ReCALL (2025), 37: 1, 22–42
doi:10.1017/S0958344024000181

RESEARCH ARTICLE

Computer-assisted pronunciation training: A systematic review

Moustafa Amrate
University of Biskra, Algeria (moustafa.amrate@univ-biskra.dz)

Pi-hua Tsai
Mackay Medical College, Taipei, Taiwan (tsaipihua@gmail.com)

Abstract

This systematic review maps the trends of computer-assisted pronunciation training (CAPT) research based on the pedagogy of second language (L2) pronunciation instruction and assessment. The review was limited to empirical studies investigating the effects of CAPT on healthy L2 learners' pronunciation. Thirty peer-reviewed journal articles published between 1999 and 2022 were selected based on specific inclusion and exclusion criteria. Data were collected about the studies' contexts, participants, experimental designs, CAPT systems, pronunciation training scopes and approaches, pronunciation assessment practices, and learning measures. Using a pedagogically informed codebook, the pronunciation training and assessment practices were classified and evaluated based on established L2 pronunciation teaching guidelines. The findings indicated that most of the studies focused on the pronunciation training of adult English learners with an emphasis on the production of segmental features (i.e. vowels and consonants) rather than suprasegmental features (i.e. stress, intonation, and rhythm). Despite the innovation promised by CAPT technology, pronunciation practice in the studies reviewed was characterized by the predominant use of drilling through listen-and-repeat and read-aloud activities. As for assessment, most CAPT studies relied on human listeners to measure the accurate production of discrete pronunciation features (i.e. segmental and suprasegmental accuracy). Meanwhile, few studies employed global pronunciation learning measures such as intelligibility and comprehensibility. Recommendations for future research are provided based on the discussion of these results.

Keywords: computer-assisted pronunciation training (CAPT); second language (L2); pronunciation; pronunciation learning; pronunciation teaching; systematic review

1. Introduction

Computer-assisted pronunciation training (CAPT) has emerged as a promising tool for enhancing second language (L2) learners' pronunciation skills, potentially overcoming some of the limitations of conventional instruction. Today, a variety of commercial and open-source CAPT systems are increasingly available on desktop and mobile devices (Bajorek, 2017). These systems promise L2 learners rich pronunciation input, self-paced practice, and immediate feedback (Neri, Cucchiarini, Strik & Boves, 2002). Input in CAPT ranges from natural speech models produced by first language (L1) speakers to manipulated speech emphasizing specific pronunciation features, as well as synthetic computer-generated speech that models human

Cite this article: Amrate, M. & Tsai, P-h. (2025). Computer-assisted pronunciation training: A systematic review. *ReCALL* 37(1): 22–42. https://doi.org/10.1017/S0958344024000181

© The Author(s), 2024. Published by Cambridge University Press on behalf of EUROCALL, the European Association for Computer-Assisted Language Learning. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

pronunciation (Wang & Munro, 2004). Such forms of input are particularly valuable when presented through high variability phonetic training (HVPT), which exposes learners to a wide range of pronunciation models presented by different speakers in different phonetic contexts (Thomson, 2012). CAPT systems are further enhanced with automatic speech recognition (ASR), which provides learners with instantaneous speech-to-text conversion, error detection, and personalized feedback (Henrichsen, 2021).

The effectiveness of CAPT has been highlighted in a number of metanalyses. Mahdi and Al Khateeb (2019), for example, reviewed 20 studies investigating the effectiveness of CAPT applications in developing L2 learners' pronunciation. The results indicated that CAPT systems are effective in enhancing L2 learners' pronunciation, particularly those at beginner or intermediate levels. In a more recent metanalysis of 15 empirical studies, Ngo, Chen and Lai (2024) found ASR-based CAPT to be more effective in developing English as a second/foreign language (ESL/EFL) learners' segmental accuracy (i.e. vowels and consonants) than suprasegmental accuracy (i.e. stress, intonation, and rhythm), with explicit feedback systems being the most effective. However, while these studies provide systematic evidence for the effectiveness of CAPT, little is known about the pedagogical practices and learning measures employed in CAPT research. The next section provides a background about the current practices in L2 pronunciation instruction, including pedagogical goals, training, and assessment.

1.1 Important concepts in L2 pronunciation teaching

An often neglected area of L2 teaching, pronunciation has witnessed increased attention in the last two decades (Thomson & Derwing, 2015). This is largely due to a shift from a "nativist" approach aiming at achieving L1-like pronunciation to teaching approaches that prioritize attainable learning goals like intelligibility and comprehensibility (Munro & Derwing, 1995). In this context, "intelligibility refers to the actual understanding of the utterance by the listener" and "comprehensibility denotes the ease or difficulty of understanding on the part of the listener" (Kang, Thomson & Moran, 2018: 117). Since these partially independent measures of pronunciation serve communicative goals, contemporary pronunciation practices increasingly focus on teaching phonetic features that contribute to intelligibility and comprehensibility. In this regard, empirical studies (e.g. Kang, 2010; Saito, Trofimovich & Isaacs, 2016) have shown that a comprehensible pronunciation requires a focus on both segmental and suprasegmental features.

