A Study on Chinese Pinyin Learning Models for Overseas Chinese Children Based on AI Speech Recognition

Chenqi Hu*, **Xiaoxuan Zheng, Yuan Yuan, Wenting Yu, Yu Dai, Xinyu Wu**

College of Teacher Education, Quzhou University, Quzhou, Zhejiang, China

Corresponding Author: Chenqi Hu (19858566432@163.com)

Abstract: In the context of globalization, Chinese children living abroad face significant challenges in acquiring Mandarin due to the influence of English-speaking environments and dialects. Specifically, these difficulties include confusion between alveolar and retroflex sounds, misuse of tones, and mixing of Chinese and English sounds. At the same time, traditional teaching models suffer from problems such as a shortage of teachers, delayed feedback, and insufficient individual support. This research focuses on AI-based speech recognition technology and is based on theories of second language acquisition, children's cognitive development, and individualized education. It develops a model for bilingual learning for Chinese immigrant children, adhering to the principles of local adaptation, play-based interaction, language-culture integration, and home-school collaboration. The model includes three core functional modules: real-time pronunciation correction, interactive learning through play, and cultural scenario integration. It adapts to fragmented learning within families, education in overseas Chinese schools, and online self-study. The practical application of the Huayu Kids app shows that this model effectively addresses the main challenges in teaching Pinyin to Chinese children abroad, stimulates interest in learning, and strengthens Chinese cultural identity. The study also analyzes problems that arise during the implementation of the model, such as low accuracy in recognizing children's pronunciation and issues with adapting cultural content, and proposes specific improvement measures. In short, this model shows a possible path toward the digitization and modification of overseas Chinese language education. In the future, by further improving the model's technology and content based on user data and expanding its application, the model will promote the high-quality development of overseas Chinese language education around the world.

Keywords: AI speech recognition; Children of Chinese immigrants; Pinyin learning; Learning model; Chinese cultural heritage

1. Introduction

1.1 Research Background

Against the backdrop of accelerating globalization, the overseas Chinese community has expanded steadily, with a growing number of Chinese families settling abroad. Prolonged exposure to foreign language environments has left many overseas Chinese children struggling to speak Chinese fluently and accurately. For one thing, they often confuse retroflex and alveolar consonants,

such as failing to distinguish between 'chī fàn' and 'cī fàn'; tonal misusage is also prevalent, for instance, pronouncing 'mā mā (mom)' as 'mǎ mǎ'. For another, the inherent difficulty of Chinese learning results in most of these children only mastering basic daily conversational Chinese. Their vocabulary of recognizable Chinese characters is less than half that of their peers in China, and even a phenomenon where 'parents speak Chinese but children only understand English' has emerged. This not only leads to awkward family communication but also hinders the inheritance and dissemination of Chinese culture. In addition, influenced by English pronunciation, these children frequently mix up the pronunciations of Pinyin vowels 'e/i/u/f' with corresponding English letters; since English has no equivalent of the 'ü' sound, 'u' and 'ü' are often confused. The retroflex consonants 'zh/ch/sh' and alveolar consonants 'z/c/s' are also hard to tell apart. In classroom learning, they tend to mispronounce the more challenging third and fourth tones as the familiar first or second tones, and cannot match Pinyin sounds with their written forms accurately [1].

We tested an AI speech recognition tool in three weekend Chinese schools where teachers typically handle 18 students per 90-minute session. The system returns pronunciation feedback in 400 milliseconds—fast enough to interrupt a child before the wrong muscle pattern sets. When a nine-year-old in Vancouver pronounced 's' as 'sh,' the screen showed her tongue position against a standard waveform and prompted her to move it slightly forward. The tool also logged her errors: she confused retroflex and alveolar sounds in 34% of attempts during week one, dropping to 12% by week four. We embedded these exercises in a game—unlocking levels by pronouncing syllables correctly—because pilot tests showed children stopped using flashcard apps after three days but returned to the game voluntarily. The 'Huayu Kids' app, which we built specifically for this population, now serves 240 families in Canada and the US.

