Retraction

Retracted: Automatic English Pronunciation Evaluation Algorithm Based on Sequence Matching and Feature Fusion

Mathematical Problems in Engineering

This article has been retracted by Hindawi following an investigation undertaken by the publisher [1].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] N. Liu, "Automatic English Pronunciation Evaluation Algorithm Based on Sequence Matching and Feature Fusion," *Mathematical Problems in Engineering*, vol. 2022, Article ID 4785355, 8 pages, 2022.

Research Article

Automatic English Pronunciation Evaluation Algorithm Based on Sequence Matching and Feature Fusion

Nian Liu

Foreign Languages College of Urumqi Vocational University, Urumqi, Xinjiang, China

Correspondence should be addressed to Nian Liu; 18407255@masu.edu.cn

This article focuses on the question answering type of automatic scoring system for large-scale spoken English examinations and scores using a method called multifeature fusion. Three types of features are extracted to score using speech recognition text as the research object. The three types of features are similarity, syntactic, and phonetic. There are nine distinct characteristics that describe the relationship between examinee responses and expert ratings. Manhattan distance is improved as a measure of similarity in the similarity feature. Simultaneously, a feature of keyword coverage based on editing distance is proposed, and the phenomenon of word variation in text recognition is fully considered, in order to provide examinees with an objective and fair score. To obtain the machine score, all extracted features were fused using a multiple linear regression model. The experimental results demonstrate that the extracted features are extremely effective for machine scoring, with the system scoring performance of the examinee as a unit equaling 98.4 percent of expert scoring performance.

1. Introduction

English instruction consists primarily of four components: listening, speaking, reading, and writing. Currently available English instruction focuses primarily on reading and writing, while English tests are generally biased toward listening, reading, and writing. As the most critical words in daily communication are frequently overlooked by individuals, certain characteristics of spoken English are frequently overlooked by Chinese English learners, resulting in these characteristics becoming a bottleneck to improving spoken English. On the one hand, oral English learners are frequently embarrassed to learn in front of others. On the other hand, because different individuals may evaluate the same oral English differently, computer-assisted oral English learning becomes the best option for these learners. The algorithm for computer-assisted spoken language learning is based on computer-assisted spoken language evaluation [1].

Spoken English includes many features. The quality of a person's oral English is usually evaluated from many aspects, and the evaluation result is the synthesis of many characteristics. The basis of synthesis is the study of individual

features. Linking is an important feature of spoken English, which reflects the coherence of learners' spoken English. Confusable sounds are another important part of spoken English, reflecting learners' pronunciation level. Based on the study of the features of linking and confusable sounds, the evaluation results of multiple features are fused into comprehensive evaluation results by using data fusion technology. The study of multifeature fusion assessment method will help learners to improve their oral English level comprehensively. Based on this idea, this paper mainly studies the multifeature fusion evaluation algorithm in spoken English.

It is evaluated mainly by the pronunciation given to a person by a computer. For example, the current spoken Mandarin test system is not only accurate but also greatly improves efficiency and saves manpower.

There are two kinds of scoring for speaking questions: one is from the perspective of pronunciation, the other is from the perspective of the text. Purely speech-based scoring involves acoustical features such as pronunciation, frequency, and prosody. However, this method can achieve better results by limiting the type of questions. But for the

open-ended questions, it is a little hard to do. If a scoring system scores only from a phonetic point of view, regardless of the grammatical structure of the content expressed, then the automatic scoring system is still not a complete ideal system. Therefore, scoring from the perspective of text will become an important supplement to the oral scoring system. Literature [2] scores the retelling questions of college English speaking test from the perspective of grammar. The data are manual copied texts, and the features used include the number of words, phrase repetition, height of grammar analysis tree, and score of probabilistic context-free grammar [3] and from the perspective of syntactic complexity, He Literature uses TOEFL test data obtained from speaking test centers, including replicated data as a training set, and sequentially replicated and recognized data as a test set. This method can effectively improve the translation speed of sentences.

