# Construction of College English Teaching Platform Based on Artificial Intelligence

Xiaqing Ma[1,*] and Xiaoyun Li[2]

[1]Basic Teaching Office, Shaanxi Fashion Engineering University, Xi'an, Shaanxi 712046, China, mxqing@snfeu.edu.cn
[2]Department of Theoretical Linguistics, University of Szeged, Szeged H-6722, Hungary, lixiao@hung.u-szeged.hu

Corresponding author: Xiaqing Ma, mxqing@snfeu.edu.cn

**Abstract.** It will strengthen the cultivation of students' basic English ability, and continuously enhance students' abilities in speaking, translation, expression, writing, etc., can they create a broader development space for students and achieve stable and stable students. The study considers the time and space constraints in the construction of the corpus, and provides different analysis efficiency applications for the classification analysis of English teaching. Through the model analysis of college English class, the specific algorithm implementation is classified structurally. The research focuses on helping students to analyze and help the corpus, so as to improve the efficiency of classroom teaching. The fundamental reason for the difficulty in improving students' writing level in English teaching is that the unity of English writing training content cannot be personalized to adapt to different students' writing levels. It is precisely because of the different levels of students, that indiscriminate writing training makes it difficult for students with low-level writing levels to grasp the essence of writing. The foundation itself is not solid, and it is of course difficult to improve later. For students with high-level writing level, the existing writing training may not be able their training, and repeated training in the past makes it difficult for their writing level to be further improved.

**Keywords:** artificial intelligence; CAD; college English; teaching platform.
**DOI:** https://doi.org/10.14733/cadaps.2023.S5.95-105

## 1 INTRODUCTION

It continues to increase, overturning the long-standing traditional foreign language teaching model and changing the original ecological environment teaching tools [1]. Konovalenko et al. [2] provides foreign language teaching. With new resources, intelligent foreign language teaching has become the norm. Laili and Nashir [3] believes that provides intelligent help for practical English teaching courses. Lucero [4] integrates teaching based on information technology and curriculum.

Based on this background, this paper starts with the traditional English teaching structure [5]. expounds the many problems existing in the mode, and analyzes technology, an intelligent teaching platform is constructed, and the It focuses on the construction of corpus and the concrete realization algorithm of spoken language evaluation. It will show a trend of diversified development [6]. Using the functions of artificial intelligence technology, the good expressions in different scenarios. and communication, to provide better support and for the cultivation of language talents There is no innovation in teaching methods, no advancement with the times, and no application of new information science and technology to teaching. The teaching method still adopts it. Such transformation of teaching environment, using artificial intelligence technology to build intelligent learning experience, teaching methods and teaching management environment for learners, and integrating artificial intelligence and computer-assisted language teaching into different aspects of English teaching. A typical AI-based assisted foreign language teaching system is mainly composed of a knowledge module, a student model module, a computer-aided learning teaching module, and a human-machine interface. The knowledge module can be constructed in the form of a corpus. Teaching module, translation teaching module and teaching effect evaluation module, etc.

## 2 STATE OF THE ART

### 2.1 Analysis of the Current College English Teaching Mode

From the nature of the curriculum, English belongs to applied linguistics, that is, the study of language problems related to real life. Qian [7] believes that classroom knowledge explanation must be combined with the cultivation of critical thinking ability. English language ability includes two dimensions: basic language knowledge and language skills. There are many problems in this respect, such as: the content of textbooks is old and unchanged, many textbooks have been used for many years, lack of new ideas, and cannot adapt to the new era. Samiei and Ebadi [8] believes that there is no innovation in teaching methods, no progress with the times, and no new information science and technology applied to teaching. The teaching method still adopts spoon feeding teaching, which has not realized the new teaching mode of student centered and teacher assisted. College English classroom is dominated by English teachers, who not only bear the teaching responsibility.