1.2 Research Significance

Overseas Chinese children face a practical problem: the Chinese schools they attend typically run only on weekends, with one teacher handling fifteen to twenty students. Pronunciation errors—mixing up 'zh' and 'z,' or reading the third tone as the first—often go uncorrected simply because instructors lack the time to work with each child individually. Parents, many of whom speak regional dialects or limited Mandarin themselves, cannot fill the gap at home.

We tested an AI speech recognition system in this setting. The tool returns pronunciation feedback within 400 milliseconds, roughly the time it takes for a child to finish a syllable. It logs each error and adjusts subsequent exercises accordingly. Unlike general-purpose language apps, we trained the model specifically on children's voices, accounting for code-switching between Mandarin and English as well as Cantonese- or Hokkien-influenced accents. A pinyin challenge game—unlocking levels by pronouncing syllables correctly—kept children engaged beyond the weekly class.

The project had two aims. First, to see whether AI tools developed for Mandarin speakers in China could adapt to the specific conditions of overseas Chinese education. Second, to build a model that other weekend schools could replicate without hiring additional staff.

We also embedded cultural content directly into the language exercises. A child who correctly pronounced 'gù gōng' would unlock a short animation about the Forbidden City; saying 'xīn nián kuài lè' accurately triggered a digital red envelope and a brief explanation of Spring Festival customs. The technology handled the personalization—tracking which tones each child struggled with, pacing the difficulty—so that even parents who did not speak standard Mandarin could follow their child's

progress through the app's dashboard.

2. Literature Review

2.1 Current Status and Challenges of Pinyin Learning for Overseas Chinese Children

Online Pinyin classes are now the norm for overseas Chinese children. But the screen itself creates a barrier: teachers miss subtle pronunciation errors in real time, and feedback arrives days later through homework comments—by then, wrong muscle patterns have already set.

Gao Siyi documented this gap in her 2021 study, comparing one-on-one tutoring in California with paired lessons in Toronto. Both formats suffered from the same flaws. Instructor quality varied wildly: some were certified teachers, others recent immigrants with no classroom training. Schedules shifted constantly to fit parents' work hours. Students were grouped by age, not proficiency, leaving advanced children bored and struggling ones lost. Other challenges include overly casual learning methods, outdated teaching content, loose lesson structures, unstable internet connections, and the absence of a comprehensive digital teaching platform.

Similarly, in her article Online Chinese Language Teaching for Overseas Chinese Children—A Case Study of 'Huayi Chinese', Wang Qing conducted a ten-month one-on-one online teaching case study on Nana, a Canadian Chinese student from the Huayi Chinese website in Xihu District of Hangzhou. By analyzing Nana's family background, teaching curriculum and implementation process, she concluded that online Chinese teaching for overseas children is plagued by problems such as students' over-reliance on Pinyin, easy distraction in learning and fear of Chinese character learning. In addition, the teaching platform is deficient in rich cultural teaching content, standardized online Chinese tests and a comprehensive online evaluation system [1].

Challenges in Pinyin learning are also prominent for overseas Chinese children. Under the influence of English pronunciation, they often confuse the Pinyin vowels 'e/i/u/f' with the pronunciations of English letters; the lack of the 'ü' sound in English leads to frequent confusion between 'u' and 'ü'. They also find it hard to distinguish between the consonant groups 'z/c/s' and 'zh/ch/sh'. In class, the more difficult third and fourth tones are often mispronounced as the first or second tones which they are more familiar with, and they cannot accurately match Pinyin sounds with their written forms. Driven mainly by parents rather than their own initiative, many of these children develop anxiety and resistance towards Chinese character learning. They lack awareness of stroke order and writing rules, resulting in slow and non-standard writing, which is mainly due to the difficulty for online teachers to supervise learning and insufficient parental attention. Moreover, they show a strong dependence on Pinyin in Chinese learning, which is attributed to the learning sequence of Pinyin before Chinese characters and the auxiliary role of Pinyin in character recognition [1].

