



Optimization of Multimedia English Teaching Computer-Aided System based on Internet of Things

Xiaoling Liu^{1,*}  and Na Li² 

¹Foreign Language Department, Anyang Institute of Technology Anyang, Henan 455000, China, 20160675@ayit.edu.cn

²Department of Computer Science and Information Engineering, Anyang Institute of Technology, Anyang, Henan 455000, China, lisa_li517@163.com

Corresponding author: Xiaoling Liu, 20160675@ayit.edu.cn

Abstract. This paper systematically discusses the development history and application status of Computer Assisted Language Learning, and proposes the construction ideas and practical application path of CALL guided by Educational model theory. Using the Computer Aided Pronunciation Training system as a reference case, the paper analyzes in detail how to guide English phonetic learning based on computer technology, and then realize the meaning understanding of language. The paper also conducts an experimental study to analyze the utility of the CAPT system, and finds that classes that have adopted the CAPT demonstrate higher language cognitive abilities. More importantly, results revealed a significant positive effect of 'use CAPT or not' on English speaking evaluation for low English capacity students. However, the effect was not significant for high English capacity students, which means CAPT is more useful for beginners. Finally, the paper suggests that the future CAPT system should focus on software design, overcome technology-driven orientation, play the role of artificial intelligence, and strengthen cross-disciplinary cooperation to fully exploit the potential of computer technology in English language teaching.

Keywords: Computer Assisted Language Learning; Computer Aided Pronunciation Training; English; Internet of Things; Rhyme

DOI: <https://doi.org/10.14733/cadaps.2022.S6.136-146>

1 INTRODUCTION

Computer technology has been used for language learning for decades. From sentence analysis of text, to multimedia-assisted human-computer interaction, to the construction of an Internet of Things cloud teaching platform, computer updates and iterations have provided more possibilities for language learning, and online education has been able to make great progress. Especially with the global spread of the new crown epidemic, offline teaching has suffered a huge impact, and English

teaching institutions with the largest market share have encountered the plight of class suspension and school closure. In response to this sudden challenge, online English education has become the primary emergency solution. However, despite the rapid development of Internet of Things-based computer-assisted teaching technology in recent years, there are few cases of construction and application in the field of practice. The key reason for this is that most teachers lack sufficient knowledge of the application value of Computer Assisted Language Learning (CALL) and the corresponding experience of using it, which leads to the mismatch between technology update and practical application [1], and it is urgent for the academic community to make a systematic demonstration. CALL technology can be traced back to the 1950s and was part of Computer Assisted Instruction (CAI) in the early days. CAI can be any subject using computer-assisted instruction and is not limited to language teaching. However, the 1983 annual meeting of the World Association of Teachers of English advanced the widespread use of technology in CALL. Since then, Computer-Assisted Language Instruction (CALI), Computer-Assisted Language Teaching (CALT), Computer Aided Language Acquisition (CALA), and Network Based Language Teaching (NBLT) and other computer-assisted language teaching concepts have emerged, but CALL has been generally recognized by the computer-assisted language teaching community for its conceptual simplicity, scientificity and operability—That is, the search for and study of application of teaching and learning [2].

CALL technology has evolved along with the update of computer network technology, and the corresponding platform construction has become increasingly complex. In general, CALL platform development requires following the basic principles of R&D, a well-structured R&D team, scientific R&D methods, and progressive R&D steps (Figure 1). However, most of the CALL platform applications lack pedagogical and linguistic foundation, and it is difficult to ensure the effectiveness of teaching applications. Therefore, this paper proposes the construction idea and application path of CALL platform with Computer Aided Pronunciation Training (CAPT) as a reference case, in order to provide experience support for the online promotion of computer-assisted English teaching.

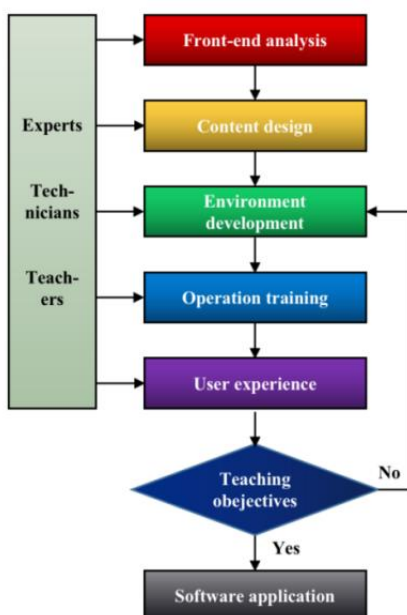


Figure 1: CALL development process.

