International Journal of Computer Information Systems and Industrial Management Applications
ISSN 2150-7988 Volume 17 (2025) pp. 1-12

https://doi.org/10.70917/ijcisim-2025-0313
Article

Construction of "Three Platforms and Six Sessions" Language Education Practice Model of Artificial Intelligence-Driven Industry-Teaching Integration

Guanguan Zeng * and Yongxing Ying

The School of Foreign Languages, Shanghai Zhongqiao Vocational and Technical University, Shanghai, 201514, China; zenggigi@163.com

Abstract: The study derives from the educational practice model of "three platforms and six links", synthesizes the multidimensional competence requirements, and uses two modules, speech recognition and language assessment, to drive the innovation of language education. The speech recognition module collects the speech of teacher-student interactions in the language teaching classroom, and realizes the adaptive recognition of lectures based on the Maximum A Posteriori Probability Reassessment (MAP) method, which combines the differential features of speakers. The language assessment module is grounded in bidirectional gated recurrent units and generative adversarial networks, and optimizes the decoding capability by improving the language audio feature encoder. In addition, the GAN with added LSTM units and fully connected layers can accomplish the discrimination of the authenticity of language evaluation more quickly. The language assessment method is able to adapt to different phoneme lengths in different language learning contexts, and when the phoneme length grows from 0 to 80, the assessment inference time only increases by less than 10ms, which has almost no effect. In the practical application, both teachers and students recognize the educational practice model of "three platforms and six links", and the students' language proficiency scores have increased to different degrees.

Keywords: three platforms and six sessions; language education; speech recognition; language assessment; MAP; GAN

1. Introduction

1.1. Background of the study

In recent years, major institutions have improved the quality of teaching, enhanced the teaching effect, and improved the teaching mode, so as to cultivate high-quality composite talents for the country and society in line with the requirements of the times. However, purely theoretical learning can no longer adapt to the high standard of cultivating practical talents, and it is necessary to further improve students' comprehensive quality and practical ability through practical teaching, and cultivate students' innovative ideas and knowledge learning ability [1-3]. Language education, as an important link in cultivating talents for cross-cultural communication and international cooperation, has been focused on its teaching practice mode [4-5]. And under the leadership of the state vigorously promoting the high-quality development of education and related documents and policies, the integration of industry and education has been clearly identified as an important direction of education reform and development, and a necessary way to promote the modernization of education and the high-quality development of education [6-7].

In different countries have emphasized the importance of the integrated integration and benign interaction between education and industry, and proposed to promote the reform of talent cultivation by integration of industry and education, deepen the integration of language education and school-enterprise

cooperation, aiming to cultivate more high-quality skilled talents in line with the market demand [8-11]. However, in the context of the integration of industry and education, the development process of language practice teaching is in a slow state due to the lack of attention to students' English education in colleges and universities, insufficient teacher strength, imperfect practical training and teaching facilities, insufficient investment by enterprises and the lack of cooperation processes [12-15]. In order to effectively implement the teaching concept of integration of industry and education, and to be able to cultivate talents needed by the country and society according to the national education policy, colleges and universities should actively implement the reform of English practical teaching, introduce advanced information technology, and promote the innovation of teaching practice by combining multiple platforms and multiple links.

With the rapid progress of information technology, artificial intelligence (AI) technology has also been effectively developed. Schools are more and more aware of the value and significance of the establishment of the practical teaching system.AI technology, as a kind of information technology, is widely used as a teaching tool in practical teaching, which promotes the application of the practical teaching system of language education for information technology [16-17]. The application of AI technology improves the original way of language teaching and enhances the effect of language teaching, and the effective integration of the language course and artificial intelligence technology is an effective way of language teaching. Effective integration is an important task of language teaching [18-20]. Under the leadership of industry-teaching integration and AI technology, it is clear that practice teaching is an important stage of language education, and practice programs are constructed so that students can improve their professional practice and innovation ability.

1.2. Status of research

With the digital transformation of education and the development of industry-education integration, the efficiency and results of language education practice can be improved through various AI tools and education platforms. The education platform is an important foundation for realizing the “three platforms and six links” of industry-education integration. Chandra et al [21] (2024) described that AI technology in English language teaching promotes the improvement and cultivation of students' autonomy, engagement, and critical thinking through natural language processing, intelligent tutoring systems, and automated assessment and feedback to achieve students' language acquisition, personalized learning, and assessment and feedback of school effectiveness. Li [22] (2025) constructed a model of English education reform and international exchange driven by AI technology, encompassing an interactive education model, cross-cultural curriculum design, an online learning platform, social media marketing, and international collaborative project practices to promote the practical aspects of international exchange in English education. Nie [23] (2024) utilized the campus information platform to establish a virtual interactive English learning community and practice pathway for students, creating a flexible and efficient practice environment for students to improve their motivation and academic performance. Akhter [24] (2024) states that targeted feedback is significantly correlated with language proficiency in AI-based language education platforms, engagement positively mediates proficiency through student motivation, rapid feedback and spaced practice help with language case analysis, and detailed feedback helps with language writing. Yao [25] (2025) established an AI-based bilingual education platform, which dynamically adjusts students' personalized learning paths based on their learning proficiency, and the automatic response system achieves timely correction and feedback, so that students' learning satisfaction and learning outcomes in bilingual education are improved.

