Malaysian Journal of Social Sciences and Humanities (MJSSH) (e-ISSN : 2504-8562)
2026, Volume 11, Issue 3, e003729
DOI: https://doi.org/10.47405/mjssh.v11i3.3729

REVIEW ARTICLE

Content Knowledge Base for AI-based Chinese Computer-assisted Pronunciation Training: An Analysis from Chinese Pronunciation Research

Dongmei Jiang[1], Goh Chin Shuang[2], Airil Haimi Mohd Adnan[3*]

[1]Academy of Language Studies, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia;
School of International Education, Hainan Medical University, 571199, Haikou, Hainan, China.
Email: jiangdongmei@muhn.edu.cn
[2]Academy of Language Studies, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
Email: gohch269@uitm.edu.my
[3]Academy of Language Studies, Universiti Teknologi MARA (UiTM), 40450 Shah Alam, Selangor, Malaysia
Email: airil384@uitm.edu.my

CORRESPONDING AUTHOR (*):
Airil Haimi Mohd Adnan (airil384@uitm.edu.my)

KEYWORDS:
Knowledge base of Chinese CAPT
Knowledge base of DDR
Chinese pronunciation
Chinese as a second/foreign language (CSL/CFL)
Guidelines

CITATION:
Jiang, D. M., Goh, C. S., & Airil Haimi, M. A. (2026). Content Knowledge Base for AI-based Chinese Computer-assisted Pronunciation Training: An Analysis from Chinese Pronunciation Research. *Malaysian Journal of Social Sciences and Humanities (MJSSH), 11*(3), e003729. https://doi.org/10.47405/mjssh.v11i3.3729

ABSTRACT

Integrating the state-of-the-art Automatic Technology (AI) technology into the Computer-Assisted Pronunciation Training (CAPT) system to assist language acquisition could address the various drawbacks of conventional pronunciation and enhance its effectiveness. However, little attention has been given to their application in Chinese language acquisition, and the lack of guidelines grounded in Second Language Acquisition (SLA) theories and pronunciation pedagogy has led to the tension between technology and pedagogy, limited the understanding of their application beyond English, and hindered their broader application. Grounded in the knowledge framework of AI-based Chinese CAPT system based on the Richey and Klein's (2007) Design and Development Research (DDR) framework, the Monitor Theory and the Speech Acquisition Model (SAM), this study aimed to establish the content knowledge base informed by Chinese pronunciation pedagogy and derive evidence-based, pedagogically sound design guidelines for the AI-based Chinese CAPT system for non-native Chinese learners. Alongside the guidelines previously derived from the SLA Interactive theories and Oral Corrective Feedback (OCF) research in the Chinese setting, these could inform future design and development of the system. Employing the qualitative method of thematic analysis, three key themes were identified from a literature review on Chinese pronunciation acquisition, leading to the formulation of pedagogical design guidelines. These findings offer valuable insights into Chinese pronunciation practices and learners' needs, providing a robust theoretical and empirical foundation for future design and development research. Finally, the study discussed its limitations and

contributions.

Contribution/Originality: The contributions of this study are threefold: it extends the scholarship on CAPT system design to the Chinese linguistic context, establishes a content knowledge base for future reference, and proposes a paradigm to address the technology-pedagogy tension inherent in integrating advanced AI into language learning.

1. Introduction

Despite the rapid increase of Chinese learners worldwide witnessed in recent decades, Chinese learners' speaking ability was far behind their reading and writing skills based on the data of HSK/HSKK results (Ding et al. 2021). Promoting Chinese speech ability through pronunciation has become a new, promising breakthrough and urgent issue (Liu et al., 2021). As a recent technological advancement, AI has been embraced by language educators to improve language learning and has demonstrated significant potential to improvement in this area (Wang et al., 2024). Integration of AI application into conventional language learning classroom could provide promising solutions to the problems encountered by the conventional instruction, such as insufficient chances for practice (Hsiao et al., 2015), lack of personalized feedback to students (Chen et al., 2022), and the feeling of anxiety when speaking a new language in front of others (Leeuwestein et al., 2021). However, its application in designing language education across diverse learning environments has been relatively uncommon (Pikhart, 2020). They need to be grounded in pedagogical and learning science theory (Luckin & Cukurova, 2019; Zawacki-Richter et al., 2019), as well as sociocultural and ethical factors (Lee et. al., 2023), and must serve the specific goals and contexts of their intended education use (Zhai et al., 2021).

The CAPT systems powered by AI have increasingly seen as a valuable method for improving the pronunciation of second language (L2) learners (Amrate & Tsai, 2024 ; Cengiz, 2023 ;) and shown favorable learning outcomes for language learners (Cengiz, 2023). However, two significant limitations characterize the current research. First, few studies are theoretically grounded (Cengiz, 2023), or explore languages other than English (Amrate & Tsai, 2024; Cengiz, 2023), especially Chinese. Second, a persistent gap exists between how CAPT systems are engineered and the pedagogical approaches proven effective for teaching pronunciation (Amrate & Tsai, 2024; Levis, 2007; Neri et al., 2002). This disconnect has fostered skepticism about CAPT's true training value and its potential for widespread adoption.

1.1. Research Objectives

This study, grounded in the knowledge framework of AI-based Chinese CAPT system previously constructed on the Richey and Klein's (2007) DDR framework by the authors, aimed to establish the content knowledge base informed by Chinese pronunciation pedagogy research, the Monitor Theory and SAM, then derive evidence-based, pedagogically sound design guidelines for the future AI-based Chinese CAPT system for non-native Chinese learners. The study will contribute significantly to the field by offering in-depth insights into Chinese pronunciation, spreading the research stream of CAPT into the Chinese setting, and enriching the theoretical and practice framework of

CAPT design-based research. Ultimately, it will offers an alternative solution for the application of AI in language learning studies.

To meet the research objective, the study was guided by the following research questions:

RQ1: What are the major research streams in the field of Chinese pronunciation training?

RQ2: What insights could be drawn from those empirical studies?

RQ3: What principles for the design and development of an AI-based CAPT system could be derived from the answers to the above research questions?

To address RQ1, some definitions were explained, providing the necessary knowledge to better understand this study. Then, a literature review on the Chinese pronunciation research was conducted based on the knowledge base of the AI-based Chinese CAPT system, which was previously constructed by the author (Jiang et al., 2025) on the grounds of the DDR framework by Richey and Klein (2007), laying the theoretical foundation of this study.

2. Literature

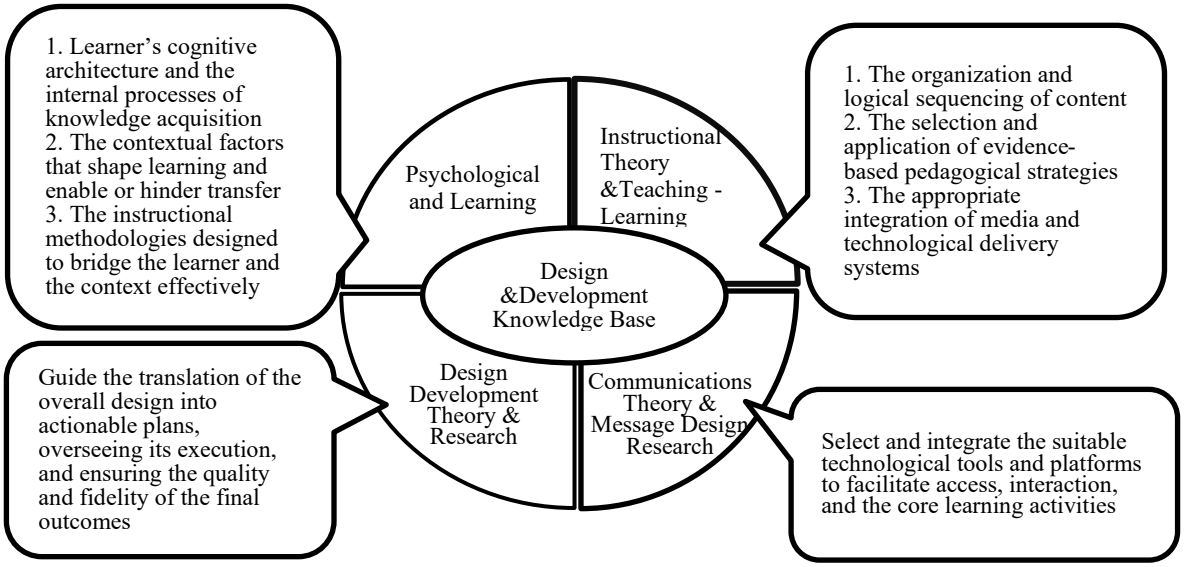
2.1. DDR and Its Knowledge Base

DDR, design and development research, refers to "the systematic study of design, development, and evaluation processes with the aim of establishing an empirical basis for the creation of instructional and non-instructional products and tools and new or enhanced models that govern their development" (Richey & Klein, 2007). Systematic and empirical in approach, this research addresses real-world problems and builds upon prior work. Its scope can include the study of products and tools, development research, and the testing of theoretical models. As outlined by Richey & Klein (2007), the knowledge base of DDR is both broad and interdisciplinary, being fundamentally shaped by both cross-disciplinary foundations and domain-specific inquiry. As shown in Figure 1, the integrated base is typically applied through at least e four key areas of research and theoretical areas: psychological learning research, instructional theory, communication theory, and the research on design and development.