The practice of pronunciation requires different types of activities depending on the target phonetic features (i.e. segmental/suprasegmental features) or phonetic skill (i.e. perception/production). Explicit pronunciation instruction, for example, focuses on increasing L2 learners' awareness of the target language features, as perception is considered a precursor to pronunciation production (Lee, Plonsky & Saito, 2020; Nagle, 2021). It relies heavily on controlled speech activities, where the target pronunciation features are predetermined and elicited through phonetic notation, listen-and-repeat, or read-aloud activities. Such forms of pronunciation practice are particularly beneficial when targeting discrete (i.e. specific) phonetic features (e.g. vowels, consonants, stress, intonation) (Immonen, Alku & Peltola, 2022). Alternatively, communicative teaching tackles pronunciation through a more implicit approach by employing activities that elicit spontaneous natural speech, such as open questions, discourse completion, or picture description tasks (Derwing & Munro, 2015: 111).

Corrective feedback (CF), defined by Lightbown and Spada (1999) as "any indication to the learners that their use of the target language is incorrect" (p. 172), is another necessary pedagogical component of pronunciation training. CF can either be implicit, where errors are subtly addressed through recasts, clarification requests, and elicitation, or it can be explicit, where errors are overtly demonstrated and corrected (Engwall & Bälter, 2007). Empirical evidence indicates that optimal pronunciation learning outcomes are achieved through a combination of implicit feedback and explicit instruction (Saito & Lyster, 2012). To align with the current pedagogical consensus,

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

feedback should prioritize addressing pronunciation errors that are most likely to impact intelligibility and comprehensibility.

Assessment constitutes another facet of pronunciation instruction pedagogy, encompassing a range of concepts, methodologies, and measures. In a meta-analysis of 77 L2 pronunciation studies, Saito and Plonsky (2019) provided a comprehensive framework of key concepts in assessment encompassing constructs (i.e. global/discrete pronunciation features), speech elicited (i.e. controlled/spontaneous), and rating method (i.e. human listeners/acoustic analysis). In line with pedagogical goals, contemporary pronunciation assessment practices ideally focus on "what matters for communication" (Derwing & Munro, 2015: 110). This can be manifested in assessment practices that prioritize evaluating global pronunciation qualities (e.g. intelligibility/comprehensibility) or specific phonetic features that contribute to it. The nature of the pronunciation assessment tasks largely depends on the pronunciation features being evaluated. For example, controlled speech activities, such as reading, are ideal when assessing phonemic accuracy because they offer control over what the learner produces. Alternatively, activities eliciting more natural spontaneous or extemporaneous speech are more suitable for assessing global pronunciation. Conversely, perception can be assessed through phonetic identification (e.g. audio recording: /naɪt/, did you hear *night* or *light*?) or discrimination tasks (e.g. are these words identical or different? /bɪt/, /biːt/).

Due to their subjective nature, global pronunciation learning measures like accent and comprehensibility tend to be evaluated using scalar ratings (e.g. 1 = *extremely easy to understand*; 9 = *impossible to understand*) (Munro & Derwing, 1995). Discrete features, on the other hand, are often evaluated through criterion-based measures, where phonemic accuracy is determined by the number of phonemic substitutions or deletions and prosodic accuracy by absence, misplacement, or misuse of stress, intonation, or rhythm (Saito, Suzukida & Sun, 2019). However, some studies (e.g. Isaacs & Thomson, 2013; Lee *et al.*, 2020) still employ scalar ratings to evaluate discrete features (e.g. 1 = *utterly inaccurate*; 9 = *perfectly accurate*). Recent pronunciation research also suggests that global pronunciation measures like intelligibility can be accurately assessed through criterion-based measures such as transcription of speech or non-sense sentences (e.g. Kang *et al.*, 2018). Pronunciation can be assessed by human raters or computer-based acoustic measures. However, despite the significant advances in automatic speech assessment, it is still far from achieving human-like assessment capabilities (Isaacs, 2013).

1.2 The pedagogy–technology conflict in CAPT

Despite technological innovation, many researchers remain skeptical about the extent to which CAPT delivers effective pronunciation training. This has originally stemmed from the observed discrepancies between CAPT systems' design and pronunciation instruction pedagogy (Levis, 2007; Neri, Cucchiarini, Strik & Boves, 2002). While modern language teaching approaches strive for more attainable goals, such as comprehensibility and intelligibility, numerous CAPT systems are still built on comparing L2 learners' pronunciation to that of adult L1 speakers (O'Brien *et al.*, 2018). This often results in ASR failures with accented L2 pronunciation (e.g. Henrichsen, 2021; Martin & Wright, 2023) and with children's speech (e.g. Gelin, Pellegrini, Pinquier & Daniel, 2021). This led to skepticism about the efficacy of ASR in assessing L2 speech and encouraged attempts at using non-ASR tools (Fontan, Kim, De Fino & Detey, 2022; Fontan, Le Coz & Detey, 2018). CAPT feedback is also often perceived as a technological innovation rather than a pedagogically informed feature, hindering its reliability in detecting L2 pronunciation issues and adapting to learners.