In addition, universities have established partnerships with industry to promote industry-education integration through cooperative projects, practical training and practice, platform education, and teaching reform, which provide guidance for the construction of language education practice models. Li [26] (2021) proposed a model of combining university English industry, education, and research based on an online education platform, in which an intelligent generation algorithm is embedded in the platform to automatically generate English test questions that are composed with industry, education, and research, and to promote the improvement of students' practical language. Zhang [27] (2022) developed an integrated platform for teaching minority languages in countries along the “Belt and Road”, which combines business and trade along the road and teaching minority languages, providing an important pillar for the cultivation of composite talents in language+trade as well as the development of trade in countries along the road. Liu et al [28] (2024) proposed a model for teaching spoken English in the context of industry-teaching integration, which enabled students' practical speaking skills to be enhanced by clarifying teaching objectives and principles, innovating teaching methods and content, integrating educational resources and building support systems. Guo [29] (2024) proposed a student-oriented foreign language education strategy for the status quo of modern university education to optimize the curriculum and reform education with resource cooperation, industrial integration, and personalization, and to

promote the practice-oriented learning model, as well as personalized, competence-oriented, and output-oriented teaching models to carry out. Fitria [30] (2024) proposes to include digital business scenarios such as e-commerce and fintech in English language teaching and learning, where students conduct business exchanges through a multilingual communication platform, which promotes the enhancement of students' language skills, communication skills, and problem-solving skills. Xu and Zhang [31] (2025) designed a vocational digital teaching material model for medical English in vocational colleges and universities, taking result-oriented education as the guiding principle, through the design of clinical workflow-related medical vocabulary integration, AI scenario simulation, situational dialogues, and adopting the “three-lane integration” teaching mode, with the help of an AI-enhanced education platform based on the dynamic updating of educational resources. It adopts a “three-line integration” teaching model, and promotes the dynamic updating of educational resources through an AI-enhanced educational platform.

In view of the fact that language education has been detached from the reality of production and life for a long period of time, this paper refers to the “three platforms and six links” teaching paradigm in radio and television education to construct a practical model of language education. According to the teaching idea of “three platforms and six links”, the core of teaching improvement is to enhance students' practical language application ability. A set of auxiliary teaching tools integrating speech recognition and automatic language assessment is constructed in the study. Through the maximum a posteriori probability reappraisal accurately recognizes the teacher's speech and students' expressions in the language teaching classroom, and obtains the educational data that can be analyzed. The data are put into a language assessment model to train a language assessment feedback mechanism that is sufficiently reflective of the students' true level. With the above technical support, students are encouraged to fully practice and design pedagogical research to subjectively test the effectiveness of language learning.

2. Design of the “three platforms and six links” language education practice model

Based on the reflection on the expected abilities of language major students and the research on the design logic of practical teaching models, this study has applied the concept of "vertical and horizontal coordination and integration" to construct a practical teaching model of "three platforms and six links". On the vertical axis, based on the "three-level advancement" type of goal ability orientation, the "Professional Practice Teaching Platform" aimed at enhancing professional practical ability, the "innovation and entrepreneurship Practice Platform" aimed at enhancing innovation and entrepreneurship ability, and the "industry and enterprise Practice Platform" aimed at enhancing integrated application ability have been constructed. The three platforms are not only in a "Sichuan" character structure relationship with their own focuses and parallel classifications, but also in an inverted "product" character structure relationship of interdependence and layering superposition. On the horizontal axis, in accordance with the "law of ability formation" and the process of talent cultivation, six key links in the cultivation of practical abilities are emphasized and coordinated, namely professional cognition, course training, joint creation, on-the-job practice, innovation and entrepreneurship, and graduation internship. Each link is interconnected and integrated, organically supporting the growth of students' abilities at different stages.

3. Technical support for practical teaching models

In the “three platforms, six links” practical teaching mode for language majors, the core cultivation goal is the students' language application ability. The article introduces the artificial intelligence method and builds a technical framework through adaptive speech recognition and automatic language evaluation to empower the practical teaching mode.

3.1. Speech Recognition Methods for Language Education

Current speech recognition systems can be categorized into two types according to their dependence on the speaker, namely, person-specific systems (SD) and non-person-specific systems (SI). The reason for the gap between the performance of person-specific and non-person-specific systems is obvious. Non-specific person systems use a very wide range of speakers' speech to train the model of the recognition system, which ensures enough data to accurately portray the various complex time-varying properties, co-articulation, etc. of the speech individually, but makes the differences between speakers negligible, thus decreasing the accuracy of the system in modeling individual speakers.

3.1.1. Differences between and within speakers

(1) Differences between speakers and talkers

For healthy people, each person's speech has its own characteristics. The speech produced by a person who is speaking is affected by many factors, such as: the length, width and physical shape of his vocal tract, age, sex, health, education, personal pronunciation habits, etc. These differences make it possible for one person's speech to be completely different from another's. There are two main areas of difference between speakers: physiological differences and differences in speaking habits.

Physiological differences are mainly due to the fact that each person's vocal organs have different shapes, sizes, and dynamic properties. Such physiological differences have a significant effect on the fundamental frequency of speech, resulting in different acoustic characteristics for different people (this is also a major factor in the differences between men and women). An extreme example of this situation is the effect of speaker gender on the spectral parameters of speech. If a model of speech production is developed that is easy to analyze, it can be found that the value of the fundamental frequency f depends on the dimensions and characteristics of the vocal folds, as well as on the tension to which they are subjected. The current study shows that males and females have significantly different resonance peak frequencies when producing vowels, with males producing vowels with a lower fundamental frequency, a narrower resonance peak bandwidth, and a flatter spectrum. This is why person-specific systems trained with male speech work poorly on female or bisexual tests.

(2) Intra-speaker differences

Even if we ignore inter-speaker differences, for the same speaker, there can be considerable differences in words that tell the same content at different times and in different psychological and physiological states. This is because there are differences in vocal tract shape and speech rate between each pronunciation. This difference is even more pronounced when a person speaks loudly or quietly due to changes in feelings. This variation in a person's own pronunciation is called intra-speaker variation. It consists mainly of the influence of such factors as speed of speech, emotional tone, and health condition. A variation in one of these factors may cause a significant degradation in the performance of this speaker's trained recognition system.