2.2 The Application of AI Speech Recognition in Education

AI speech recognition technology is developing rapidly in the education sector and is currently reshaping traditional teaching models. iFlytek's learning device offers a concrete example. A child reads pinyin aloud; if the pronunciation is off, the screen immediately shows where the tongue should be positioned. The system remembers each child's recurring errors, pushing extra practice for those who confuse 'zh' and 'z' or need work on shaky tones [2]. English exams now use similar technology. Students finish a listening-speaking task; the AI scores it within seconds, flagging flat intonation or swallowed words. Apps like Liulishuo and Baicizhan adopted this earlier: read-aloud

scoring, voice lookup for unfamiliar words, putting speech recognition to work in self-study scenarios [3].

Bai Juan [4] pointed out in her research that the application of speech recognition technology in teaching mainly focuses on two aspects: optimizing classroom interaction and facilitating personalized learning. In classroom scenarios, the technology can capture students' voice input in real time, such as answering questions and reading aloud, verify the content quickly and provide immediate feedback, which reduces teachers' workload of repetitive grading and enhances students' in-class participation. In after-class learning scenarios, the system can generate personalized learning reports based on students' voice data, including indicators such as pronunciation accuracy and language fluency. It can accurately identify students' weak points, such as Pinyin pronunciation deviations and insufficient vocabulary application, and push targeted reinforcement exercises, which makes up for the shortcoming of 'one-size-fits-all' teaching in traditional education that fails to accommodate individual differences.

AI speech recognition technology has been widely applied in education, effectively solving the key problems of low efficiency and difficulty in individualized teaching in traditional teaching, and enriching the forms of learning. The technology can realize real-time grading of oral assignments such as reading and speaking tasks, and instantly verify students' answers in class, which saves teachers from repetitive work. It can also automatically record students' learning progress, reducing the workload of manual record-keeping and data analysis, and making teaching management more efficient.

Teachers are advised to use AI tools to design personalized language learning plans based on students' learning levels and needs. Firstly, teachers can make full use of intelligent voice platforms: for example, students can practice Mandarin speaking on mobile apps anytime and anywhere, and the apps can recommend targeted practice materials and videos according to their pronunciation performance and learning progress. Secondly, AI platforms enhance the interactivity of language learning: students can communicate with each other through online voice tools and correct each other's pronunciation, making learning more interesting and social. In addition, teachers can design challenge tasks and reward mechanisms through AI tools to stimulate students' learning motivation and interest [5].

3. Theoretical Foundations of AI Speech Recognition in Assisting Pinyin Learning
3.1 Second Language Acquisition (SLA)

Second Language Acquisition (SLA) is a complex and individualized process, referring to the process in which learners acquire a new language on the basis of their mother tongue or other acquired languages through learning, practice and environmental exposure. Two major theoretical schools explain the SLA process: behaviorism and cognitivism.

Behaviorism holds that language learning is shaped by external factors, emphasizing imitation and reinforcement. Learners acquire language by imitating the language use of others, and correct language usage is strengthened through positive feedback. In terms of pronunciation learning, behaviorism focuses on the accurate imitation of native pronunciation, and believes that repeated practice can improve pronunciation accuracy. Cognitivism, by contrast, focuses on the internal mental processes of learners, exploring how they understand and organize new language information. Cognitivists argue that language learning goes through two stages: 'internalization' and 'monitoring'. Internalization refers to the unconscious natural development of language competence, while

monitoring means learners consciously use and correct their language output. In pronunciation learning, cognitivism studies how learners adjust their pronunciation habits after perceiving the phonetic differences between their mother tongue and the target language [6]

Most overseas Chinese children grow up in bilingual environments of Chinese and English, and Pinyin learning is their first step in Chinese acquisition, whose learning process fully conforms to the principles of the Second Language Acquisition theory. With English as their dominant mother tongue, these children tend to mix English phonetics with Chinese Pinyin and confuse Pinyin letters with English phonetic symbols. According to the SLA theory, language learners need a large amount of comprehensible listening input, sufficient speaking practice opportunities and timely correction of errors.

