Innovative Practice of Mandarin Oral Expression and Teaching in the Context of Intelligent Teaching and Learning

Zhaozhao Yang[1] and Huixu Li[1],*
[1]Hunan First Normal University, Changsha, Hunan, 410002, China.
Corresponding authors: (e-mail: lihuixu7820@163.com).

Abstract The cultivation of oral expression ability requires Mandarin training, and Mandarin training is an auxiliary tool for students' oral expression, and the study of Mandarin should be fully utilized to enhance students' oral expression ability in teaching practice. This paper explores the application of artificial intelligence technology in the teaching of Mandarin oral expression, constructs a Mandarin oral pronunciation assessment corpus based on speech recognition under artificial intelligence technology, and applies the corpus in the actual students' Mandarin oral expression training. The Mandarin spoken corpus, after the filtering process of the speech signal, is calculated using the MFCC parameter to complete the speech static feature recognition. On this basis, DCT frequency transformation is introduced to describe the dynamic characteristics of speech through differential parameters, and the processed speech signal is output using HMM model. In the analysis of the effect of intelligent spoken Mandarin teaching, the average score interval of the 10 indexes evaluation of the subject class is between [6.53,7], which is much higher than that of the control class. After intelligent teaching, the average scores of answering questions, conversational communication and independent expression of the subject class are 2.978, 3.336 and 2.945 respectively, which are higher than the results of the oral test before teaching, and the students' oral expression ability is further improved.

Index Terms speech recognition, pronunciation assessment corpus, mfcc parametric calculation, hmm model, putonghua

I. Introduction

Putonghua is a skill that students must master, and it is an essential and important tool in the process of communication and exchange between students and others, and the students' Putonghua ability and level will, to a certain extent, directly affect others' impression of them [1], [2]. Therefore, students must pay full attention to Mandarin pronunciation and improve their oral expression ability and level [3]. Teachers should fully analyze the current situation of students' oral expression, find out the reasons for the problems in the process of oral expression, and help students improve in a targeted way, so as to help students make better progress [4].

Using Putonghua is also a very common thing, but many students do not pay enough attention to Putonghua, thinking that Putonghua is just a tool for speaking [5], and they do not pay enough attention to whether their pronunciation is standardized and correct, and whether their speed of speech is reasonable, and so on, and so there are more problems when they use Putonghua. In addition, some students feel that it is a waste of time to dedicate time to study and practice Putonghua, and their motivation and interest in learning Putonghua are not high [6], [7].

In the process of teaching Putonghua, it is necessary to make a reasonable design of the content of teaching, and the designed teaching difficulty should have a certain gradient, and gradually design the teaching content, in order to let the students get more progress [8]. At present, some Mandarin teaching courses are not reasonably designed, the design of teaching difficulty is not well grasped, and some teachers do not fully understand the students' level of Mandarin, and simply teach according to the syllabus materials, for some students whose foundation of Mandarin is weak, their basic skills will only get worse and worse, and in the long run it will also make the students lose their interest in learning [9].

With the extensive global regional cooperation and exchange, Putonghua as the official language of our country is more and more emphasized by the whole society and even the whole world [10]. And the teaching of spoken Mandarin is a heavy task and a long way to go, so we have to continuously strengthen the exploration and innovative research on the teaching of spoken Mandarin, improve the teaching method, create a scientific, easy and extensive environment for spoken Mandarin, improve the learning enthusiasm of spoken Mandarin, and improve the students' ability of oral expression and application of spoken Mandarin [11].

Literature [12] suggests that dialect differentiation and dialect polarization have positive effects on both per capita income and economic growth. Adjusted dialect differentia-

tion and peripheral heterogeneity have a positive effect on economic growth, which is robust when the effects of Mandarin level and inmigration are taken into account. Literature [13] describes that the development of the tourism industry has entered a new stage after the "Belt and Road" initiative, and Mandarin has become an important prop for standardization and efficient communication in tourism services. Literature [14] found that language influences migrants' decision-making through different channels, including imposing communication barriers, establishing social identity, and enhancing skill exchange, with Mandarin as the background. Literature [15] suggests that many Mandarin teachers are native speakers of Mandarin and do not receive formal training before starting their careers, and therefore may not have the relevant knowledge to teach Mandarin. Literature [16] considers this phenomenon in the context of the geopolitical background of China's economic rise and its soft power promotion, as well as the market forces that reward Mandarin competence in contemporary Indonesia. Literature [17] proposes a data-driven approach to constructing a prosodic grammar for Mandarin read-aloud speech. The validity of the approach is confirmed by a meaningful interpretation of the outliers of the main prosodic and target syntactic patterns derived from the inference rules. Literature [18] investigated the acoustic characteristics of Mandarin clear and Mandarin flat conversations. Linear mixed-effects regression modeling was used to explore how these cues vary in the distinction between clear and flat tones for each.