2 RELATED STUDIES

2.1 Educational Model Theory

Behaviorist learning model. Behaviorist theory is one of the founding theories of teaching and learning, which believes that learning follows a stimulus-response model that requires students to repeatedly train, reinforce, and consolidate knowledge in order to develop correct learning behaviors. This theory has led to the emergence of a series of teaching principles that follow behaviorism, emphasizing the teacher's dominant role and the student's role as a passive recipient of knowledge, with the language learning model consisting of the teacher's gradual output of knowledge, the student's active response, and the mutual feedback and improvement of the teacher and the student. However, this model greatly ignores students' initiative and creativity, and language learning is not a one-way input and output. Although students in this model have solid basic skills, their learning pattern is still stuck in the simple repetition and memorization stage, and they lack the ability to cultivate critical thinking and innovative thinking.

Cognitivist learning model. Cognitivist theory believes that human factors dominate the stimulus-response framework, and that individual perception, observation, analysis, reasoning and information processing all play a decisive role in learning ability. Accordingly, English teaching under the cognitivist model of learning emphasizes that students are the human subjects and should be the primary, teacher-assisted, student-led model requiring meaningful excavation of language, understanding the logic of language in everyday interactions, and thus a contextualized model of teaching. The model also assumes that the process of learning English has three stages: cognition, comprehension, and application, i.e., the gradual appreciation of the inner rules of language under the guidance of the teacher, and the formation of a certain degree of linguistic competence through conscious imitation and contextualization, followed by communicative training away from the textbook, and finally the evolution of language into an internal, conscious behavior of the individual. This model emphasizes individual motivation more than the behaviorist learning model.

Constructivist learning model. Constructivist theory, the dominant theory in the field of cognition and learning, views language learning as a process in which the learner interacts with the environment and builds an understanding of the world through language. Learners need to follow their individual agency to improve their understanding of things. Specifically, students need to communicate and interact linguistically in the appropriate sociocultural contexts and to construct meaning with the assistance of others. The model emphasizes student-centeredness, with the teacher playing a guiding role in providing an environment that stimulates students' initiative to construct meaning in their language knowledge. More importantly, the main way to construct meaning in the logic of language learning is to grasp the rhythm of language.

Teaching model under the theoretical logic (Figure 2) emphasizes the active role of the students and the simulation of the situation, while the teacher only acts as a supporter, spectator and guide [3]. Among them, the students' active role emphasizes imitation and meaning understanding, which need to be deepened in the context of the situation. In fact, ELT has long focused on the value of imitation, but the challenge of how to achieve contextualized imitation is a pressing one. English educators in China have adopted incremental reforms and attempts to establish a sound language learning environment. the advent of CALL technology has brought about a turnaround. the 2004 curriculum reform advocates the adoption of a computer- and classroom-based English teaching model that makes full use of modern information technology and improves the single teacher-led teaching model. Instead, a student-centered, personalized and independent learning orientation should be formed, with the goal of fostering students' lifelong learning abilities and truly transforming from English learners to English users, thus realizing the purpose of teaching English for learning to use.