Overall, the factors contributing to acoustic variation across speakers (inter-speaker variation) are subtle and extensive, and are much larger and more difficult to capture and characterize than the factors contributing to acoustic variation within a specific speaker (intra-speaker variation). In some recognition systems, it is necessary to distinguish between inter-speaker variation and intra-speaker variation. For example, recognizing the speech of a particular person from the speech of many people takes into account inter-speaker differences and mitigates intra-speaker differences. However, for non-specific person speech recognition systems, both aspects have to be taken into account, whether it is a change in the speaker or a change in the pronunciation conditions. Unfortunately, it has not yet been possible to build a more accurate model for this purpose, so the only way to reduce the involvement of personal characteristics is to resort to statistical methods and to obtain some kind of “average sense” of the information through extensive training. However, it is also the deliberate weakening of personal information that causes the system to be less effective in recognizing a particular person. In order to solve this problem, speaker adaptation techniques have been developed.

3.1.2. Adaptive techniques for speech recognition

MAP improves the adaptive effect by introducing a priori knowledge to maximize the a posteriori probability. So in contrast to the maximum likelihood (ML) reestimation method, this method is called the maximum a posteriori probability (MP) reestimation method.MAP is currently one of the main methods for model parameter tuning, i.e., Bayesian Adaptation.

Adaptation to new speakers for discrete HMM model parameters was the first introduction of MAP re-estimation methods for self-adaptation, which was later extended to adaption to continuous HMM model parameters, and then later, research gave ways to transform HMM model parameters with mixed Gaussian output distributions. At present, most of the various HMM-based recognition systems, especially the larger ones, use the MAP method for adaptation. In the following, we look at the principle and implementation of MAP, starting with the difference between MAP and ML.

Since speaker adaptation is expected to use a small amount of adaptive data from new speakers to adapt the system, the problem of sparse training data is often encountered. The current HMM model parameter training generally uses the classical Baum-Welch maximum likelihood (ML) re-estimation method, and this ML algorithm can only be optimal with a large and sufficient amount of corpus training. Therefore, the standard ML training method for HMM model parameters does not work well with sparse data. So in order to solve the problem of insufficient adaptive data, the a priori information of the HMM model was introduced into the model training process, and the maximum a posteriori probability method (MAP) was developed.

The principle of MAP revaluation is very similar to that of ML revaluation, and the fundamental

difference lies in whether or not the prior distribution of parameters is used in the revaluation process.

Suppose $O = \{o1, o2, ..., oT\}$ is a series of observations with a probability density function (p.d.f.) of $p(o)$, and λ is the set of parameters defining the distribution. The reestimation problem can be viewed as the process of estimating λ given a sequence O of training data. This process we can realize by finding the following equation:

$$\lambda_{estimate} = \arg\max_{\lambda} p(\lambda \mid o) \tag{1}$$

Apply the Bayesian criterion, where $p(\lambda)$ is the prior distribution of the HMM parameters:

$$p(\lambda \mid O) = \frac{p(\lambda \mid O)p(\lambda)}{p(O)} \tag{2}$$

Get:

$$\lambda_{estimate} = \arg\max_{\lambda} \frac{p(\lambda \mid O)p(\lambda)}{p(O)} \tag{3}$$

In the traditional estimation formula for ML, the model parameter λ is considered to be unknown but fixed, while the probability density function $p(o)$ is independent of the model parameter, so that the denominator $p(O)$ in the fractional equation and $p(\lambda)$ in the numerator are ignored, and $p(o \mid \lambda)$ is simply maximized. That is, the

$$\lambda_{ML} = \arg\max_{\lambda} p(o \mid \lambda) \tag{4}$$

And the most important feature of the MAP revaluation method is that it introduces the consideration of the prior distribution $p(\lambda)$ of the HMM parameters, i.e., it is considered that the model parameter λ is a random variable that conforms to the prior distribution $p(\lambda)$. Therefore, only the denominator part of the formula is ignored in the revaluation process. That is, the

$$\lambda_{\text{map}} = \arg\max_{\lambda} p(\lambda \mid o) = \arg\max_{\lambda} p(o \mid \lambda)p(\lambda) \tag{5}$$

Comparing Eq. (4) and Eq. (5), we can consider ML reestimation as a special case of MAP reestimation in the case of ignoring the assumption on the prior distribution of the parameters, whereas MAP introduces a prior distribution $p(\lambda)$ on the HMM parameters λ into the ML reestimation.

It is worth noting that the use of prior knowledge (e.g., prior distributions of the parameters to be estimated) is crucial for the success of MAP. And playing an important role in the simplification of MAP revaluation is a concept called the family of distributional conjugates. The conjugate prior of a random vector is defined as the prior distribution of the parameters of the probability density function (p.d.f.) of this vector, such that both the posterior distribution $p(\lambda \mid o)$ and the prior distribution $p(\lambda)$ belong to the same family of distribution functions. For example, the conjugate prior of the mean of a Gaussian probability density function is also a Gaussian density.

3.2. Language Assessment Module

The language assessment model based on the attention mechanism and semantic embedding is an improved model based on the language assessment model based on long and short-term memory neural networks and the language assessment model based on bi-directional gated recurrent units and generative adversarial networks. The encoder of the optimized language assessment model is more accurate in extracting linguistic and audio features, where an attention mechanism is applied to improve the accuracy of the encoded audio feature vectors corresponding to the decoded natural language text. The decoder outputs more accurate and flexible reviews of speech and audio because the decoder not only applies LSTM neural network but also incorporates the semantic embedding method Bert.