This paper starts from artificial intelligence technology to explore the application of unsupervised learning, neural network and speech recognition technology under artificial intelligence technology in the education of Mandarin oral expression. Based on speech recognition technology, a corpus is constructed for students' Mandarin oral expression training. Through the pre-emphasis filter, the speech signal is preprocessed, and the processed speech signal is subjected to MFCC parameter calculation to improve the speech recognition ability of the corpus. The dynamic characteristics of speech are described by differential parameters, and the HMM model is constructed to calculate the output probability and other indexes. The model constructed in this paper is applied to the actual teaching of Mandarin oral expression, and through the real-time detection of classroom status and pre- and post-testing, the effect of intelligent teaching on Mandarin oral expression is jointly measured from the aspects of teaching effect in the classroom and students' oral expression ability.

II. Teaching Mandarin Oral Expression Based on Artificial Intelligence

A. Related Technical Concepts

1) Unsupervised Learning

Let's say that a two-year-old child suddenly realizes that he knows the word "flower" when no one has ever told him what a flower is before. This is actually another kind of unsupervised process, because there is no one to supervise, no one to tell him the label "flower", he just learns it through countless pictures. This is called unsupervised learning. One possible unsupervised method is to do statistics on some features. Hidden Markov models, for example, can be used to automatically find new words in a large vocabulary that have never appeared before. Consider model $f(x) = a$ (constant) and loss function $C = E\left[(f(x) - x) \wedge 2\right]$, where minimizing the current loss produces a value equal to the average of the data.

2) Neural Networks

A neural network is an acyclic graph of interconnected neurons. The outputs of neurons in the previous layer are used as inputs to the neurons in the next layer. The arrangement of neurons is usually regular, and they are constructed as layers of connected structures, each containing multiple neurons. A common neural network structure is called a fully connected layer, which means that neurons in two neighboring layers are connected two by two, and neurons in the same layer are not connected to each other.

The first layer of a neural network is the input layer and the last layer serves as the output layer with numerous hidden layers in between. When using unsupervised layer-by-layer training, the first layer is trained first, and then the nodes that have been trained in the first layer are used as input nodes for the next layer, and then the next layer is trained, and so on. After each layer is trained, the entire neural network is then trained using the bp algorithm to realize the mapping between the input target and the output target, each convolutional layer contains multiple feature maps, each feature map is a plane with one or more neurons.

Neural networks can solve complex problems better. Commonly used activation functions are Sigmoid function, Tanh function and ReLU function etc. the Sigmoid function formula is:

$$f(x) = \frac{1}{1 + e^{-x}} \tag{1}$$

Its advantages are that it can map the output between 0 and 1, monotonically continuous, optimization stable and easy to derive, the disadvantage is that the output is not centered on 0, easy to saturate and thus produce gradient disappearance leading to problems in training.Tanh function, i.e., hyperbolic tangent function, is given by Eq:

$$\tanh(x) = \frac{1 - e^{-2x}}{1 + e^{-2x}} \tag{2}$$

Relative to the Sigmoid function it has the advantage of fast convergence, and the output is centered on 0, able to compress the data between -1 and 1, but there will still be a disappearance of the gradient.The ReLU function, i.e., the linear rectifier function, is given by Eq:

$$f(x) = \max(0, x) \tag{3}$$

3) Speech Recognition

Artificial intelligence algorithms in human-computer communication can be used in the field of teaching and learning, and speech recognition can be used to correct students' pronunciation and give them help in language training. The application of speech recognition in teaching can then make it easy for students to look up some words and sentences when learning English. There is also the use of speaking websites to correct students' pronunciation to grade their pronunciation when they are learning to speak.

B. Application of Artificial Intelligence in Mandarin Oral Expression Education

(1) Selection and utilization of teaching media
Artificial intelligence can better analyze the current teaching media, and then compare it with the cognitive style of the learner to select the most suitable learning resources for the learner, so as to improve ? teaching, stimulate the interest of learners in learning, and promote efficient learning.

(2) Analyzing learners' learning characteristics
You can analyze the content time and other situations of learners browsing the website on the Internet to find out the most suitable learning mode for learners, which will have a great guiding effect on the course design.

(3) Designing learners' independent learning styles
It allows learners to understand themselves better and recognize themselves, and then through some applications that provide designated learning methods, learners can develop appropriate learning plans, learning tasks, learning methods to improve the quality of teaching.

(4) Advice on learning resource design
Artificial intelligence can analyze the current situation based on existing data, and then use related deep learning technology to predict the future situation, so as to find out the most suitable learning resources for learners, and provide personalized learning services for learners.

(5) Guiding role in the design of teaching objectives
Through the formative and summative evaluation of learners, it can scientifically analyze which parts of the teaching are not in place for learning and which parts are well mastered by students, and it can provide suggestions for teachers to check the gaps and make up for the omissions, so as to improve the quality of teaching.

III. Mandarin Speaking Corpus Establishment

A. Pronunciation Assessment Mandarin Corpus Creation

Based on the speech recognition function of artificial intelligence technology, this paper establishes a corpus for students' Mandarin speaking training.