Given this pedagogy–technology conflict, it is necessary to map pronunciation teaching practices in empirical studies investigating the effectiveness of CAPT. This is particularly important because although previous reviews have demonstrated evidence for the effectiveness of CAPT (e.g. Mahdi & Al Khateeb, 2019; Ngo *et al.*, 2024), they did not shed enough light on the

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

pedagogical practices. Therefore, it remains unclear how CAPT studies approached pronunciation training and measured learning. Further uncertainty arises about findings in the literature due to insufficient details about the methodology and CAPT systems used. This study presents a pedagogically informed systematic review that aims to categorize and evaluate the methodology, CAPT systems, pronunciation training scopes and approaches, and assessment practices in the CAPT literature. This review seeks to answer the following research questions:

1. What are the most researched L2 communities in CAPT research?
2. What are the methodological designs employed in CAPT research?
3. What systems are used in CAPT research?
4. What are the pedagogical scopes of pronunciation teaching in CAPT research?
5. How is pronunciation practiced in CAPT research?
6. What are the pedagogical pronunciation assessment practices in CAPT research?
7. How is pronunciation learning measured in CAPT research?

2. Method

2.1 Inclusion and exclusion criteria

This systematic review employed a set of inclusion and exclusion criteria to filter the relevant sources (see Table 1). The review was limited to peer-reviewed journal articles on the topic of CAPT. To ensure a minimum quality standard for the retrieved sources, the following academic databases were used: Education Resources Information Center (ERIC), ProQuest, Scopus, and PubMed. The search was also limited to the articles published between 1999 and 2022. This is primarily because the pedagogical criteria used to extract and classify the instruction and assessment practices were established during the late 1990s and early 2000s. Moreover, the studies conducted before 1999 mostly relied on CAPT systems that are vastly different from those available on the market today. Given that most research on CAPT is published in English, only studies written in this language were reviewed.

In terms of research design, this review was limited to experimental and quasi-experimental studies investigating the effectiveness of CAPT in improving L2 pronunciation. Non-experimental studies, such as viewpoint articles and reviews, were excluded. The review focused on the studies involving commercial, free open-source, or prototype CAPT systems that are specifically designed for L2 pronunciation training. The search was limited to studies involving healthy language learners with no exclusion criteria for context, age, L1, or target language. However, studies with speech- or hearing-impaired participants were not included, as they go beyond the scope of the current review. As for data collection, the review was limited to studies that measured participants' pronunciation learning after a CAPT treatment. To assess the consistency of the inclusion and exclusion criteria, an interrater reliability test (Cohen's kappa) was conducted with a second coder, who made decisions on the inclusion or exclusion of 25 publications. The results showed substantial agreement ($\kappa = .73$, percentage agreement = 88%) with the main author.

2.2 Search process

To find the relevant sources, in addition to the academic search engines ERIC, ProQuest, Scopus, and PubMed, a manual search was also conducted using the search engines of seven major journals in the field of technology and language learning: *Computer Assisted Language Learning*, *CALICO Journal*, *The JALT CALL Journal*, *Language Learning & Technology*, *ReCALL*, *Speech Communication*, and *System*. The review employed three main search keywords to generate relevant search results (see Table 2). The keywords were used to identify studies investigating the effects of CAPT systems on L2 learners' pronunciation.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 1. Inclusion and exclusion criteria

Criteria		Included	Excluded
Publication	*Year of publication*	From 1999 to 2022	
	Publication language	English	
	Publication index	Scopus, ProQuest, ERIC, and PubMed databases	
	Publication type	Journal articles	
	Review status		Not peer-reviewed
	Topic of study	Computer-assisted pronunciation training (CAPT)	
Method	*Study design*	Experimental/quasi-experimental studies	
	Training system	Computer/mobile applications designed for pronunciation training	
	Data collection	Studies that measured pronunciation learning after a CAPT treatment	
Context and participants	*Profile data*		Missing key details about the context or participants
	Participants	Second/foreign language learners	
	Speech/hearing acuity		Speech-/hearing-impaired participants

Table 2. Search keywords

Search terms
"computer-assisted" AND "pronunciation" AND "training"
"pronunciation" AND "teaching" AND "technology"
"pronunciation" AND "learning" AND "technology"

To ensure that the review includes all of the possible relevant publications, the search process was carried out on four different occasions. The first search was conducted on 14 November 2019, the second on 4 October 2020, the third on 10 December 2021, and the final search on 18 December 2022. The search process yielded 256 publications, 180 of which were generated using academic search engines, while 76 publications were the result of a manual search in reputable CALL journals. As a first step, 81 publications were deleted due to duplication. This was followed by a title and abstract screening that resulted in the removal of 100 publications due to their incompatibility with the inclusion criteria of the review. Finally, an in-depth reading of the remaining 75 sources resulted in the exclusion of 45 publications, which were either irrelevant or missing key information, leaving 30 publications for the main review and data extraction.

