directly and automatically. For open-ended oral test tasks such as answering questions and oral composition, the system needed to be calibrated first. The calibration was based on the scores of experts on 200 candidate data; see Figure 3.

This work extracted similarity features, syntactic features, and phonetic features from the examinee samples, added the corresponding features of all subjects of each examinee, and compared them with the expert score. The expert score here was the average value of the score of the same sample. The performance of scoring was described by correlation coefficient and difference. The so-called correlation coefficient described the correlation degree between two vectors, and the value range of the correlation coefficient was $[-1,1]$. The correlation coefficient was used as the evaluation index to compare the consistency of scores

between experts; that is, the higher the correlation coefficient, the more consistent the scores of the two experts on the sample; see Table 1.

Table 1 was a list of correlation coefficients between all features mentioned in the text and expert scores. It could be seen from the table that, in addition to the high correlation coefficient between similarity features and expert scores, syntactic features and phonetic features also had relatively high correlation coefficients, indicating that these features were highly consistent with expert scores.

For all the features extracted above, we could only observe the correlation coefficient between them and the expert score, and we could choose whether all these features together play a positive role in the score.

4. Experiment and Analysis

4.1. Comparison of Man-Machine Consistency Results. The comparison of the average scores of machine evaluation and human evaluation of the four tasks of this oral test is shown in Figure 4 (U^* represents machine; $U1 \sim U4$ represents 4 manual raters). It could be seen from the figure that there was a great difference in the average score between machine evaluation and human evaluation, and the difference in the average score between machine evaluation and human evaluation was greater than that among the three raters. The difference between machine evaluation and human evaluation in the other two tasks was small, and the difference between machine evaluation and human evaluation was less than that between people. Specifically, the machine scores reading aloud was higher than the three raters. The difference between the machine and $U3$ was the smallest ($MD = 0.66$) and the difference between the machine and $U2$ was the largest ($MD = 3.30$). The score of the retelling task machine was low, the difference between machine and $U2$ was the smallest ($MD = 0.21$), and the difference between machine and $U3$ was the largest ($MD = 1.73$). The score difference between machine and $U1$ on oral composition was the smallest ($MD = 0.45$), and the difference between machine and $U2$ was the largest ($MD = 1.70$). The result of the machine evaluation was close to the average score of the three raters ($MU^* = 5.57$, $MU1 \sim U4 = 5.16$). In combination with the consistency rate of machine evaluation and human evaluation (Table 1), the complete consistency rate and proximity score (difference less than 2) consistency rate of machine evaluation of reading task and $U4$ were much higher than those of the other two raters, the complete consistency rate and proximity score consistency rate of machine evaluation of retelling task and $U2$ were the highest, and the consistency rate of machine evaluation of oral composition with $U1$ and $U4$ was relatively high; see Figure 4.

The correlation coefficient of the scoring results of the three tasks by the machine and four manual raters is shown in Table 2: the correlation coefficient between the machine score of reading aloud and the manual score was low, ranging from 0.279 to 0.469, which was statistically significant. There was a significant medium to high correlation between machine and manual scoring results of retelling and

