Research on Regression Model of Multi-parameter English Pronunciation Quality Evaluation Based on Computer Technology

Jiangnan Zhang(✉)

Jiangxi Industry Polytechnic College, Nanchang 330095, Jiangxi, China
2425096836@qq.com

Abstract. The quality of English pronunciation determines the level of spoken English. Students learning English often have trouble judging whether their pronunciation is correct. The multi-parameter-based English pronunciation quality evaluation model proposed in this paper can provide students with an easy way to use computers to independently evaluate their English pronunciation quality. Based on the previous speech recognition algorithm, this paper extracts the Mel-frequency cepstrum feature parameters in the speech signal, and uses the feature parameters as the central structure of the model to construct a model based on deep learning neural network. Within the model, the system refers to the speech rate, intonation, rhythm and intonation in the speech as evaluation treatments, and constructs an evaluation model for the quality of spoken English pronunciation based on how far regression. The model proposed in this paper can help students carry out autonomous learning and improve their learning efficiency, which is of great significance to English teaching.

Keywords: English Pronunciation · Evaluation Model · Regression Model

1 Introduction

With the increasingly close communication between countries, English has become a compulsory course for students in many countries. The learning of spoken English occupies an important position in English teaching, and the ability of spoken English determines the efficiency of English communication. Based on the English teaching curriculum in China, this paper proposes a quantitative evaluation of English pronunciation using a computer algorithm model. With the rapid development of computer technology and artificial intelligence, computer-assisted English learning has become a trend. Most of the current computer-assisted language learning systems focus on word memory, grammar learning and text reading, and are less used in oral language learning. Because it is very difficult for a computer to judge the quality of speech from the received spoken pronunciation. However, in recent years, with the application of speech recognition technology to computer-assisted language learning systems, the oral English

learning mode that collects the tester's spoken speech to evaluate the speech quality in real time and correct the pronunciation has become increasingly popular. Computer systems that recognize English speech are very effective for English learning. At present, the key technology of English speech recognition system is the quality evaluation of spoken English pronunciation. In this paper, English speech recognition is carried out based on multi-parameters, and the accuracy of English speech recognition is improved. The regression model used in this paper improves the efficiency of evaluating English speech.

## 2   Evaluation of the Quality of Spoken English Pronunciation

Research on the evaluation of the quality of spoken English pronunciation has started in the 1990s. In the 1990s, L.Neumeyer designed a set of voice interactive language learning system VILTS (Voice Interactive Language Training System) [10]. The speech recognition model of Hidden Markov Model is utilized in this system. The system automatically scores the speech from several aspects, such as the accuracy of the speech piece, the period of the speech piece and the speed of speech. After this system, various research institutions have begun to study the model of speech recognition. In the thirty years since the development of automatic speech recognition systems, many models have been applied to speech recognition systems, and good results have been achieved [12].

Nowadays, there are still many controversies about the evaluation of the pronunciation quality of spoken English in the world, and there is no unified model to evaluate the pronunciation quality of English accents. The current research focuses on two issues. First of all, there are many countries and regions in the world, and the pronunciation habits and pronunciation characteristics of these countries and regions are different [3]. This means that the quality evaluation system of spoken English pronunciation in other countries is not suitable for English teaching in China. Secondly, the existing computer-assisted language learning systems at home and abroad pay more attention to the learning of English words and English grammar. There is currently no system on the market that focuses on evaluating the quality of spoken English pronunciation [11]. In order to build a complete evaluation system for the quality of spoken English pronunciation, it is necessary to analyze and compare from multiple dimensions, including pronunciation content, intonation, rhythm, pronunciation intonation, fluency, etc. [2].

In recent years, with the development of computer technology, big data technology, cloud computing technology, and artificial intelligence technology have become popular information technologies, and breakthroughs have occurred in many industries. In this paper, the deep neural network algorithm is used to improve the intelligence of speech recognition, pronunciation evaluation, and intelligent speech interaction, which makes the evaluation of spoken English easier [5].

3   Voice Data Collection

The data in the model is collected from an audio corpus on the Internet, which contains more than 8,000 sets of English speech data, including the pronunciation of 10 English
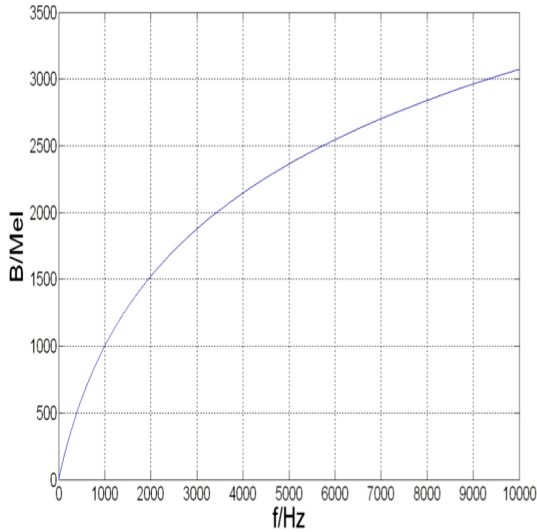
Fig. 1. Relationship between Linear Frequency and Mel Frequency.

words by people of different ages. This dataset is mainly used to train speech recognition models [8].