The pronunciation unit model and grading scoring model of computer-aided pronunciation learning model are both obtained through corpus training the model usually requires class of language cores standard pronunciation corpus and non-standard pronunciation corpus. In general, the standard pronunciation corpus is mainly used to train the pronunciation unit model. The corpus that should be trained should be as consistent as possible with the learner's pronunciation characteristics. The content of the corpus should depend on the way the basic pronunciation processing unit is employed in the future model. It can be phonemes, syllables or words. The non-standard pronunciation corpus used to manually train the graded scoring model by experts and to test the model performance should be broadly representative. The content of the corpus varies according to the function of the model, as the focus of the different learning phases is different and the target of the grading judgment is different. To this end, we established a Chinese corpus consisting of 13 male and 11 female students, including 412 Chinese pronunciations, 1319 Chinese syllables with tones, and 668 commonly used isolated words, as the standard pronunciation corpus.

B. Mandarin Speech Recognition

1) Flow of Calculation of Mfcc Parameters

The human hearing model is a special nonlinear model, which responds to different frequency signals with different sensitivities, basically a logarithmic relationship. In recent years, a parameter that can make full use of this special perceptual property of the human ear has been widely used, which is the Mel scale cepstrum parameter, or Mel frequency cepstrum parameter. A large number of studies have shown that the MFCC coefficients can improve the model recognition performance better than the LPCC coefficients.

The MFCC coefficients are computed using the "bark" as its frequency reference, and the conversion relationship between it and the linear frequency is:

$$f_{mel} = 2595 \lg \left(1 + \frac{f}{700}\right) \tag{4}$$

The MFCC parameter is also calculated on a frame-by-frame basis. The first step is to obtain the power spectrum $S(n)$ of the signal at that frame by FFT, which is converted to the power spectrum at the Mel frequency. This requires a number of bandpass filters in the spectral range of the speech prior to the calculation, i.e:

$$H_m(n)(m = 0, 1, \cdots, M - 1; n = 0, 1, \cdots, N/2 - 1) \tag{5}$$

Where M is the number of filters, usually taken as 24-26, N is the number of points in a frame of speech signal, and N is usually taken as 256 for the convenience of calculating the FFT. filters are simple triangles in the frequency domain, with a center frequency of f_m, and they are uniformly divided on the Mel frequency axis. At linear frequencies, when m is small, the neighboring f_m are spaced very small, and as m increases, the neighboring f_m are gradually spaced apart. Also in the lower frequency region, there is a segment between

f_m and f that is linear the parameters of the bandpass filter are calculated in advance and can be used directly in the calculation of the MFCC parameters.

2) Preprocessing of Speech Parameter Computation

Before the speech parameter is calculated, the speech signal is generally made to pass through a pre-emphasis filter, viz:

$$y(n) = x(n) - \alpha x(n-1) \qquad (6)$$

The filter is actually a first-order high-pass filter for suppressing the low-frequency components of the speech signal, which can effectively improve the performance of the speech recognition model. And after that, it is also necessary to add windows to the subframed speech signal, such as the Hamming window:

$$\omega(n) = \begin{cases} 0.54 + 0.64 \cos\left\{\left[\left(\frac{2n}{N-1}-1\right)\pi\right]\right\}, & (n = 0, 1, \cdots, N-1), \\ 0, & (n = others). \end{cases} \qquad (7)$$

The add-window operation is actually simply a multiplication operation, i.e:

$$s_w(n) = s(n)w(n), (0 \le n < N). \qquad (8)$$

When framing the speech signal, care should be taken to have a certain length of overlap between two adjacent frames. This is in order to have a better continuity of the speech parameters. For example, when the sampling frequency is 8000Hz, the frame length is taken as 20ms, which corresponds to 160 sampling points, and if the frame interleaving is 10ms, the interleaved part corresponds to 80 sampling points. The pre-emphasis filter combined with the utilization of short time energy and zero crossing rate is better for endpoint detection of speech signals.

3) Post-Computation Processing of Phonetic Parameters

After calculating the Kst order MFCC coefficients for each frame of data, it is common to also multiply each of these K coefficients by a different weighting factor, effectively a short window:

$$\begin{cases} \hat{c}_m = w_m c_m \\ w_m = 1 + \frac{K}{2} \sin\left(\frac{\pi m}{K}\right) (1 \le m \le K) \end{cases} \qquad (9)$$

4) Differential Cepstrum Parameters

The standard MFCC parameters only reflect the static characteristics of speech parameters, however, the human ear is more sensitive to the dynamic characteristics of speech, how to obtain parameters that reflect the dynamic changes of speech? Usually, differential parameters are used to characterize such dynamics. The calculation of differential parameters is adopted:

$$d(n) = \frac{1}{\sqrt{\sum_{i=-k}^{k} i^2}} \sum_{i=-k}^{k} ic(n+i) \qquad (10)$$