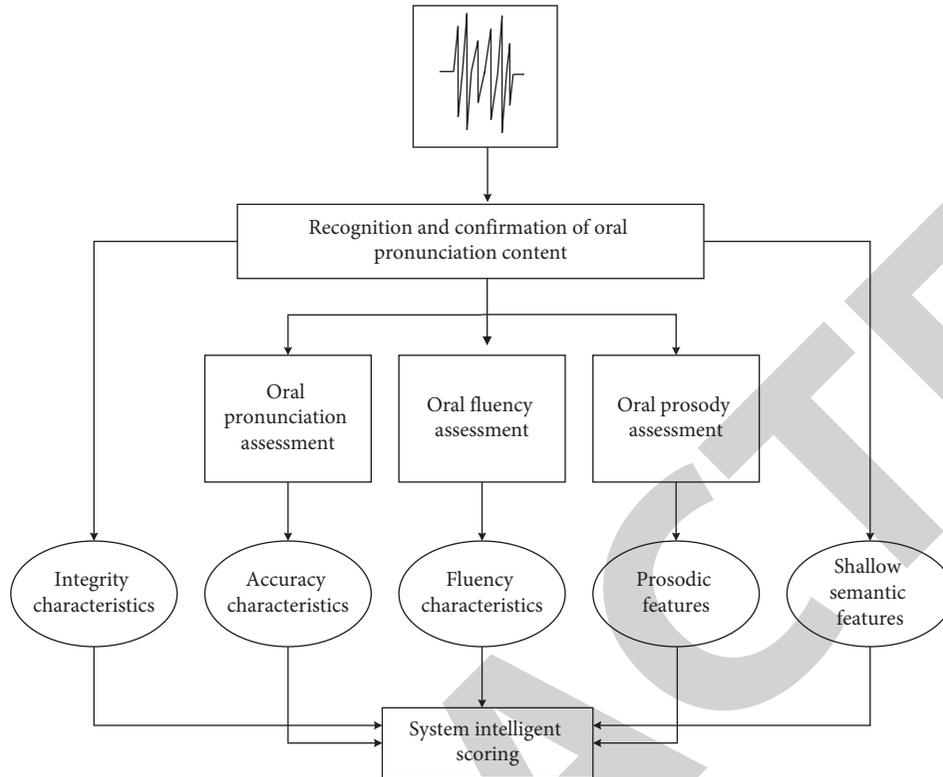


FIGURE 2: Oral test pronunciation comprehensive evaluation.

oral composition, and the correlation coefficient was between 0.600 and 0.703. On the whole, the correlation between machine evaluation and human evaluation was lower than the consistency coefficient between human evaluation, but in some tasks, the correlation coefficient between machine and individual raters was higher than that between people. For example, the correlation coefficient between machine evaluation and U3 of oral composition ($r = 0.703$) was higher than that between two of the four raters of the task ($r_{1/2} = 0.663$, $p < 0.01$; $r_{1/3} = 0.653$, $p < 0.01$; $r_{2/3} = 0.619$, $p < 0.01$); see Table 2.

Based on Table 2, it could be seen that the correlation coefficient between U3 and the machine, which had the smallest difference with the machine and the highest proportion of consistency with the adjacent score, was not as high as that between U2 and the machine, which had a large difference with the machine and a small consistency rate with the adjacent score. This might be due to the fact that the adjacent score was both high and low. However, the correlation only depended on the consistent trend, that is, the proportion of both high or both low.

4.2. Sequence Matching Dynamic Optimization. The software development and design of the automatic scoring system for the oral English computer test were realized under the multilayer B/S architecture system. The embedded development and module design of the automatic scoring system for the oral English computer test were carried out under the development environment of Multigen Creator 3.2. The

information processing center of the automatic scoring system for oral English computer test took DSP as the core [22–24]; the central centralized controller of the automatic scoring system of oral English computer test was constructed to realize the communication and information sharing between the automatic scoring system of oral English computer test and the computer network. The JTAG debugging interface was used for real-time program reading and writing and AD conversion control of the automatic scoring system for the oral English computer test. The SEL1 level was controlled by DSP to realize the clock sampling and ad bus control of the automatic scoring system for the oral English computer test [25–27]. The B/S architecture system was used to realize the hardware development and design of the automatic scoring system for the oral English computer test [28–29]; see Figure 5.

In the frequency domain, the spectral components of speech signals are mainly concentrated in the range of 300~3400 Hz. Using this feature, we can use an antialiasing band-pass filter to take out the frequency component of the speech signal in this range and then sample the speech signal according to the sampling rate of 8 kHz to obtain the discrete speech signal. Taking the data sequence of oral English pronunciation source collection results as the test object, as shown in Figure 5, combined with the sequence dynamic selection method, the speech correlation characteristics of oral English computer test were analyzed, the automatic feature matching model of oral English computer test was established, and the block matching and template matching methods were used to realize the sequence correlation