In order to reduce the influence of the ambient sound on the speech recognition effect, after the speech recording is completed, the system firstly preprocesses the speech. In the preprocessing, the original speech should be pre-homed, windowed, noise filtered and end-point detected. The speech preprocessing operation is mainly completed in the MATLAB platform, which has digital filter tools, which can realize all the above-mentioned operations [13].

In preprocessing, we set the input speech signal as x(n), and set the output digital speech signal after quantization as y(n). The pre-emphasis of the speech signal can emphasize the high-frequency part of the speech, remove the lip-touching sound that occurs when people read, and improve the system's high-frequency resolution efficiency of speech. Windowing is to perform window processing on the speech waveform in the digital speech signal to make the speech waveform more clear. Voice outage detection can detect the beginning and end of speech [7]. This system uses the double-threshold front-end power failure detection short algorithm. In this algorithm, the time domain parameters and short-term zero-crossing rate of the calculated speech are used. After these processes are performed, the characteristic parameters of the digital speech signal are extracted [4] (Figs. 1 and 2).

This system uses the characteristic parameter of Mel frequency cepstral coefficient when extracting the characteristic parameter of digital speech signal. This parameter transforms the speech from the time domain to the cepstral domain for feature extraction [1] (Fig. 3).

The FFT used in the process is the fast-read Fourier transform. The log. It is the log operation. The DCT is a discrete cosine transform. The MFCC is the feature parameter of the output.
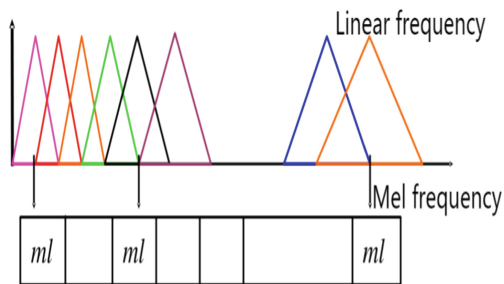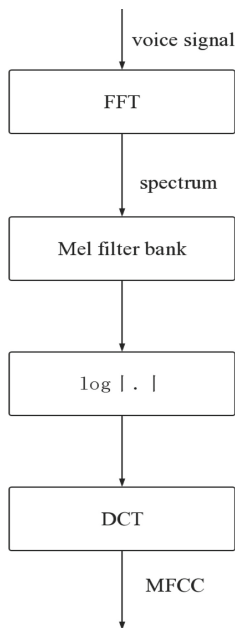
Fig. 2. Mel filter bank.

Fig. 3. Extraction of Mel frequency cepstral coefficients.

4 English Accent Speech Recognition Model

In the process of recognizing English accent speech, the model uses three machine learning methods: deep neural network, support vector machine and BP neural network. In the spoken English speech recognition model, the MFCC feature parameter is the input volume, which includes four evaluation criteria of speech intonation, speech rate, pitch, and rhythm [6].

The deep neural network model will not lose information in the process of machine learning and has good performance. The ResNet neural network model was used in this study. After the neural network model is trained, the improved ResNet neural network model is trained on the new cross-feature image training set, and finally the machine training results are optimal.
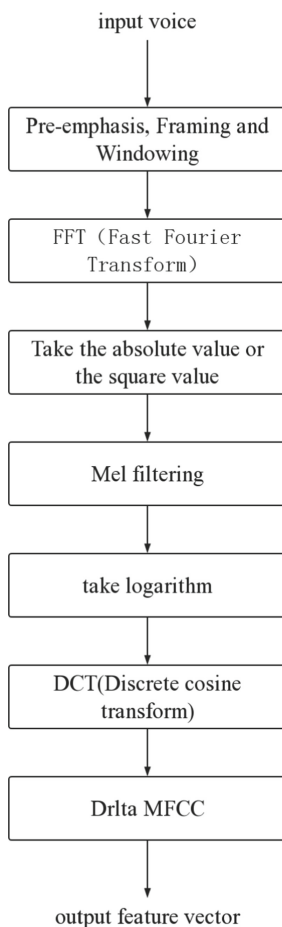
Support vector machine is a supervised machine learning algorithm, which can learn the characteristics of different kinds of known samples, and classify and predict the location samples according to the learning results. In this study, a large amount of English corpus information is used as a training dataset to train the SVM. After training, parameter recognition and verification are performed on the test data set speech [9].