3.2.1. Audio Feature Encoder

The basic idea of audio feature encoder in deep learning based language assessment model is to input the original audio MFCC feature sequence $\vec{a} = (a_1, a_2, ..., a_n)$, and generate the audio coding feature vector $\vec{a}'_t = (a'_1, a'_2, ..., a'_n)$ and hidden vectors $\vec{h}_n$, $\vec{h}'_n$. The audio feature encoding generated by the encoder forms a correspondence with the rubric, which is learned by the model. The audio feature

encoder and the text feature decoder improve the accuracy of the correspondence between the audio feature vectors and the part of the rubric through the attention mechanism.

The audio feature coding in the audio feature encoder part uses Bi-GRU neural network, which can process the audio sequence sequentially in the time dimension in both sequential and inverse order to obtain the feature representation at each moment, using Bi-GRU not only in relation to the output of the current moment, but also in relation to the state of the subsequent moments.The specific design of the Bi-GRU structure takes into account the speech features as well as the time sequence length and other aspects. The application of the attention mechanism can improve the probability of the correspondence between the speech features and the rubrics in the rubric generation decoder, and improve the accuracy of generating the corresponding rubrics for the speech features.

The audio feature encoder inputs the MFCC feature sequence $\vec{a} = (a_1, a_2, ..., a_n)$. After the gated loop unit is fed with the feature sequence, a hidden vector is finally generated, and Eqs. (6) to (9) show the computation process.The Bi-GRU neural network performs a bidirectional computation, where the input sequence is computed from left to right to get a hidden vector $\vec{h}_n$, while the other gated loop unit is computed in the opposite direction to get another hidden vector $\vec{h}'_n$. The overall information of the forward and backward audio sequences can be obtained by the Bi-GRU neural network, and the final output audio feature vector $\vec{a}' = (a'_1, a'_2, ..., a'_n)$, and get two hidden vectors $\vec{h}_n$ and $\vec{h}'_n$.

$$\vec{r}_t = \sigma\left(W_r \cdot [\vec{h}_{t-1}, \vec{a}_t]\right) \tag{6}$$

$$\vec{z}_t = \sigma\left(W_z \cdot [\vec{h}_{t-1}, \vec{a}_t]\right) \tag{7}$$

$$\tilde{\vec{h}}_t = \tanh\left(W_{\tilde{h}} \cdot [\vec{r} \otimes \vec{h}_{t-1}, \vec{a}_t]\right) \tag{8}$$

$$\vec{h}_t = (1 - \vec{z}_t) \otimes \vec{h}_{t-1} + \vec{z}_t \otimes \tilde{\vec{h}}_t \tag{9}$$

The two hidden vectors $\vec{h}_n$ and $\vec{h}'_n$ obtained in the Bi-GRU neural network layer are spliced, and then go through the fully connected layer to obtain the final $\vec{K}$, which will be used as the initial hidden layer of the decoder. The formula for $\vec{K}$ is shown in (10), where W_K represents the weight matrix and $\vec{b}_K$ represents the deviation vector. The audio feature vector $\vec{a}' = (a'_1, a'_2, ..., a'_n)$ and the comment word vector $\vec{y} = (y_1, y_2, ..., y_n)$ as input to the attention mechanism network. The audio feature vector $\vec{a}' = (a'_1, a'_2, ..., a'_n)$ and comment word vector $\vec{y} = (y_1, y_2, ..., y_n)$ input, the attention network first determines the similarity to get the attention score, and then gets the attention weight α_{ij} by Softmax function, and the attention weight formula is shown in equation (11). Finally the attention network obtains the reconstructed content vector y'_j, i.e., the word vectors in the rubric, by learning the obtained attention weights α_{ij} and the audio coding features $\vec{a}'$. The formula for the attention mechanism is shown in Equation (12).

$$\vec{K} = \vec{h}_n \oplus \vec{h}'_n\left(W_K \cdot \vec{a}_t + \vec{b}_K\right) \tag{10}$$

$$\alpha_{ij} = Soft\max\left\{pool\left[y_t \cdot \left(a_t'^T\right)\right]\right\} \tag{11}$$

$$y'_j = \sum_{i=1}^{T} \alpha_{ij} \vec{a}' \tag{12}$$

3.2.2. Evaluation Models for Generative Adversarial Networks

The underlying model based on recurrent neural networks has the ability to generate understandable and more relevant evaluations under some scenarios. However, since it only uses the loss of generated evaluations versus real evaluations, the output is too constrained and rigid for some audios. Generative Adversarial Networks are considered to solve the above problem.

Compared with the neural network training method of data compression and reconstruction, GAN is no longer limited to the calculation of the distance between the real data and the generated data to judge

the authenticity of the generated data. It constructs a $real / fake$ classifier to supervise the generation effect of the generator.The advantage of the GAN method is that it can make the data generated by the generator get closer and closer to the real data through generative confrontation, and gradually reach infinitely close to the real data, but not exactly the same as the real data. Through the GAN method, the model can generate evaluations that are close to the real but different from the evaluations given by the experts, thus solving the problem based on the existence of a recurrent neural network underlying the model.