In listening training, teachers mostly use recording equipment to repeatedly play the established learning materials to improve students' listening level. The training mode is too single to arouse the interest of students. And only a few questions are used to test whether students understand the listening materials. There is a great possibility that the test results will be distorted. Students may only rely on grasping a few words in the listening materials to answer the questions, but they do not really understand the words. what exactly was said.

Shaukat et al. [9] believes that an obvious problem in traditional English teaching is the lack of English pronunciation training, and standardized English pronunciation is actually the key to English learning. You can only listen, read, write and not speak, which is equivalent to learning dumb English. The lack of pronunciation training in traditional teaching is mainly due to the inability to create a face-to-face communication environment for native English pronunciation. The pronunciation of college English teachers is actually uneven, and the pronunciation training environment created by different teachers is also very different. Shen [10] believes that this is the lack of high-quality training for students. Secondly, the topics designed in the pronunciation training materials may not be consistent with the students' actual spoken English level, and it is difficult to stimulate students' interest and actively participate in discussion and communication.

### 2.2 Overview of Artificial Intelligence Technology

The AI was used in expert systems, and then my country began this technology. The purpose of its research is to more comprehensively grasp the essence of intelligence through the research, simulation and development of AI, so as to hope to use relevant machines or equipment to simulate human intelligence. Comparable brand-new intelligent machines. Image recognition, speech recognition, machine translation, speech synthesis, and natural language understanding have been applied to English teaching. The artificial intelligence technology is briefly described as follows:

it will require new data processing methods to have very strong.

(2) Computer vision technology. In general, computer vision can in real time, and then scientifically measure, and then make some specific graphics processing, which can complete the processing work, which is more efficient than using It is more convenient.

(3) Artificial intelligence development so far is speech recognition.

(4) It is necessary to use language and thoughts and emotions that the language wants to express.

(5) More learning knowledge and professional capabilities, reintegrate optimize and improve. The various performances are here.

### 2.3 Application of Artificial Intelligence in College English Teaching

In Figure 1, it us through these AI entities, and students can also understand their own learning status through these AI entities.

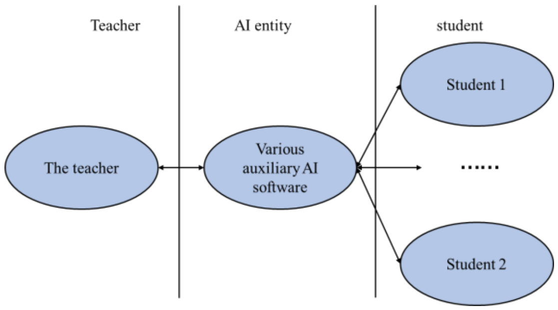


**Figure 1**: Role orientation in college English teaching based on artificial intelligence.

One is to use college English teaching practice, the second is college oral English teaching based on speech recognition technology. Speech recognition is a comprehensive technology covering many aspects, including signal processing technology, computer science technology, etc. it is the one. As the most critical content in English learning, oral English is the way for students to use communication tools proficiently. Diversified dialogue scenarios can also be created for different learning needs, and in dialogue exercises, students' thinking time and actual communication conditions are analyzed, and vocabulary prompts and pronunciation corrections are provided to do it.

The third two aspects of human language, namely speech and text. Text can be divided into multiple levels, including five levels of phonetics, morphology, syntax, meaning and pragmatics.

The fourth is the teaching evaluation and rectification system based on big data. The construction of the teaching evaluation and rectification system is to highlight the tracking and updating of the quality concept of continuous improvement, so that the talent training program can run effectively.

# 3 METHODOLOGY

## 3.1 Construction of College English Corpus Based on Rule Engine

English learning will be continuous accumulation process. The learning of words, phrases and sentences is the foundation. A complete English teaching system should provide a rich and comprehensive word library, phrase library and sentence library as a source of materials for students' daily learning. At present, the recognition of English phrases mainly focuses on the recognition of noun phrases. Generally, the recognition, corpus-based methods and methods that integrate various strategies.