2.3 Data extraction

A codebook was created to manually extract and classify the necessary information from the relevant studies (see the supplementary material for the complete codebook). The codebook is

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

divided into four main sections, namely: (1) Methodology, (2) CAPT System, (3) Training, and (4) Assessment. The first section was used to extract ethnographic information, including participants' educational level, target language, language proficiency, and age range. This helped in methodically classifying the participants of the various studies. The methodology section was used to classify the studies' experimental designs, particularly with regard to the sampling approach, group design, and pre-/post-treatment testing.

The CAPT section of the codebook was used to categorize the systems based on their access type (i.e. commercial, open source, or prototype) and their technological basis (ASR/non-ASR based). CAPT input was classified into three different types of speech: natural speech, which refers to unaltered human pronunciation; manipulated speech, which is edited to emphasize certain pronunciation features; and synthetic speech, which is artificial computer-generated speech. As for the scope of training, studies were categorized based on the target phonetic level (i.e. segmental/suprasegmental), phonetic skill (i.e. perception/production), and whether the activities elicited controlled speech (e.g. listen and repeat) or spontaneous speech (e.g. picture description). The feedback was classified as implicit, in cases of simple spectrograms without error detection, or explicit, in cases where systems visually highlight specific pronunciation errors and provide a correction.

The codebook was also employed to classify the pronunciation learning assessment practices in the CAPT literature based on L2 pronunciation research consensus. This allowed coders to categorize the pronunciation production elicitation tasks into controlled speech or spontaneous speech tasks, along with discerning whether they were rated by human listeners or acoustic measures. In studies investigating participants' perception, the scheme was used to classify the assessment tasks into identification or discrimination activities. Finally, the codebook was used to classify whether the studies employed discrete pronunciation learning measures targeting phonological accuracy or global measures such as comprehensibility or accent.

The primary data extraction and coding was carried out by the main author of the study. To address the research questions, data concerning various features within each category (i.e. Methodology, CAPT System, Training, and Assessment) were classified and reported in the form of frequencies and percentages in an Excel spreadsheet. Through systematic categorization and quantification of these features, key trends, patterns, and relationships were identified in the field of CAPT research (see section 3. Results). This analysis enabled the identification of insights into extensively studied L2 communities, methodological designs, CAPT system types, the scopes of pronunciation training, assessment practices, and measures of pronunciation learning.

To evaluate the dependability of the coding scheme, an interrater reliability test was performed with a second researcher, who coded 12 out of the 30 studies. The results showed a significant agreement with the main author ($\kappa = .75$, percentage agreement = 82.20%). On an item level, the two researchers reached an agreement percentage of 82% ($\kappa = 0.75$) in coding the methods data, 81% ($\kappa = 0.74$) in coding the CAPT system data, 88% ($\kappa = 0.77$) in coding the training data, and 79% ($\kappa = 0.64$) in coding the assessment data. While disagreements among coders were resolved through discussion, readers are advised to interpret findings with caution, as coding is subject to individual variation and minor discrepancies are inevitable.

3. Results

In this section, the information extracted using the codebook is displayed in a categorical format and conveyed through measures of frequency. Table 3 provides a data extraction summary of the studies reviewed in terms of the sample size, CAPT systems, target languages, training durations, scopes of training and assessment, and pronunciation learning measures. To facilitate readability, the studies in the table are arranged based on target language, scope of training, and assessed skills.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 3. Data extraction summary

Study	*n*	Educational level	Target language	Language proficiency	Experimental design	CAPT system	Treatment duration	Training scope	Assessed skill	Pronunciation learning measure
Cucchiarini *et al.* (2009)	30	Higher education	Dutch	Beginner	Control group design	Software prototype (ASR)	4 weeks	Segmental	Production	Discrete *(Vowels, Consonants)*
Neri, Cucchiarini & Strik (2008)	30	Higher education	Dutch	Beginner	Control group design	Software prototype (ASR)	4 weeks	Segmental	Production	Discrete *(Vowels, Consonants)*
Liao (2010)	123	Higher education	English (EFL)	Unspecified	One group design	Software prototype (non-ASR)	32 weeks	Segmental	Perception	Discrete *(Vowels)*
Qian *et al.* (2018)	32	Higher education	English (EFL)	Unspecified	One group design	Software prototype (non-ASR)	70 minutes	Segmental	Perception	Discrete *(Vowels, Consonants)*
Thomson (2012)	26	Non-student participants	English (EFL)	Beginner	Comparison group design	Software prototype (non-ASR)	3 weeks	Segmental	Perception	Discrete *(Vowels)*
Neri, Mich, *et al.* (2008)	28	Primary education	English (EFL)	Beginner	Control group design	Software prototype (ASR)	4 weeks	Segmental	Production	Discrete *(Vowels, Consonants)*
Tejedor-García *et al.* (2020)	18	Higher education	English (EFL)	Intermediate	Control group design	Software prototype (ASR)	4 weeks	Segmental	Production	Discrete *(Vowels)*
Fouz-González (2020)	52	Higher education	English (EFL)	Intermediate	Control group design	English File Pronunciation (non-ASR)	2 weeks	Segmental	Perception & production	Discrete *(Vowels, Consonants)*
Lai *et al.* (2009)	120	Primary education	English (EFL)	Unspecified	Control group design	Software prototype (ASR)	12 weeks	Segmental	Perception & production	Discrete *(Vowels, Consonants)*
Amrate (2022)	18	Higher education	English (EFL)	Intermediate	Control group design	Tell Me More (ASR)	6 weeks	Suprasegmental	Production	Discrete *(Stress, Intonation)* & Global *(Comprehensibility)*
Bozorgian & Shamsi (2020)	5	Higher education	English (EFL)	Intermediate	One group design	My English Tutor (ASR)	8 weeks	Suprasegmental	Production	Discrete *(Stress, Timing, Intonation)*
Tsai (2015)	90	Higher education	English (EFL)	Unspecified	Control group design	My English Tutor (ASR)	10 weeks	Suprasegmental	Production	Discrete *(Intonation, Timing)* & Global (*Unspecified*)