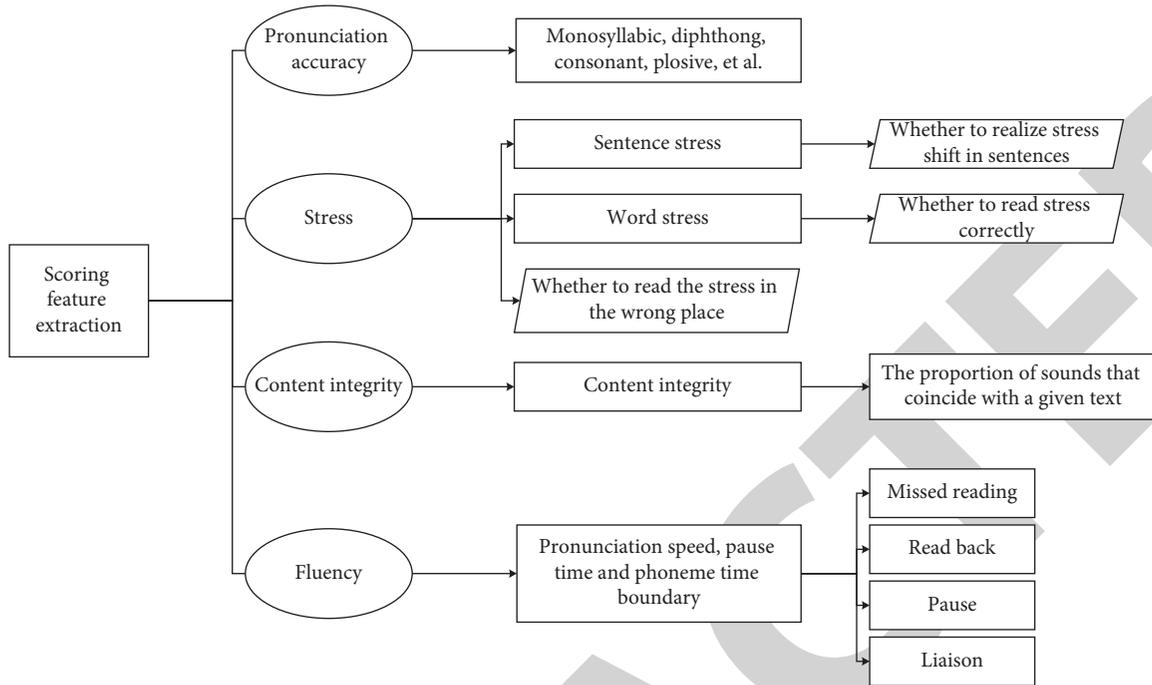


FIGURE 3: Scoring characteristics adopted by the system.

TABLE 1: Correlation coefficient between all characteristics and expert score.

Features	Correlation coefficient with expert score
Number of words per unit time	0.501
Syntax tree depth	0.295
Normalized syntax tree depth	0.338
Score of the whole syntax tree	-0.414
Manhattan similarity	0.634
Dice coefficient	0.692
Cosine similarity based on word frequency	0.716
Pearson coefficient	0.792

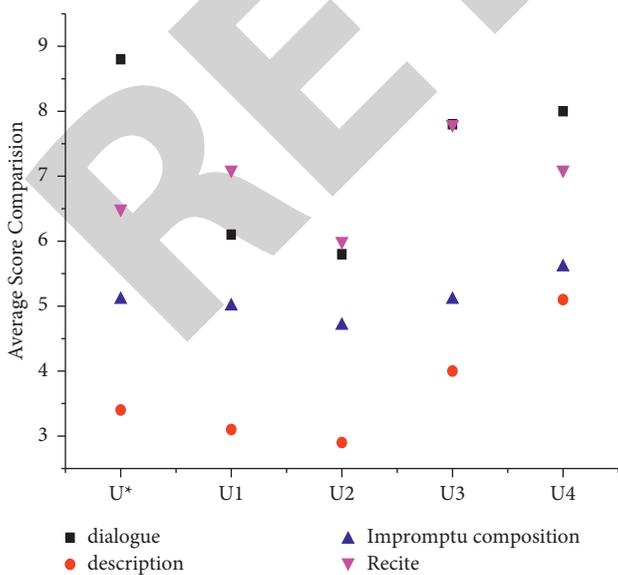


FIGURE 4: Dot graph of average score comparison between machine evaluation and human evaluation.

selection and association rule mining of automatic scoring of oral English computer test. The optimization results of sequence dynamic selection were obtained, as shown in Figure 6.

The results showed that this method could effectively realize the dynamic selection and clustering of oral English pronunciation sequences and improve the recognition ability of oral pronunciation. The commonly used speaker clustering evaluation index is speaker clustering error rate (DER), which compares the reference speaker tag segment with the tag segment predicted by the system to obtain DER.