To better understand how these knowledge bases could support and inform the DDR in this study, an analysis of the interrelationships among these components was conducted. These four components can be grouped into expertise knowledge base, technology knowledge base and procedure knowledge base. The "expertise knowledge base" compasses the first two components, focusing on the specific knowledge and practice required to design and develop a product, ensuring a strong theoretical and empirical foundation within the field where the tool will be used. The "technology knowledge base" refers to the third component of technology which can be utilized to facilitate the pedagogical purpose and promote the pedagogical effect. The "procedure knowledge base" is the fourth component concerning the processes related to product design and development, focusing the management of design procedures to ensure rigor and high quality. While the first two knowledge bases are relatively independent and may overlap, the fourth component integrates and supervises the entire design and development process, incorporating multiple interactive cycles and ongoing evaluations. Therefore,

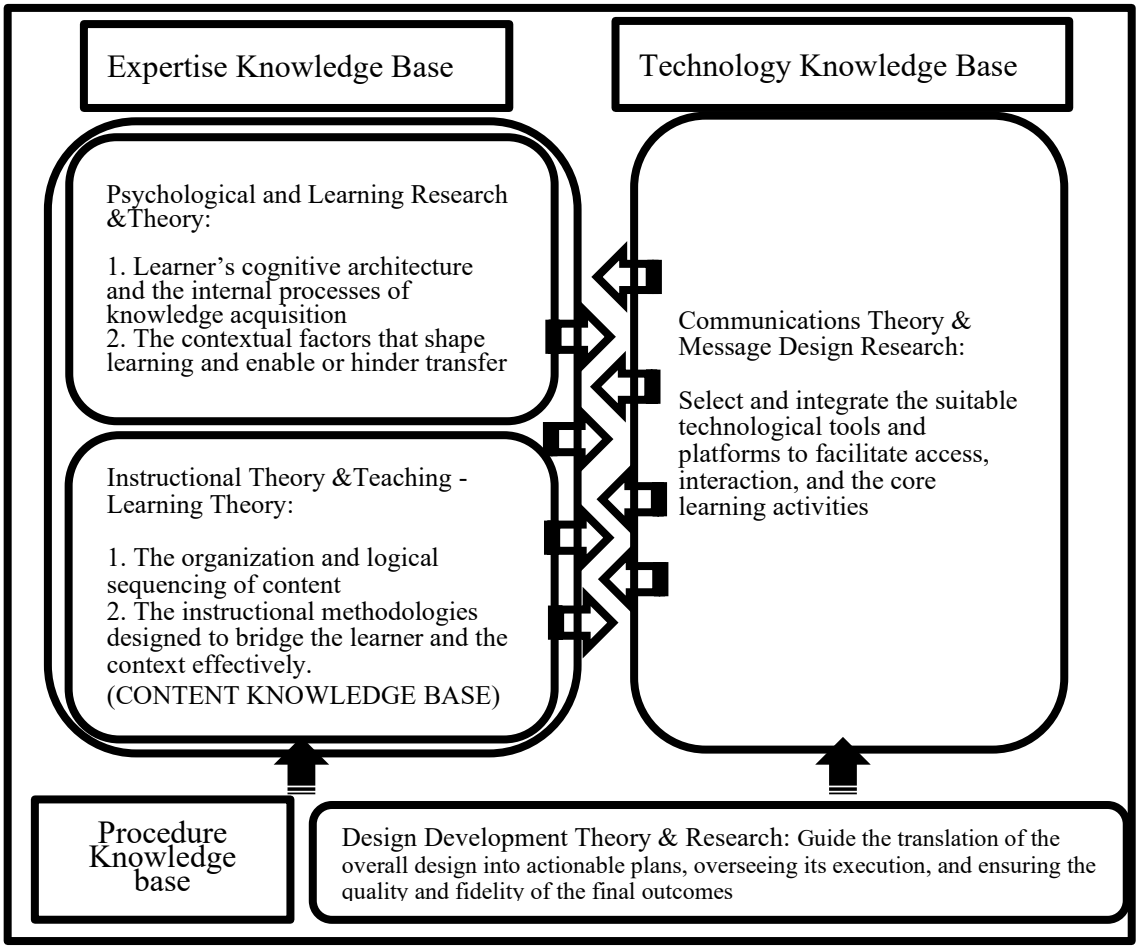
the Richey and Klein's (2007) framework of the DDR knowledge base has been modified as Figure 2.

Figure 1: The Framework of DDR Knowledge Base by Richey and Klein (2007)



Source: Adapted from Richey and Klein (2007)

Figure 2: The Modified DDR Knowledge Base Framework

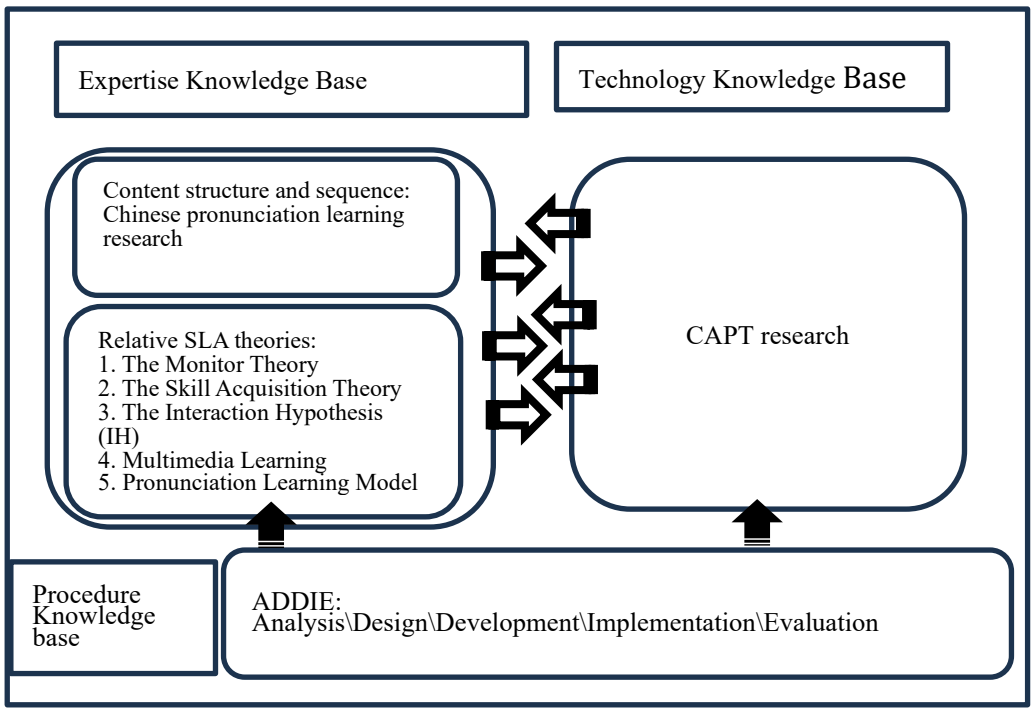


Source: Adapted from Richey and Klein (2007)

2.2. The Knowledge Framework of Developed AI-based Chinese CAPT

Based on the above modified knowledge framework of DDR, the knowledge base of the developed AI-based Chinese CAPT system were constructed, as shown in Figure 3, which serves as the foundation of the literature review of Chinese pronunciation research in this study.