(*Continued*)

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 3. (*Continued*)

Study	*n*	Educational level	Target language	Language proficiency	Experimental design	CAPT system	Treatment duration	Training scope	Assessed skill	Pronunciation learning measure
Yenkimaleki & van Heuven (2019)	48	Higher education	English (EFL)	Unspecified	Control group design	Accent Master (non-ASR)	4 weeks	Suprasegmental	Production	Discrete *(Stress)* & Global *(Accent, Comprehensibility)*
Liu *et al.* (2020)	40	Secondary education	English (EFL)	Unspecified	Control group design	Pronunciation Power 2 (non-ASR)	12 weeks	Suprasegmental	Production	Discrete *(Intonation, Rhythm)*
AbuSeileek (2007)	50	Higher education	English (EFL)	Intermediate	Control group design	Mouton Interactive (non-ASR)	12 weeks	Suprasegmental	Perception & production	Discrete *(Stress)* & Global *(Communicative competence)*
Gao & Hanna (2016)	60	Secondary education	English (EFL)	Intermediate	Control group design	New Oriental Pronunciation (non-ASR)	7.5 hours	Segmental & suprasegmental	Production	Discrete *(Vowels, Consonants, Stress, Rhythm, Intonation, Linking)*
Hincks (2003)	26	Professional participants	English (EFL)	Advanced	Control group design	Talk to Me (ASR)	10 weeks	Segmental & suprasegmental	Production	Discrete *(Vowels, Consonants, Intonation)*
Mehrpour *et al.* (2016)	30	Higher education	English (EFL)	Unspecified	Control group design	Accent Master (non-ASR)	10 weeks	Segmental & suprasegmental	Production	Discrete *(Consonant, Vowel, Stress, Intonation, Linking)*
Seferoğlu (2005)	40	Higher education	English (EFL)	Unspecified	Control group design	Pronunciation Power (non-ASR)	3 weeks	Segmental & suprasegmental	Production	Discrete *(Vowels, Consonant, Stress, Intonation, Linking)*
Lan (2022)	63	Higher education	English (EFL)	Beginner	Control group design	English Pronunciation Tutor (ASR)	8 weeks	Segmental & suprasegmental	Production	Discrete *(Vowels, Consonants, Stress, Intonation)*
Wang & Chen (2009)	80	Higher education	English (EFL)	Unspecified	Comparison group design	My English Tutor (ASR)	14 weeks	Segmental & suprasegmental	Production	Discrete *(Vowels, Consonants, Pitch, Timing, Stress)*
Elimat & AbuSeileek (2014)	64	Primary education	English (EFL)	Beginner	Control group design	Tell Me More (ASR)	8 weeks	Segmental & suprasegmental	Perception & production	Discrete & Global *(Communicative competence)*
Wang & Munro (2004)	16	Higher education	English (ESL)	Advanced	Control group design	Software prototype (non-ASR)	8 weeks	Segmental	Perception	Discrete *(Vowels)*

(*Continued*)

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 3. (*Continued*)

Study	*n*	Educational level	Target language	Language proficiency	Experimental design	CAPT system	Treatment duration	Training scope	Assessed skill	Pronunciation learning measure
Thomson (2011)	22	Non-student participants	English (ESL)	Beginner	Comparison group design	Software prototype (non-ASR)	3 weeks	Segmental	Production	Discrete *(Vowels)*
Ding *et al.* (2019)	15	Higher education	English (ESL)	Unspecified	One group design	Software prototype (non-ASR)	3 weeks	Segmental & suprasegmental	Production	Global *(Comprehensibility)*
Walker *et al.* (2011)	5	Higher education	English (ESL)	Intermediate	One group design	Software prototype (ASR)	1 hour	Segmental & suprasegmental	Production	Global *(Intelligibility)*
Kawai & Hirose (2000)	5	Higher education	Japanese	Unspecified	One group design	Software prototype (ASR)	32 minutes	Segmental	Production	Discrete *(Double-mora phonemes)*
Hew & Ohki (2004)	132	Higher education	Japanese	Beginner	Control group design	Software prototype (non-ASR)	45 minutes	Segmental & suprasegmental	Production	Global *(Unspecified)*
García *et al.* (2020)	76	Higher education	Spanish	Beginner	Control group design	iSpraak (ASR)	15 weeks	Segmental	Production	Global *(Accent, Comprehensibility)*
Teeranon (2020)	40	Higher education	Thai	Unspecified	Comparison group design	Thai Tone Application (ASR)	5 weeks	Segmental	Perception & production	Discrete *(Vowel tones, Consonant tones)*

Note. EFL = English as a foreign language; ESL = English as a second language; CAPT = computer-assisted pronunciation training; ASR = system employs automatic speech recognition; Non-ASR = system does not employ automatic speech recognition.