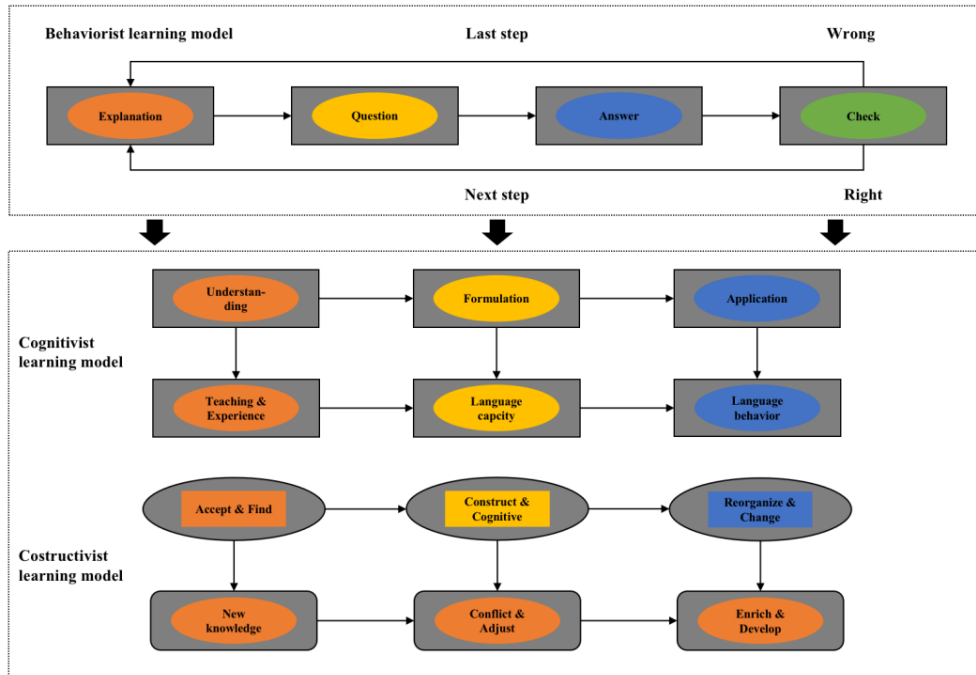


Figure 2: Educational model theory.

2.2 Computer Aided Pronunciation Training

Neurocognitive science has found that adults are no less able than children to learn the syntax and semantics of a foreign language, but the ability to learn phonology and intonation declines with age. Scholars have further found that newborn infants have the ability to discriminate phonological sounds in different languages, and their ability to discriminate phonological sounds in their native language increases after the first year of life, while their ability to discriminate phonological sounds in non-native languages decreases. To this end, the Native Language Commitment Theory (NLNC) has been proposed by Kuhl et al. [4], which assumes that infants' brain mechanisms have the ability to learn by "commitment" and that they will stimulate their native speech system according to their native language environment as they mature, but the maturation period is only one year. However, the maturation period is only one year, and after one year, the plasticity of the brain's processing of non-native speech sounds is significantly reduced and solidified by commitment, and more importantly, the non-native speech sounds are processed and modified to match the native speech representation pattern. However, recent research has shown that the impairment in adult non-native language learning is not an age-related physical limitation, but secondly due to self-reinforcing perceptual interference. Based on this finding, Zhang et al. [5] developed corresponding phonological learning software based on the pattern of language learning, which successfully enabled Japanese people to discriminate /r-l/ syllables other than those previously unchangeable, and observed by EEG that the neural sensitivity and efficiency of the subjects' brains were significantly enhanced after phonological training. In other words, adults' foreign language phonological learning is plastic, and the corresponding assistive software can reshape the neural mechanism of language learning in adults and effectively change the non-native pronunciation. Along with the development of CALL technology, the Computer Aided Pronunciation Training (CAPT) system has emerged as one of the most promising branches of CALL technology, providing students with a large amount of speech materials for repeated practice and imitation, and providing timely and personalized feedback.

3 CAPT SYSTEM APPLICATION

The CAPT system was first applied to the teaching of vowels and consonants. Since CAPT can provide multiple pronunciations of the same audio, high variability phonetic training is possible, and it is possible to construct phonological categories necessary for foreign language learning, thus improving the accuracy of vowel and consonant pronunciation. It is important to emphasize that feedback is a necessary process in phonetic intonation modeling. Before CAPT, the traditional English phonological learning model used tape recordings to follow along and identify phonological intonation errors, but lacked scientific evaluation criteria. However, CAPT can provide timely feedback that can visualize the relevant acoustic parameters of the recording. Specifically, CAPT can provide waveforms to represent the two-dimensional information of rising breaking ability over time, and further generate spectrograms to represent the three-dimensional information of speech frequency over time. Figure 3 and Figure 4 show that the vertical axis represents frequency, the horizontal axis represents time, and the strength of frequency is characterized by frequency spectrum. spectrograms contain The spectrograms contain a large amount of information on speech features, and are able to visualize speech intonation through the frequency height of resonance peaks, as well as the underlying frequency profile and the duration of syllables. Through combined visual and auditory analysis and long-term training, subjects are able to better identify vowels and consonants in English, gradually moving away from native stereotyped tones and towards the pronunciation of native English speakers, improving the effectiveness of international communication.