The discriminator part of the model is designed to consist of an embedding layer, an LSTM unit and a fully connected layer. The purpose of training the discriminator is to maximize the judgment of whether the input evaluations are real or generated evaluations, and thus monitor the effectiveness of the generator. The inputs to the discriminator are generated pseudo-expert reviews and real reviews, and the output is the result of judging the evaluation data $real / fake$. The discriminator part first embeds the fake or real reviews into the matrix $E = (\overrightarrow{e_1}, \overrightarrow{e_2}, ..., \overrightarrow{e_n})$, and then the LSTM unit takes the embedding $\overrightarrow{e_i}$ of each word as input and the output hidden state $\vec{h}_t$. Finally, we feed the last hidden state vector into a fully connected layer with an activation function and output the probability that the input comment is true.

$$\vec{f}_t = \sigma\left(W_f \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_f\right) \tag{13}$$

$$\vec{i}_t = \sigma\left(W_i \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_i\right) \tag{14}$$

$$\vec{\tilde{c}}_t = \tanh\left(W_C \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_C\right) \tag{15}$$

$$\vec{C}_t = \vec{f}_t \otimes \vec{C}_{t-1} + \vec{i}_t \otimes \vec{\tilde{c}}_t \tag{16}$$

$$\vec{o}_t = \sigma\left(W_o \cdot [\vec{h}_{t-1}, \vec{x}_t] + \vec{b}_o\right) \tag{17}$$

$$\vec{h}_t = \vec{o}_t \otimes \tanh\left(\vec{C}_t\right) \tag{18}$$

The loss function of the discriminator can judge the ability of the discriminator, the loss function of the discriminator is defined as shown in equation (19). In Eq. (19), $G(A, \vec{z})$ denotes the review generated by the generator; $D_{real/fake}(x)$ denotes the result of the discriminator's judgment of x; and $\vec{T}$ denotes the real sample.

$$L^D_{real/fake} = -E\left(logD_{real/fake}\left(T\right)\right) + E\left(1 - log\left(D_{real/fake}\left(G\left(A, \vec{z}\right)\right)\right)\right) \tag{19}$$

The generator part of the model follows the structure in the base model based on recurrent neural networks. The generator is divided into an extraction part and a generation part. First, the extraction part extracts the compressed knowledge from the input audio features. Then, Gaussian distributed random vectors $\vec{z}$ are created and connected to the extracted knowledge. Finally, the processed knowledge is fed into the generation part to generate comments in the form of vectors.

One of the criteria for the strength of the generator is the ability to generate sufficiently realistic reviews. Therefore, the reviews generated by the generator need to have the ability to deceive the discriminator. The $real / fake$ loss function is designed to evaluate this ability of the generator. The $real / fake$ loss function is shown in equation (20) below.

$$L^G_{real/fake} = E\left(-logD_{real/fake}\left(G\left(A, \vec{z}\right)\right)\right) \tag{20}$$

Relying on the supervision of the discriminator alone would leave the generation results with a great deal of uncertainty. Therefore, the distance between the generated evaluations and the true evaluations is evaluated as well, which constrains the generator's generation effect even further. Equation (21) represents this loss function, where α denotes the balancing factor.

$$L^G_{dis} = \alpha \times \left(G\left(A, \vec{z}\right) - \vec{T}\right)^2 \tag{21}$$

Equation (22) represents the loss function of the generator.

$$L^G = L^G_{real/fake} + L^G_{dis} \tag{22}$$

4. Technical validation of language education practices

4.1. Speech Recognition Effect

4.1.1. Audio Coding Noise Reduction

The captured classroom video files are converted to mono audio files in wav format with a bit rate of 125kbps and a sampling rate of 24k. Then it was subjected to noise reduction, and the time domain comparison analysis before and after noise reduction is shown in Fig. 1 and Fig. 2. After noise reduction, the classroom voice is more discrete and independent in time domain distribution, and is no longer a continuous audio signal on the time axis, which naturally sounds clearer in practice. Meanwhile, in the direction of the vertical axis, the amplitude (decibels) of the sound is more concentrated after noise reduction. The distribution range is gathered from the original (60.4dB,79.7dB) to (62.3dB,77.8dB), eliminating excess noise and making it more comfortable to listen to.

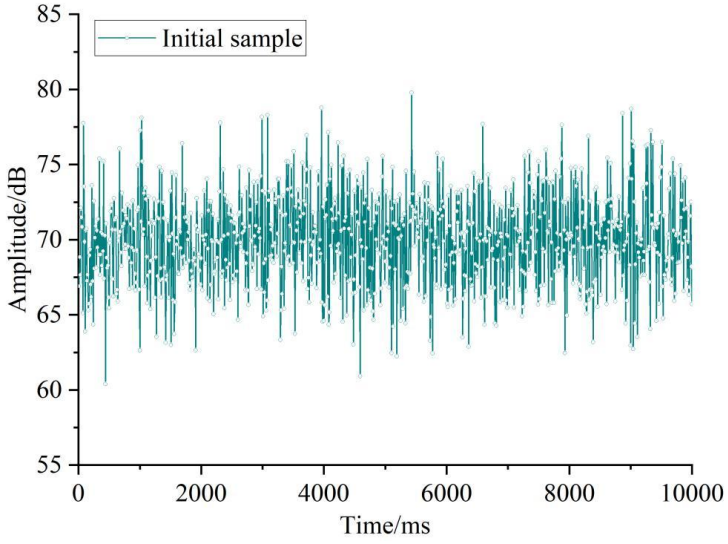


Figure 1. Time-domain expansion of the speech before noise reduction.