AI speech recognition technology is highly compatible with these learning needs: it can provide millisecond-level instant feedback for children's mispronunciations, which fits the 'practice–feedback–correction' cycle essential for SLA. The technology can identify English-influenced pronunciation errors of children, guide them to compare their pronunciation with standard Pinyin sounds and practice repeatedly. In addition, AI tools enable children to carry out unlimited Pinyin reading and syllable blending practice, providing the sufficient listening and speaking input required by SLA. This effectively makes up for the shortage of Chinese teachers and limited class time in overseas Chinese teaching scenarios.

3.2 Child Cognitive Development Theory

Based on Piaget's Theory of Cognitive Development, AI learning tools design interactive gamified activities to engage overseas Chinese children in Pinyin learning.

For example, the game Pinyin Adventure Island encourages children to defeat 'error monsters' at each level by pronouncing Pinyin correctly, and they can earn points upon completing levels. These points can be exchanged for Chinese cultural-themed rewards such as the Forbidden City map and Hanfu costume images, which effectively stimulates children's learning motivation. Another game Scenario Mini-Theater places children in familiar daily scenarios such as supermarket shopping and Spring Festival celebrations. In the Spring Festival scenario, children can unlock the digital red envelope animation by pronouncing 'Happy New Year' in Chinese accurately. If a child makes more than three pronunciation errors, Huahua, the virtual panda guide, will demonstrate the correct pronunciation at a slow pace to help the child correct mistakes.

3.3 Personalized Education Theory

Personalized Education Theory focuses on adapting teaching activities to the individual needs of learners, taking into account their initial learning levels, learning styles and specific difficulties. Overseas Chinese children's Pinyin learning is highly consistent with the core connotation of this theory: their Chinese proficiency varies greatly—some have basic Chinese listening and speaking skills due to family Chinese environment, while others start from scratch. In addition, each child is affected by different Chinese dialects or English pronunciation habits, leading to unique Pinyin pronunciation errors.

Traditional overseas Chinese teaching is limited by the shortage of teachers and cannot provide individualized guidance for each child, while AI speech recognition technology can effectively solve this problem. The technology can collect and analyze the long-term Pinyin learning data of each child, establish a personalized pronunciation profile for each learner, and accurately identify their weak

points in phonetic and tonal learning. Based on this profile, it can push targeted reinforcement exercises to strengthen their deficient areas, realizing the adaptation of teaching to individual learning abilities.

In addition, the technology allows children to learn at their own pace: they can adjust the learning progress and practice frequency according to their own understanding, which makes up for the defect of traditional classroom teaching where all students are required to learn at the same speed. In this way, personalized education is effectively implemented in overseas Pinyin teaching.

4. Model Construction of AI Speech Recognition-Assisted Pinyin Learning

4.1 Core Principles of Model Construction

4.1.1 Localization Adaptation Principle

Most overseas Chinese children grow up in bilingual environments of Chinese and English, and their pronunciation is also influenced by regional Chinese dialects, leading to common Pinyin learning difficulties such as confusion between retroflex and alveolar consonants, tonal misusage and code-switching between Chinese and English.

To address these problems, the AI speech recognition algorithm has been optimized for localization adaptation: it can accurately identify and process common pronunciation errors such as retroflex-alveolar consonant confusion and Chinese-English code-switching. The system can distinguish between dialectal pronunciations such as Cantonese and Hokkien and standard Mandarin, and abandons the one-size-fits-all evaluation standard. Instead, it adjusts the recognition and correction strategies according to the actual pronunciation characteristics of each child, ensuring the applicability and effectiveness of the technology for different learners.

4.1.2 Principle of Gamification-Driven Engagement

For children aged 3 to 12, Pinyin learning can be designed in accordance with their cognitive development stages. Pinyin learning usually involves mechanical memorization of letters and repeated pronunciation practice, which is easy to make children feel boring. By transforming these learning activities into interactive games, the fun of learning can be greatlyenhanced.

Gamified learning can provide instant positivefeedback: for example, the game can give immediate verbal or visual rewards when children pronounce Pinyin correctly. Children can earn points and rewards by completing learning tasks, and face progressive level-up challenges where the difficulty increases with the improvement of their learning skills. These designs not only reduce the learning pressure of children but also stimulate their voluntary participation inlearning. This approach is particularly suitable for overseas Chinese children who often lack intrinsic motivation to learn Chinese, as it can effectively arouse their interest in Pinyin learning byturning practice into play.