Although CAPT has unique advantages in teaching vowel and consonant pronunciation, the suprasegmental segment is more important for the semantic construction of language, and Anderson-Hsieh et al. [6] found that the suprasegmental segment plays an important role in the evaluation of non-native speakers. As a result, ELT has begun to focus on incorporating suprasegments in the development of CAPT systems. Suprasegments are qualities that go beyond the audio level, such as rhythm, stress, intonation, tempo, and tone quality. However, suprasegmental segments cannot be represented without vowels and consonants; for example, the prominence of several points also changes the resonance peaks of vowels, the duration of consonants, etc. In general, rhyme is an effective way to characterize these higher-level features. More importantly, rhyme is the key to learning English in English language teaching. Tone, rhythm, and stress are the three dimensions that characterize the rhythm of a phrase or sentence, and all three interact to form the rhythm of a language. For example, there is a significant difference in intonation variation between tonal and non-tonal languages. Similarly, the rhythmic variation patterns of accented metrical languages and syllabic metrical languages also show different characteristics.

In the development of CAPT, the visualization of rhythm has been a long-standing difficulty in updating ELT. Jo et al. [7] develop a program to visualize rhythm, providing a software basis for the development of visualization of rhythm. Figure 3 and Figure 4 show the rhythmic variation of the pronunciation of "I was really looking forward to seeing you again." by native English speakers and English learners, respectively. The program splits the sentences in syllables, with each circle representing a syllable. The height of the circle represents the position of the pitch and the diameter represents the length of the syllable. The rise and fall of the height represents the intonation. It is easy to see that the rhythm of native speakers has a "declarative" tone with a lot of fluctuations. However, the pronunciation of English learners is characterized by "alternating small and large waves", with little fluctuation. In other words, the rhythm of English learners' pronunciation is difficult to escape from the Chinese phonetic quality and lacks rhythmic changes. Chai et al. [8] further found that rhyme parameters can not only systematically identify rhyme biases of Chinese English learners through machine learning technology, but also assist in pronunciation teaching.

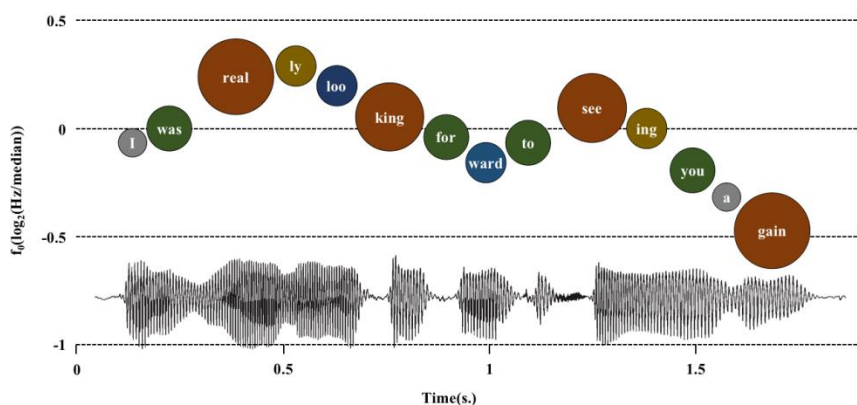


Figure 3: Rhyme of the native speaker's tongue pronunciation.