Table 4. Summary of profile information in the reviewed CAPT literature

Data categories	Labels	*n*
Education level	Primary education	3
	Secondary education	2
	Higher education	22
	Professional participants	1
	Non-student participants	2
	Total	**30**
Age group	Children (6–12)	3
	Teenagers (13–17)	2
	Young adults (18–33)	22
	Middle-aged adults (34–59)	3
	Total	**30**
Target language	English	24
	English (EFL)	*20*
	English (ESL)	*4*
	Dutch	2
	Japanese	2
	Spanish	1
	Thai	1
	Total	**30**
Language proficiency level	Beginner	9
	Intermediate	7
	Advanced	2
	Not specified	12
	Total	**30**

3.1 Methodological characteristics

Table 4 provides a general summary of the contexts and participants in the reviewed CAPT literature. The great majority of studies were conducted with adult learners of English at a higher educational level. While a few studies were conducted with language learners in primary ($n = 3$) and secondary ($n = 2$) schools, many more were conducted in higher education institutions ($n = 22$). Furthermore, very few CAPT studies were conducted with professional participants ($n = 1$) or non-student participants ($n = 2$). This means that the great majority of studies were conducted with young adults aged between 18 and 33 years old ($n = 22$), while fewer studies involved children, teenagers, or older adults.

In terms of the target language in these CAPT studies, English was by far the most frequent ($n = 24$). Meanwhile, the few remaining studies targeted other languages like Dutch ($n = 2$), Japanese ($n = 2$), Spanish ($n = 1$), and Thai ($n = 1$). As for the language proficiency of the participants, many studies did not specify the level of the participants ($n = 12$). Of the remaining 18 studies, nine were conducted with beginners and seven with intermediate learners, while only two studies were conducted with advanced learners.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

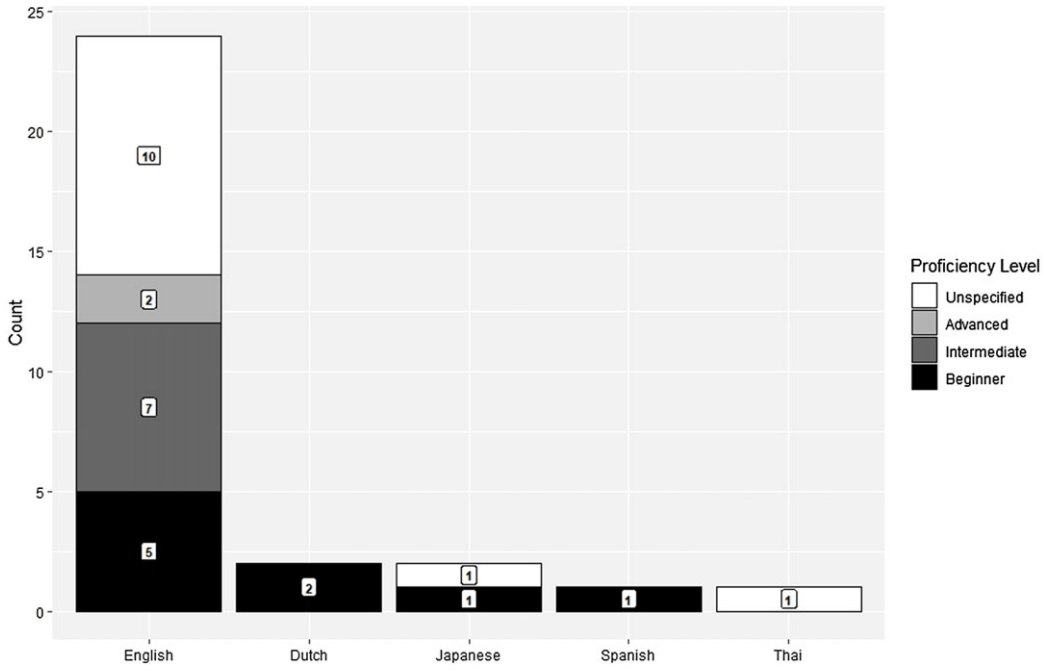


Figure 1. Target language vs. language proficiency level in CAPT research.

Figure 1 provides an overview of the target languages and participants' language proficiency levels in the studies reviewed. As noted, the majority of studies were conducted with learners of English and with intermediate-level learners ($n = 7$). The studies conducted with other languages were either with beginners or did not specify the level.