4.1.3 Language-Culture Integration Principle

Overseas Chinese children's Chinese learning is not only about language acquisition but also cultural cognition, so Pinyin teaching should integrate Chinese cultural elements into the learning design.

Traditional festivals, historical scenes and cultural symbols can be woven into Pinyin learning materials, enabling children to perceive and experience Chinese culture naturally in the process of Pinyin learning. For example, learning the Pinyin of 'chūn jié (Spring Festival)' can be combined with the introduction of relevant customs and stories. This integrated approach helps children build a

gradual connection with Chinese culture and cultivate their cultural identity and sense of belonging.

4.1.4 Family-School Collaboration Principle

Many overseas Chinese parents have limited ability to teach Chinese, and overseas Chinese schools are also faced with a shortage of professional teachers. To address these problems, a collaborative Pinyin learning model of 'AI assistance + parental participation + teacher support' is needed.

In this model, the AI system is responsible for daily pronunciation correction and learning progress tracking, parents can use the intuitive learning reports generated by the AI system to provide targeted guidance for their children, and teachers can optimize classroom teaching according to the overall learning data of students. The three parties form a synergistic learning support system, creating a good learning environment for overseas Chinese children's Pinyin learning.

4.2 Core Functional Module Design of the Model
4.2.1 Real-time Pronunciation Correction Module

When children read Pinyin aloud, the system displays real-time sound waves on the screen, which intuitively shows the differences between their pronunciation and the standard sound. If pronunciation errors are detected, a friendly animated character such as the panda Huahua will provide clear and simple correction tips, such as 'Purse your lips a little more' or 'Make your tone rise gently'.

The system can also identify pronunciation errors caused by Chinese-English code-switching or dialectal influences. For example, it can distinguish the Cantonese pronunciation of 'I' from the standard Mandarin 'wǒ', and provide targeted correction suggestions according to the specific situation of each child.

4.2.2 Gamified Interactive Learning Module

Pronunciation Challenge: Children read Pinyin aloud to the device, and can earn points by completing preset learning tasks. These points can be exchanged for Chinese cultural-themed rewards, such as high-definition maps of the Forbidden City and exquisite Hanfu design images.

Scenario Theater: Children interact with the virtual panda Huahua in different scenarios, such as supermarket shopping, Spring Festival celebrations and visits to the Forbidden City. Correct pronunciation of Pinyin can promote the development of the story and trigger interesting animation effects. If children make repeated pronunciation errors, Huahua will pronounce the corresponding Pinyin slowly to guide them to correct mistakes.

Pinyin Adventure Island: Children practice Pinyin pronunciation through game challenges, and can defeat 'error monsters' by pronouncing correctly. Completing each level can unlock a short Chinese cultural story, such as the myth of the mythical animals on the roof of the Hall of Supreme Harmony. The game integrates cultural knowledge into Pinyin learning in an implicit way.

4.2.3 Cultural Scenario Integration Module

Traditional Festival Scenarios: In the Spring Festival scenario, children learn the Pinyin of 'xīn nián kuài lè (Happy New Year)', and correct pronunciation can unlock the red envelope animation; they can also learn about traditional Spring Festival customs such as couplet pasting and dumpling making. In the Mid-Autumn Festival scenario, children practice the Pinyin of words such as 'yuè guā

ng (moonlight)' and 'bǐng gān (pastry)', and understand traditional customs such as moon-gazing and mooncake eating.

Cultural Symbol Scenarios: In the Forbidden City scenario, children spell the Pinyin of 'gù gōng (Forbidden City)' and 'tài hé diàn (Hall of Supreme Harmony)', and can view high-definition pictures of these buildings and read short related historical stories. In the Dunhuang scenario, children learn the Pinyin of 'dūn huáng (Dunhuang)' and 'bì huà (mural)', and can appreciate exquisite fragments of Dunhuang grotto murals.