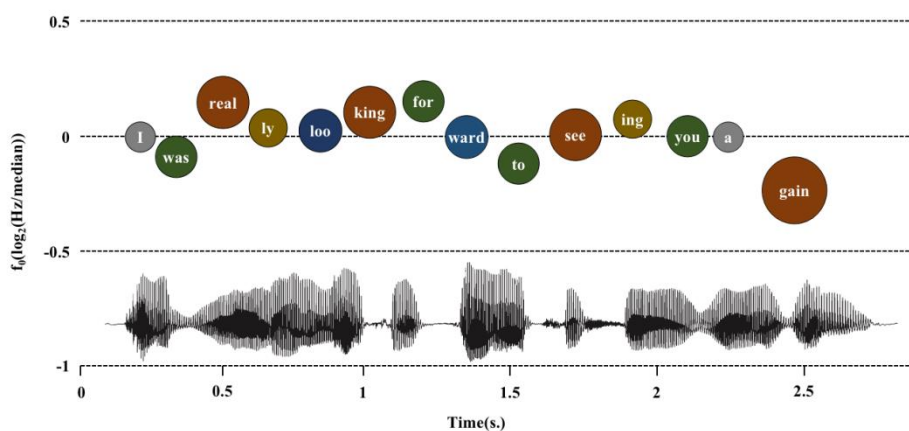


Figure 4: Rhyme of the non-native speaker's tongue pronunciation.

The important value of rhyme in foreign language learning has been widely recognized by the academic community, and the CAPT system has made great progress in this regard. Subsequently, some scholars further analyzed the user experience of the CAPT system and found that CAPT can detect 86% of phonetic errors and help learners reduce 23% of pronunciation errors [9], which is significantly effective for pronunciation improvement and has high user satisfaction [10]. It should be emphasized that the biggest obstacle to the development of rhyme visualization in the CAPT system is the difficulty of teaching the principles of speech rhyme, and if the principles of rhyme cannot be described concisely to learners, the value of rhyme parameter visualization will be greatly weakened; therefore, the development of rhyme visualization requires the joint efforts of the academic community of the industry. Figure 5 illustrates the general path of integrating a traditional classroom with a CAPT system.

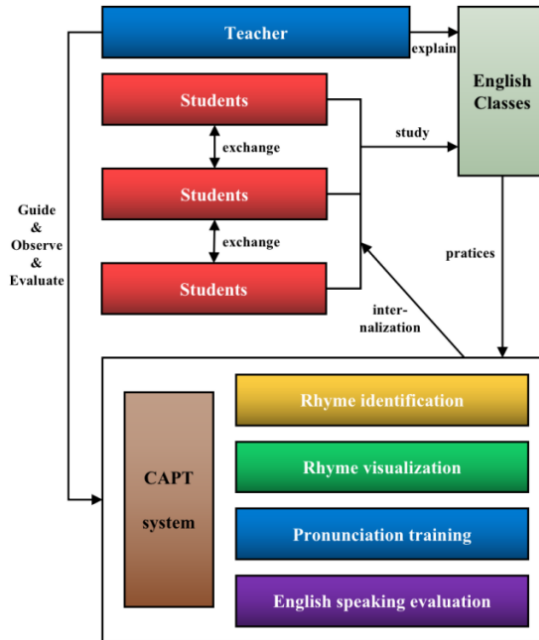


Figure 5: Pathway for CAPT embedding in the classes.

Notably, although the CAPT system has made some achievements, especially in speech recognition, and the emergence of artificial intelligence technology has greatly improved the accuracy of CAPT for rhyme analysis, the lack of cross-professionals has led to high user learning costs for the software, and there are still major limitations in practical applications and many pressing problems to be solved. Therefore, the following experimental analysis is used to identify the effectiveness of CAPT in improving students' oral English expressions by requiring teachers to explain rhyme principles to their classes and then using the CAPT system to train students for a month.

4 EXPERIMENTAL STUDY

To further explore the advantages of the CAPT system, we used an experimental approach to explore the causal effects of the CAPT system on learners' language cognitive abilities. We concluded that the learning of rhyme not only improved pronunciation accuracy, but more importantly, significantly improved English learners' oral expression ability, allowing English to be truly internalized as a tool for learners. In addition, we introduced a moderating variable-English proficiency. That is, the subject's current level of English. Therefore, we asked the teacher to evaluate the English proficiency of the experimental participants before using CAPT. This is because we hypothesized that individuals with lower English proficiency would be more effective in using the CAPT and have higher ratings of English oral expression. English oral expression skills were scored by a special teacher who administered the test. The ratings of English proficiency and English oral expression were scored from 1 to 10, with higher scores representing higher proficiency.

Figure 6 shows the rhyme analysis and scores of two classes of students (N = 144) using the CAPT system over time for non-native learners of different levels of English, and it is easy to see that although the tones and rhythms are essentially the same, the fluctuations are clearly greater for beginners.

