

Computer Assisted Pronunciation Training (CAPT): A Systematic Review of Studies from 2012 to 2021

Xu CHEN^a Jie MU^{a*} & Tingting ZHANG^b

^a*Center for Research on Technology-Enhanced Language Education,*

School of Humanities, Beijing University of Posts and Telecommunications, China

^b*School of Computer Science, Beijing University of Posts and Telecommunications, China*

*mujie@bupt.edu.cn

Abstract: Learners' skills of pronunciation are crucial linguistic and communicative competence in second or foreign language learning. Computer-assisted pronunciation training (CAPT) emerged since the early 1970s and was studied by researchers from both linguistic and non-linguistic field. CAPT has been increasingly implemented in the field of language teaching and learning due to the advantages of Web 2.0 technologies. CAPT systems are used to improve learners' pronunciation in target language. A systematic review of empirical studies over the recent years is called forth in order to arouse broadened knowledge and provide new insights in this area. This study systematically reviewed 54 empirical articles on CAPT published in nine representative SCI, SSCI, or CSSCI-indexed journals from 2012 to 2021 in the field of applied linguistics and computer science. This research revealed the yearly publications, research sites, target languages, and research settings of CAPT and further analyzed the research methodologies of the selected articles. The present study revealed that the majority of studies employed quantitative research method and only one article adopted qualitative method. Finally, it summarized the common systems used in the process of pronunciation learning and found mobile app-based systems are most-widely used in recent years. This systematic literature review contributed to the knowledge of the current research status, and it also provided some insights for future research and development of CAPT.

Keywords: Computer-assisted pronunciation training, language learning, pronunciation learning, systematic literature review

1. Introduction

Language teachers have faced the dual challenge when helping learners produce target-like L2 pronunciation. One is that the perception of sounds is a subjective experience. Speech sound is transient and invisible, making the instruction of pronunciation difficult (Liu & Tseng, 2019). Another is that the teaching of pronunciation to language learners is in a less significant position since the curriculum is still dominated by other aspects of the language competence and teachers themselves usually do not have adequate training in phonetics (Pennington & Rogerson-Revell, 2019). Therefore, efforts should be made to apply technology to help language teachers instruct pronunciation.

Computer-assisted pronunciation training (CAPT) is a crucial component of Computer Assisted Language Learning (CALL) and it concentrates on the evaluation of pronunciation proficiency or correcting pronunciation errors. Technologies like computerized visualization feedback have usually been employed for pronunciation training. Recently, digital technologies have offered great potential for pronunciation training. The rapid evolution of technologies makes it possible to integrate pronunciation training with advanced technologies including automated speech recognition (ASR) and artificial intelligence (AI). Employing ASR, CAPT can provide user-friendly visualizations for pronunciation features and feedback on learners' pronunciation.

In essence, CAPT resources have the potential to provide an individualized and self-paced learning environment with a wide range of multimodal materials as well as opportunities for immediate, customized feedback (Revell-Rogerson, 2021).

2. Literature Review

Computers have been used for teaching since the 1960s and educational software was developed ever since. Computer systems used for improving language learners' pronunciation occurred in 1970s (Agarwal & Chakraborty, 2019). Kalikow and Swets (1972) designed the first software tool for the instruction of pronunciation named automated pronunciation instructor system in the early 1970s. CAPT systems received considerable attention in applied linguistics community since its existence because such systems give learners the opportunity to practice pronunciation on their own pace and significantly reduce teachers' pressure in instructing pronunciation (Richards, 1986). Students receiving pronunciation training in classroom tend to be afraid of having their pronunciation corrected in front of their peers. CAPT systems in this respect can help reduce students' anxiety and provide them with personalized feedback or suggestions and allow students to avoid face-to-face training with tutors which might be embarrassing for them.