Since the question type of this paper is question answering, the question answering type based on pure text is also of reference significance to our scoring, while the purely text-based question and answer scoring is mainly based on the question and answer questions in the written examination of students [4]. For example, short-answer questions require students to write their own correct answers according to the questions. The data used by the system are either copied from the examinee's test paper data or obtained directly from the computer test electronic text. These data are usually preprocessed before use, that is, spelling correction, for further analysis. These systems include ETS's CRater, which maps a candidate's answers to a template of reference answers and grades them for accuracy. The UCLES system studied by Oxford University in the UK uses two method diagrams. The first one uses information extraction to extract features from candidates' answers and matches them with artificially constructed answer templates. The second method uses text classification technology to classify the answers of ungraded examinees and determine which grade of the answer is closest to the artificially constructed reference answer and then determine the score of that grade.

This paper attempts to use the text data after speech recognition as the research object and adopts the method of multifeature fusion, in which multiple features are mainly extracted from three aspects: similarity related to reference answers, features, syntactic features related to grammar, and features related to speech. The fusion uses multiple linear regression model. Finally, the machine score is compared with the expert score to analyze the system performance.

The method of speech signal processing is used to enhance the English oral examination [5, 6]. The algorithm for designing an automatic scoring system and time series analysis are used to improve the oral English of voice signal acquisition, the acquisition of spoken English pronunciation sequence decision feedback equalization regulation method for voice mixing processing, and the extraction of spoken English pronunciation sequence statistics characteristics. Correlation function feature matching is used to improve the evaluation and comparison of oral English pronunciation sequence states, to extract oral English pronunciation sequence spectrum characteristics, to extract oral English

pronunciation characteristics and standard pronunciation characteristics match analysis, and to realize the automatic test of spoken English and ratings based on the comparison of the differences. The algorithm for matching the pronunciation sequences of spoken English is loaded into the hardware module, the system's hardware design is combined with DSP, and the validity conclusion is obtained.

2. Related Work

2.1. Development of Oral English Assessment at Home and Abroad. With the development of speech recognition and natural language processing technologies, virtual language teachers who evaluate students' oral English quality in real time and make suggestions for improvement are becoming a new research hotspot in the fields of natural language processing and speech recognition. Status of foreign research.

Some evaluation systems use speed to evaluate spoken language, which has led to ways of cheating the system by using a higher speed. This problem can be solved by using the similarity ratio [5]. The assessment of spoken language at the phoneme level has also made great progress. S. M. Witt's assessment of spoken language at the level of phonemes can accurately locate the errors of users' pronunciation, evaluate the similarity between users' spoken language and target speech, and find out the differences through comparison [8].

Machine learning techniques are used to aggregate features associated with human recognition in the nonacoustic domain [9]. Additionally, advances have been made in the fields of stress detection [10], speech error detection [11], and prosody [12]. Oral English learning via the web has also grown in popularity [13], as has hardware based on speech evaluation technology [14].

Domestically, this area of research began late but made some strides. The majority of existing research employs speech signal processing and hidden Markov models to analyze and evaluate the similarity between users' spoken language and standard speech [6]. Oral evaluation has also incorporated speech verification and signal cutting [15]. The literature [16] discusses methods for developing robust speech models and segmenting phonemes.

Oral evaluation has also been conducted using nonlinear analysis, wavelet analysis, and other techniques [17]. In learning data analysis and mining, soft computing technologies and data mining methods based on fuzzy sets, rough sets, and neural networks can be introduced [18].