Rule-based methods label phrase boundaries and phrase types according to manually written or automatically acquired grammatical rules. Phrase boundaries are inserted when the phrases in the input sentence satisfy the rules. When there is a conflict, according to the longest matching principle, the longer one. With the emergence and rapid development of corpus technology and the establishment of large-scale corpora, it is possible to obtain knowledge from corpus, and some statistical methods based on corpus have also been produced.

The B1 algorithm expresses more similar they are. The distance equation is calculated as follows:

$$\Delta(x,y)=\sum_{i=1}^{n}\delta(x_i,y_i) \tag{1}$$

X and Y are the feature vectors to be compared, and the $i^{th}$ value in these two vectors is represented as follows:

$$\delta(x_i,y_i)=\begin{cases}0 & if\,(x_i=y_i)\\ 1 & otherwise\end{cases} \tag{2}$$

The IB1-IG algorithm assigns an information gain IG to each feature, and when the eigenvalue is known, the uncertainty of the solution corresponding to the eigenvalue is reduced by the information gain IG.

$$sim(X,Y)=\sum_{i=1}^{n}G(f_i)\delta(x_i,y_i) \tag{3}$$

$$H(D)=-\sum_{p_i}p_i b p_i \tag{4}$$

$$H(D_{|f1|})=-\sum_{v_i\in V}H(D_{|f=w|})\frac{|D_{|f=w|}|}{|D|} \tag{5}$$

$$G(f)=H(D)-H(D_{|f|}) \tag{6}$$

The test template and the reference template. The problem of finding the best path boils down to finding the best path function while satisfying the constraints,

The problem that the acoustic characteristic parameters are not equal in comparison time. The grid point that the path passes through is the frame number for calculating the distance

Such the existing writing training may not be able their training, and repeated training in the past makes it difficult for their writing level to be further improved.

It is of course difficult to improve later. For students with high-level writing level, In this method, for each test item, its eigenvalues must be compared with all relevant eigenvalues in the training set, so the time complexity is high. The mutual information method is a method based on boundary statistics. M.

$$MI(X,Y) = lb\frac{P(X,Y)}{P(X)\times P(Y)} \tag{7}$$

If X and Y appear together more often than they do randomly. The higher the mutual information value, the more likely it is that X and Y form a phrase, and the lower the mutual information value, the more likely there is a phrase boundary between x and Y. Since a phrase does not necessarily consist of only two words, the mutual information inside an n-gram (combination of n symbols) needs to be calculated. The calculation formula of generalized mutual information is:

$$GMI(x_1,\cdots,x_i;y_1,\cdots,y_i) = \sum\frac{1}{\sigma XY}MI(X,Y) \tag{8}$$

Rule method and statistical method have their own advantages and disadvantages, so in practice, these two methods are often used in combination and complement each other. After years of attempts, more and more researchers tend to combine multiple methods and apply different language models to identify noun phrases. This method combines multiple knowledge and multiple methods, and can solve different problems in a targeted manner to obtain better accuracy. Corpus, also known as linguistic material, is a collection of naturally occurring linguistic material. A corpus is a database for research and use that is integrated with a large amount of language information used in real situations. It collects naturally occurring continuous languages according to certain language principles and uses random sampling methods, and uses text or discourse fragments to generate A large-scale electronic text library with a certain capacity is built.

### 3.2 Spoken English Features and Pitch Extraction Methods

Spoken English refers to the English pronunciation of the speaker. The pronunciation waveform of spoken English is irregular for a long period of time, and the waveform parameters change with time, so from a macro perspective, the pronunciation of spoken English is in a non-stationary state. For example, spoken English words are composed of vowels and consonants, and the consonants have no regularity, while the vowel part of English syllables is quasi-periodic, so the base sound can be extracted. The fundamental sound is a periodic feature expressed by vowel sounds in spoken English pronunciation.