BP neural network has good parallelism, nonlinearity and fault tolerance. Compared with other machine learning models, BP neural network also has excellent pattern recognition and classification capabilities. In this study, the MFCC eigenvalues are input into the BP neural network model, the activation function is used in the hidden layer, and the output information is set as the speech recognition information. In the BP neural network model, this study also requires the model to record relevant information such as speech rate and intonation. In this study, the expected error value of the model was set to 0.001, and the number of iterations was set to 200.

5   Multi-parameter English Pronunciation Quality Evaluation Regression Model

In this study, the correlation coefficient between the MFCC feature parameters of standard sentences and the MFCC features output by the speech recognition model is used as a quantitative index of pitch to judge whether the pronunciation is clear and accurate. Speech rate evaluation is quantified by the ratio of the standard sentence duration to the test sentence duration. For the rhythm evaluation, the Pairwise Variability Index (PVI) proposed by Nanyang Technological University in Singapore was used to calculate the rhythm correlation between standard sentences and input sentences (Fig. 4).

On the basis of checking the credibility of the above evaluation indicators, and further considering the multi-parameter indicators such as intonation, speech rate, rhythm and intonation and their weights, the multi-parameter statistical analysis method is used to construct a multi-parameter English pronunciation quality evaluation model. Taking the comprehensive score of spoken English pronunciation quality as the dependent variable, and taking intonation, speech speed, rhythm and intonation as independent variables, a multiple linear regression model was established between the comprehensive score of pronunciation quality and the pronunciation evaluation indicators.

input voice

Pre-emphasis, Framing and Windowing

FFT (Fast Fourier Transform)

Take the absolute value or the square value

Mel filtering

take logarithm

DCT(Discrete cosine transform)

Drlta MFCC

output feature vector

Fig. 4. MFCC process.

6 Conclusion

This paper utilizes speech signal processing techniques and machine learning algorithms to construct a computer-based SVM model for intelligent evaluation of English. This model takes spoken English pronunciation, extracts the plum frequency inversion coefficient based on human ear and hearing, and focuses on the characteristic parameters of speech signal. The model constructed in this paper has an important reference value in identifying the English speech quality, which can improve the students' English learning efficiency and reduce the workload of teachers.

References

1. Bian Ya, Zheng Zhanguo. Research on the mixed teaching mode of English pronunciation for non-English majors from the perspective of the core theory of lingua franca [J]. English Teacher, 2021, 21(10): 8-12.
2. Chen Juying. Research on the influence of English phonological awareness training on the vocabulary reading and writing ability of students in junior high vocational classes [D]. Minnan Normal University, 2021.
3. Gao Xia. Effective strategies to improve students' English pronunciation ability [J]. English for Middle School Students, 2021(48):86-87.
4. Gao Zhenfeng. Automatic recognition method of robot English speech based on automatic coding [J]. Automation Technology and Application, 2021, 40(07): 96–100+160.
5. Huang Qiuwen, Gu Libostan·Ai Maiti. Research on the relationship between college students' English pronunciation output anxiety, critical thinking ability and grades [J]. Science and Education Wenhui (late issue), 2021(09):183–185.DOI: https://doi.org/10.16871/j.cnki.kjwhc.2021.09.058.
6. Li Li. Research on the influence of humanistic education on English phonetics teaching in teachers' colleges in the professional sense [J]. Overseas English, 2021(10):92-93.
7. Liu Xirui. Research on the construction and effect of communicative competence-oriented English phonetic teaching model [J]. Foreign Language Teaching, 2021(04):107–113+16.
8. Tang Linwei. Research on English pronunciation and intonation of medical college students [J]. Modern Vocational Education, 2021(36):178-179.
9. Wang Chen. An empirical study on the analysis of common errors in English pronunciation of high school students from the perspective of interlanguage theory [D]. Shenyang Normal University, 2021. DOI: https://doi.org/10.27328/d.cnki.gshsc.2021.000510.
10. Wang Fang. Current Situation and Countermeasures of English Phonetics Teaching in Senior Primary Schools [J]. Literacy and Math, 2022(03):144-146.
11. Yang Enhua. Comparison and Interference Research between English Phonetics Teaching and Chinese Phonetics Teaching [J]. English Square, 2021(35):122-124.DOI:https://doi.org/10.16723/j.cnki.yygc.2021.35.035.
12. Yu Rui, Gong Jian. Research on the Influence of Chinese Dialect Phoneme Merging on English Speech Perception [J]. English Square, 2021(36):37-39.DOI:https://doi.org/10.16723/j.cnki.yygc.2021.36.011.
13. Zhang Fengqin. The design of picking robot speech recognition system based on English vocabulary integration and WAP platform [J]. Agricultural Mechanization Research, 2022,44(05):223–226+231.DOI:https://doi.org/10.13427/j.cnki.njyi.2022.05.041.