Although CAPT as a branch of CALL existed in 1970s, the actual progress in CAPT for pedagogical purpose happened only since the beginning of the 21st century. Since digital technologies such as computers, laptops and smartphones are ubiquitous nowadays, self-study with the help of these technological tools becomes possible. Some of these tools are becoming technologically sophisticated, and CAPT systems integrated with ASR and AI are available to teach and test pronunciation (Revell-Rogerson, 2021). These tools detect and diagnose mispronunciations in learners' speech, and then help them to correct their pronunciation (Agarwal & Chakraborty, 2019). The field of CALL has developed massively in recent decades, and interest specifically in CAPT has grown similarly, with a proliferation of research focusing on web- and mobile-based pronunciation learning (eg., Dai & Wu, 2021). In this literature review, the following three questions are used to guide the analysis:

- ★ What are the general research trends of the reviewed studies about CAPT from 2012 to 2021?
- ★ What are the research methodologies employed in the reviewed publications?
- ★ What are the types of systems used in the process of pronunciation learning within CAPT context?

3. Methodology

3.1 The process of identifying journals

Because of the multi-disciplinary nature of CAPT, the researchers first searched studies from Web of Science (WoS) and IEEE *Xplore*. To explore research conducted in CAPT by Chinese scholars, we also included China National Knowledge Infrastructure (CNKI) as our database. After preliminary literature search, nine journals were selected from China and abroad. Four high-impact international journals were selected from WoS, namely *Computer Assisted Language Learning* (CALL), *ReCALL*, *Language Learning & Technology* (LLT), and *System*, all of which are indexed by Social Science Citation Index (SSCI). Two journals indexed by Science Citation Index (SCI) were retrieved, that is, *IEEE Access*, and *IEEE/ACM Transactions on Audio, Speech, and Language Processing*. The other three journals are Chinese journals indexed by Chinese Social Science Citation Index (CSSCI), namely *Technology Enhanced Foreign Language Education* (TEFLE), *Foreign Language Learning Theory and Practice* (FLLTP), and *e-Education Research* (EER). The present research adopted the systematic content analysis as the major method for reviewing studies.

3.2 The process of data coding and data analysis

Firstly, the author manually screened publications in recent ten years from the above nine journals by referring to the titles and the abstracts. Since all the selected journals are related to CALL, we first screened titles and the abstracts containing “pronunciation learning”, “pronunciation training” as their key words and excluded review articles or commentaries. After reviewing the abstract and key words, we got 64 empirical articles. Second, we read the full articles to exclude ten articles either unrelated to pronunciation training or irrelevant to technology. Finally, 54 empirical studies concerning CAPT were identified for further analysis. Following the analytical framework proposed in previous studies (Chai, Koh, & Tsai, 2013), a coding scheme was defined to systematically review the general publication trends, current issues, research methodologies and CAPT tools in this field. The coded data for the systematic review consisted of quantitative and qualitative data. The social science analysis software, SPSS 22.0, was used to analyze the quantitative data. The qualitative data were categorized and summarized with the help of qualitative data analysis software, NVivo 11.0.

4. Results

4.1 General publication trends

4.1.1 Number of the studies published from 2012 to 2021

There are 54 empirical studies in total concerning CAPT published from nine journals in recent ten years. Of all 54 published articles, 44 were from English journals and ten were from Chinese journals. CALL (16 publications) published the largest number of studies in the past ten years, followed by LLT (11 publications), *System* (6 publications), and *ReCALL* (3 publications).