At the same time, speech verification and speech signal cutting are also introduced into the oral evaluation to assist oral learners in English learning. Speech verification makes use of the credibility of speech verification and rejects incorrect statements to be scored based on this. Speech signal cutting provides a method to cut the speech signal into each phoneme time segment, using pretrained English pronunciation acoustic model as the cutting basis, and then using speech recognition technology to cut the correct pronunciation segment from the appropriate acoustic model; English speech scoring uses the similarity of standard speech and rated speech to score from four aspects: volume
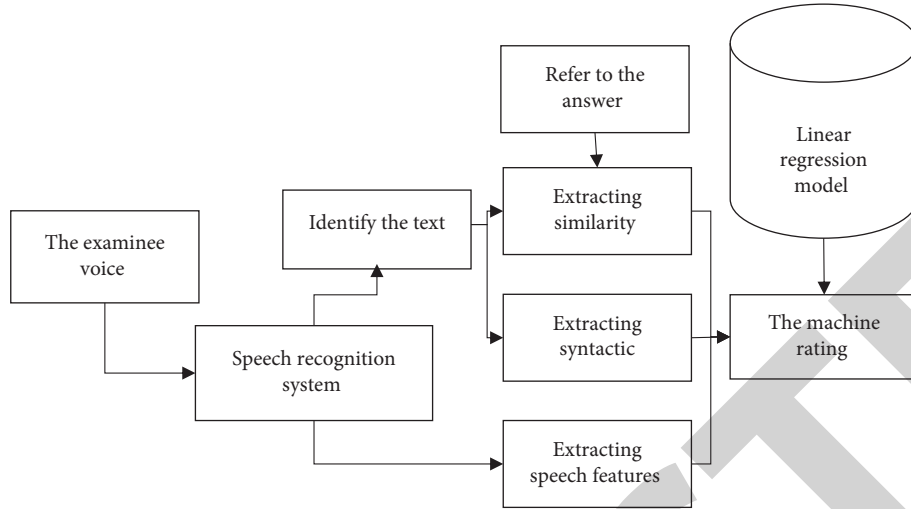
FIGURE 1: Scoring system framework.

intensity curve, fundamental frequency trajectory curve, sharp and slow change of speech, and HMM logarithmic probability difference. Finally, a comprehensive score is given for each aspect, with different weights for each aspect [19].

Literature [20] mainly proposes methods of training robust speech models and phoneme segmentation. For the former, the phoneme set of TIMIT was mapped to the phoneme set of CMU, and the phoneme set was reduced from 60 to 40, which greatly improved the robustness of the model under the condition of insufficient training data. In the aspect of phoneme segmentation, Viterbi decoding is used to segment the phoneme of the statement. Dynamic insertion method is adopted, forcing alignment based on short-pause model and silence model adopted, respectively, and then, the largest probability of the two is selected to determine the insertion sequence. The problem of increasing its own period due to using silence model alignment is successfully solved. This method achieved a correlation of 0.66 at the sentence level with artificial grading [21].

While domestic and international research on oral assessment has produced some promising results, they have primarily focused on the acoustic characteristics of pronunciation and have rarely involved the application of specific grammar or the resolution of specific pronunciation problems for English learners of specific native languages. While waveform comparison or quantitative score feedback can be provided to language learners, providing professional and useful assessment results that enable language learners to improve a specific pronunciation feature of the target language is rarely possible. Indeed, while a variety of factors influence the performance of speech evaluation algorithms, two phonemes stand out, namely, the randomness of natural language and the instability of existing speech recognition systems, which have emerged as the primary impediments to breaking through speech evaluation. Simultaneously, the method for modeling the stability of speech recognition systems and the randomness of natural language is rarely used in speech evaluation systems.

2.2. Introduction to the Scoring System Framework. The whole scoring system is shown in Figure 1, including three parts.

3. Oral English Evaluation Algorithm Based on Sugeno Integral

3.1. Linking Evaluation Algorithm Based on Sugeno Integral. Evaluation results of the three grades of the linking evaluation algorithm based on Sugeno integral: excellent, good, and bad.

(1) Fuzzy measure of belonging to or better than good.

$$\mu(A) = \begin{cases} 0.0, |A| = 0 \\ \left[1 + \left(\dfrac{|A| - \alpha}{\beta}\right)^{-2}\right]^{-1}, 1 \leq |A| \leq L \end{cases}, \quad (1)$$

where L represents the maximum length of the linking cluster in the training corpus, A represents a set of subattributes composed of link placeholders, and α and β are hyperparameters.

(2) Fuzzy measure of degree of excellence

$$\mu(A) = \begin{cases} e^{\dfrac{(|A| - c)^2}{2 \times \lambda^2}}, 1 \leq |A| \leq c \\ 1.0, c \leq |A| \leq L \, L < |A| \end{cases}, \quad (2)$$

where L represents the maximum length of the linking cluster in the training corpus and A represents the set of subattributes composed of link placeholders.