In order to further prove the effectiveness of the proposed method, the normalized root mean square error of the proposed method was compared with that of literature [18] and literature [22], so as to test the accuracy of automatic scoring of oral English computer test by different methods. The calculation process of normalized root mean square error was as follows:

$$R_e = \sqrt{\frac{\sum d_i^2}{n}} \tag{14}$$

TABLE 2: Average difference and consistency results of machine evaluation and human evaluation.

Types of oral test questions	Human evaluation and machine evaluation	Dialogue	Description	Impromptu composition	Recite
Complete consistency between human evaluation and machine evaluation/%	U* and U1	0.34	21.24	12.23	34.71
	U* and U2	43.91	3.49	3.14	81.31
	U* and U3	43.55	2.41	43.01	21.08
	U* and U4	28.71	56.23	86.26	19.75
Proximity consistency between human evaluation and machine evaluation/%	U* and U1	14.91	74.24	73.45	29.12
	U* and U2	24.34	50.12	81.34	43.42
	U* and U3	81.4	38.23	48.01	34.11
	U* and U4	4.12	47.13	74.85	42.22
Correlation coefficient between human evaluation and machine evaluation/%	U* and U1	65.78**	29.48**	14.47**	21.89**
	U* and U2	85.67**	92.9**	74.17**	21.87**
	U* and U3	14.1**	18.75**	65.32**	30.24**
	U* and U4	40.11**	31.77**	39.09**	42.87**

** $p < 0.001$.

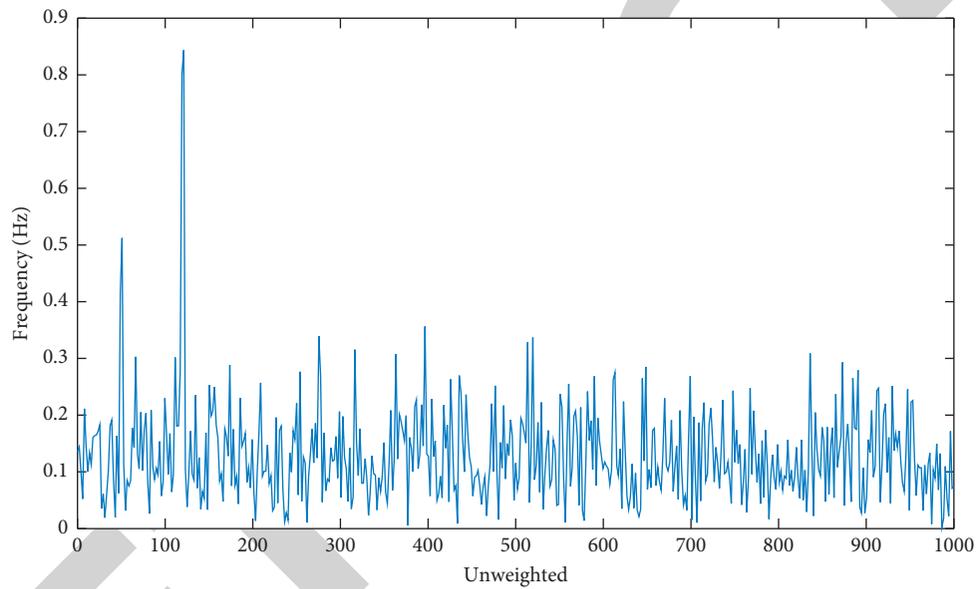


FIGURE 5: Unweighted spectrum.