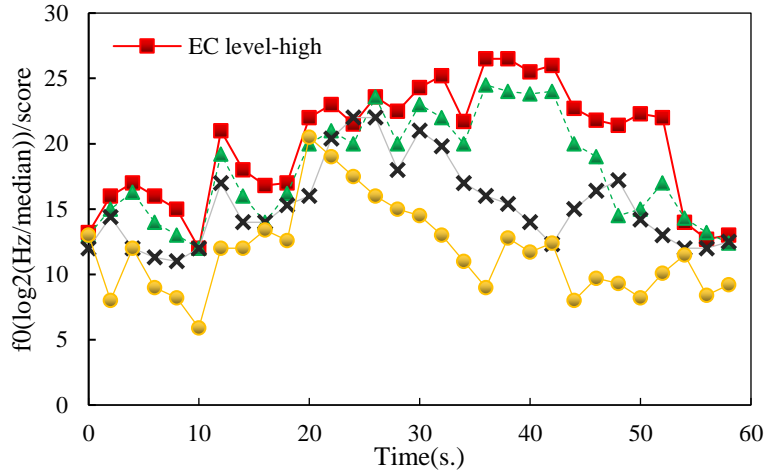


Figure 6: Rhyme analysis and scores for non-native speakers.

To exclude confounding variables, we randomly selected two middle school classes of the same grade and divided them into an experimental group and a control group, where the number of the experimental group was 71 and the number of the control group was 73, and the average age of all subjects was 13.5 years. For the experimental group, we asked the classroom teacher to use CAPT for the class study for one month, while for the control group there was no intervention (Figure 7). It should be emphasized that English proficiency was tested at the beginning of the experiment and the difference between the English proficiency ratings of the students in the experimental and control groups was not significant ($M_{no\ capt} = 4.22$, $M_{capt} = 4.24$, $t(143) = -0.271$, $p = 0.892$), so the manipulation test can be considered successful.

An evaluation of English verbal skills was conducted one month later. An English teacher was invited to conduct a simple, but standardized, daily oral conversation with each student and to give each student a rating. An independent samples t-test found that students who received training in the CAPT system had significantly higher English verbal skills than those who did not receive training in the CAPT system ($M_{no\ capt} = 4.65$, $M_{capt} = 5.88$, $t(143) = -12.370$, $p < .001$), indicating that the CAPT system had a significant positive effect on English learners' cognitive language skills.

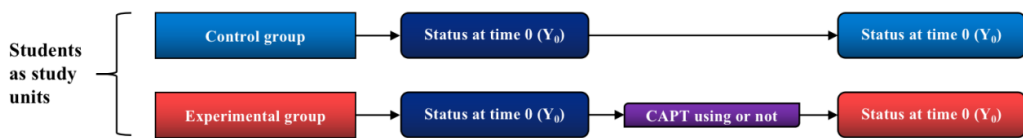


Figure 7: Experimental design.

A two-way ANOVA (Figure 8) revealed a significant main effect ($F(1, 143) = 6.231$, $p < .001$) and a significant interacting effects between familiarity and food neophobia ($F(1, 143) = 7.215$, $p < .001$). Figure 6 shows students with low English capacity in the experimental group gained higher English speaking evaluation than did those in the control group ($M_{no\ capt} = 4.51$, $SD = 1.33$ vs. $M_{capt} = 5.33$, $SD = 1.27$, $t(143) = 5.61$, $p < 0.001$). However, students with high English capacity showed no difference across the two groups ($M_{no\ capt} = 4.22$, $SD = 0.92$ vs. $M_{capt} = 4.24$, $SD = 1.22$, $t(143) = 0.33$, $p = 0.811$).