When a link is recognized by the system, the credibility of the link is defined as

$$f(x) = c[Cat(x)]. \quad (3)$$

When a link is missed by the system, the probability of the link actually being pronounced by the trainee depends on A. Therefore, in this case, its credibility is defined as

$$f(x) = \sim \gamma[Cat(x)] = 1 - \sim \gamma[Cat(x)]. \tag{4}$$

3.2. HDP Evaluation Algorithm Based on Sugeno Integral.

The HDP evaluation algorithm based on Sugeno integral gives four evaluation levels: excellent, good, normal, and bad, which are defined as follows:

(1) Fuzzy measure of belonging to or better than normal

$$\mu(A) = \begin{cases} e^{\dfrac{(|A| - c)^2}{2 \times \lambda^2}}, & 1 \leq |A| \leq L \text{ and } p(|A|) < c, \\ 1.0, & p|A| \geq c \text{ or } L < |A| \end{cases} \tag{5}$$

where L represents the length of the HDP cluster and A represents the set of subattributes composed of HDP placeholders.

(2) Fuzzy measure of belonging to or better than good

$$\mu(A) = \begin{cases} e^{\dfrac{(p(|A|) - c)^2}{2 \times \sigma^2}}, & 1 \leq |A| \leq c \text{ and } p(|A|) < c \\ 1.0, & p|A| \geq c \text{ or } L < |A| \end{cases} \tag{6}$$

(3) Fuzzy measure of belonging to excellent.

$$\mu(A) = \begin{cases} e^{\dfrac{(p|A| - c)^2}{2 \times \sigma^2}}, & 1 \leq |A| \leq L \text{ and } p(|A|) < c \\ 1.0, & p(|A|) \geq c \text{ or } L < |A| \end{cases} \tag{7}$$

When phoneme X is correctly identified by the system, the credibility of the phoneme is defined as

$$f(x) = r(x). \tag{8}$$

When phoneme X is identified as the ith phoneme in the same HDP set, then the confidence of the phoneme is defined as

$$f(x) = e_x^i. \tag{9}$$

The specific assessment methods are as follows:

First, calculate the conditional probability average of each phoneme in HDP.

$$P = \frac{1}{n} \sum_{m=1}^{n} P(a_m \mid b_m), \tag{10}$$

where n is the number of phonemes in HDP.

First, determine the reference points. HDP evaluation is divided into four categories, so four reference points are selected. On the other hand, since every phoneme in HDP is

equivalent, each dimension of the reference point should be equal. Therefore, the reference point is selected as follows:

$$B_i = (b_{i1}, b_{i2}, \ldots, b_{in}). \tag{11}$$

In the formula, B_i represents the reference point.

The distance between the point $(h_1, h_2, \ldots, h_n)$ represented by the current HDP and each reference point is then calculated.

$$d_i = \sqrt{(h_1 - b_{i1})^2 + (h_n - b_{in})^2 + \ldots + (h_n - b_{in})^2}, \tag{12}$$

where d_i represents the distance from the ith standard point.

Firstly, the multiple features are evaluated and multiple evaluation results are obtained. These results are then quantified, and the quantification criteria are obtained through systematic training in advance. Then, the weighted average of these quantified results is carried out to obtain the quantified results of the comprehensive evaluation.

$$S = \sum_{i=1}^{n} a_i S_i, \tag{13}$$

where a_i is the weight of each feature.

Linking is to evaluate each linking group and give the evaluation results. Comprehensive assessment requires an assessment of the entire sentence, and a sentence may contain multiple linking groups. Therefore, how to convert the assessment results of multiple linking groups into the assessment results of the whole sentence becomes a problem. Specific methods are as follows:

(1) Evaluate each linking group and determine the number of linking groups n

(2) Quantify the assessment results of each linking group

(3) Use some method to synthesize the quantitative results (such as arithmetic average) to obtain the quantitative evaluation result of the whole sentence

$$S_L = \frac{1}{n} \sum_{i=1}^{n} S_i, \tag{14}$$

where Si is the quantitative evaluation result of the ith linking group.