Figure 3 : The Knowledge Framework of the Developed AI-based Chinese CAPT



Source: Created by the researcher

2.3. Pronunciation Assessment Criteria

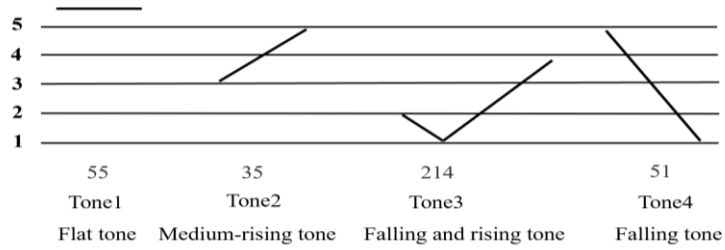
Based on the communicative goals, the contemporary pronunciation teaching approach prioritized intelligibility and comprehensibility (Munro & Derwing, 1995), indicating the listener's understanding of learners' utterances and meaning (Kang et al., 2018). While phonetic features are generally acknowledged to enhance intelligibility and comprehensibility and have thus gained scholarly attention, research specifically links supra-segmental features to perceived comprehensibility (Saito et al., 2016). Consequently, a comprehensive approach to pronunciation instruction must address both segmental and supra-segmental elements (Kang, 2018; Saito et al., 2016). As a tonal language, Mandarin Chinese treats tone as a fundamental supra-segmental feature, the accurate production of which is critical for intelligible speech (Li & Lin, 2018).

2.4. Chinese Pronunciation

As a foundational component in the acquiring Chinese, pronunciation is typically the first skill learners engage with. It should be developed progressively across elementary, intermediate and advanced levels, beginning with Hanyu Pinyin, an alphabetic system that main includes 28 initials, 48 finals, and 4 basic tones. Initials refer to the beginning letter of a pinyin, which are usually consonants, while finals refer to the letters of a pinyin after initials, mostly vowels. Tones refer to variations in pitch produced by the

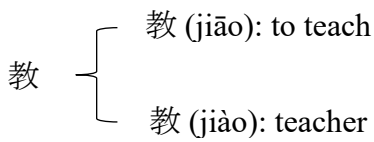
vibration frequency of the vocal cords when pronouncing a Chinese syllable. There are four basic tones: First Tone, Second Tone, Third Tone, and Fourth Tone (Huang, 2021), which can be represented by the numerical values 55, 35, 214, and 51 based on a 5-scale developed by Zhao Yuanren, as shown in Figure 4, and are characterized by level, rising, rising-then-falling and falling pitch contours, respectively (Huang, 2021). Chinese tone can alter the meaning of the character (Chen, 2020). A character with the same pinyin but different tones mean different. For example, the character 教 (jiào) can be read as the first tone (jiāo), meaning "to teach", and the fourth tone (jiào), meaning "teacher", shown in Figure 5.

Figure 4: Four Basic Tones in Mandarin Chinese Based on Five-level Contour Tone Marks by Zhao Yuanren



Source: Created by the researcher. Adapted from Huang, B., & Liao, X. (2021). *Modern Chinese* (6th ed.). Higher Education Press.

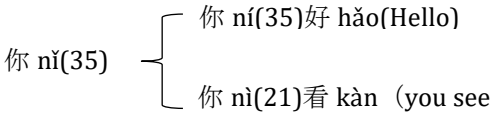
Figure 5: The Chinese character "教" with different



Source: Created by the research

The tone of a Chinese character may change when in the flow of a sentence due to the influence of the tones of neighboring characters, which is called sandhi. For example, the third tone of a character, a falling and rising tone with a value of 214, will change to a rising tone valued as 35 when followed by another third tone character, or a falling tone valued as 21 when followed by another character with tones other than the third tone, as illustrated in Figure 6.

Figure 6: Sindhi of the Third Tone of the Chinese



Source: Created by the researcher

Within the domain of Chinese language acquisition, tone is widely regarded as one of the most particularly challenging elements for learners to master (Chen, 2020; Shen, 1989; Song, 2009; Wang, 2016; Yang, 2021; Yi, 2021). Many learners still struggle with serious pronunciation issues that hinder communication (Chen, 2020; Deng & Lin, 2017; Ding et al., 2021; Song, 2009) after years of Chinese learning. According to a preliminary and

systematic investigation of the characteristics and patterns of 25 L1 learners, Ran (2019) presented that the proportion of tone errors is much higher than that of initials and finals among 528 error items of typical Chinese speech errors. The average numbers of participants who committed tone, initial and final errors were 27.2, 7.23 and 5.83, respectively. This indicated that some learners committed errors with initial and final, while almost all learners made tone errors. Tone errors were unlikely to be naturally corrected or acquired as learners do with other elements (Chen, 2020; Chen & Wang, 2008), such as vocabulary and grammar, especially after the critical period. If learners lack a solid phonetic foundation and do not pay attention to correction and practice, these errors are likely to persist (Chen, 2020). As a result, helping learners acquire Chinese tones is a critical issue in pronunciation acquisition (Lin, 2016).

Due to the unique role of tone in Chinese language, its high likelihood of making errors, and the hot discussion sparked among scholars, its pedagogical practice research could imply some common rules of Chinese language pronunciation training and provide deep insights to inform the way of integrating AI technology into Chinese pronunciation training and the DDR of AI-based Chinese CAPT systems.

3. Research Methods

A qualitatively systematic and thematic analysis method was adopted. Firstly, a thorough search was performed across China National Knowledge Infrastructure (CNKI) and Chinese Social Sciences Citation Index (SSCI) databases. Secondly, a primary screen was conducted based on inclusion criteria through abstracts and titles, deleting the duplicate studies. Thirdly, the selected studies were read throughout and carefully. Fourthly, identify the main themes through comparison, contrast and analysis. Lastly, synthesized and presented the result. The inclusion criteria are that the study is about the Chinese tone acquisition in non-native Chinese learners indexed by CNKI and SSCI. The thematic analysis is based on the six steps: be acquaint with data, develop initial codes, search for themes, review themes, define and label themes, create report (Braun & Clarke, 2006).

4. Results

Through scrutinizing, analysis and synthesis, three themes were finally identified. They are (1) types of learners' tone errors, (2) negative transfer from the native language, and (3) the effectiveness of perceptions towards tone acquisition. Three sub-themes were identified about the last theme: (1) the establishment of learners' sense of tone awareness, (2) the correlation between perception and production, and (3) facilitating output through perception. Study methods involved auditory listening, manual recording, contrastive analysis, and experiments. Alongside the development of the cutting-edge AI technology, Automatic Speech Recognition (ASR) could extract acoustic information and present clear and predefined units (hertz, milliseconds) of the given speech, allowing an objective comparison between the target speech and learners' speech. It has been employed in Chinese pronunciation teaching and research to identify learners' tone errors and test the effectiveness of facilitating pronunciation acquisition through enhancing perception.

4.1. Types of Tone Errors

The identification of tone errors has been one of the main topics in the field of Chinese pronunciation learning. It enables us to understand how the learners' inter-language of Chinese tone developed during the process of tone acquisition (Ran, 2019). According to Krahsen's Natural Order Hypothesis of the Monitor Theory, learners acquire the specific forms of language in predictable sequences, guided by their innate language acquisition faculty. This process occurs independently of instructional sequences and the complexity of acquired structures. Therefore, predicting the difficulty order of acquisition based on tone error identification has been one of the main research streams since the early 1980s. The higher error rates indicate the higher difficulty level in acquisition.

Tone errors were categorized into two groups according to the learners' native languages: learners speaking non-tonal native languages and the other speaking tonal native languages.

4.1.1. Tone errors committed by learners speaking non-tonal native languages

The learners speaking non-tonal native languages involved those from America, Japan, Korea, and Russia etc. Both Shen (1989), Gui (2000) and Wang (2016) examined tone errors committed by American learners. Shen (1989) analyzed learners' tone errors in a specific passage, while Gui (2000) studied tone errors in disyllabic and trisyllabic words and phrases. Besides American learners, Wang (2016) also studied tone errors committed by Japanese and Korean learners. Based on the error rates, Shen (1989) and Ran and Yu (2019) predicted different orders of tone difficulty from the most challenging to the least, as shown in Table 1.