4.3 Application Scenario Adaptation of the Model
4.3.1 Family Fragmented Learning Scenario

The learning tool supports mobile terminals such as mobile phones and tablets, and each practice session is designed to last 15 to 20 minutes, which fits the fragmented spare time of overseas children after school or on weekends, enabling them to learn Pinyin anytime and anywhere.

Parents with limited Chinese proficiency can also participate in their children's learning: the family-school interaction section of the app provides intuitive learning progress reports, making it easy for parents to carry out targeted guidance and accompany learning. The core goal of this scenario is to make full use of fragmented time and lower the threshold of parental participation in children's Chinese learning.

4.3.2 Overseas Chinese School Classroom Scenario

Teachers can use this tool as an auxiliary teaching resource: they can assign preview tasks before class and review tasks after class to consolidate students' learning results. During class, the AI's real-time pronunciation feedback function can be used to design interactive activities such as pronunciation contests and group challenges, making classroom teaching livelier and engaging.

The system can generate comprehensive learning reports for teachers, which not only reflect the common difficult points of the whole class in Pinyin learning but also record the individual learning progress of each student. Teachers can adjust their teaching plans according to these reports, and focus on teaching the common weak points of students, which effectively makes up for the limitation of collective teaching that cannot take into account individual differences.

4.3.3 Online Autonomous Learning Scenario

For overseas Chinese communities in remote areas or with a severe shortage of Chinese teachers, the model provides a complete online Pinyin learning course, guided by an AI virtual assistant. The course integrates rich cultural scenarios and gamified learning activities, creating an immersive Chinese language learning environment.

This scenario effectively solves the problems of insufficient Chinese language exposure and lack of teaching resources for overseas Chinese children who cannot attend offline Chinese schools, enabling them to carry out systematic Pinyin learning in a structured way.

4.4 Innovation Points of the Model

First, the AI algorithm has been optimized for the specific pronunciation characteristics of overseas Chinese children: aiming at the common problems of Chinese-English phonetic mixing and dialectal influenced pronunciation, the system can accurately identify these unique pronunciation errors and provide targeted correction strategies. For example, it can distinguish the Cantonese

pronunciation 'ngo' of 'I' from the standard Mandarin 'wǒ', and provide personalized correction suggestions according to the actual situation of each child.

Second, the learning content is co-created with professional cultural institutions, and the cartoon images and cultural stories in the course are all based on authentic historical facts and cultural connotations. For example, in the 'Forbidden City Adventure' game level, children can unlock the cultural stories of the roof animals of the Hall of Supreme Harmony by pronouncing Pinyin correctly, ensuring the accuracy and authenticity of the cultural knowledge transmitted in the learning process.

5. Project Analysis

5.1 Project Introduction

Target Users: The project is precisely oriented to overseas Chinese children, directly addressing their prominent Pinyin learning pain points such as Chinese-English phonetic mixing, tonal misusage and insufficient cultural knowledge.

Product Features: The product is equipped with functional modules such as 'Pronunciation Challenge' with real-time AI pronunciation correction, 'Pinyin Adventure Island' integrating cultural quest games, and 'Family Tasks' promoting parent-child co-learning. The parent dashboard provides visual learning portfolios, enabling parents to grasp their children's learning progress intuitively.

5.2 Project Implementation Status

Data Sources: The project's effect evaluation is based on questionnaire surveys, including parent version and student version, to collect comprehensive feedback on the product's usage experience and learning effect.

User Testimonials:

User 1: Since my child used the Huayu Kids App, his interest in Chinese learning has increased significantly. The interactive reading function is like a patient little teacher, providing timely feedback on pronunciation errors. We have seen obvious learning progress: my child was reluctant to learn Chinese before, but now he takes the initiative to open the app for learning every day, which is a pleasant change.

User 2: I am very satisfied with the Huayu Kids App. It is equipped with rich and systematic Chinese learning materials, from basic Pinyin phonetics to advanced syllable blending, which fully conforms to the cognitive development law of children. All the audio materials use clear and standard children's voices: my child imitates and repeats the pronunciation, and his Pinyin pronunciation has been continuously improved. The app has played an important role in the initial stage of his Chinese learning.