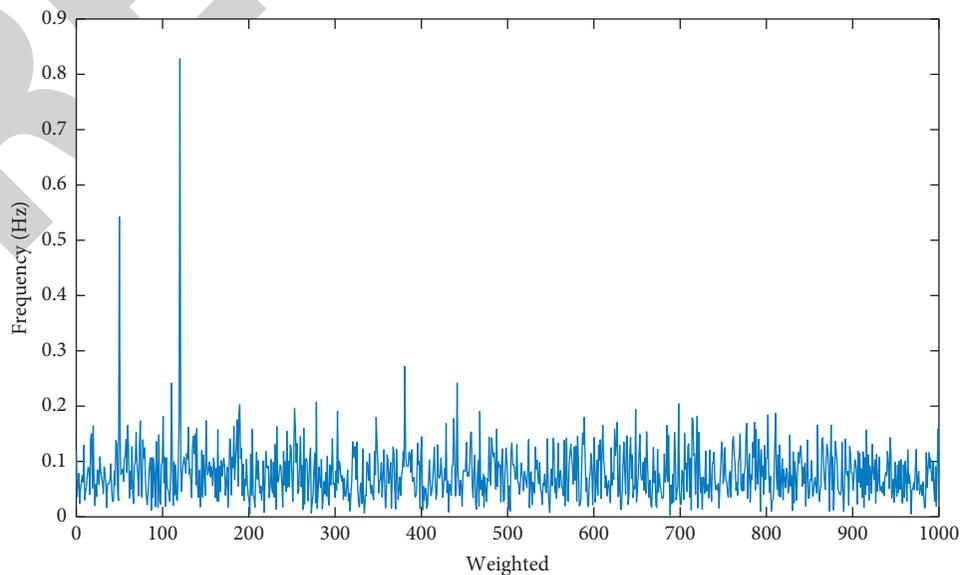


FIGURE 6: Sequence dynamic selection optimization spectrum.

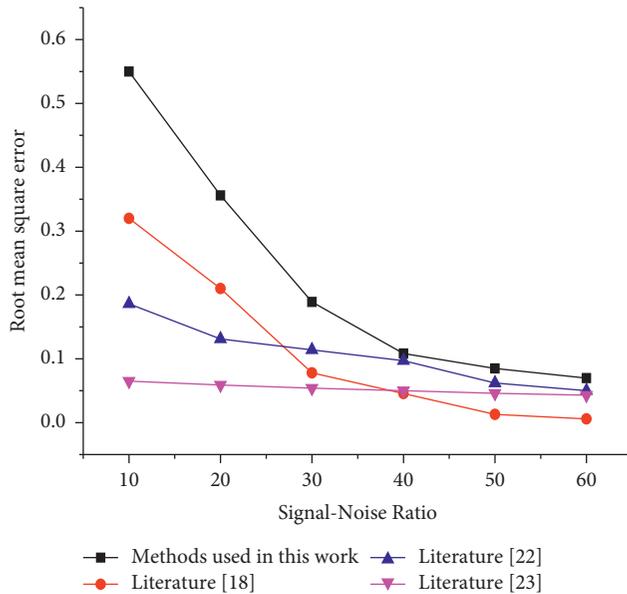


FIGURE 7: Root mean square error test results.

In the above, n was the number of measurements, and d_i was the deviation between a set of measured values and true values. The comparison results are shown in Figure 7.

According to the analysis of Figure 7, with the continuous increase of signal-to-noise ratio, the normalized root mean square error of different methods was also decreasing. The normalized root mean square errors of the methods in literature [18], literature [22], and literature [23] were close, but both were larger than the proposed methods. Therefore, it could be proved that the proposed method had low error and good overall scoring performance.

4.3. Accuracy Test Results of Automatic Scoring. In order to test the system performance, the pronunciation sequence signal analysis and scoring test of oral English test were carried out in MyEclipse 8.5 environment [24]. Firstly, the oral pronunciation sequence was collected. The pronunciation sequence signal duration of the oral English test was 500 s. The frequency modulation bandwidth of oral sequence detection was 120 kHz, the initial frequency of detection was $f_0 = 20$ kHz, and the number of candidate pitches per frame was 2000. The marking interval was 0.25 s, and the marked oral pronunciation melody pitch interval was 0.35 s. According to the above simulation parameter settings, the automatic scoring analysis of the oral English test was carried out to obtain the original voice signal acquisition results, and the automatic scoring of the oral English test was carried out to obtain the scoring accuracy test results, as shown in Figure 8.

According to the analysis of Figure 8, the system could effectively realize the automatic scoring of oral English, the scoring results were accurate and reliable, and the stability of the system was great.