Table 1 : Predicted Different Difficulty Order of Chinese Basic Tones

Nationality of learners	Difficult order of Chinese basic tones	Researchers
American learners	First Tone>Fourth Tone> Second Tone >Third Tone	Shen Xiaonan (1989)
25 countries learners	First Tone>Third Tone> Second Tone >Fourth Tone	Ran Qibin and Yu Shuang (2019)

Source: Created by the researcher

ASR technology was employed to assist in identifying learners' tone errors since the 2010s. Guo and Xuan (2013) and Gao and Li (2018) utilized software of Praat and Mini-Speech Lab to investigate the tone errors committed by American beginners and Russian two-year learners, respectively, through measuring and comparing the differences of native Chinese speakers' and learners' pronunciation spectrum. Gao and Li (2018) found that the duration order of Chinese tones was stable, with the order from the longest to the shortest being Third Tone>First Tone>Second Tone >Fourth Tone. Whereas Russian learners exhibited shorter and more unstable tone duration. Their tone ranges were comparatively narrower, with starting points generally lower than those of native Chinese speakers. Regarding tone patterns, the Second Tone was most likely mispronounced and confused with the Third Tone.

4.1.2. Tone errors committed by learners speaking tonal native languages

The learners speaking tonal-language native languages were mainly Vietnamese and Thai. Wu and Hu (2004) found that Vietnamese learners made Fourth Tone errors the most (62%), followed by the Third Tone (39%), the First Tone (21%), and the Second Tone (19%). Li (1995) and Xu (2009) studied Thai learners' tone errors. Li (1995) noted that Thai learners made more errors with the Chinese First Tone and Fourth Tone, which is similar to the reports of American Chinese learners by Shen (1989). The First Tone produced by Thai learners is typically lower than that of native Chinese speakers, while the Fourth Tone produced by Thai learners is often longer than that of native speakers.

4.2. Negative transfer from the native language

According to behavioral structuralism, the First language (L1) that a learner has acquired before learning L2 is one main factor influencing the internalization of the new L2 behavior, which is referred to as transfer. Transfer can be beneficial or obstructive to SLA, depending on how different the construction of L2 is from that of L1. The more different a particular L2 pattern is from the learner's L1, the more challenging it will be to acquire, and the more likely errors are to occur. The contrastive analysis approach was the most frequently used approach to explain why errors occurred through comparing the differences between two languages. In acquiring Chinese tones, learners tend to rely on familiar linguistic patterns: they often use native-language tone as a reference if they speak a tonal language, or native intonation if they speak a non-tonal language.

According to Wu and Hu (2004), Vietnamese Chinese learners usually exhibited the highest error rates with the Chinese Fourth Tone, because Vietnamese lacks a falling tone similar to the Chinese Fourth Tone. And due to the overall tone range of Vietnamese being narrower and not including a tone with the value of 55 of the Chinese First Tone, they usually uttered the Chinese First tone at a value of 44. Li (1995) also posited that Thai Chinese learners often committed the error of uttering the Chinese First Tone at a lower starting point than that of native Chinese speakers, as the starting point of the first tone in the Thai language is lower than that of the Chinese First Tone.

Whereas non-tonal native languages learners, such as English and Russian, were often influenced by the intonation of their native language. Gui (2000) reported that American Chinese learners tended to produce Chinese tones using the corresponding intonation patterns of American English. Similarly, Russian learners also attempted to identify similar phonetic elements in Russian to aid their Chinese tones learning. They often struggled with the Chinese Third Tone, and the duration of tones they pronounce is typically shorter than those pronounced by Chinese native speakers. This discrepancy is attributed to the negative transfer from Russian intonation, which is characterized by short, fast, and light (Gao & Li, 2018).

4.3. The Effectiveness of Auditory Perception towards Tone Acquisition

Besides the negative transfer from L1, more cognitive and psychological factors, such as motivation, the amount of practice a learner engages in, and the potential impact of missing the CAH for language acquisition (Gui, 2000), particularly perception, were also found to affect tone acquisition. From a psychological and cognitive perspective, the

acquisition of pronunciation is a sound-centered process within the vocal-auditory modality (Anderson et al., 2022). This process involves a series of interconnected mental activities, encompassing the transmission and processing of language information, as well as its encoding, storage, and retrieval in memory. Effective communication helps to bridge any confusion or gaps between the learner and the speaker, linking language forms to meaningful messages.

Researchers have established a consensus, both theoretically and empirically, on the interconnection between perception and production (Saito & Plonske, 2019). Nagle (2018) argued that the complementary roles of perception and production were essential for effective L2 speech acquisition. Phonetic perception was the precursor to pronunciation production (Fledge, 1999, Speech Learning Model, Lee et al., 2020; Nagle, 2021), and has been proved to positively facilitate production (e.g. Nagle, 2021; O'Brien et al., 2018), helping learners recognize the exact vocal forms of a language. Failure to perceive and understand each tone's distinctive features can lead to wrong utterances.

Concerning perception and production, researchers mainly studied: (1) the establishment of learners' sense of tone awareness, (2) the correlation between perception and production, and (3) facilitating output through perception.

4.3.1. The establishment of learners' sense of tone awareness

According to linguistics research findings, even though various people may have varied absolute tone levels, they all share the same comparative pitch and pattern, which are distinguishing characteristics, which is the most fundamental element of phonological structure that distinguishes on sound from another within a language. Each Chinese basic tone possesses a unique characteristic: the First Tone is high and level, the Second Tone is a rising tone commencing at the mid-level, the Third Tone is a falling-rising tone, and the Fourth Tone is a falling tone.

In the pedagogical practice, phonetic awareness is mainly developed through structured speech activities, where the specified pronunciation features are elicited and practiced through a variety of exercises, such as phonetic transcription, listen-and-repeat, or read-aloud (Amrate & Tsai, 2024). This approach is considered especially effective for focusing on discrete phonetic features, like vowels, consonants, stress, and intonation (Immonen et al., 2022). Tone awareness refers to an individual's ability to respond to and control tonal features during the processes of tone production and perception. It reflects not only one's metalinguistic competence but also one's linguistic ability (Chen, 2008). The accuracy of phonetic perception is a crucial factor influencing the accuracy of speech production (Chen, 2020; Deng & Lin, 2017). When the accuracy rate of perception reaches 80%, learners are considered to have established a sense of Chinese phonetic awareness (Yi, 2021). This implies that they can distinguish tones even when variations occur due to the influence of adjacent tones in conversational flow.

Yi (2021) argued that, after one year of learning, Malaysian Chinese learners could establish a sense of tone awareness for single characters. Both their accuracies of tone recognition with true and false characters were more than 80%. However, their sense of tone awareness for dual characters has not been built, as only the tone perception of true dual words met the acquisition standard. Wang (2016) also reported that students from the United States, Japan, and Korea committed fewer tone perception errors with single characters than with dual words.

4.3.2. Perception and production development

Some researchers examined the relationship between perception and production by contrasting their accuracy rates through experiments. Chen (2020) investigated the error rates of the perception and production among learners after receiving three months of instruction, while Huang (2022) did this among the intermediate-level Chinese learners.

Regarding the difficulty order of perception and production, Chen (2020) and Huang (2022) reported quite different results. Chen (2020) found the most difficult tones for perception and production were Tone two and Tone Three, respectively (shown in Table 2). Huang Bin (2022) presented that Tone Third was the most difficult one both for perception and production. Table 3 presents his findings both from the most difficult to the least.

Table 2: Tone Difficult Order for Perception and Production Identified by Chen Hong (2020)

	Difficult order
Perception	Second tone (74.2%)>First tone (84.19%)>Fourth tone (89.49%)>Third tone (90.89%),
Production	Third tone (51.7%)>Second tone (56.64%)>Fourth tone (76.17%)>First tone (90.49%)

Source: Created by the researcher

Table 3 : Tone Difficult Order for Perception and Production Identified by Huang Bin (2022)

	Difficulty order
Perception	Third tone>Second tone>First tone>Fourth to
Production	Third tone>First tone>Second tone>Fourth tone

Source: created by the researcher

Despite the mixed picture for the difficulty orders of perception and production, both Chen (2020) and Huang (2022) found an overall higher accuracy of tone perception than that of production. In addition, Huang (2022) noticed the difference between the error rates of perception and production of single characters and dual words. As to perception, the error rates of Second Tone and Third Tone were higher than the other two tones. The error rates of single characters were 19.62% and 5.38%, respectively, whereas those of dual words were 22.36% and 27.28%, respectively. When comes to production, the error rates of Third Tone and First Tone exhibited higher error rates compared to the other two tones. The error rates of single characters were 35% and 3.37%, respectively, and those of dual words were 24.52% and 16.38%. Table 4 presents the findings of Huang Bin (2022).