Here c and d both represent a frame of speech parameters, k is a constant, usually taken as 2, then the difference parameter
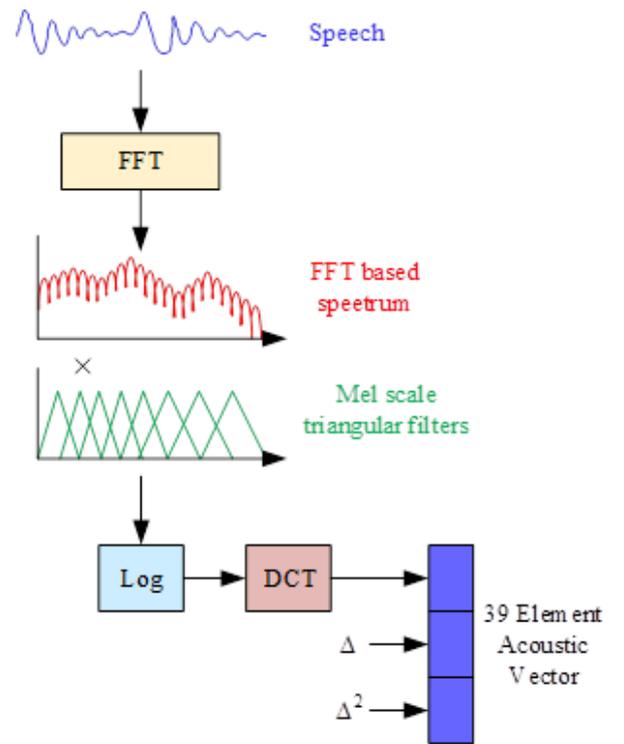

Figure 1: Based on MFCC front-end processing

is called a linear combination of the first 2 frames and the last 2 frames of the current frame. The above formula is the first-order MFCC differential parameter of the differential parameter, and the second-order differential MFCC parameter can be obtained by using the same formula for the first-order differential parameter. In practice, the MFCC parameter and each order difference parameter are combined into one vector as the parameter of a frame of speech signal.

Figure 1 shows the flowchart of the MFCC-based front-end processing. For better pattern matching, we transformed the speech waveform into a sequence of 39 acoustic vectors, which were obtained from log-smoothed spectra spaced at 10-ms intervals. The improved form is obtained by using a discrete cosine transform (DCT) based on a nonlinear Mel frequency scale. Next the uncorrelated signals with interactions are assumed to be statistically independent. Finally, the first- and second-order difference components of the signal are added to the dynamic information of the signal.

C. Hidden Markov Model-based Information Output for Mandarin Corpus

1) Hidden Markov (Hmm) Output Probability Calculation

Let given a sequence of observation vectors $O = (o_1, o_2, \cdots, o_T)$ and an HMM model $\lambda(A, B, \pi)$, the probability that the HMM model λ outputs an observation sequence O as a sequence of state transfers q is if the sequence of state

transfers $q = (q_1, q_2, \cdots, q_T)$ is known:

$$P(O|q) = \prod_{t=1}^{T} P(o_t|q_t, \lambda) = b_{q_1}(o_1) b_{q_2}(o_2) \cdots b_{q_T}(o_T) \tag{11}$$

The probability that HMM model λ outputs sequence q is:

$$P(q|\lambda) = \pi_{q_1} a_{q_1 q_2} a_{q_2 q_3} \cdots a_{q_{T-1} q_T} \tag{12}$$

And what is needed here is the probability $P(O|\lambda)$ for all possible occurrences of the probabilistic state transfer sequence q and the model output observation sequence O. Out the full probability formula can be obtained:

$$\begin{aligned} P(O|\lambda) &= \sum_{a/q_q} P(O|q, \lambda) P(q|\lambda) \\ &= \sum_{q_1, q_2, \cdots, q_T} \pi_{q_1} b_{q_1}(o_1) a_{q_1 q_2} b_{q_2}(o_2) \\ &\qquad \cdots a_{q_T q_T} b_{q_T}(o_T) \end{aligned} \tag{13}$$

The equation requires $2TN^T$ computation, which is not affordable in practice. To reduce the computational complexity, forward and backward algorithms can be used.

2) Hmm Model Forward and Backward Probabilities

First define the forward probability of an HMM as $\alpha_t(i) = P(o_1, o_2, \cdots, o_T, q_t = i|\lambda)$ denoting the probability that a partial observation sequence $(o_1, o_2, \cdots, o_T)$ is in state i at moment t given the HMM model parameter λ.

The forward probability $\alpha_t(i)$ can be calculated using the following recursive formula.

(1) Initialization:

$$\alpha_1(i) = \pi_i b_i(o_1) \, (1 \le i \le N) \tag{14}$$

(2) Iterative calculations:

$$\alpha_{i+1}(j) = \left[\sum_{t=1}^{N} \alpha_i(i) a_{ij} \right] b_j(o_{t+1}) \tag{15}$$
$$(1 \le t \le T-1, 1 \le j \le N)$$

(3) Termination of calculations:

$$P(0|\lambda) = \sum_{i=1}^{N} \alpha_t(i) \tag{16}$$

In contrast to forward probability, there is backward probability. Define the backward probability as $\beta_t(i) = P(o_{t+1}, o_{t+2}, \cdots, o_T, q_t = i|\lambda)$.