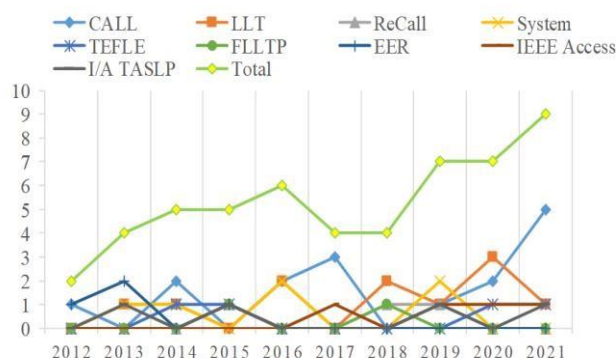


Figure 1. The numbers of reviewed publications in the 9 journals (2012 to 2021)

Over the past ten years, there has been a gradual rise in the total number of publications, and 2021 witnessed the largest number of publications (9 articles). Although CAPT is not a new topic and has developed for almost 50 years, the research about CAPT in specific is still limited. Thanks to the increasing intelligence of smartphones, more researchers focus on mobile assisted pronunciation learning, which explains why CAPT-related research has increased in recent two years. Due to the COVID-19 pandemic, many schools adopted online teaching, therefore providing the opportunity to improve students' pronunciation with online resources during school closures. For example, Jiang and Chun (2021) found that web-based training on English discourse intonation can effectively scaffold and benefit their English pronunciation.

4.1.2 Research sites and research settings of the empirical studies

It should be noted that all the all CSSCI-indexed articles (10 publications) were conducted by Chinese scholars in China, which probably explains why China ranked number one considering research sites of all the studies. Apart from ten CSSCI-indexed articles, only four international studies were conducted in China. As indicated in Figure 2, we found that scholars from USA (11 publications) and Taiwan,

China (9 publications) published many studies related to CAPT in recent ten years. In addition, scholars from Spain, Canada, Japan and Poland also showed growing interest in this field.

As shown in Figure 3, the majority (34 articles) of studies were conducted in higher education settings, which shows the prominence of CAPT in higher education. College students are advanced language learners and have more sophisticated digital literacy compared with students in other degree.

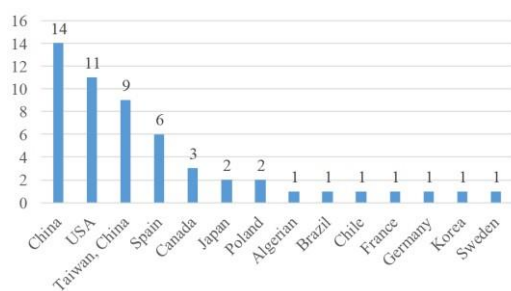


Figure 2. The research sites of the studies

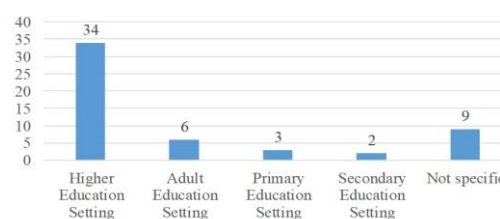


Figure 3. The research settings of the studies

Hence, they become desirable participants. There are six studies focusing on adult language learners who participated in online or distance language courses. Less attention has been paid on primary school students, with only three studies enrolling pupils as their participants. Interestingly, all of the three studies were conducted in elementary school in Beijing, and two of the studies were conducted by the same group of researchers. For example, Sun and his colleagues (2017) integrated a social media app named Papa into first-grade English classes and found that the use of such an app improved young learners' pronunciation in terms of fluency.

4.1.3 Target languages studied in the reviewed publications

As indicated in Figure 4, more than 60 percent (39 papers) of the reviewed articles targeted at the English pronunciation learning, indicating that English is the dominant target language in the reviewed articles. Since English is a foreign language in China, learners often face problems related to English pronunciation. As for CSSCI-indexed articles, eight out of ten were concerned about English pronunciation learning. Similarly, nine research conducted in Taiwan, China also focused on English pronunciation learning with the help of technology. The discrepancy between Chinese and English phonetics hinders Chinese students' accuracy in English pronunciation. Pronunciation is traditionally considered as a difficult part in EFL classrooms for both teachers and students in China, which explains the reason why a large number of studies conducted in China considered English as the target language.