Table 4 : Tone Difficult Order for Perception and Production Identified by Huang Bin (2022)

	Single-character word or Dual-character word	**The 1st Tone**	**The 2nd Tone**	**The 3rd Tone**	**The 4th Tone**

Perception of the tone	single	*	19.62	5.38	*
	dual	*	22.36	27.28	*
Production of the tone	single	3.37	*	35	*
	dual	16.38	*	24.52	*

Source: created by the researcher

4.3.3 Facilitating production through perception

Based on the SAM, perception is the precursor of production. Experiment studies have demonstrated that perceptual training can facilitate Chinese pronunciation (Deng, 2017; Wang et al., 1999; Yang, 2021). Failure to perception would affect production. During a two-month experiment, Deng and Lin (2017) investigated how Chinese tone perceptual training affected the tone generation of six non-tonal native language learners from European. The results indicated that their improvements in Chinese tone perception significantly enhanced their pronunciation of Chinese tones in both the short and long term. Yang (2021) also found that audio-visual perceptual training could significantly facilitate the pronunciation learning of untrained words.

Moreover, the use of automated tonal feedback has been demonstrated to significantly enhance learners' perception of Chinese tones, thereby facilitating better pronunciation. A study by Chen (2022) involving a four-week training focused on word found that the group receiving Praat-assisted feedback made much greater progress compared to the control group. In the afterword interview, participants expressed a preference for replicate the corrected tones and comparing their own tone contours with those of the corrected ones. The findings underscore the value of incorporating research-driven technologies and automated tonal feedback into the pedagogy of Chinese as a foreign language.

5. Implications

This section addresses RQ2 and 3 to identify the implications and outline the general guidelines for the design of the AI-based Chinese CAPT system, informed by the insight of literature review of Chinese pronunciation research and discussion grounded in the relevant SLA theories. From the literature review, this study identified three themes: types of learners' tone errors, negative transfer from the native language, and the effectiveness of perceptions towards tone acquisition. As to the last theme, another three sub-themes were identified: (1) the establishment of the sense of tone awareness, (2) the interaction effect between the phonetic perception and its production.(3) facilitating pronunciation production through the enhancement of perception. These results imply the ongoing study trends about the Chinese pronunciation: (1) AI technology has been variously involved in Chinese pronunciation acquisition since the late 1980s; (2) Facilitating the pronunciation acquisition through its perception has become a main study stream and proved effective. The following parts illustrate the enlightenment given by each theme, then the guidelines derived for the designing of AI-base Chinese CAPT.

In terms of the types of tone errors and the root of error commitment, researchers found that both learners, either speaking tonal or non-tonal native languages, committed tone errors in various materials, including single characters, dual characters, trisyllables, phrases and passages. Though aiming to identify the difficulty learning order to predict the acquisition sequence of tone based on the tone errors according to Krahenmann's

Natural Order Hypothesis, researchers did not find the absolutely same order. Even within either of the learning groups, learners presented different difficulty orders. However, their errors shared the common characteristics: (1) Most of their utterance showed the narrower tone ranges and lower starting points compared to those of native Chinese speakers, which was proved both by the traditional way of auditory listening and manual recording, and with the assistance of AI technology. (2) Most of the errors could be accounted for by the negative transfer from the tone and intonation of learners' L1.

The findings from the first two themes imply: (1) Tone errors were not limited to certain layers of language constructs, but need to be involved in almost all layers of language constructions, including single characters, dual characters, phrases and passages. (2) ASR technology could effectively assist learners to find errors accurately in different ways, such as visualization and immediate feedback. (3) Knowing the gap between learners' utterance and the target speech through contrastive analysis could help learners know what the target sound exactly looks like.

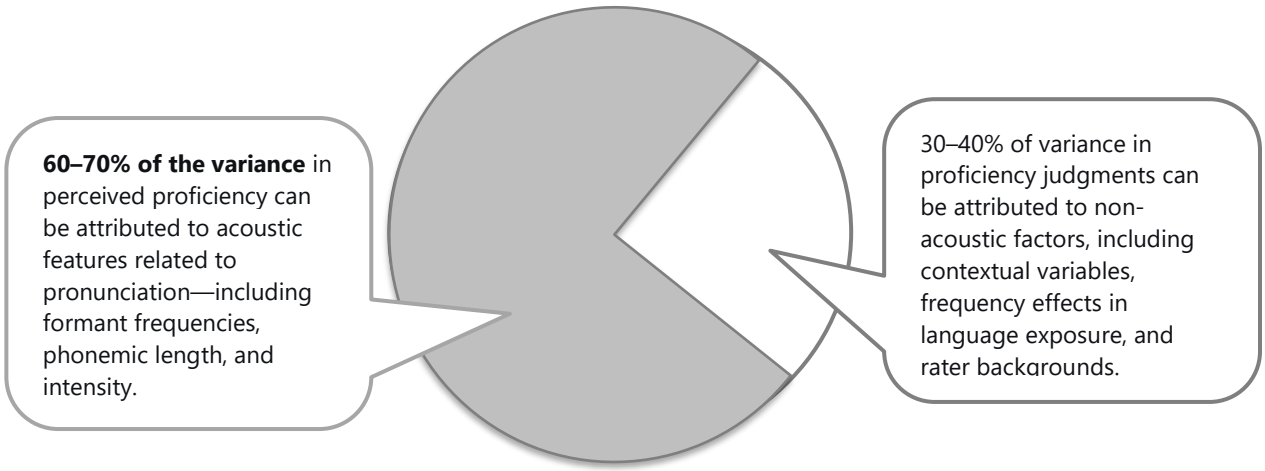
Given the hypothesis that perception is a facilitative element for production from the cognitive and psychological angle, researchers discovered that learners could build a sense of tone categories for single characters after a period of learning. However, that of dual words could not be established in the same period. Similar to the previous result of mixed difficulty tone learning order presented in the type of errors, researchers also identified different difficulty orders of perception and production in their experiments. The differing results may be attributed to various factors, such as the participants' differing proficiency levels and the nature of the listening and reading materials used—whether they were presented in pinyin alone or with accompanying characters. The common findings that deserve paying attention to are (1) the accuracy rates of perception, both of elementary learners and intermediate learners, were higher than those of their production, and (2) the error rates of single characters were generally much lower than those of dual words. These imply that (1) the establishment of the sense of tone of the single character doesn't mean that of the dual words, even the sentence's; (2) the perception is the precursor of production.

Lastly, not least, regarding the study of how phonetic perception facilitates production, researchers discovered that perception training could significantly enhance production. The automatic tonal visual feedback, aided by technology, dramatically improved learners' production. Additionally, participants stated that they prefer for the imitating corrected tones in their voices and for directly comparing their tone contours with the corrected tones in the follow-up interviews. The findings gave the enlightenment that (1) training perception facilitates the improvement of production and pronunciation acquisition. (2) AI-based CAPT holds promise for aiding pronunciation acquisition and is well-received by learners.

Input is essential for SLA, which provides the correct forms of target language to listeners through auditory or visual means, so that learners can create the hypothesis which could be tested in the later language acquisition processes of negotiation and production. However, not all input information would be acquired. The input forms must be processed by the learner's cognitive faculties, which involve principles such as comprehension, attention, working memory, and prior experience, before they can be produced as output. Generally, only the information that learners notice and truly acquire can be accurately produced. This is why production accuracy is usually lower

than that of perception. Though AI has been proven to facilitate pronunciation acquisition, relying solely on acoustic data is insufficient to replicate human speech evaluation. Human raters efficiently process speech signals for word recognition while also integrating contextual factors (Broersman & Cutler, 2008). Figure 7 illustrates the correlation between expert ratings and acoustic measurements across subjective and objective scoring approaches Therefore, scholars should clearly align their methodology—whether employing acoustic analysis, human judgment, or a combination of both—with their specific research aims (Saito & Plonsky, 2019).