Denotes the probability that given the HMM model parameter λ, the observation sequence is in state i at moment t, and the model outputs a partial observation sequence $\{o_{i+1}, o_{i+2}, \cdots, o_T\}$.

IV. Analysis of Intellectualized Mandarin Oral Expression Ability and Teaching Effectiveness

A. Intelligent Mandarin Speaking Teaching Effectiveness

In order to understand the subjective feelings of the students in the observed classes about the teaching of intelligent Mandarin oral expression, a classroom evaluation form was borrowed, and based on previous research, the classroom evaluation form was combined with a questionnaire to generate a student classroom evaluation questionnaire.The questionnaire survey was conducted among the students in the two Mandarin speaking classes, Class A, the class that used intelligent Mandarin oral expression teaching, and Class B, the class that used traditional Mandarin oral expression ability teaching class. After the investigation of classroom specific reality and teaching details, the 10 indicators were rated from 1 to 7, with 1 indicating the lowest and 7 the highest, and the students rated the actual classroom situation as well as the teacher's performance according to the descriptions of the questions.After the investigation of the classroom specific reality and teaching details, the overall evaluation of the communicative nature of classroom teaching was obtained, and the classroom observation was further examined from the students' point of view to determine the validity of the classroom evaluation form. Therefore, the classroom evaluation form can truly reflect the measure of the communicative classroom, and is also important for the quantitative study of classroom interactivity in this paper.The total number of students in the two samples is 100, and there are 50 students in each class.

1) Analysis of Classroom Evaluation Results

Figure 2 shows the evaluation of the specific practical and teaching details of the Mandarin speaking classroom, in which 1-10 represent the learning of speaking knowledge, teacher patience, learning interest, classroom atmosphere, classroom instructions, classroom task effectiveness, classroom activities, and teaching aids and materials. From the overall situation of the ten specific scoring indicators of classroom actuality and teaching details, the overall scores of Class A are higher than those of Class B, and their average scores are maintained between 6.53 and 7, while the average scores of Class B are not so satisfactory, and their average scores are maintained between 5.12 and 6.42, which indicates that the students of Class B do not yet recognize and are satisfied with the classroom arrangement or the specific teaching actuality, and the reason for this is believed to have something to do with the The relationship between online classroom interactivity and teaching effectiveness is further analyzed in terms of the 3-8 indicators in the Mandarin speaking classroom.The 3-8 indicators in the Mandarin speaking classroom are validated in terms of students' interest in learning, classroom atmosphere, classroom instructions, effectiveness of classroom tasks, classroom activities, and teaching aids and materials, and through the comparison of the average scores of the two classes, it is found that the students in class B are not satisfied with the classroom arrangement or the actual teaching. Class mean scores, it is found that the mean scores of class B in the above six aspects are lower than those of A. The differ-
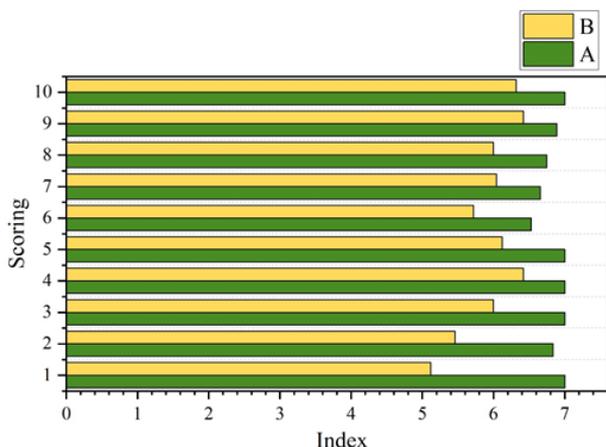
Figure 2: The actual and teaching details of the oral English language

ences between the two classes are 1, 0.58, 0.88, 0.81, 0.62, 0.75, 0.47 and 0.68, respectively, of which the item with the smallest difference is the one about the indicator of classroom activities.However, the overall mean of the two classes differs by 0.905 points, and the speaking instruction in class A makes students more satisfied.

2) Analysis of the Dynamic Curve of Interaction in the Mandarin Speaking Classroom

In the analysis of teacher-student speech rate and classroom silence rate, using the Chinese spoken language corpus designed in this paper, the basic data, the teacher-student speech rate per minute and the classroom silence rate to be calculated, and then refer to the results of the calculations, to generate a dynamic curve graph of the horizontal coordinate on behalf of the classroom time, the range of 1~80 (unit: minutes), the vertical coordinate represents the teacher speech rate, the student speech rate, the classroom silence rate per minute, the three with time change will be shown in a coordinate graph. With the way of coordinate graph, the change of these three over time is revealed, as shown in Figure 3, can be more intuitive to analyze the teacher-student verbal interaction of the special.