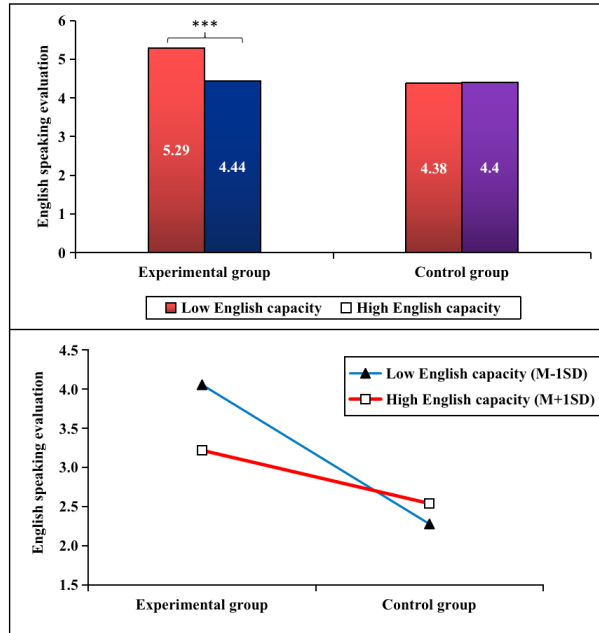


Figure 8: Moderating effect of English capacity.

To further analyze the moderating effects of English capacity, a bootstrapping procedure with 5,000 iterations (PROCESS Model 1) using 'use CAPT or not' as the independent variable, English speaking evaluation as the dependent variable revealed a significant positive effect of 'use CAPT or not' on English speaking evaluation for low English capacity students ($\beta = 0.24$, $SE = 0.25$, 95% CI, LLCI = 2.44, ULCI = 0.85). However, the effect was not significant for high English capacity students ($\beta = 0.03$, $SE = 0.33$, 95% CI, LLCI = -0.27, ULCI = 0.28). Furthermore, a Johnson-Neyman technique (Figure 9) revealed a significant effect of 'use CAPT or not' on English speaking evaluation for students with lower levels of English capacity.

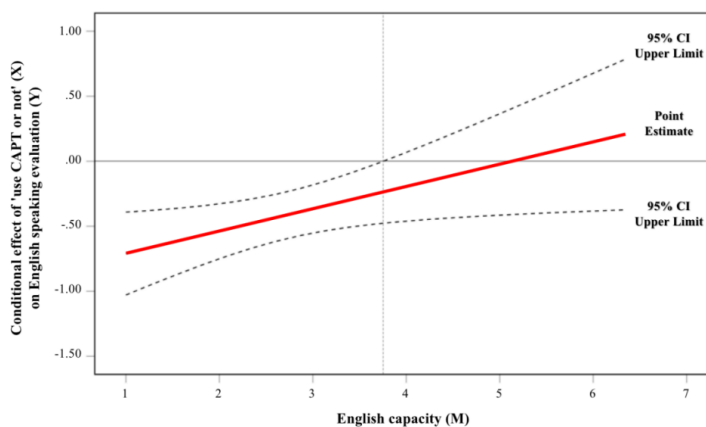


Figure 9: J-N analysis.

5 CONCLUSIONS

This paper mainly focuses on the development history and current situation of CALL technology and systematically discuss the educational model theory. We distinguish and identify the differences and similarities between the behaviorist, cognitivist, and constructivist learning models and point out that the focus of CALL development is to build a student-centered and teacher-assisted learning model so as to give full play to students' subjective initiative, tap into their imitation ability for language learning, truly understand the meaning construction behind the language, and internalize English as their own language tool. After that, the paper uses CAPT as a reference case to emphasize that rhyme is one of the key factors in improving spoken English expression, and that learning the principles of rhyme helps learners tap into the socio-cultural meaning of language. The paper further analyzes the differences in rhyme representation between native and non-native speakers of English, pointing out that it is difficult for non-native speakers to escape from the phonological intonation habits of their own native language and to delve into the construction of meaning in the language. However, the development of CAPT makes visual analysis of rhyme possible, allowing English learners to intuitively understand the acoustic parameters of their own recordings, thus better guiding and nurturing language learning towards native speakers, internalizing language as an everyday tool, and making every day international communication the norm. Finally, this paper uses an experimental method to analyze the effect of CAPT software on students' English-speaking skills, and finds that CAPT helps improve students' English oral expression skills, especially for students whose own English capacity is low, using CAPT can significantly improve their English pronunciation. These findings are important for guiding the future development direction of CAPT system. However, most computer-aided language software only provides pronunciation correction services, and users cannot identify where their own errors are, so it is difficult to get good improvement, which makes the effectiveness of software assistance greatly reduced. However, the experiments in this paper applied the CAPT system with rhyme analysis and found significant improvement of English oral expression ability. This requires future foreign language teaching experts and software eco-developers to focus on the teaching and analysis of rhyme. In addition, big data has become an important trend of thought that leads a new round of technological change after cloud computing and the Internet of Things, and has a broad application prospect in all industries and has a profound impact on the production and life style of human society. The CAPT system also needs to integrate and build a cloud platform, widely adopt machine learning and big data analysis technology, collect user pronunciation information and build a corpus, which can not only dynamically identify users' pronunciation misunderstandings, but also provide more accurate expert ratings. It can also systematically analyze and identify pronunciation errors of non-native learners, and thus provide pronunciation correction practice modules applicable to all learners. In conclusion, the CAPT system still has great potential, which requires linguists, educators, and IT personnel to participate in the software ecology development to promote the improvement and progress of the CAPT system.