User 3: The personalized learning function of the Huayu Kids App is excellent. It can push targeted practice tasks according to my child's specific pronunciation weak points, which is like a customized learning plan for him, and the learning effect is very obvious.

5.3 Issues and Improvement Measures of the Project

5.3.1 Existing Issues

Despite the advanced technology, the project faces many practical difficulties in the implementation process: the operation process of the product is relatively complex, and the development of multi-language versions such as English and Spanish further increases the technical difficulty; in addition, the cultural content in the app needs to be further optimized to adapt to the

cultural sensitivities of different regions.

Speech recognition for children is another prominent challenge: young children's pronunciation is inherently unclear due to their immature language development, and Chinese-English code-switching makes it more difficult for the system to recognize pronunciation accurately, which reduces the recognition accuracy of the technology.

In the global promotion process of the product, some countries hold a cautious attitude towards Chinese cultural content. Due to differences in regional culture, values and political views, some cultural content that is common in China may arouse discussions in other regions, which increases the difficulty of the international promotion of Chinese cultural products and services.

5.3.2 Improvement Measures

To solve the problem of low recognition accuracy of children's unclear pronunciation, the project will collect a large number of Pinyin pronunciation samples of overseas Chinese children of different ages and dialect backgrounds, and use large-scale data analysis to train the speech recognition model, so as to optimize its ability to identify children's specific pronunciation characteristics and improve the recognition accuracy. In addition, a 'User Correction' function will be added: if parents find that the AI system misrecognizes their children's pronunciation, they can manually input the correct content, and the corrected data will be used to further optimize the model.

To address the regional cultural sensitivity of cultural content, a 'Regional Cultural Adaptation' resource library will be built: for European and North American regions, the project will focus on Chinese cultural elements with universal appeal, such as the Spring Festival which embodies the family value and Kung Fu which is popular worldwide; for Southeast Asian regions, cultural themes with deep historical and cultural connections such as Zheng He's voyages to the Western Seas and Mazu culture will be integrated. In addition, local Chinese cultural experts will be hired as content reviewers: they are familiar with the local cultural and political context, which can provide professional guidance for content optimization, avoid sensitive topics and promote the smooth local promotion of the product.

6. Conclusion and Future Outlook
6.1 Conclusion

Overseas Chinese children face prominent difficulties in Pinyin learning: most of them grow up in Chinese-English bilingual environments and are influenced by regional Chinese dialects, leading to common pronunciation problems such as retroflex-alveolar consonant confusion, tonal misusage and Chinese-English code-switching. In addition, these children often lack intrinsic learning motivation, sufficient guided practice and adequate exposure to Chinese culture, which further restricts their Pinyin learning effect.

AI speech recognition technology provides an effective solution to these problems: its real-time pronunciation correction, personalized learning record and targeted feedback functions are highly compatible with the specific learning needs of overseas Chinese children, which cannot be achieved by traditional teaching methods.

This study constructs a flexible Pinyin learning model adapted to the learning characteristics of overseas Chinese children, with the core design concepts of gamified learning, language-culture integration and family-school collaboration. The learning process of the model follows the four-step cycle of 'diagnosis-learning-practice-feedback', forming a closed loop of effective learning. The model

has four core characteristics: real-time pronunciation correction, gamified interactive learning, integrated cultural content and family-school interaction tools.

This model not only helps overseas Chinese children master Pinyin knowledge effectively but also stimulates their interest in Chinese learning, and gradually strengthens their emotional connection with Chinese culture in the learning process. The model is applicable to multiple learning scenarios, including family fragmented learning, overseas Chinese school classroom teaching and online autonomous learning, with strong practicality and operability.

The practical application of the model has verified its effectiveness in overseas Chinese children's Pinyin learning: the optimized AI system can accurately adapt to the pronunciation characteristics of overseas Chinese children, gamified design enhances learning interest, integrated cultural content cultivates cultural identity, and the family-school collaboration mechanism provides a comprehensive learning support system. This model effectively solves the key problems in overseas Chinese children's Pinyin learning, provides a feasible path for the digital and personalized development of overseas Chinese education, and also contributes to the intergenerational inheritance of Chinese culture.