The teacher's speech ratio was basically stable between 25% and 45%, except for occasional greater than 50%, and the student's speech ratio was stable between 40% and 60%, the student's speech ratio was greater than the teacher's speech ratio, which indicated that the overall classroom structure was student oriented. In the case of the teacher, the overall speech ratio was less than 50%, which indicated that the teacher moved away from the traditional lecture-style classroom. Furthermore, from the perspective of fluctuation amplitude, the teacher demonstrated even pause times during lectures without monotonous, excessive one-sided teacher explanations. Throughout the course of teaching, the peaks of the teacher's speech appeared about 8 times, and the troughs appeared about 5 times.The analysis shows that each peak was at the

beginning of the classroom or at the end of the classroom.This is due to the fact that the teacher began the classroom by This is due to the fact that the teacher focuses on topic-oriented questions at the beginning of the class and on summarizing students' expressions or explaining other supplementary knowledge at the end of the class, while the teacher's troughs are mostly concentrated in the middle of the class, when the students tend to engage in free speech, so the teacher's rate of speech appears to be in a low ebb.

As far as the students are concerned, the whole speech rate of the students is relatively stable without big fluctuations, which shows that the frequency of students' speech in the classroom is in a stable state without any breaks, and the peak area of students' speech tends to be concentrated in the middle of the classroom, and the trough is concentrated in the front and back of the classroom, which is staggered from the peak of the teacher's speech, which is in accordance with the law.In addition, there are certain intervals and small ups and downs in students' speech rate in the classroom, which also reflects the benign interactive relationship between students and teachers. In addition, the students' speech rate in the classroom has a certain interval, small fluctuation, which also reflects the relationship between the students and the teacher's question and answer, dialog, and cooperation of the healthy interaction.

The variability and complexity of the classroom constitute the silent and chaotic parts of oral classroom teaching. In the graph, the rate of classroom silence fluctuates, with about seven peaks, all of which are above 8%. combining the classroom performances, it can be seen that three of the pre-classroom silences are in the preclass introduction part of the teacher's lesson, and four of the mid-lesson silences are due to the teacher's questioning of the students, and the students' time to reflect on the questions. in addition, observing the The degree of change of the peak, it can be obtained that the peak did not appear a long time of static state, indicating that there is no large-scale white space or chaos in the classroom, and combined with the previous data found that the silence in the classroom appeared in the study is mostly to help the silence in the classroom, which can be seen that, through a certain amount of silence, the teacher in the teaching and learning to give the students a certain amount of time to inspire and think, so that the students will enter into the mode of meaningful learning.

In summary, from the analysis of teacher-student speech rate and classroom silence rate, it can be seen that in the study of interactive teaching of Mandarin speaking learners, the teacher gets rid of the traditional lecture-style classroom in the classroom, and the overall classroom structure is student-oriented, the students maintain a stable frequency of speeches in the classroom, and the teacher gives the students time to think through a certain amount of silence in teaching and then guides the students to carry out the meaning of the Learning.

B. Analysis of Mandarin Oral Expression Ability

According to the content of the test and scoring standards, an oral test was conducted for class A. There were 50 students
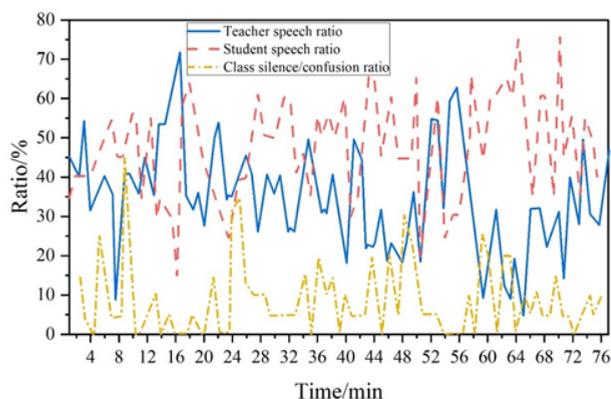
Figure 3: Teacher and student speech and classroom silence ratio

in class A. In the pre-study period, two weeks of teaching practice was carried out, and there was a certain degree of familiarity and understanding between teachers and students. two students were divided into a group, and the author utilized the time outside the classroom to carry out the first test for each student's Putonghua oral expression ability in the quiet and empty office, and the students skimmed through each topic and thought about it for two minutes before The students skimmed each topic and thought for two minutes before starting the speaking test, to improve the objectivity and scientificity of the testing process as much as possible.