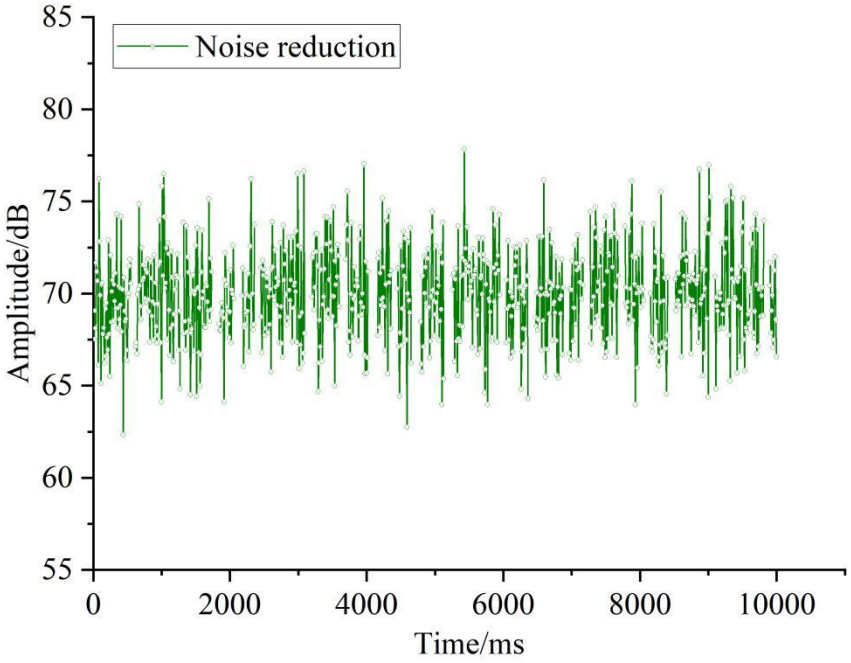


Figure 2. Time-domain expansion of the noise-reduced speech.

4.1.2. Speech recognition error rate

For the same speech data, ASRT, KDDI and Baidu AI were used for speech recognition, and the final text was compared by string comparison. The final text was compared by string comparison. The differences in recognition were based on the principle of majority rule, and the content recognized by the majority of the speech recognition tools was chosen as the final result. The word error rate is calculated for the final result and the word error rate statistics are shown in Table 1. The word error rate of the MAP-based adaptive speech recognition method is better than that of using a single speech recognition tool in all six time dimensions, especially when the speech data is less than 10 min, the sub-error rate of speech recognition reaches an amazing 3.95%. Speech recognition models using multi-method fusion can effectively compensate for some of the word error rates generated by using a single method for recognition.

Table 1. Comparison results of character error rates in speech recognition (%).

	iFLYTEK	Baidu AI	ASRT	This method
<10min	4.43	5.97	6.3	3.95
10-20min	5.13	8.97	8.12	3.37
20-30min	5.9	9.76	10.88	4.68
30-40min	8.08	9.95	12.18	6.82
40-50min	10.21	11.12	15.55	10.17
50-60min	12.02	13.45	18.91	10.76

Of course, the above data does not mean that the performance of the present model has fully surpassed the mature speech recognition systems on the market. Rather, for the research scope of this paper, the special speech domain in language classroom this model has unique advantages.

4.2. Application of the automatic language assessment module

In language teaching, speech recognition has been validated as an important data collection detection tool for the teaching task. Next, the study continues to examine the feasibility of language assessment technology, which is the practical tool that can more directly touch the teaching and learning activities in language education.

4.2.1. Comparison of evaluation effects

In this chapter, the following spoken language evaluation models are used for comparison: the mainstream phoneme-level speech recognition model (ASR), the textual a priori based phoneme-level speech recognition model (TC-ASR), and the textual a priori based end-to-end spoken language evaluation (TC-Direct). The results of the language review are shown in Table 2. For the language evaluation task, the model needs to strike a balance between finding mispronunciations and accepting correct pronunciations. Therefore, this paper uses F1 as the main evaluation metric for performance. After that, the model is further categorized into True Acceptance,False Rejection,False Acceptance and True Rejection according to whether the predicted error states are consistent with the judgment. Meanwhile, False Rejection Rate (FRR) and False Acceptance Rate (FAR) are equally important for the speaking assessment task.

The F1 score of the ASR model has improved considerably to 0.2047. Comparing the TC-ASR model and the TC-Direct model, since the TC-Direct adopts the target text as an additional input condition, the decoding process is implicitly limited to the similar phonemes of the target text. As a result, the rejection rate FRR has been reduced significantly and the accuracy of speech recognition has been improved. However, the False Recognition Rate FAR has also increased significantly at the same time, so its F1 score is not improved. For the model proposed in this paper, the model is directly optimized for the evaluation state, and thus achieves the best F1 score of 0.689 in the “original phoneme” annotation group.

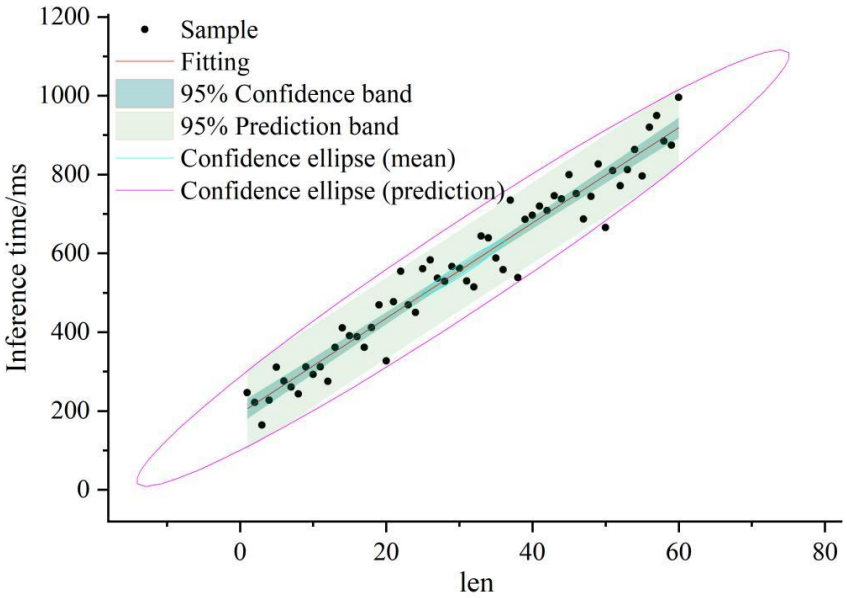
Table 2. Language evaluation results.