4. Experimental Results and Analysis

Oral English comprehensive evaluation system, in order to advance parameter optimization and testing, provides an experimental platform. This section will first introduce the experimental method then introduce the training and testing based on probability mean and probability space distance. Finally, this chapter will introduce and analyze the training and testing methods for an oral English evaluation algorithm based on multifeature fusion.

The experiment is divided into two parts, training and testing. The training corpus consists of three groups: the standard spoken recording (T1), the spoken recording of

students with good English pronunciation (T2), and the spoken recording of students with average English pronunciation (T3). Each group contains the same 50 sentences. The test corpus consists of three groups: the standard spoken recording (TE1), the spoken recording of students with good English pronunciation (TE2), and the spoken recording of students with average English pronunciation (TE3). Each group contains the same 57 sentences.

In Figure 2, L is the size of the training or test set; E/L reflects the performance of training or testing, and experimental method then introduce the training and testing based on probability mean and probability space distance. From the experimental results, the performance of the test results is 16.4% lower than that of the training results, indicating that the stability of the link evaluation algorithm based on the Sugeno integral is relatively good [22].

The mean value of the objective function in training and test is 19.64 and 9.84, respectively. In this paper, a group of good parameters is selected as system parameters. The objective function $E = 20$ in the training, and the statistics of evaluation results are shown in Figure 3.

In this paper, the most direct feature selection method is adopted to improve the feature selection based on the system performance. Of all 12 features f, only the keyword coverage based on editing distance is selected. 1 is added to the system in sequence according to the absolute value of correlation coefficient from high to low to judge the influence of features on system performance. The experimental results are shown in Figure 4, and the abscissa in Figures 4(a) and 4(b) represent the number of features added to the system in sequence. The ordinate represents the correlation coefficient and differential of system performance, respectively. As can be seen from the figure, the overall trend of the correlation coefficient between machine score and expert score is increasing, while the overall trend of difference is decreasing. However, after the 3rd, 10th, and $L1$ features were added into the system, the system performance basically did not change, indicating that these features did not make positive contributions to the system. These three features are removed and the process is repeated until the system is stable. Finally, the system deleted the cosine similarity based on word frequency, the whole syntax tree score and the depth of the syntax tree, and the remaining 9 features. Their impact on system performance is shown in Figure 5. It can be seen that the influence of these 9 features on the system is monotonically rising or decreasing, whether from the correlation coefficient or the difference, that is, redundant features have been removed.

Finally, the system performance obtained before and after feature selection is compared with the expert rating performance as shown in Table 1. It can be seen that the system performance is improved after redundant features are deleted. Finally, the machine scoring performance of the nine features reached 98.4% of the expert scoring performance.

In this paper, the above methods are applied to extract similarity features, syntactic features, and phonetic features of examinee samples, and the corresponding features of all
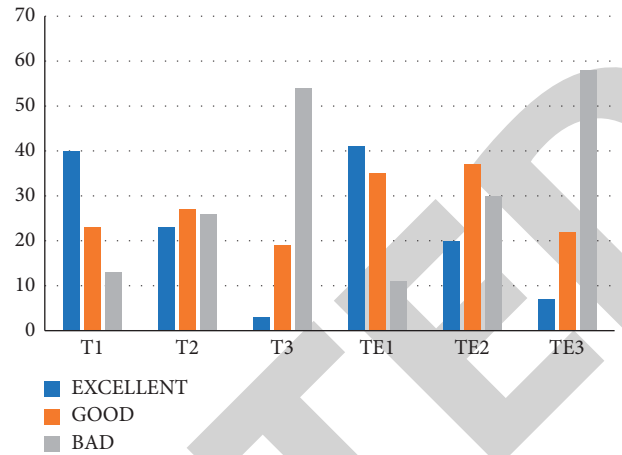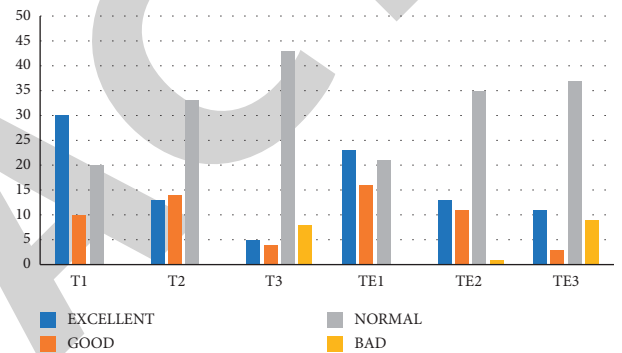

FIGURE 2: Results of linking experiment.