1) Results of the First Speaking Test
First, after the first speaking test, the author organized and analyzed the video recordings of each student according to the scoring criteria, assessed the scores, and recorded the problems of the students' oral expressions. Secondly, the author recorded the scores obtained from each question of the first oral test of 50 students in Class A into SPSS and analyzed the results.
Table 1 shows the scores of the first oral test of Class A. In the first oral expression ability test of Class A, the mean value of answering questions is higher than that of dialogic communication and autonomous expression, which is 2.415 points, and the mean value of dialogic communication and autonomous expression is around 2.1 points. The standard deviations of answering questions, conversational communication and autonomous expression are 0.795, 0.725 and 0.764 respectively, indicating that answering questions, conversational communication and autonomous expression are more unstable, among which answering questions has a larger standard deviation and is the most unstable. Therefore the specific analysis of Class A is as follows:

(1) The highest distribution of scores for answering questions in the test of Putonghua oral expression in Class A is 2, with a percentage of 60%, indicating that it shows that more than half of the students meet the standard, answering questions basically fluently, with language errors, but the content is basically correct. The lowest distribution is 1 point,

and only 3 students answered the questions less fluently, with more language errors and incorrect content.

(2) At the level of conversational communication, 56% of the students scored 2 points, indicating that they could use everyday language correctly and communicate relatively fluently, with some language errors but not affecting communication. The lowest distribution is 4 points with only 4 students, indicating that few students can use everyday language appropriately and communicate fluently, with some language errors, but not affecting communication.

(3) Similar to the situation of dialogic communication, only 4 students got full marks of 4 at the level of independent expression, and more students concentrated on 2 marks, indicating monotonous and incoherent expressions, more language errors, and insufficient fluency.

According to the results of the first oral test, it can be concluded that students answered questions better than dialogic communication and independent expression. In answering questions, most students answered in English with basic fluency but with linguistic errors, and few students answered questions with more linguistic errors and less fluency. For dialogic communication, most students communicated relatively smoothly with language errors, but a few students could not communicate fluently with more language errors. In self-expression, most students expressed themselves in a monotonous manner, incoherently and not fluently, with more language errors.

2) Results of the Second Speaking Test
After 15 weeks of intelligent Mandarin speaking instruction for Class A in response to the problems in Class A, the test class was tested again for speaking, and Table 2 shows the second oral expression ability test for Class A.
In the second oral expression ability test of Class A, the average of answering questions, dialogic communication, and autonomous expression are 2.978, 3.336, and 2.945 respectively, and according to the scoring criteria, it can be obtained that answering questions and autonomous expression are in the range between up to the standard and better, and close to better answering questions, and dialogic communication are more unstable than autonomous expression.

In conclusion, according to the comparison between the second oral test and the first oral test, it is known that overall the results of the second oral test are better than the first one, and the students have made a greater degree of progress after a semester of intelligent oral expression practice, and most of the students' fluency and logic in answering the questions have been improved, and the ability of oral expression has been improved to some extent.

V. Conclusion
Based on the speech recognition function of artificial intelligence, this paper establishes a Mandarin spoken language corpus, which is practically applied to the Mandarin spoken expression teaching classroom. The practical application effect of intelligentized spoken Mandarin teaching is analyzed from

/	N	Min	Max	Mean	SE	SD
Answer questions	50	1	4	2.415	0.155	0.795
Conversational communication	50	1	4	2.163	0.138	0.725
Autonomous expression	50	1	4	2.119	0.105	0.764
Valid case number	50	/	/	/	/	/
Answer questions						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	3	6.00%	6.00%	6.00%	
	2	30	60.00%	60.00%	66.00%	
	3	11	22.00%	22.00%	88.00%	
	4	6	12.00%	12.00%	100.00%	
Conversational communication						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	7	14.00%	14.00%	14.00%	
	2	28	56.00%	56.00%	70.00%	
	3	11	22.00%	22.00%	92.00%	
	4	4	8.00%	8.00%	100.00%	
Autonomous expression						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	9	18.00%	18.00%	18.00%	
	2	26	52.00%	52.00%	70.00%	
	3	11	22.00%	22.00%	92.00%	
	4	4	8.00%	8.00%	100.00%	

Table 1: Class a first oral test score

/	N	Min	Max	Mean	SE	SD
Answer questions	50	1	4	2.978	0.154	0.764
Conversational communication	50	2	4	3.336	0.163	0.715
Autonomous expression	50	1	4	2.945	0.078	0.605
Valid case number	50	/	/	/	/	/
Answer questions						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	1	2.00%	2.00%	2.00%	
	2	16	32.00%	32.00%	34.00%	
	3	23	46.00%	46.00%	80.00%	
	4	10	20.00%	20.00%	100.00%	
Conversational communication						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	8	16.00%	16.00%	16.00%	
	2	27	54.00%	54.00%	70.00%	
	3	11	22.00%	22.00%	92.00%	
	4	4	8.00%	8.00%	100.00%	
Autonomous expression						
In effect	/	Frequency	Percentage	Effective percentage	Cumulative percentage	
	1	3	6.00%	6.00%	6.00%	
	2	5	10.00%	10.00%	16.00%	
	3	35	70.00%	70.00%	86.00%	
	4	7	14.00%	14.00%	100.00%	