Figure 7：Expert Ratings and Acoustic Properties Scoring Methods on Specific L2 Pronunciation Proficiency



Source: created by the researcher

Based on the above discussion, we identified the general guidelines for the design and development of AI-based Chinese CAPT in Table 5.

Table 5：Guidelines Derived from the Pedagogy-Informed Knowledge Basis of Chinese Pronunciation

Theme	Finding	Implication	Guideline
The types of tone errors	1. Different difficult ratings of difficulty. 2. ARS technology provided an objective comparison between the auditory features of the target sound and the learners'. 2. Shorter and more unstable tone duration. 3. Tone ranges were comparatively narrower.	Tone error patterns are systematic and measurable. These errors can be objectively analyzed with technological tools, enabling tailored feedback and focused training.	P1. Integrate automatic tone analysis and visualization features to compare learner tone contours with native models, highlighting instability and range issues.
Reasons for committing tone errors	1. L1 transfer is one main factor to affect the tone acquisition. 2. The tone and intonation of learners' native languages potentially influence the Chinese tone acquisition.	1. Learners from different language backgrounds may commit different types of tone errors due to native language interference. 2. The system serves as a supplement of formal classroom teaching, 3.Teachers need to know the basic knowledge of learners' native language, and the differences between it and the Chinese language in order to take suitable strategies to deal	P2. Design diagnostic tools and adaptive modules that identify L1-related tone error patterns and offer targeted corrective feedback.

		with learners' errors.	
Perception& production	1. Perception and production dimensions are interrelated. 2. Phonetic perception was the precursor to pronunciation production. 3. Learners presented higher accuracy of tone perception than that of production after a period of learning Chinese. 4. Chinese learners committed fewer tone perception errors with single characters than with dual words. They established a sense of tone awareness for single characters after one year of study, but not for dual characters. 5. Learners showed different learning difficulty order. 6. Perceptual training can significantly facilitate Chinese pronunciation. Audio-visual perceptual training could significantly facilitate the pronunciation learning. 7. Technology and automated tonal feedback greatly facilitate the pronunciation acquisition. Participants prefer the imitation of correct tones in their voices.	1. Perception training is a foundational step in tone acquisition. 2. Technology-assisted, multimodal (audio-visual) training can enhance both perception and production, especially if feedback is personalized and learner-controlled. 3. The practice should cover different layers of language constructions, including single characters, dual characters and sentences.	P3: Integrate perception-first progression: Sequence tasks from perception to production. Include single-character tone perception exercises before progressing to disyllabic or sentence-level tasks. P4: Employ multimodal feedback. Provide audio-visual training materials with real-time feedback. P5: Support learner voice modeling. (1) Enable learners to hear and imitate their own voice compared with model pronunciations. (2) Allow learners to track their progress via feedback tools.

Source: Created by the researcher

7. Conclusion

With the goal to establish the knowledge base informed by Chinese pronunciation research in China, providing the resolution to address the tension between technology and pedagogy when integrating advanced AI-technology into Chinese CAPT to assist language acquisition, and grounded in the CAPT knowledge framework constructed on the Richey & Klein's DDR knowledge framework, this study derived the general guidelines based on the literature review of Chinese pronunciation in China. This not only provided a deep insight into Chinese pronunciation research, but also laid a content knowledge base and robust theoretical foundation for the design and development of AI-based Chinese CAPT.

However, this study has some limitations. The guidelines are context-based and not suitable to be generalized to other contexts. Besides the knowledge base drawn from the fields of SLA theories and Chinese pronunciation pedagogical research, laying the relative theoretical foundation, the real challenges faced by the state holders when they are learning Chinese pronunciation and current CAPT systems also need to be analyzed to triangulate and enrich the results based on the DDR process. The future research

should explore the real needs of the state holders in Chinese pronunciation and challenges encountered in the application of the CAPT systems, providing a more comprehensive understanding of the Chinese CAPT system.

Ethics Approval and Consent to Participate

The researchers used the research ethics provided by the Research Ethics Committee of Universiti Teknologi MARA (UiTM). All procedures performed in this study involving human participants were conducted in accordance with the ethical standards of the institutional research committee. Informed consent was obtained from all participants according to the Declaration of Helsinki.

Acknowledgement

The manuscript is extracted from the doctoral thesis submitted to Universiti Teknologi MARA (UiTM), Shah Alam, Selangor, Malaysia.

Funding

This study was supported by the Centre for Language Education and Cooperation of China (Grant No. YHJXCX23-012), the Hainan Provincial Higher Education Teaching Reform Research Project (Grant No. HNJG2024ZC-60), and the Hainan Medical University Educational Research Project (Grant No. HYYB202320).

Conflict of Interest

The authors declare no conflict of Interest.

References

Amrate, M. & Tsai, P-h., (2024). Computer-assisted pronunciation training: A systematic review. *ReCALL*, 37(1), 22–42. https://doi.org/10.1017/S0958344024000181

Anderson, C., Bjorkman, B., Denis, D., Doner, J., Grant, M., Sander, S., & Taniguchi, A. (2022). *Essentials of Linguistics*（2nd ed.). eCampusOntario. https://openlibrary-repo.ecampusontario.ca/jspui/bitstream/123456789/497/1/Essentials-of-Linguistics-1539375869.pdf

Braun, V., & Clarke, V. (2006). *Using thematic analysis in psychology. Qualitative Research in Psychology, 3(2), 77–101.* https://doi.org/10.1191/1478088706qp063oa

Broersma, M., & Cutler, A. (2008). Phantom word activation in L2. *System*, 36, 22-34. DOI: https://doi.org/10.1016/j.system.2007.11.003

Cengiz, B. C. (2023). Computer-Assisted Pronunciation Teaching: An Analysis of Empirical Research. *Participatory Educational Research, 10*(3). 72-88. https://doi.org/10.17275/per.23.45.10.3

Chen, C. H., Koong, C. S., & Liao, C. (2022). Influences of integrating dynamic assessment into a speech recognition learning design to support students' English speaking skills, learning anxiety and cognitive load. *Educational Technology & Society*, 25(1), 1–14. https://www.jstor.org/stable/48647026.

Chen, H. (2020). New understanding of Chinese tone teaching based on cognitive theory. *Journal of Tianjin University* (Social Sciences), 22(04), 361–367.

https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk3aQQg3mh3ZtOQ7zHW5fGb-LLPraP_uNirJIQfNG_RWKVm5QkIflCXIOzV5hK3taoeshOxIS0bs5a3BVyX3JEhafMV7--XDgriOcfyU8DmLW_h_suko5gzSiuK6kOH6C5st-w8JorfkiCT-NENNiYsOmOOTmauCGWk=&uniplatform=NZKPT

Chen, M. (2022). *Computer-aided feedback on the pronunciation of Mandarin Chinese tones: Using Praat to promote multimedia foreign language learning. Computer Assisted Language Learning, 37(3), 363-388* https://doi.org/10.1080/09588221.2022.2037652

Chen, M., & Wang, J. Q. (2008). Initial Development of Chinese Monosyllabic Tone Awareness of Non-native Chinese Speakers: a Case Study. *Nankai Linguistics.* 02:63-70+165. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk34prxhfAxBOaqQnU82RhfQ_RgRUdX5yTHkbk408oQ0ZTgmgQE-zvoTEdaUvejjg1XZ9fECcuKP4qAjMI3_yXxluEs0WsYttU5iyy0mCTKku9A-cuJink2QjgtYCcQVy3rjiceWhAvJNHNxRuwFQCl1cmk6jExn6gc=&uniplatform=NZKPT

Deng, D., &Lin, Y.Q. (2017). Application of perceptual training to Chinese tone teaching. *Journal of Yunnan Normal University (Teaching& Studying Chinese as a Foreign Language Edition) 15* (03), 21-27. Doi:10.16802/j.cnki.ynsddw.2017.03.003. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk03_8NL0obQ1--s6agu5Q36e85AgdCL8Os5RKkpHaXDmVWwlQmV7hT8HPFg7GrTKlIOYQSZUifIH7XcP3gi5PeFLC7luJ8Lh0C8cfFDFwuCQajmSX8Fh7jjlEsRlyUr_AfPLGnXwqBQCw3DpcIKxVdEp5T2-6C3LEQ=&uniplatform=NZKPT