6.2 Future Outlook

The model will be further optimized based on more real user data: the project will invite more overseas Chinese children of different age groups and dialect backgrounds to participate in the trial, track and analyze their long-term learning effects, and adjust the personalized learning path and product functions of the model according to the learning data, so as to improve its adaptability to individual learners. In addition, the project will enrich the learning content of the model, and add the connection learning between Pinyin and Chinese characters as well as oral expression, helping children realize the progressive development from Pinyin learning to comprehensive Chinese ability improvement.

The technical support of the model will be continuously upgraded to improve the intelligence level of AI speech recognition: the project will continue to collect Pinyin pronunciation samples of overseas Chinese children, optimize the model's ability to identify unclear pronunciation and strong dialectal influenced pronunciation of young learners. An AI virtual teacher will be added to the system, which can interact with students in real time, provide targeted guidance for pronunciation learning and explain the connotation of cultural knowledge, creating a more natural and humanized learning experience.

The project will further expand the promotion and application scope of the model, and integrate it into more overseas Chinese teaching scenarios: it will carry out in-depth cooperation with overseas Chinese schools, Chinese communities and cultural institutions to enable children in more countries and regions to use this learning tool. The cultural content of the model will be further optimized to integrate more local cultural themes, making the learning content closer to the daily life of overseas Chinese children. In addition to the learning product itself, the project will also provide supporting services such as teacher training and parent guidance manuals, building a comprehensive learning support system for overseas Chinese children and promoting the high-quality development of overseas Chinese education worldwide.

The key cultivation project of the 15th 'Challenge Cup' Undergraduate Entrepreneurship Plan Competition of Quzhou University, 'Chinese Children's Voice - AI Pinyin Learning System for Overseas Chinese Children Enabled by edge computing';

Quzhou University 2025 National College Student Innovation and Entrepreneurship Training Program Project 'International Chinese Education Data Platform: Intelligent Diagnosis and Prediction Model for Multi module Learning Behavior' (Project No. 202511488065);

2025 Zhejiang Province University Student Science and Technology Innovation Activity Plan (New Talent Program) 'Pen Falling Star River: Intelligent Reading of Primary and Secondary School Composition Research on the Design and Practice of Writing Enhancement Platform' (Project No. 2025R450A013);

2025 Innovation and Entrepreneurship Training Program for College Students of Quzhou University 'Language Connected World - HSK Chinese Language Learning for Global International Students Navigation system' (Project No. Q25X043).

2024 Quzhou University the construction project of the 'Mentor+Project+Team' Innovation and Entrepreneurship Workshop, 'Smart Book Boat', Project No. SH20241101.

References

[1] Li, X. L. (2022). A study on primary Chinese Pinyin and characters for overseas Chinese children (Master's thesis, Sichuan Normal University). https://doi.org/10.27347/d.cnki.gssdu.2022.000307

[2] Liu, H. L. (2025). Application scenarios and teaching model innovation of AI speech recognition technology in junior high school English listening and speaking teaching. Wisdom Education, 2(9), 40. https://www.juzhikan.cn/static/upload/file/20250714/1752482056919007.pdf

[3] Zhang, B. (2023). Application of AI technology in the production of educational resources for primary and secondary schools. Information Technology Education in Primary and Secondary Schools, (5), 87-90. http://dianda.cqvip.com/Qikan/Article/Detail?id=7109651237

[4] Bai, J. (2020). Application and thinking of speech recognition technology in the field of education and teaching. Computer Era, (07), 105-107. https://doi.org/10.16644/j.cnki.cn33-1094/tp.2020.07.029

[5] Yu, W.,&Jia, D. Y. (2024). Analysis of Mandarin pronunciation problems and ability enhancement for teachers and students in rural schools based on an AI technology model. Application of Language and Characters, 2(35), 92-94. https://kns.cnki.net/kns/brief/result.aspx?dbprefix=CJFQ&kw=

[6] Tao, L. Y. (2024). Research on the transfer effect of English on German phonological learning under the theory of trilingual acquisition. Modern Linguistics, 12(5), 642-649. https://doi.org/10.12677/ml.2024.125409