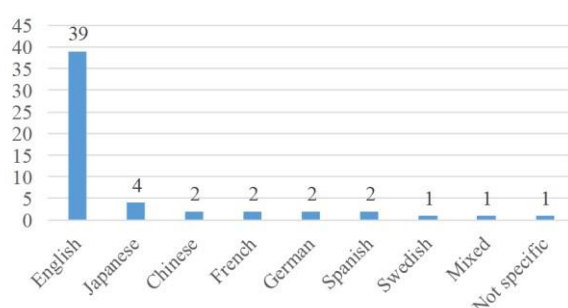


Figure 4. The target languages of the studies

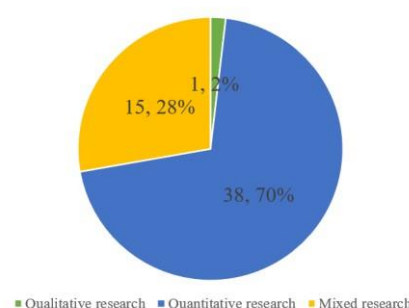


Figure 5. The research methods of the studies

4.2 Analysis of research methodologies

Among the empirical studies we reviewed, 38 studies (70%) employed quantitative research method, and 15 studies (28%) employed mixed method design. As indicated in Figure 5, only one inquiry adopted purely qualitative method. Drawing upon qualitative approach, Tsai (2019) investigated how students feel about the mediated assistance of CAPT, their difficulty and attitudes toward using technology for pronunciation learning.

Besides, most of the quantitative and mixed-method research adopted experimental- or quasi-experimental design with pre- and post-tests aimed at measuring the learners' pronunciation before and

after the training. Experimental and quasi-experimental research designs examine whether there is a causal relationship between independent and dependent variables (Rogers & Revesz, 2019). It is especially appropriate in CAPT research. The mixed research design is advantageous in that students' learning effects and perceptions were captured by quantitative and qualitative data, therefore allowing a comprehensive account of both outcomes and experiences of the CAPT process

4.3 Systems and tools used to enhance pronunciation learning

Based on the framework of Agarwal and Chakraborty (2019), we further classified the CAPT systems into five categories, namely, app-based systems, visual simulation-based systems, AI-based systems, comparative phonetics-based systems, and game-based systems.

As indicated in Table 1, we found that app-based systems (16 articles) were the most frequently used systems of all 54 empirical studies. The ubiquitous accessible and flexible practices of mobile learning might contribute to the proliferation of research that adopted mobile apps as its tool for pronunciation training. For instance, Fouz-Gonzalez (2017) investigated the potential of Twitter-based pronunciation instruction and found that learners' pronunciation improved through the intervention based on Twitter.

Visual simulation-based systems (14 articles) provide computerized visualization feedback that helps learners better visualize, analyze and produce various aspects of L2 sound (Liu & Tseng, 2019). Hence, visual-based systems ranked second frequently adopted systems with sub-categories, such as speech waveforms and spectrograms, video feedback, and Praat (eg., Liu & Tseng, 2019;).

In terms of AI-based systems, ASR seemed to be the most prevailing tool since six papers used ASR. Although ASR was not originally developed for language learning, the advantage of ASR for pronunciation learning was identified. Dai and Wu (2021) investigated the effectiveness of ASR feedback in WeChat and explored the affordances of ASR in mobile-assisted pronunciation learning.

Due to the importance of segmental features to pronunciation learning, some researchers attempt to develop perceptual phonetic training system and found that such a system was effective in enhancing participants' perception of phonemic contrasts. The ongoing popularity of game-based tools like 3D virtual worlds have been successfully integrated into pronunciation courses with positive learning outcomes (eg., Grimshaw & Cardoso, 2018).