Ding, A. Q., Cheng, H. T., Yu Y., Ding, H, & Cheng, W. J. (2021). The present situation and problems of global Chinese learners' oral performance: Based on big data analysis on HSKK. *Language Teaching and Linguistic Studies* (6), 13-23. https://kns.cnki.net/kcms2/article/abstract?v=A-1EuXenf_rmlqHUdsUe0-HUeFr7T3AKW3RgC0W72Ym4EZCBmVk5AmEh__B3rfH3117R6zBUT5OvNdl-dlIcGFtUyBcT7BXI-2T99UlREsEKrIVKUILQY6UAHCulPRH33uviB_5us-dZ_2-bpygW5s7cjiAsd7V18a1_q9XbnFI=&uniplatform=NZKPT

Flege, J. E. (1999). The relation between L2 production and perception. In *Proceedings of ICPhs99 (San Francisco*). https://www.internationalphoneticassociation.org/icphs-proceedings/ICPhS1999/papers/p14_1273.pdf

Gao, Y. J., &Li, B. G. (2018). An experimental study on the errors made by Russian students in learning Chinese tones. *Journal of Yunnan Normal University (Teaching &Studying Chinese as a Foreign Language Edition) 16 (05)*, 18-26. Doi:10.16802/j.cnki.ynsddw.2018.05.003. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk0QKMJCjSjVarJh4XV8ZiOxOa8GVeyJsYWR_MHHZzI0WJ2LAeV26F5FaiohAXLPEVK9SRvA2Srzssk4ecKujviJv1DIqre03finCINCwifKRpSMS7DK-MwIgL3qQ1WJ2yr6UhJeHI8_58MZILZ1VBusR7Y33H7boWw=&uniplatform=NZKPT

Gui, M. C. (2000). The interference of English intonation to Mandarin tones perception: the linguistic analysis and the empirical solutions. *Chinese Teaching in the World (01)*, 89–93. Doi:10.13724/j.cnki.ctiw.2000.01.013. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2AsgMfWa6LKUGjN-AdCbwp4BVDrFR5MTku2d203o5GIwY_RoMGC2vO6mc3gGp6uJCQO3ALwSHSnoNS-

UOifxD6MMgEa6d7fLvtttDAiJizw5lFDX8FJzYAII84E07sL1GvP3xTW3XAJFzXIjrEiMbx&uniplatform=NZKPT

Guo, L., & Xuan, W. H. (2013). American students' acquisition of Chinese tones at the beginning stage. *Journal of Yunnan Normal University (Teaching & Studying Chinese as a Foreign Language Edition 11*(02), 25–31. Doi: 10.16802/ j.cnki. ynsddw. 2013.02.005. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk0JoOhBniGX8b4OWJG5QyfXeHI0nUVZllU9OVl2Jzcdydknvz6n_Aic-Kwk1kM76B5RsHsO_7YaNTCZPzEI2FwjP0NLgx9ODUH731sgZiBgEZFY-PbMfzUtwSHC8bFN4gggbZuwZfrD9xvGjxgdyuhV&uniplatform=NZKPT

Huang, B. (2022). The investigation about tonal learning of multinational learners. *Journal of Social Science of Harbin Normal University, 12* (2), 96–100. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk0-0GFxD20BScBlv30GMRAt6GjWic-_LYwWAhBnlBQl5pEC9hzt4EJhJq9hikYEsZB4u0uL52-OryLJ01kzzRi2rsaFZ_hBLURRbhr7r01uIsmDta5q0v1lg7jre08IEX_kUV27xRf0lyPlv7r4pjcaN2F-BaYKbAw=&uniplatform=NZKPT

Huang, B., & Liao, X. (2021). The *Modern Chinese* (7th ed.). Higher Education Press. https://www.scribd.com/document/657889462/%E7%8E%B0%E4%BB%A3%E6%B1%89%E8%AF%AD%E7%AC%AC%E5%85%AD%E7%89%88%E4%B8%8A%E5%86%8C

Hsiao, H.-S., Chang, C.-S., Lin, C.-Y., & Hsu, H.-L. (2015). "iRobiQ": The influence of bidirectional interaction on kindergarteners' reading motivation, literacy, and behavior. *Interactive Learning Environments, 23(3)*, 269–292. https://doi.org/10.1080/10494820.2012.745435

Immonen, K., Alku, P. & Peltola, M. S. (2022). Phonetic listen-and-repeat training alters 6–7-year-old children's non-native vowel contrast production after one training session. *Journal of Second Language Pronunciation*, 8(1): 95–115. https://doi. org/10.1075/jslp.21005.imm

Jiang, D. M., Goh, C. S., & Airil Haimi, M. A. (2026). Interactive Knowledge Base for Designing an AI-based Chinese Computer-assisted Pronunciation Training: A Systematic Review, *WSEAS Transactions on Computer Research,*14. 223-236. https://doi. org/10.37394/232018.2026.14.20.

Kang, O., Thomson, R. I. & Moran, M. (2018). Empirical approaches to measuring the intelligibility of different varieties of English in predicting listener comprehension. *Language Learning*, 68(1), 115–146. https://doi.org/10.1111/lang.12270

Lee, B., Plonsky, L. & Saito, K. (2020). The effects of perception- vs. production-based pronunciation instruction. *System*, 88, 102185. https://doi.org/10.1016/j.system.2019.102185

Lee, D., Kim, Hh. & Sung, SH (2023). Development research on an AI English learning support system to facilitate learner-generated-context-based learning. *Educational Technology Research and Dev*elopment, 71, 629–666. https://doi.org/10.1007/s11423-022-10172-2

Leeuwestein, H., Barking, M., Sodacı, H., Oudgenoeg-Paz, O., Verhagen, J., Vogt, P., ... Leseman, P. (2021). Teaching Turkish-Dutch kinder gartners Dutch vocabulary with a social robot: Does the robot's use of Turkish translations benefit children's Dutch vocabulary learning? *Journal of Computer Assisted Learning*, 37(3), 603–620. https://doi.org/10.1111/jcal.12510

Levis, J. (2007). Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics* 27, 184–202. https://doi.org/10.1017/S0267190508070098

Li, Y. H. (1995). Pronunciation error committed by Thai Chinese learners. *Chinese Teaching in the World,* (02), 66-71. DOI:10.13724/j.cnki.ctiw.1995.02.013. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2jMaHnq0ENj9_B4jpokT3rsvx8uHSb7CZ9mGHa4ifnlmhQYmZkMtM_y_G4-4o0EHB6rJZpnVeofEWZ2SdF6ehBVXcKpkIHxGXSZYB-yOv0eukeoiQEfQbh_owrMrJmD0KD_9dp9EGOvO14qLL6g5yv&uniplatform=NZKPT

Li, Z. Q., &, Lin, M. C. (2018). Phonetic Issues in CFL Tone and Intonation Teaching. *Journal of International Chinese Teaching*, 3 , 26-36. https://xueshu.baidu.com/ndscholar/browse/detail?paperid=124g02h05s650ga0x00y0g102k780546
https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk3wuoMtaYV-yCPopzfMf1pLaqsj1YDIn8krtZpgA5TYa3SnjLPhO-A3s2qX6jCldNBFqE1SjkKP-E4G257-u-0dgzboLp0tfVvgY4dsZrSYKyW6jasGPW0aiNqjZyF1XYNr6lTksfdjxNiVpgbuF4e5CkN14TdiSdQa_yoBULAIDw==&uniplatform=NZKPT&language=CHS

Lin, J. P. (2016). Positioning and key points on teaching Chinese phonetics as an international language. *International Journal of Chinese Studies*, 7(02), pp.1–7..https://doi.org/10.14095/b.cnki.jics.2016.02.001.
https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2LXyHshOx8j-59J8IjCZskX4_kWPXc1HF3E9Th9uRY3ZOUjo9iyDXf4lIsSWpGDBy7DAmZr2tser7AYn6-LppgZOvn9sZFg-LkPyS8N_-zHoMkayXdHngyQMVAnqUql5S--Imyax81dBjI1ai_JI_bxnkCS4ul0GI=&uniplatform=NZKPT