Table 5 details the experimental designs employed in the 30 studies. Since most employed a quasi-experimental approach, the research design criteria were categorized based on the group design, measurement of effects, and sampling approach. Most CAPT studies employed a pre-test/post-test control group design with random recruitment of participants ($n = 8$). Overall, the control group design was the most common design in CAPT studies ($n = 20$), while fewer studies employed a comparison group design ($n = 4$) or a single group design ($n = 6$). As for measuring effects, most of the studies made use of a pre-test/post-test design, while fewer studies used a post-test only or a time series design. As for sampling, more studies reported using random sampling approaches ($n = 10$) than those reporting non-randomized sampling ($n = 6$). Many studies, however, did not specify their sampling approach ($n = 14$). Table 6 provides descriptive statistics about the sample sizes and treatment durations in the studies reviewed.

The sample sizes in the studies reviewed ranged from a minimum of five participants to a maximum of 132 participants, and the overall average sample size was 47.47. As for the duration of interventions, training ranged from a minimum of one week to a maximum of 32 weeks, while the average study lasted seven weeks.

3.2 CAPT systems

Table 7 summarizes key details about the CAPT systems used in the studies reviewed. The systems were categorized according to their access type (i.e. commercial, free open-source, software prototype) and their use of ASR technology.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 5. Experimental designs in CAPT research

Group design	Measurement of effects	Sampling approach			Total
		Non-random	Random	Unspecified	
One group design	Post-test only design			1	1
	Pre-test/post-test design	1		3	4
	Time series design	1			1
	Total	**2**		**4**	**6**
Comparison group design	Post-test only design				
	Pre-test/post-test design		2	1	3
	Time series design			1	1
	Total		**2**	**2**	**4**
Control group design	Post-test only design				
	Pre-test/post-test design	4	8	8	20
	Time series design				
	Total	**4**	**8**	**8**	**20**
Total		**6**	**10**	**14**	**30**

Table 6. Overview of sample sizes and treatment durations in CAPT research

	M	*SD*	Min	Max
Sample size	47.47	35.21	5	132
Treatment duration (weeks)	7	6.30	1	32

Table 7. Type and technology basis of systems in CAPT research

Software type	ASR	Non-ASR	Total
Commercial software	9	7	16
Software prototype	7	7	14
Total	**16**	**14**	**30**

The studies reviewed used commercial CAPT applications ($n = 16$) as well as software prototypes that are designed for specific L2 populations ($n = 14$) almost equally. As for the use of ASR technology, just over half of the studies reviewed employed ASR-based CAPT systems ($n = 16$), while 14 studies employed non-ASR-based CAPT systems.

3.3 Pronunciation training

Table 8 details the pronunciation training scopes and approaches adopted in the studies reviewed.

Overall, Table 8 shows that most of the studies reviewed focused on the practice of segmental features ($n = 14$) rather than suprasegmental features ($n = 6$), with a specific focus on pronunciation production ($n = 15$) rather than perception ($n = 6$). Natural speech was the most frequently used input modeling tool ($n = 23$), while very few studies made use of HVPT,

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 8. Pronunciation training scopes and approaches in CAPT research

Aspect of training		*n*
Phonetic level	Segmental features	14
	Suprasegmental features	6
	Segmental & suprasegmental features	10
	Total	**30**
Phonetic skill	Perception	6
	Production	15
	Perception & production	9
	Total	**30**
Input modeling	Natural speech	23
	Manipulated speech	2
	Synthetic speech & natural speech	1
	High variability phonetic training (HVPT)	3
	Orthography	1
	Total	**30**
Speech practiced	Controlled practice	30
	Spontaneous practice	0
	Total	**30**
Feedback	Explicit feedback	11
	Implicit feedback	6
	Implicit & explicit feedback	11
	No feedback	2
	Total	**30**

manipulated speech, synthetic speech, or orthography. As for the type of speech practiced, all of the studies reviewed employed a controlled speech practice ($n = 30$), mostly through listen-and-repeat activities. Furthermore, most of the studies used explicit feedback ($n = 11$), where the system specifically highlighted pronunciation errors. Conversely, fewer studies employed implicit feedback ($n = 6$), by using speech visualization spectrograms without error detection. Interestingly, many studies used a combination of implicit and explicit feedback types ($n = 11$), while very few studies did not provide any type of feedback ($n = 2$).

Figure 2 visualizes the scopes of training in the studies reviewed. Pronunciation training in most of the studies targeted the production ($n = 6$) and perception ($n = 6$) of segmental features. Conversely, suprasegmental features were explored in production more than in perception. When targeting both phonetic levels, CAPT studies particularly focused on production ($n = 7$).