Table 1. *Systems and tools used for pronunciation learning*

Systems	Specific Tools	Number
App-based systems (17)	Website	3
	WeChat	3
	Twitter	2
	Papa	2
	Online podcast	2
	Facebook	1
	Skype	1
	Tell Me More	1
	MyET	1
	Multimedia resources	1
Visual simulation-based systems (14)	Waveform and spectrogram	6
	Video feedback	5
	Praat	3
AI-based systems (11)	ASR tool	6
	TTS (text-to-speech) tool	3
	Intelligent assistant	2
Comparative phonetics-based systems (5)	Segmental perceptual tool	3
	Pronunciation scoring tool	2
Game-based systems (4)	Self-designed game	2
	3D virtual world game	1
	Spaceteam mobile game	1

5. Conclusion

At present, the great potential of technology for pronunciation learning is undeniable. Employing the content analysis method, this study systematically reviewed 54 empirical studies published by nine Chinese and international journals focusing on the affordances of technology for pronunciation. A coding scheme and analytic framework was postulated to analyze the publication trends of CAPT. Besides, we analyzed research methodologies of the studies and found most of the studies employed quantitative research method. Finally, we summarized the popular CAPT systems and found that mobile apps were the most popular type of tools in recent years. This systematic literature review might enhance our understanding of current research status in CAPT, and provide some insights for future research and development of CAPT.

Acknowledgements

This research is funded by the Education and Teaching Reform Project of Beijing University of Posts and Telecommunications 2022, Research on the Path and Practice of Cultivating English Majors in the Context of the New Liberal Arts (2022JXYJ-B02). We would like to acknowledge the insightful suggestions of Prof. Chunping Zheng, director of the Center for Research on Technology-Enhanced Language Education in Beijing University of Posts and Telecommunications.

References

- Agarwal, C., & Chakraborty, P. (2019). A review of tools and techniques for computer aided pronunciation training (CAPT) in English. *Education and Information Technologies*, 24(6), 3731-3743.
- Chai, C. S., Koh, J. H. L., & Tsai, C. C. (2013). A review of technological pedagogical content knowledge. *Journal of Educational Technology & Society*, 16(2), 31-51.
- Dai, Y., & Wu, Z. (2021). Mobile-assisted pronunciation learning with feedback from peers and/or automatic speech recognition: a mixed-methods study. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2021.1952272>
- Fouz-Gonzalez, J. (2017). Pronunciation instruction through Twitter: the case of commonly mispronounced words. *Computer Assisted Language Learning*, 30(7), 631-663.
- Grimshaw, J., & Cardoso, W. (2018). Activate space rats! Fluency development in a mobile game-assisted environment. *Language Learning & Technology*, 22(3), 159-175.
- Jiang, Y., & Chun, D. (2021). Web-based intonation training helps improve ESL and EFL Chinese students' oral speech. *Computer Assisted Language Learning*. <https://doi.org/10.1080/09588221.2021.1931342>
- Kalikow, D. N., & Swets, J. A. (1972). Experiments with computer-controlled displays in second language learning. *IEEE Transactions on Audio and Electro Acoustics*, 20(1), 23-28.
- Liu, Y.-T., & Tseng, W.-T. (2019). Optimal implementation setting for computerized visualization cues in assisting L2 intonation production. *System*, 87, 102145. <https://doi.org/https://doi.org/10.1016/j.system.2019.102145>
- Pennington, M. C., & Rogerson-Revell, P. (2019). *English pronunciation teaching and research: Contemporary perspectives*. Palgrave Macmillan, London.
- Revell-Rogerson, P. M. (2021). Computer-Assisted Pronunciation Training (CAPT): Current Issues and Future Directions. *Relc Journal*, 52(1), 189-205.
- Rogers, J., & Revesz, A. (2019). Experimental and quasi-experimental designs. In *The Routledge handbook of research methods in applied linguistics* (pp. 133-143). Routledge.
- Sun, Z., Lin, C.-H., You, J., Shen, H. J., Qi, S., & Luo, L. (2017). Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning*, 30(3-4), 304-324.
- Tsai, P.-h. (2019). Beyond self-directed computer-assisted pronunciation learning: a qualitative investigation of a collaborative approach. *Computer Assisted Language Learning*, 32(7), 713-744.