Liu, Y. L., Ma J. F., &Zhao, G. CH. (2021). *Chinese Proficiency Grading Standards for International Chinese Language Education: Application and Interpretation,* 2021, pp3-12*.* Beijing Language and Culture University Press*. https://www.blcup.com/EnPInfo/index/10867#001*

Luckin, R., & Cukurova, M. (2019). Designing educational technologies in the age of AI: A learning sciences-driven approach. *British Journal of Educational Technology*, 50(6), 2824–2838. https://doi.org/ 10.1111/bjet.12861

Munro, M. J. & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1): 73–97. https://doi.org/10.1111/j.1467-1770.1995.tb00963.x

Nagle, C. L. (2018b). Motivation, comprehensibility, and accentedness in L2 Spanish: Investigating motivation as a time-varying predictor of pronunciation development. *The Modern Language Journal*, 102(1), 199–217. https://doi.org/10.1111/modl.12461

Nagle, C. L. (2021). Revisiting perception–production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning*, 71(1), 243–279. https://doi.org/10.1111/lang.12431

Neri, A., Cucchiarini, C., Strik, H. (2002). Feedback in computer assisted pronunciation training: technology push or demand pull? Proc. *7th International Conference on Spoken Language Processing (ICSLP 2002)*, 1209-1212. Doi: 10.21437/ICSLP.2002-246

O'Brien, M. G., Derwing, T. M., Cucchiarini, C., Hardison, D. M., Mixdorff, H., Thomson, R. I., Strik, H., Levis, J. M., Munro, M. J., Foote, J. A. & Levis, G. M. (2018). Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation*, 4(2): 182–207. https://doi.org/10.1075/jslp.17001.obr

Pikhart, M. (2020). Intelligent information processing for language education: The use of artificial intelligence in language learning apps. Procedia Computer Science, 176, 1412–1419. https://doi.org/10.1016/j.procs.2020.09.151

Ran, Q.B. & Yu (2019). Characteristics and patterns of Chinese speech errors: analysis and exploration into the error item database from 25 L1 backgrounds. *Chinese Teaching in the World, 33* (03), 417–432. https://doi.org/10.13724/j.cnki.ctiw.2019.03.012.https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk297Bo36bPErFK8blRrh-tCrE2B7uHXhq9T0pTfZkp3tmOYwF6v8UCvkWRAGoZxx6noUYwvjptK0amhH0LiKiWYdXZarOuT13U9pBvD1dAy8DgB0XhTN324DEzfEuGt97ht3vlpGS5-8jiHSoyH0DDOo1DzZ5JL-0c=&uniplatform=NZKPT

Richey, R. C., & Klein, J. D. (2007). *Design and Development Research: Methods, Strategies, and Issues (1st ed.)*. Routledge. https://doi.org/10.4324/9780203826034

Saito, K., Trofimovich, P. & Isaacs, T. (2016). Second language speech production: Investigating linguistic correlates of comprehensibility and accentedness for learners at different ability levels. *Applied Psycholinguistics*, 37(2): 217–240. https:// doi.org/10.1017/S0142716414000502

Saito, K. & Plonsky, L. (2019) Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, 69(3): 652–708. https://doi.org/10.1111/LANG.12345

Shen, X. N. (1989). American learn Chinese language tone. *Chinese Teaching in the World* (03), 158-168. https://doi.org/10.13724/j.cnki.ctiw.1989.03.006. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2_4UqBrcEq_J3PNAwTQSBZlWuV-zT8gKA5UGRCZs6QCuWF2I-C41aMk8Hkn_V5XxsWoRmYeMwmi8hsDX5N-NiBmrVsxXc-IQm53eGBbqKnj3auvTkM_5JmmLr2D9N6_OusgzB2kWiDtsHjC-urcxOo&uniplatform=NZKPT

Song, Y. D. (2009). Strategies on teaching tone in Chinese as a foreign language. *Language Teaching and Linguistic Studies* (03), 48-53. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk1seTLsXAtOudk1165yDRbEeIBV7iANlw_t6A7MvYk4M5TimtSzjgdv-W8gH7FJiQsESMPejB3z6sjjpyxyQD0FFqu1Die4F6bOi4_G-mhavoi_D4kXBEbvi_iYXtbTYUfI4sASuPsg8Iy4RSdlPfyn&uniplatform=NZKPT

Wang, F., Cheung, A. C. K., & Chai, C. S. (2024). Language learning development in human-AI interaction: A thematic review of the research landscape. *System, 125(000)*, 24. https://doi.org/10.1016/j.system.2024.103424.

Wang, Y. (2016). Study on the strategy of teaching Chinese tone to intermediate and advanced-level learners from Korea, American and Japanese. *Language Planning 18*, 21-22. Doi:10.16412/j.cnki.1001-8476.2016.18.012. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk1GeANakpa8svhrw4TlvCAZZilC96fj5y86QHMmZn26xECi9cE2W47F5KEgWHIxDQCs5l3H1SaSAb-avTYY0VoXpB1ov4hXxejJrRulpr1NsTetNv4l9o2uR1TnWqPaa3Mnt9r2iYTVbYqOVo0SzXZCI9LPTOalSRM=&uniplatform=NZKPT

Wu, M. J., & Hu, M.G. (2004). Explore the reason of tone errors committed by Vietnamese. *Chinese Teaching in the World 02*, 81-87+4. DOI:10.13724/j.cnki.ctiw.2004.02.009. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk1UFTLVytYpFGmg82qpvTZn-Da-ejrrLsCM2rUcqYjOmVyrjjmIxHXuQ5cTF3w-Sk3z1_C4mZPqEG20_SKUTLNh3wAfeUxhNeuQaeS18ZWgc5nZyxf0AC4nCw_Px45SBgdqzROlwf0SuzHkQFOzCwlM&uniplatform=NZKPT

Xu, Y. Y. (2009). On difficulties for Thai students to Learn Chinese pinyin and teaching methods. *Journal of Guangxi Teachers Education University* (Social Science Edition) 30 (3), 139-142. DOI:10.16601/j.cnki.issn1002-5227.2009.03.018. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2NfXfyFHojaFlZ2PHbv7oMQmq7Qs2JgYjPe2-HLCSo4878mqjVHRkdCpI_ADOtZ7HN7aX81-

pkO9Ar0oJIFiZ_L5elLXpqnuME9pC_xbbrnVThhhb4jLsshceed8SXY5W7PbhuR9KhMSZXK1nA_S92&uniplatform=NZKPT

Yang, B. (2021), Audio-visual training for tone production: considering form and meaning. *Language Teaching and Linguistic Studies 3*, 31-44. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk0f8jQ5zZC6yPUe0yXP79dZ6oW-DAEvkB8M1Rxr0rPuiA7xtL1dnSxtzE9V9Y6d5SgY1SsHIXsT-muKMCsiQmrlhDYRJBb0FesVAjODsitxXW82Cge66iefqHFIxVaea2rPmRQPH7uSmtw5cHL76lcox-keiij4x9E=&uniplatform=NZKPT

Yi, B. (2021). Study on perception of Chinese tone by the Chinese learner from Malaysia. *Nankai Linguistics 2*, 52-60. https://kns.cnki.net/kcms2/article/abstract?v=OLEU9YGVhk2HubR81a9ziUXHTn-eKGxdukiyY_63ZbxQ51ekigLIAYTf8R3-tWoToKAp7Yx8DzAai2N3NkbIYPJOqmtEoZwcWkpFscBz4IDFKVr4I5tnFe_OgZxeFjwutLM_sBZWo0l0J09M7XuzmOYY3TtRkarJB_ikd0pV7-0=&uniplatform=NZKPT

Zawacki-Richter, O., Marín, V. I., Bond, M., & Gouverneur, F. (2019). Systematic review of research on artificial intelligence applications in higher education–where are the educators? *International Journal of Educational Technology in Higher Education*, *16(1)*, 1–27. https://doi.org/10.1186/s41239-019-0171-0

Zhai, X., Chu, X., Chai, C. S., Jong, M. S. Y., & Li, Y. (2021). A review of artificial intelligence (AI) in education from 2010 to 2020. *Complexity, 6,* 1–18. https://doi.org/10.1155/2021/8812542