FIGURE 3: Confusable sound training results based on probability mean.

five questions of each examinee are added up for correlation comparison with the expert score, which refers to the average score of six examinees on the same sample. Table 2 shows the improvement of two characteristics with the traditional method of correlation coefficient of the list; as can be seen from the table in this paper, the improved similarity characteristics than the traditional Manhattan cosine similarity based on word frequency characteristics of the correlation coefficient is high, and based on edit distance proposed in this paper, keywords coverage is better than the traditional keywords coverage. Table 2 is the list of correlation coefficients between all features mentioned in this paper and expert ratings. It can be seen from the table that in addition to the high correlation coefficients between similarity features and expert ratings, syntactic features and phonetic features also have relatively high correlation coefficients, indicating that these features are highly consistent with expert ratings.

In this case, true matches is the actual match point verified manually. Selected matches is the ratio threshold of the nearest neighbor distance and the second neighbor distance and the geometrically verified feature point. The matching accuracy and time complexity are listed in Tables 3 and 4.
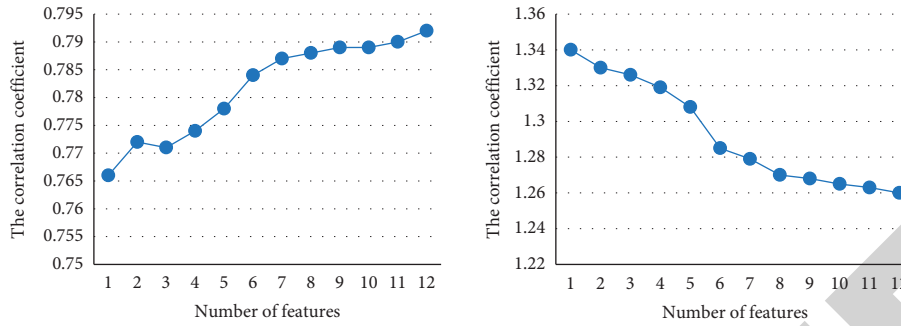
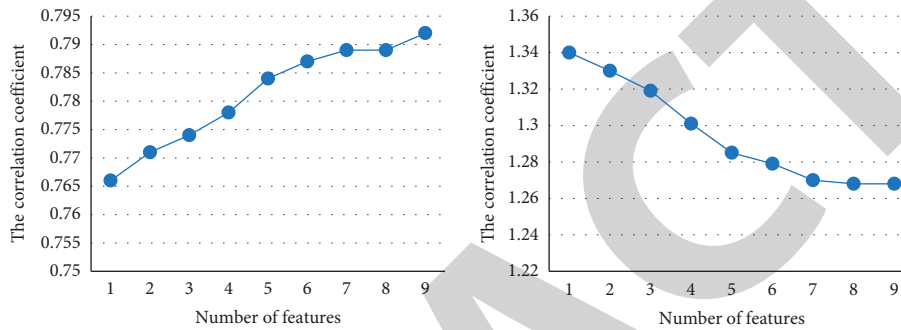FIGURE 4: Performance distribution of the system before feature selection.

FIGURE 5: Performance distribution of the system after feature selection.

TABLE 1: Comparison of machine scoring performance and expert scoring performance.

	The correlation coefficient	Difference of average
The machine scores 12 features	0.7918	1.2611
The machine scores 9 features	0.792	1.2607
Expert rating performance	0.805	1.221

TABLE 2: Correlation coefficients between all features and expert ratings.

Characteristics of the	Correlation coefficient with expert rating
Cosine similarity based on word frequency	0.753
Jaccard similarity	0.739
Dice coefficient	0.767
Manhattan similarity	0.762
Keyword coverage based on edit distance	0.646
The entire syntax tree is scored	−0.304
Syntactic tree score by sentence normalization	−0.381
Syntactic tree score normalized by word	0.051
Normalized syntax tree depth	0.386
The depth of the syntax tree	0.305
Sentence confidence score	0.631
The number of words per unit time	0.52

TABLE 3: Comparison table of matching accuracy.