	ASR	TC-ASR	TC-Direct	This method
PER	0.2902	0.4102	0.6737	0.429
PRE	0.2911	0.3823	0.8136	0.6232
ACC	0.2047	0.5174	0.8814	0.3169
REC	0.0351	0.5269	0.8302	0.6088
FRR	0.2902	0.4102	0.1023	0.429

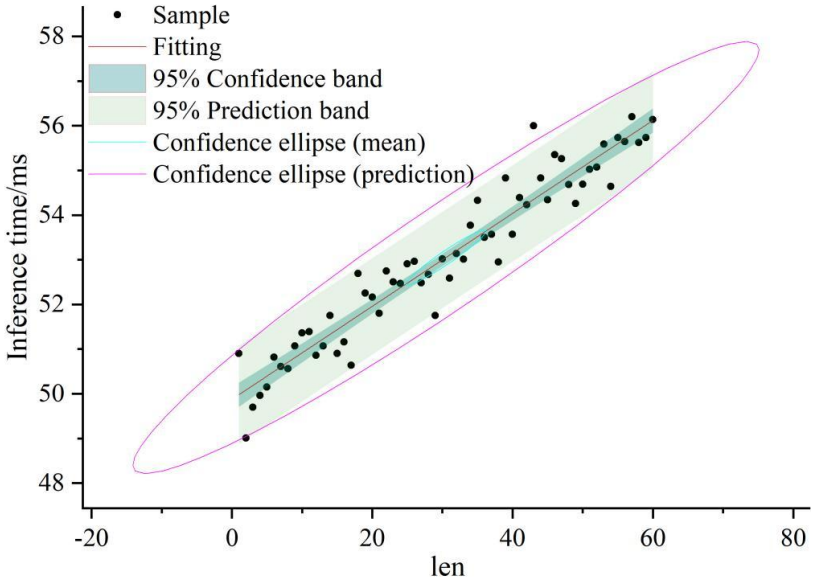
FAR	0.2911	0.3823	0.8136	0.6232
F1	0.2047	0.5174	0.524	0.689

4.2.2. Real-time testing

We performed latency evaluation on the same servers. When limiting the data loading function to a single corpus (batchsize=1), the inference length of each model is shown in Figure 3. After counting the inference durations of the models for different values of len, the length of each target phoneme, curve fitting was performed on the inference durations. For the test set, each sentence contains about 25 target phonemes on average. This indicates that on average about 25 forward operations are required for the original base model to obtain each decoded phoneme in turn. As a comparison, the textual prior-based approach proposed in this paper requires only one forward computation and thus does not vary much with the target phoneme length len. Therefore, the present language assessment model can be adapted to different language learning contexts and provide an effective assessment of students' real language proficiency.



(a) TC-Direct



(b) This method

Figure 3. The inference time of the language assessment model.

5. Validation of the application of the language education practice model

This paper is based on actual classroom observation, and the target students are all intermediate and advanced Chinese learners in the third class of the School of Chinese Language and International Education of a university. The study started on November 1 and lasted for two months with the “Three Platforms, Six Sessions” practical teaching model.

In the course of the study, students' self-assessment of language proficiency and teachers' evaluation were conducted every six days for both students and teachers (10 evaluations were conducted in total). The results of the language proficiency evaluation are shown in Table 3, which was analyzed by recovering and analyzing the teacher's evaluation scales and the students' self-assessment scales. The mean score of teacher's evaluation of learners' classroom performance increased from 75.4 out of 100 at the beginning to 88.6, and learners' evaluation of their own performance increased from 72.1 out of 100 at the first time to 81.6, with a more significant increase in the mean score of the ratings of both scales. From the “teacher-student” dual-perspective performance evaluation of student performance, it can be seen that both teachers and students believe that the appropriate and reasonable use of AI technology can positively promote learners' Chinese language learning and classroom performance, and can achieve the good effect of assisting learning. The results of this study are summarized as follows

Table 3. Language proficiency evaluation results.

Sorting	Teacher	Student
1	75.4	72.1
2	75.5	73.8
3	81	75.6
4	82.1	76.4
5	82.8	77
6	82.9	77.9
7	85.1	79.4
8	87.4	79.9
9	87.8	80.9
10	88.6	81.6

6. Conclusion

The “three platforms and six links” language education model constructed in this paper breaks the previous situation of single content and lack of learning surfaces in language teaching, and strengthens the core objective of practical language use. Among them, the MAP-based speech recognition method provides a quantitative research method that has always been missing in language education, and the suberror rate of only 3.95% (when the segment is <10min) can fully guarantee the accurate digitization of the classroom corpus for language learning. The language assessment mechanism built by audio feature coding and improved GAN can realize the objective evaluation of students' language expression, and the F1 value of comprehensive performance achieved 0.689, which is verified to be scientifically valid. The teaching experiment in the School of Chinese International Education of a university further confirms that this paper has successfully explored a new path of language education that emphasizes practical ability by integrating artificial intelligence technology and the “three platforms and six links” education theory.

References

1. Artyukhina, M., Dorokhova, T., Vyguzova, Y., & Nachernaya, S. (2018). Practical oriented training as formation conditions of professional communication. European Proceedings of Social and Behavioural Sciences, 51.
2. Zhang, D. (2023). Practical Education and the Cultivation of Professional Competence: Reform Strategies for Vocational Education. Advances in Vocational and Technical Education, 5(13), 16-22.
3. Xu, Y. (2022). Exploration of practical teaching reform based on high-quality international talent cultivation. ICCCM Journal of Social Sciences and Humanities, 1(3), 38-48.
4. Siegel, A., Vance, M., & Nilsson, D. (2025). Military English language education: a scoping review of 30 years of research. Innovation in Language Learning and Teaching, 19(5), 433-451.
5. Misechko, O. Y. (2017). Multiple impact of international cooperation in the language education of children in Ukraine after 1991. Multidisciplinary Journal of School Education, 6(2 (12)).
6. Ji, Y., Li, J., Gao, D., & Sun, Y. (2025, June). the Background of Industry Education Integration. In Proceedings of the 2025 3rd International Conference on Language, Innovative Education and Cultural Communication (CLEC 2025) (Vol. 938, p. 297). Springer Nature.