Table 2: A class a second oral expression ability test

two aspects: spoken Mandarin expression ability and teaching effect analysis. The average scores of the 10 dimensions in Class A, which was taught through intelligent Mandarin oral expression teaching, were maintained between 6.53 and 7, which was higher than that of the class without intelligent Mandarin oral expression teaching application. In the analysis of the dynamic curve of Mandarin speaking classroom interaction, the teacher's classroom speech ratio was stable between 25% and 45%, while the students' classroom speech ratio was higher than the teacher's speech ratio between 40% and 60%. Meanwhile, the peaks of silence and disorganization in the classroom appeared seven times, accounting for more than 8%. Combined with the classroom performance, the phenomenon of silence appeared reasonably, and the teacher-student speech rate and classroom silence ratio showed that intelligent Mandarin speaking teaching got rid of the constraints of the traditional classroom and left more time for students to express themselves orally. In the pre- and post-test analysis of Mandarin oral expression ability, the three ability scores of the students in the tested class were centrally distributed at 2, and the mean values of the three abilities were 2.415, 2.163, and 2.119, respectively, and there were not many students who were able to obtain full scores. After the intelligent Mandarin oral expression ability teaching, the mean values of the three abilities increased to 2.978, 3.336 and 2.945 respectively, and the oral expression ability was improved to a certain extent.

References

[1] Xu, H. (2019). Putonghua as "admission ticket" to linguistic market in minority regions in China. Language Policy, 18(1), 17-37.

[2] Gómez-Cram, R., & Grotteria, M. (2022). Real-time price discovery via verbal communication: Method and application to Fedspeak. *Journal of Financial Economics, 143*(3), 993-1025.

[3] Qi, Y., & Wang, J. (2020). A talent cultivation model for improving the innovation ability of college students in scientific research. International Journal of Emerging Technologies in Learning (iJET), 15(18), 151-164.

[4] Sauro, S., & Sundmark, B. (2019). Critically examining the use of blog-based fanfiction in the advanced language classroom. *ReCALL, 31*(1), 40-55.

[5] Wong, A. D. (2019). Authenticity, belonging, and charter myths of Cantonese. *Language & Communication, 68,* 37-45.

[6] Du, X. (2017). Rethink about heritage language learning: a case study of children's mandarin chinese learning at a community language school in ontario, canada. (1).

[7] Du, Y., Sheng, L., & Tekinbas, K. S. (2020, June). " Try your best" parent behaviors during administration of an online language assessment tool for bilingual Mandarin-English children. In *Proceedings of the Interaction Design and Children Conference* (pp. 409-420).

[8] Budianingsih, T. (2019, March). Expert Perception on Quantum Approach Teaching Material Model for Speaking Mandarin. In *International Symposium on Social Sciences, Education, and Humanities (ISSEH 2018)* (pp. 54-58). Atlantis Press.

[9] Sung, K. Y., & Tsai, H. M. (2019). *Mandarin Chinese Dual Language Immersion Programs* (Vol. 119). Multilingual Matters.

[10] Goh, Y. S., & Wu, Y. (2017). *Teaching Chinese as an International Language: A Singapore Perspective*. Cambridge University Press.

[11] Pace, A., Lü, C., Guo, L. X., & Zhou, J. (2024). Narrative Skills in Mandarin–English Dual Language Immersion Learners. *Applied Linguistics, amae013*.

[12] Choi, T. H., & Adamson, B. (2020). China's belt and road initiative: Opportunities and linguistic challenges for Hong Kong. Multilingualism and Politics: Revisiting multilingual citizenship, 261-284.

[13] Liu, Y., Jiao, Y., & Xu, X. (2020). Promoting or preventing labor migration? Revisiting the role of language. *China Economic Review, 60,* 101407.

[14] Su, Y., Naigles, L. R., & Su, L. Y. (2018). Uneven expressive language development in Mandarin-exposed preschool children with ASD: Comparing vocabulary, grammar, and the decontextualized use of language via the PCDI-Toddler Form. *Journal of Autism and Developmental Disorders, 48*(10), 3432-3448.

[15] Lam, S., Diamantidaki, F., & Huang, A. (2019, May). Challenging Mandarin teachers' pedagogical approaches to the teaching of Chinese characters. *In International Conference on Education.*

[16] Hoon, C. Y., & Kuntjara, E. (2019). The politics of Mandarin fever in contemporary Indonesia: Resinicization, economic impetus, and China's soft power. *Asian Survey, 59*(3), 573-594.

[17] Yu-Siang, H., Chen-Yu, C., Yih-Ru, W., & Sin-Horng, C. (2023). An approach to constructing prosodic grammar for mandarin read speech. *The Journal of the Acoustical Society of America, 153*(4), 2406-2425.

[18] Tupper, P., Leung, K. W., Wang, Y., Jongman, A., & Sereno, J. A. (2021). The contrast between clear and plain speaking style for mandarin tones. *The Journal of the Acoustical Society of America, 150*(6), 4464-4473.

• • •