Category	AS	AS + RSP	PCA	KPCA (cosine)	Our method
All Souls	92.2	84.2	95.5	26.2	100
Ashmolean	84	100.0	82.4	30.0	100
Balliol	100.0	100	100.0	100.0	99.7
Bodleian	8.9	100.0	57.8	100	100.0
Christ Church	88.6	99.1	86.8	62.2	98.6
Cornmarket	100.0	100.0	95.7	56.7	84.6
Hertford	100.0	100.0	100.0	100	100.0
Magdalen	72.7	84	76.7	27.3	72
Pitt Rivers	100	100.0	100	100.0	l00.0
Radcliffe Camera	89.0	90.3	100.0	40.8	100.0

method and then conduct training and testing separately using the Sugeno integral, the probability of mean, and an easy-to-confuse sound evaluation algorithm based on probability space distance. Finally, this chapter will describe and analyze the training and testing procedures for an algorithm for evaluating oral English using multifeature fusion.

Oral English comprehensive evaluation system provides an experimental platform for parameter optimization and testing. This chapter will introduce the experiment

TABLE 4: Time complexity comparison table.

Category	AS	AS + RSP	PCA	KPCA (cosine)	Our method
All Souls	7.3	7.73	7.63	15.66	9.53
Ashmolean	6.16	6.54	6.08	10.87	7.38
Balliol	7.43	8.04	7.98	20.71	12.76
Bodleian	8.36	8.81	8.67	20.93	13.1
Christ Church	7.21	7.03	7.47	18.12	10.73
Cornmarket	4.41	4.48	4.58	11.93	7.68
Hertford	7.89	8.34	8.09	18.99	12.27
Magdalen	0.7	0.76	0.77	0.71	0.7
Pitt Rivers	6.76	6.36	6.6	14.82	8.74
Radcliffe Camera	4.23	4.09	4	9.04	6.1

5. Conclusion

Oral pronunciation is a critical component of oral English learning and also serves as a barometer of one's English proficiency. Linking and confusable sounds are critical oral pronunciation skills. If they are mastered proficiently, they will aid English-language learners in improving their oral comprehension. The popular method of CALL computer-assisted language learning (Computer Assisted Language Learning) is to use an automatic speech recognition system to evaluate oral English pronunciation. The method's primary drawbacks are natural language's randomness and the instability of the current automatic speech recognition system. These issues also complicate the process of developing a satisfactory oral evaluation system. The same issue arises when assessing linking in spoken English. As a result, this paper focuses on the evaluation of linking and confusable sounds in spoken English. While this article focuses exclusively on the assessment of linking and confusable sounds in spoken English, there are numerous other aspects of spoken English to which learners should pay attention, including prosody and stress. Simultaneously, the application of the speech evaluation algorithm to a network-based distributed speech evaluation system is also a worthwhile research direction.

Data Availability

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

Conflicts of Interest

The author declares that he has no conflicts of interest.

References

[1] R. Hinkcs, "Speech recognition for language teaching and evaluating: a study of existing commercial products," in *Proceedings of the ICSLP*, pp. 733–736, Stockholm, Sweden, January 2002.

[2] S. Seneff, C. Wang, and J. Zhang, "Spoken conversational interaction for language learning," in *Proceedings of the In-STIL Symposium on Computer Assisted Language Learning*, pp. 151–154, Venice, Italy, June 2004.

[3] D. K. Jonathan, A. S. Mark, A. W. Robert, and J. S. Robert, "The minitary language tutor (MILT) program: an advanced authoring system," *Computer Assisted Language Learning*, vol. 11, no. 3, pp. 265–287, 1998.

[4] L. Neumeyer, H. Franco, M. Weintraub, and P. Price, "Automatic TextIndependent pronunciation scoring of foreign language student speech," *Proc. of ICSLP*, vol. 96, pp. 1457–1460, 1996.