The pitch extraction of spoken English signal is mainly divided into time domain method, frequency domain method and time-frequency hybrid method. The time domain method is to directly estimate the spoken English signal to analyze the periodic peak value of the waveform, mainly including the short-term autocorrelation method and the short-term average amplitude difference function method. The human ear's perception of spoken English does not vary purely on a linear frequency scale, but can be expressed in a finite series of frequency bands called Buck's critical frequency bands. Therefore, spectrum bending is required to simulate the characteristics of human hearing when extracting feature parameters.

Before extracting the characteristic parameters of the non-uniform linear prediction cepstrum, the spoken English signal is preprocessed and the calculation steps are as follows:

(1) After each frame of spoken English signal, use the P-order linear prediction method to calculate the channel all-pole transfer function H(z):

$$H(z) = \frac{G}{A(z)} = \frac{G}{1-\sum_{k=1}^{p} a_k z^{-k}} \tag{9}$$

Among them, G is the gain constant, ak represents the k-th order linear prediction coefficient, and P takes the value of 12.

(2) Bark frequency scale approximated by bilinear transformation, the relationship between it and the original linear frequency is:

$$t_k = \arctan\left[\frac{e^{j2\pi k/M} + c}{1 + ce^{j2\pi k/M}}\right] \tag{10}$$

Among them, C is the frequency bending factor determined by the sampling frequency. Considering that there are 22 frequency bands in the Bark spectral domain, M is taken as 88 in this paper.

(3) The LPC spectrum after bending is

$$P(k) = \frac{G^2}{|A(k)|^2} \tag{11}$$

The q-order AR model is used to approximate P(k) to obtain the corresponding autocorrelation function in the channel time domain:

$$r(n) = \frac{1}{M}\sum_{k=0}^{M-1} P(k)\cos(2\pi kn / M) \tag{12}$$

(4) According to the Levinson-Durbin algorithm, a new set of linear prediction coefficients can be obtained according to r(n), from which the L-order cepstral coefficients can be obtained, and this parameter is the NLPC parameter. Considering that the high-value parameters contain more speaker-dependent feature information, the feature parameters of the 5th-order non-uniform linear prediction cepstrum are used to describe the spoken language features when extracting the feature parameters of the spoken English signal.

### 3.3 Evaluation Method of College Spoken English Based on Feature Comparison

During oral English training, there will be a standard pronunciation, which is usually the standard spoken English pronunciation. The spoken language entered by the reader during the follow-up reading is used as the reference spoken language, and the similarity between it and the standard spoken language is compared as the basis for judgment. Specifically, in operation, the feature comparison algorithm is mainly used. The acoustic characteristic parameters of the reader's spoken language are compared with those of the standard spoken English, and the degree of difference between the reader and the standard spoken English is calculated, and the reader's spoken language is scored accordingly.

Such the extraction of feature parameters, which mainly carries out pitch extraction and MFCC extraction for standard spoken language and reader spoken language. The second part compares the features, and uses the commonly used DTW algorithm to calculate the similarity according to the acoustic features extracted above. The third part uses the function mapping method to map the calculated difference distance vector to the objective score, and calculates the correlation coefficient between the score obtained by the algorithm and the foreign language teacher to verify the reliability of the algorithm.

The dynamic time warping algorithm (DTW) is a nonlinear optimization method of pattern matching, which successfully solves the problem that the acoustic characteristic parameters are not equal in comparison time. The grid point that the path passes through is the frame number for calculating the distance between the test template and the reference template. The problem of finding the best path boils down to finding the best path function while satisfying the constraints, so that the following distance is the smallest:

$$\sum_{i=1, j=x(i)}^{N} D[i, j] = \min \sum_{i=1, j=x(i)}^{N} D[i, j] \tag{13}$$

In this way, the distance vector calculation can be performed between the reader's spoken English and the standard spoken English to reflect the difference in characteristics between the two.