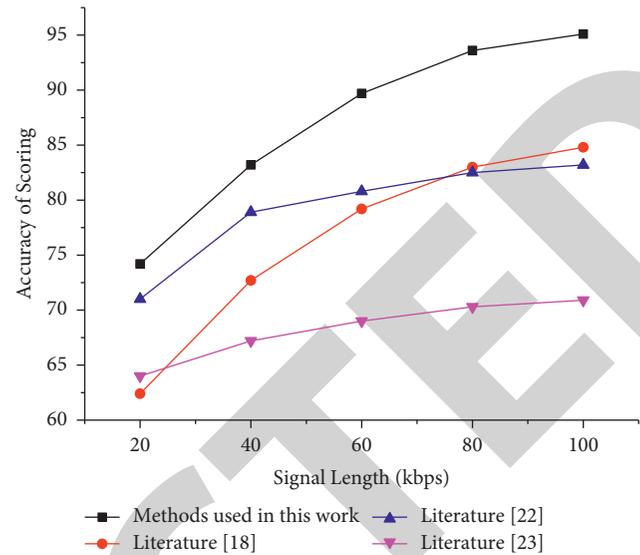


FIGURE 8: Accuracy results of automatic scoring in oral English test.

5. Conclusion

Based on the principle of sequence matching, this study optimized the design of the automatic scoring system for the oral English test, improved the ability of oral English automatic test and scoring, and also improved the objectivity of the automatic scoring system. The time series analysis method was used to collect the spoken English speech signal, and the spoken English speech feature extraction method was used to extract the features of the collected spoken English pronunciation sequence. Based on the pronunciation sequence distribution signal, the collected signal model was modeled by the continuous digital speech recognition method. According to the characteristics of the signal model, an automatic scoring system model based on sequence matching was designed to realize the automatic scoring sequence of the oral English computer test and improve the automatic scoring ability of the oral English computer test. Based on the results of machine evaluation and human evaluation, it could be seen that this method had advantages in improving the accuracy of the automatic test score of oral English.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

No conflicts of interest exist concerning this study.

Acknowledgments

This paper was not funded by any organization.