[5] K. Yoon, H. Franco, and L. Neumeyer, "Automatic pronunciation scoring of specific phone segments for language instruction," *Eurospeech'97,Rhodes, Greece September*, vol. 22-25, pp. 645–648, 1997.

[6] M. Mohler, R. Bunescu, and R. Mihalcea, "Learning to grade short answer questions using semantic similarity measures and dependency graph alignments," in *Proceedings of the 49th annual meeting of the association for computational linguistics: Human language technologies*, pp. 752–762, Portland, Oregon, June 2011.

[7] H. Franco, L. Neumeyer, Y. Kim, and O. Ronen, "Automatic pronunciation scoring for language instruction," *Proc. of ICASSP*, vol. 97, pp. 1471–1474, 1997.

[8] S. M. Witt and S. J. Young, "Phone-level pronunciation scoring and assessment for interactive language learning," *Speech Communication*, vol. 30, no. 2-3, pp. 95–108, 2000.

[9] H. Franco, L. Neumeyer, V. Digalakis, and O. Ronen, "Combination of machine scores for automatic grading of pronunciation quality," *Speech Communication*, vol. 30, no. 2-3, pp. 121–130, 2000.

[10] J.-C. Chen, J.-S. Roger Jang, J.-Yi Li, and M.-C. Wu, "Automatic pronunciation assessment for Mandarin Chinese. Multimedia and expo," *ICME*, vol. 04, pp. 1979–1982, 2004.

[11] J.-C. Chen, J.-L. Lo, and J. S. R. Jang, "Computer assisted spoken English learning for Chinese in taiwan," *Chinese Spoken Language Processing*, in *Proceedings of the 2004 International Symposium*, pp. 337–340, Hong Kong, China, December 2004.

[12] A. C. Lindgren, M. T. Johnson, and R. J. Povinelli, "Joint frequency domain and reconstructed phase space features for speech recognition," in *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing*, no. 1, pp. 533–536, Montreal, CA, USA, May 2004.

[13] S. Mitra, S. K. Pal, and P. Mitra, "Data mining in soft computing framework: a survey," *IEEE Transactions on Neural Networks*, vol. 13, no. 1, pp. 3–14, 2002.

[14] J. Llinas and D. L. Hall, "An introduction to multi-sensor data fusion," *Proceedings of the 1998 IEEE International Symposium on*, vol. 85, pp. 537–540, 1998.

[15] Y. He, C. s Zhang, X. m Tang, and X.-j.J.-h. Chu, "Coherent integration loss due to pulses loss and phase modulation in passive bistatic radar," *Digital Signal Processing*, vol. 23, no. 4, pp. 1265–1276, 2013.

[16] L. Ferrer, E. Shriberg, and Andreas Stolcke, "A prosody-based approach to end-of-utterance detection that does not require speech recognition," vol. 1, pp. 608–611, 2003.

[17] D. Yang and Y. Yuzo, "Multi-sensor data fusion and its application to industrial control," in *Proceeding of the 39th SICE Annual Conference*, pp. 254–561, Iizuka, Japan, July 2000.

[18] M. Sugeno, "Theory of fuzzy integrals and its application," *Doctoral Thesis*, , pp. 5–20, Tokyo Institute of Technology, 1974.

[19] M. A. Govoni, H. Li, and J. A. Kosinski, "Range-Doppler resolution of the linear-FM noise radar waveform," *IEEE*

Transactions on Aerospace and Electronic Systems, vol. 49, no. 1, pp. 658–664, 2013.

[20] W. Tushar, D. Smith, T. A. Lamahewa, and J. Zhang, "Non-cooperative power control game in a multi-source wireless sensor network," in *Proceedings of the 2012 Australian Communications Theory Workshop (AusCTW)*, pp. 43–48, IEEE, Wellington, New Zealand, January 2012.

[21] G. Zhang, P. Liu, and E. Ding, "Energy efficient resource allocation in non-cooperative multi-cell OFDMA systems," *Journal of Systems Engineering and Electronics*, vol. 22, no. 1, pp. 175–182, 2011.

[22] P. Li, H. Zhang, and S.-B. 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, pp. 1–10, Article ID 3018285, 2021.