Xiaoling Liu, <https://orcid.org/0000-0002-9783-9069>

Na Li, <https://orcid.org/0000-0001-5730-5399>

REFERENCES

- [1] Golonka, E.-M.; Bowles, A.-R.; Frank, V.-M.; Richardson, D.-L.; Freynik, S.: Technologies for foreign language learning: A review of technology types and their effectiveness, *Computer assisted language learning*, 27(1), 2014, 70-105. <https://doi.org/10.1080/09588221.2012.700315>
- [2] Bailin, A.: Intelligent computer-assisted language learning: A bibliography. *Computers and the Humanities*, 29(5), 1995, 375-387. <https://doi.org/10.1007/BF02279528>
- [3] Salamakhin, P.-M.; Chasovnikov, A.-D.: The concept of development of computer aided design programs for bridges or other engineering structures according to the specified criterion of

- optimality, *Structural Mechanics of Engineering Constructions and Buildings*, 15(5), 2019, 345-352. <https://doi.org/10.22363/1815-5235-2019-15-5-345-352>
- [4] Kuhl, P.-K.; Conboy, B.-T.; Coffey-Corina, S.; Padden, D.; Rivera-Gaxiola, M.; Nelson, T.: Phonetic learning as a pathway to language: new data and native language magnet theory expanded (NLM-e), *Philosophical Transactions of the Royal Society B: Biological Sciences*, 363(1493), 2008, 979-1000. <https://doi.org/10.1098/rstb.2007.2154>
- [5] Zhang, Y.; Kuhl, P.-K.; Imada, T.; Iverson, P.; Pruitt, J.; Stevens, E.-B.; Nemoto, I.: Neural signatures of phonetic learning in adulthood: A magnet oence phalography study, *Neuroimage*, 46(1), 2009, 226-240. <https://doi.org/10.1016/j.neuroimage.2009.01.028>
- [6] Anderson-Hsieh, J.; Johnson, R.; Koehler, K.: The relationship between native speaker judgments of nonnative pronunciation and deviance in segmentais, prosody, and syllable structure, *Language Learning*, 42(4), 1992, 529-555. <https://doi.org/10.1111/j.1467-1770.1992.tb01043.x>
- [7] Jo, J.; Ko, S.; Kim, D.-W.; & Lee, J.: Effective computer-assisted pronunciation training based on phone - sensitive word recommendation, *Concurrency and Computation: Practice and Experience*, 33(2), 2021, e5151. <https://doi.org/10.1002/cpe.5151>
- [8] Chai, Y.; Liu, T.-T.; Marrett, S.; Li, L.; Khojandi, A.; Handwerker, D.-A.; Bandettini, P.-A.: Topographical and laminar distribution of audiovisual processing within human planum temporale, *Progress in Neurobiology*, 205, 2021, 102121. <https://doi.org/10.1016/j.pneurobio.2021.102121>
- [9] Agarwal, C.; Chakraborty, P.: A review of tools and techniques for computer aided pronunciation training (CAPT) in English, *Education and Information Technologies*, 24(6), 2019, 3731-3743. <https://doi.org/10.1007/s10639-019-09955-7>
- [10] Saleh, A. J.; Gilakjani, A.-P.: Investigating the impact of computer-assisted pronunciation teaching (CAPT) on improving intermediate EFL learners' pronunciation ability, *Education and Information Technologies*, 26(1), 2021, 489-515. <https://doi.org/10.1007/s10639-020-10275-4>