The fundamental frequency trace can reflect the stress, intonation and other aspects of spoken English. The reader's spoken English pronunciation is not proficient enough to master the accent and intonation of standard spoken English. Therefore, from the pitch curve, it often shows a change trend that is different from the pitch of standard spoken language. The pitch reflects the vibration frequency information of the human vocal cords. Since the pitch ranges of different readers' oral pronunciation are not the same, it is not reasonable to directly compare the pitch. The mean difference of pitch is calculated between the standard spoken English and the spoken English to reflect the difference in the characteristics of the two, so that the pronunciation of spoken English is more objective and accurate.

The word level uses the following pitch normalization method:

$$\begin{cases} P_{Sa}(i) = \dfrac{1}{N(i)} \sum_{j=1}^{N(i)} (P_S(j) - P_{Sm}(i)) \\ P_{Ra}(i) = \dfrac{1}{M(i)} \sum_{j=1}^{M(i)} (P_R(j) - P_{Rm}(i)) \end{cases} \tag{14}$$

Sentence-level pitch normalization is performed, as shown in the following formula:

$$\begin{cases} P_{Sa} = \dfrac{1}{N} \sum_{j=1}^{N} (P_S(i) - P_{Sm}) \\ P_{Ra} = \dfrac{1}{M} \sum_{j=1}^{M} (P_R(i) - P_{Rm}) \end{cases} \tag{15}$$

Psm(i) the $j^{th}$ point of the $i^{th}$ word in standard spoken language. Ph(i) represents the maximum value of the pitch period of the $i^{th}$ word in standard spoken language. The average pitch difference score between the reader's spoken and standard spoken words at the word level is defined as:

$$S_w = \frac{1}{L} \sum_{i=1}^{L} \left| P_{Sa}(i) - P_{Ra}(i) \right| \tag{16}$$

Since the obtained DTW distance cannot be directly used as an objective score, to map the distance to the 0-100 range, the formula for score mapping is as follows:

$$score = \frac{100}{1 + a(dis)^b} \tag{17}$$

The values of a and b are both greater than 0, and dis represents the obtained characteristic distance. The curve of the distance-to-score mapping expressed by this formula is shown in Figure 2 below, so as to realize the evaluation result of spoken English.

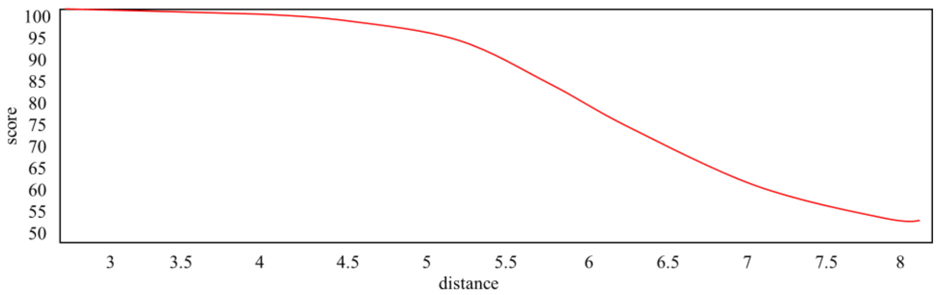


**Figure 2**: Score Mapping Curve.

# 4 RESULT ANALYSIS AND DISCUSSION

## 4.1 Collection and Screening of Experimental Samples

College oral English training mostly adopts the mode of "speaking, reading, and scoring". The sounds played are generally standard American or British pronunciations. Therefore, we use the TIMIT corpus as the training library for standard spoken English. The experiment adopts the HTK toolkit provided by Cambridge University as the research platform of the non-native language adaptive algorithm for the recognition and acquisition of spoken English. The 61 phoneme maps in the TIMIT library are called the 41 phonemes of the CMU dictionary, and the mono-phone model and the state-bound triphone model are trained respectively. Finally, 41 monophonic sub-models are obtained, with a total of 120 states, and each state has 16 Gaussian components of the diagonal covariance matrix. The real tritone model has 1185, a total of 491 states, each state has 8 Gaussian components of the diagonal covariance matrix.