References

- [1] N. Lewandowski and M. Jilka, "Phonetic convergence, language talent, personality and attention," *Frontiers in Communication*, vol. 4, p. 18, 2019.
- [2] O. Dušek, J. Novikova, and V. Rieser, "Evaluating the state-of-the-art of end-to-end natural language generation: the e2e nlg challenge," *Computer Speech & Language*, vol. 59, pp. 123–156, 2020.
- [3] L. Guillou, C. Hardmeier, E. Lapshinova-Koltunski, and S. Loáiciga, "A pronoun test suite evaluation of the English–German MT systems at WMT 2018," in *Proceedings of the Third Conference on Machine Translation: Shared Task Papers*, pp. 570–577, Belgium, Brussels, October 2018.
- [4] D. Fogerty, R. Madorskiy, J. B. Ahlstrom, and J. R. Dubno, "Comparing speech recognition for listeners with normal and impaired hearing: simulations for controlling differences in speech levels and spectral shape," *Journal of Speech, Language, and Hearing Research*, vol. 63, no. 12, pp. 4289–4299, 2020.
- [5] H.-T. D. Huang, S.-T. A. Hung, and L. Plakans, "Topical knowledge in L2 speaking assessment: comparing independent and integrated speaking test tasks," *Language Testing*, vol. 35, no. 1, pp. 27–49, 2018.
- [6] W. Pouw, J. P. Trujillo, and J. A. Dixon, "The quantification of gesture–speech synchrony: a tutorial and validation of multimodal data acquisition using device-based and video-based motion tracking," *Behavior Research Methods*, vol. 52, no. 2, pp. 723–740, 2020.
- [7] A. Gardiner, "Motivating EFL students with conversation data," *Journal of Language Teaching and Research*, vol. 9, no. 6, pp. 1138–1146, 2018.
- [8] V. P. Karunaimathi, D. Gladis, and D. Balakrishnan, "Estimation of voice perturbation measures using signal processing algorithms," *International Journal of Engineering Research and Technology*, vol. 13, p. 3806, 2020.
- [9] Y. S. Moon and I. C. Choi, "Salient linguistic features of EFL learner spoken corpus elicited by a computerized speaking test," *Multimedia-Assisted Language Learning*, vol. 22, no. 3, pp. 54–58, 2019.
- [10] M. A. Yeo, "Conversation with aek phakiti on calibration, online writing assessment, automated scoring and quantitative research," *RELC Journal*, vol. 50, no. 3, pp. 483–490, 2019.
- [11] R. A. Naqvi, M. Arsalan, A. Rehman, A. U. Rehman, W.-K. Loh, and A. Paul, "Deep learning-based drivers emotion classification system in time series data for remote applications," *Remote Sensing*, vol. 12, no. 3, p. 587, 2020.
- [12] M. Floegel, S. Fuchs, and C. A. Kell, "Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control," *Nature Communications*, vol. 11, no. 1, pp. 2839–2912, 2020.
- [13] C. Oh, S. Han, and J. Jeong, "Time-series data augmentation based on interpolation," *Procedia Computer Science*, vol. 175, pp. 64–71, 2020.
- [14] L. Chen, K. Zechner, S. Y. Yoon et al., "Automated scoring of nonnative speech using the speechrater sm v. 5.0 engine," *ETS Research Report Series*, vol. 2018, no. 1, 31 pages, Article ID 12198, 2018.
- [15] L. Gu, L. Davis, J. Tao, and K. Zechner, "Using spoken language technology for generating feedback to prepare for the TOEFL iBT test: a user perception study," *Assessment in Education: Principles, Policy & Practice*, vol. 28, no. 1, pp. 58–76, 2021.
- [16] I. R. Katz, "Current research in cognitive science at educational testing Service (ETS)," in *Proceedings of the CogSci, 2016*, Philadelphia, PA, USA, August 2016.
- [17] B. K. Iwana and S. Uchida, "An empirical survey of data augmentation for time series classification with neural networks," *PLoS One*, vol. 16, no. 7, Article ID e0254841, 2021.
- [18] J. J. Moreno Escobar, O. Morales Matamoros, R. Tejeida Padilla et al., "Biomedical signal acquisition using sensors under the paradigm of parallel computing," *Sensors*, vol. 20, no. 23, Article ID 6991, 2020.
- [19] S. Dong, Z. G. Wu, Y. J. Pan, H. Su, and Y. Liu, "Hidden-Markov-model-based asynchronous filter design of nonlinear Markov jump systems in continuous-time domain," *IEEE transactions on cybernetics*, vol. 49, no. 6, pp. 2294–2304, 2018.
- [20] S. Lim, S. J. Kim, Y. J. Park, and N. Kwon, "A deep learning-based time series model with missing value handling techniques to predict various types of liquid cargo traffic," *Expert Systems with Applications*, vol. 184, Article ID 115532, 2021.
- [21] P. Kuwalek, "AM modulation signal estimation allowing further research on sources of voltage fluctuations," *IEEE Transactions on Industrial Electronics*, vol. 67, no. 8, pp. 6937–6945, 2019.
- [22] A. Cherif, "DSP real-time implementation of DOST algorithm used for speech enhancement," in *Proceedings of the 8th International Conference on Sciences of Electronics, Technologies of Information and Telecommunications (SETIT'18)*, vol. 147, no. 2, p. 77, Maghreb, Tunisia, December 2019.
- [23] V. Arora, A. Lahiri, and H. Reetz, "Phonological feature-based speech recognition system for pronunciation training in non-native language learning," *Journal of the Acoustical Society of America*, vol. 143, no. 1, pp. 98–108, 2018.
- [24] N. Mercer, P. Warwick, and A. Ahmed, "An oracy assessment toolkit: linking research and development in the assessment of students' spoken language skills at age 11-12," *Learning and Instruction*, vol. 48, pp. 51–60, 2017.
- [25] T. Grubljesic, P. S. Coelho, and J. Jaklic, "The shift to socio-organizational drivers of business intelligence and analytics acceptance," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 37–64, 2019.
- [26] L. Z. Zhang, M. Mouritsen, and J. R. Miller, "Role of perceived value in acceptance of "bring your own device" policy," *Journal of Organizational and End User Computing*, vol. 31, no. 2, pp. 65–82, 2019.
- [27] A. Shahri, M. Hosseini, K. Phalp, J. Taylor, and R. Ali, "How to engineer gamification: the consensus, the best practice and the grey areas," *Journal of Organizational and End User Computing*, vol. 31, no. 1, pp. 39–60, 2019.
- [28] Z. Cai and X. Zheng, "A private and efficient mechanism for data uploading in smart cyber-physical systems," *IEEE Transactions on Network Science and Engineering*, vol. 7, no. 2, pp. 766–775, 2020.
- [29] H. Li, T. Wang, W. Pan et al., "Mining key classes in java projects by examining a very small number of classes: a complex network-based approach," *IEEE Access*, vol. 9, pp. 28076–28088, 2021.