7. Putra, R. C., Barliana, M. S., Komaro, M., & Hamdani, A. (2025). A Systematic Literature Review of Integrated Learning Models for Skills Development in Industry-Academia Partnerships: Preparing Workforce for Industry 4.0. VANOS Journal of Mechanical Engineering Education, 10(1), 45-61.
8. Zhuang, T., Oh, M., & Kimura, K. (2025). Modernizing higher education with industrial forces in Asia: A comparative study of discourse of university-industry collaboration in China, Japan and Singapore. Asia Pacific Education Review, 26(1), 195-210.
9. Guo, Z. (2023). Evaluation on the Prospects of School Enterprise Cooperation and the Integration of Industry and Education in Vocational Education in the 5G Era. The Frontiers of Society, Science and Technology, 5(13).
10. Ye, R. (2024). Research on the Cultivation of Business English Majors with "Double Creation" Under the Background of School-enterprise Cooperation. The Educational Review, USA, 8(8).
11. Li, H. (2019, August). Analysis of School-enterprise Cooperative Training Path for Translation Major from the Perspective of Application Transformation. In 5th International Conference on Arts, Design and Contemporary Education (ICADCE 2019) (pp. 1056-1059). Atlantis Press.
12. Li, S. (2024). Application of OBE Educational Philosophy in English Listening and Speaking Courses in Vocational Colleges under the Context of Industry-Education Integration. International Journal of New Developments in Education, 6(7).
13. Dyorina, N. V., Antropova, L. I., & Zalavina, T. Y. (2018). Integration Processes in Successful University and Corporate Professional Training in Foreign Languages. Arab World English Journal, 9(4), 200-210.
14. Liao, M. (2023). Integration of industry and education: A Study on the multimedia innovation of English Curriculum in Private Higher Vocational Colleges in Guangdong. In SHS Web of Conferences (Vol. 168, p. 01029). EDP Sciences.
15. Sun, L., & Li, Z. (2019, October). Research on the Training of Language Service Professionals Based on the School-Enterprise Cooperation. In 2019 International Conference on Advanced Education Research and Modern Teaching (AERMT 2019) (pp. 280-283). Atlantis Press.
16. Kuddus, K. (2022). Artificial intelligence in language learning: Practices and prospects. Advanced analytics and deep learning models, 1-17.
17. Zou, S. (2017). Designing and practice of a college English teaching platform based on artificial intelligence. Journal of Computational and Theoretical Nanoscience, 14(1), 104-108.
18. Negrila, A. M. C. (2023). The new revolution in language learning: The power of artificial intelligence and education 4.0. Bulletin of" Carol I" National Defence University (EN), 12(02), 16-27.
19. Wu, T., & Yu, Z. (2024). Bibliometric and systematic analysis of artificial intelligence chatbots' use for language education. Journal of University Teaching and Learning Practice, 21(6), 174-198.
20. Du, J., & Daniel, B. K. (2024). Transforming language education: A systematic review of AI-powered chatbots for English as a foreign language speaking practice. Computers and Education: Artificial Intelligence, 6, 100230.
21. Chandra, K. R., Muthumanikandan, M., Kathyayini, S., Akhila, H. G., Pathak, P., & Shivaprakash, S. (2024). The impact of artificial intelligence tools and techniques for effective English language education. Nanotechnology Perceptions, 20(S7), 897.
22. Li, F. (2025). Changes in English Education Based on Artificial Intelligence and the Construction of a New Mode of International Communication. Mediterranean Archaeology & Archaeometry, 25(1).
23. Nie, L. (2024). Construction and Practice of an Interactive Virtual English Learning Community on Campus Information Platforms. Journal of Modern Educational Theory and Practice, 1(1).
24. Akhter, E. (2024). AI-BASED LANGUAGE EDUCATION PLATFORMS: A SYSTEMATIC ANALYSIS OF EDTECH TOOLS FOR ENGLISH PROFICIENCY. Journal of Sustainable Development and Policy, 3(04), 01-31.
25. Yao, B. (2025). AI-based Thai-English Bilingual Education Platform: Personalized Learning and Automatic Feedback System. Journal of Criminal Investigation and Criminology, 76(2).
26. Li, S. (2021, May). Mode of Combination of Production, Teaching and Research of College English Based on Online Education Platform. In 2021 2nd International Conference on Computers, Information Processing and Advanced Education (pp. 1553-1556).
27. Zhang, L., Xie, X., & Chen, S. (2022). On the Establishment of an Integrated Minority Language Teaching and Trade Platform in Countries along the Belt and Road Initiative. JOURNAL OF SIMULATION, 10(1), 119.
28. Liu, X. (2024). Research on Oral English Teaching Model in Universities under the Background of Industry-Education Integration. Journal of Modern Educational Theory and Practice, 1(1).
29. Guo, J. (2024, September). Optimizing Data Intelligence Empowerment and Industry-Education Integration in University Foreign Language Teaching. In 2024 3rd International Conference on Science Education and Art Appreciation (SEAA 2024) (pp. 149-154). Atlantis Press.
30. Fitria, T. N. (2024). Teaching Digital Business: Integrating Digital Business Materials into English Language Teaching (ELT). JURNAL ILMIAH EDUNOMIKA, 8(3).
31. Xu, J., & Zhang, J. (2025). Exploration on the Innovative Development Model of Digital Teaching Materials in Higher Vocational Colleges in China Under the Background of Industry-education Integration: Taking Digital English Teaching Materials in Higher Vocational Medical Colleges as an Example. The Educational Review, USA, 9(7).