In the algorithm verification of spoken English, the English teacher's score is used to compare the pronunciation with the students' audio data, and the corpus is selected from the TIMIT corpus for pronunciation test.

## 4.2 Experimental Results and Analysis

First, the algorithm test of pitch extraction is carried out according to the algorithm designed in Section 3, and the correlation coefficient between the average pitch difference score and the score given by the teacher is calculated as shown in Figure 3 below.

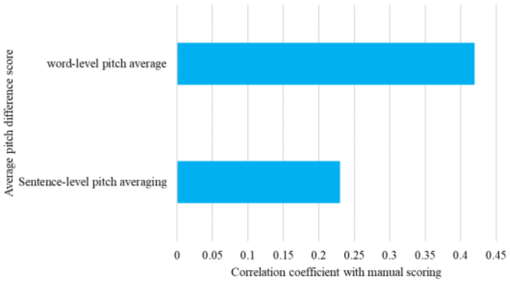


**Figure 3**: The correlation coefficient between the average pitch and the teacher's score (changed to a percentage bar chart).

From Figure 3, it is between the average pitch difference within a word and the teacher's score is significantly higher than the correlation coefficient at the sentence level. This is because the sentence level reflects the overall difference on a macro level, while the word level method A global average is performed to capture sentence-level differences.

Such formula (17) can map the distance to the score between 0 and 100. In order to obtain the unknown parameters a and b in the above formula, a certain score and the obtained distance must be known. The fundamental frequency trace extracted from the standard spoken pronunciation Can I have your order is shown in figure 4:

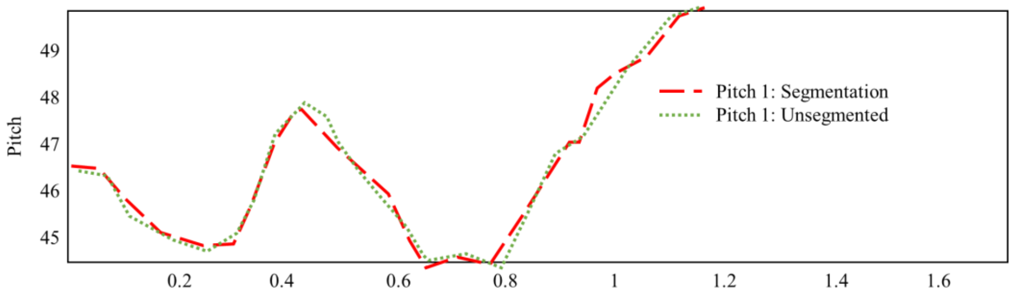


**Figure 4**: The fundamental frequency trace of standard spoken Can I have your order.

The fundamental frequency trace extracted by the author after reading the oral pronunciation Can I have your order extracted from this test oral language is shown in Figure 5 below:

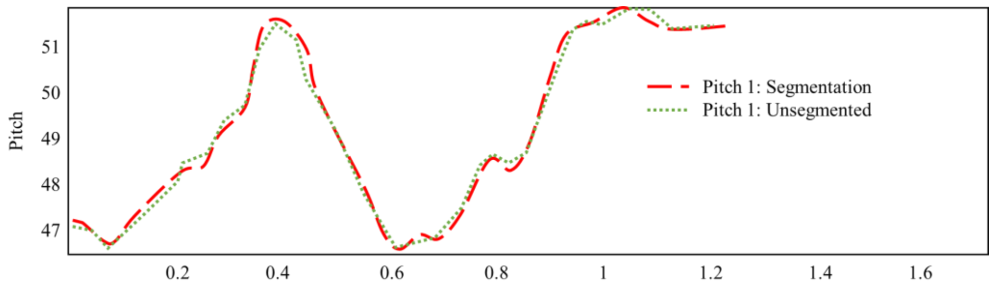


**Figure 5**: Follow-up oral Can I have your order fundamental frequency trajectory.

We asked a foreign language teacher to rate the spoken language of my follow-up reading. Since the tone of the follow-up reading is basically the same as the standard oral language and the pronunciation is relatively fluent and correct, the score given is 88 points. In addition, a classmate was selected for fundamental frequency extraction and the teacher's score was obtained. With these two parameters, a and b can be calculated, so that the mapping relationship between distance and score can be obtained. Even if the obtained distance is larger or smaller than the preset distance, it can be reasonably mapped to the score interval 0-100. In the same way, the score is calculated in the same way for MFCC.

Selected 40 independent English texts, and asked 2 male and 2 female students to record 160 sentences of oral test. At the same time, Mr. Liu from the Department of Foreign Languages was invited to compare the 160 sentences of test oral English with the standard oral English and grade them into three grades: poor (0-59), moderate (60-80), and good (81-100).

The horizontal axis represents the grading level of Mr. Liu from the Department of Foreign Languages, the vertical axis represents the grading level of the scoring algorithm in this paper, and the data in the table represents the number of graded tests spoken languages in the relevant grades. They are sentences on the diagonal is significantly larger than that in the same row and column. This means that the grading algorithm in this paper has a significant positive correlation with the foreign language teachers' grading.

The two speech files in the TIMIT corpus were used as the grading corpus, and the teachers of the foreign language department were invited to manually grade the two English corpora. The relationship between the resulting score and the algorithm score is shown in Figure 6 below.

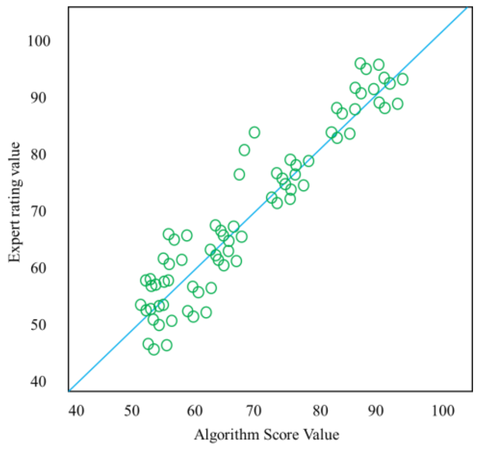


**Figure 6**: Follow-up oral Can I have your order fundamental frequency trajectory.

At the same time, the 20 sentences of spoken English entered by a classmate of the laboratory English pronunciation standard were scored by the algorithm, and the DTW algorithm was also used to score this corpus. We carried out algorithm scoring and DTW algorithm scoring for each sentence of spoken English, and invited Mr. Liu from the Department of Foreign Languages to score the spoken pronunciation. The comparison chart of the obtained algorithm score and manual score and the comparison chart of DTW algorithm and manual score are shown in Figure 7 respectively:

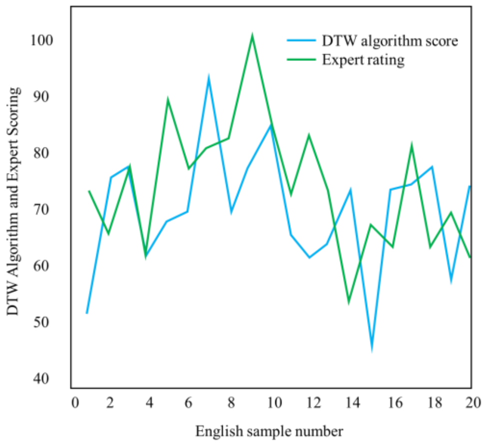


**Figure 7**: Comparison of DTW algorithm scoring and manual scoring.

## 5 CONCLUSION

This paper compares English teachers' scores with students' audio data, and selects a corpus from TIMIT corpus for pronunciation test. The algorithm is designed to test the pitch extraction algorithm, and the correlation coefficient between the average pitch difference and the score given by the teacher is calculated. The results show that the difference between the average pitch of words and the teacher's score is significantly higher than the correlation coefficient at the sentence level. This shows that the sentence level reflects the overall difference on the macro level. Finally, two speech files in TIMIT corpus are used as graded corpora, and teachers of foreign language departments are invited to grade the two English corpora manually. DTW algorithm is used to score the corpus. We gave each sentence of spoken English an algorithm score and a DTW algorithm score. The average pitch difference between standard spoken English and spoken English is calculated to reflect the differences between the two characteristics, so as to make spoken English pronunciation more objective and accurate.

*Xiaqing Ma*, https://orcid.org/0000-0003-0328-8153
*Xiaoyun Li*, https://orcid.org/0000-0003-4064-8728

## REFERENCES

[1] Chandra, T.-B.; Singh, B.-K.; Jain, D.: Integrating patient symptoms, clinical readings, and radiologist feedback with computer-aided diagnosis system for detection of infectious pulmonary disease: a feasibility study, Medical & Biological Engineering & Computing, 60(9), 2022, 2549-2565. https://doi.org/10.1007/s11517-022-02611-2

[2] Konovalenko, T.-V.; Yivzhenko, Y.-V.; Demianenko, N.-B.; Romanyshyn, I.-M.; Yemelyanova, Y.-S.: The possibilities of using distance learning in the professional training of a future foreign language teacher, Linguistics and Culture Review, 5(S2), 2021, 817-830. https://doi.org/10.21744/lingcure.v5nS2.1423

[3] Laili, R.-N.; Nashir, M.: Higher education students' perception on online learning during Covid-19 pandemic, Edukatif: Jurnal Ilmu Pendidikan, 3(3), 2021, 689-697. https://doi.org/10.31004/edukatif.v3i3.422

[4] Lucero, R.: Effects of instructional materials in multimedia computer-assisted instruction in teaching folk dance, Edu Sportivo: Indonesian Journal of Physical Education, 2(1), 2021, 40-50. https://orcid.org/0000-0002-2285-8548

[5] Pang, X.; Zhao, Z.; Weng, Y.: The role and impact of deep learning methods in computer-aided diagnosis using gastrointestinal endoscopy, Diagnostics, 11(4), 2021, 694. https://doi.org/10.3390/diagnostics11040694

[6] Park, Y.-S.; Yi, Y.: The education model of liberal arts to improve the artificial intelligence literacy competency of undergraduate students, Journal of The Korean Association of Information Education, 25(2), 2021, 423-436. https://doi.org/10.14352/jkaie.2021.25.2.423

[7] Qian, L.: Research on college English teaching and quality evaluation based on data mining technology, Journal of Applied Science and Engineering, 26(4), 2022, 547-556. https://doi.org/10.6180/jase.202304_26(4).0011

[8] Samiei, F.; Ebadi, S.: Exploring EFL learners' inferential reading comprehension skills through a flipped classroom, Research and Practice in Technology Enhanced Learning, 16(1), 2021, 1-18. https://doi.org/10.1186/s41039-021-00157-9

[9] Shaukat, A.; Lichtenstein, D.-R.; Somers, S.-C.; Chung, D.-C.; Perdue, D.-G.; Gopal, M.; Hunsberger, C.: Computer-aided detection improves adenomas per colonoscopy for screening and surveillance colonoscopy: a randomized trial, Gastroenterology, 163(3), 2022, 732-741. https://doi.org/10.1053/j.gastro.2022.05.028

[10] Shen, L.: Data mining artificial intelligence technology for college English test framework and performance analysis system, Journal of Intelligent & Fuzzy Systems, 40(2), 2021, 3489-3499. https://doi.org/10.3233/JIFS-189386