3.4 Pronunciation assessment

Table 9 summarizes key information about the pronunciation assessment scopes, tasks, rating methods, and learning measures in the studies reviewed. As with training, pronunciation assessment mainly focused on production ($n = 21$) rather than perception ($n = 4$). Another

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

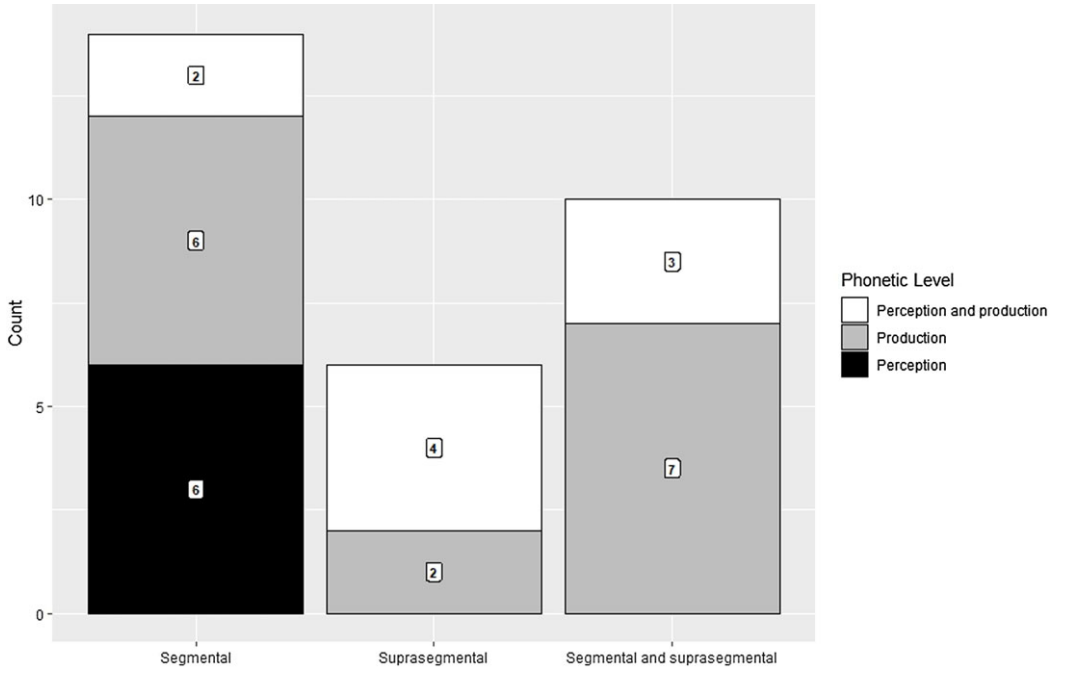


Figure 2. Scopes of pronunciation training in CAPT research.

similarity with the training trends was evident in the choice of production tasks, where most studies opted for controlled speech tasks ($n = 19$). Other studies used a combination of controlled and spontaneous speech elicitation activities ($n = 5$) and only a few studies relied solely on spontaneous speech elicitation ($n = 2$). Perception activities, on the other hand, mostly used identification tasks ($n = 5$) or a combination of identification and discrimination tasks ($n = 4$).

Table 10 shows the scope of pronunciation assessment in relation to the rating methods and learning measures in the studies reviewed.

Overall, most of the studies reviewed employed discrete pronunciation learning measures ($n = 21$) rather than global pronunciation learning measures ($n = 4$). In other words, researchers were mainly interested in evaluating learners' segmental or suprasegmental accuracy, rather than broader measures such as accent, intelligibility, or comprehensibility. In both cases, human listeners' ratings were prioritized over acoustic measures in evaluating both discrete and global measures. In Figure 3, the proportion of specific discrete and global pronunciation learning measures in the data set is represented by cell size.

Figure 3 shows that the discrete measures employed in the studies were more directed at assessing L2 learners' segmental quality ($n = 12$) than suprasegmental quality ($n = 7$). When it comes to global measures, only a small number of studies measured comprehensibility, accentedness, communicative competence, and intelligibility. Notably, two studies evaluated pronunciation globally without specifying the criteria used for assessment.

4. Discussion

4.1 Methodological trends in CAPT research

Overall, the studies reviewed showed that CAPT research is mostly conducted with adult learners of English at a higher education level. Such results are in line with Mahdi and Al Khateeb's (2019) review showing the predominance of English as the target language in empirical CAPT studies.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

Table 9. Pronunciation assessment in CAPT research

Aspects of assessment		*n*
Scope of assessment	Perception	4
	Production	21
	Perception & production	5
	Total	**30**
Production assessment task	Controlled speech	19
	Spontaneous speech	2
	Controlled & spontaneous	5
	Total	**26**
Perception assessment task	Identification task	5
	Discrimination task	0
	Identification & discrimination tasks	4
	Total	**9**
Production rating method	Human listeners	19
	Acoustic measures	7
	Total	**26**
Learning measures	Discrete measures	21
	Global measures	4
	Discrete & global measures	5
	Total	**30**

Table 10. Assessment scopes vs. ratings methods vs. learning measures

		Pronunciation learning measures			
Scope of assessment	Rating method	Discrete	Global	Discrete & Global	Total
Perception	Test scores	4			4
	Total	**4**			**4**
Production	Acoustic measures	5	1		6
	Human listeners	9	3	3	15
	Total	**14**	**4**	**3**	**21**
Perception & production	Acoustic measures	1			1
	Human listeners	2		2	4
	Total	**3**		**2**	**5**
Total		**21**	**4**	**5**	**30**

The emphasis on the English language is unsurprising, considering its growing popularity and significant role as a lingua franca. However, this does not justify the significant lack of CAPT research for the learning of other languages. Currently, little is known about how the capabilities of ASR in pronunciation training can be harnessed for languages other than English. Such a

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

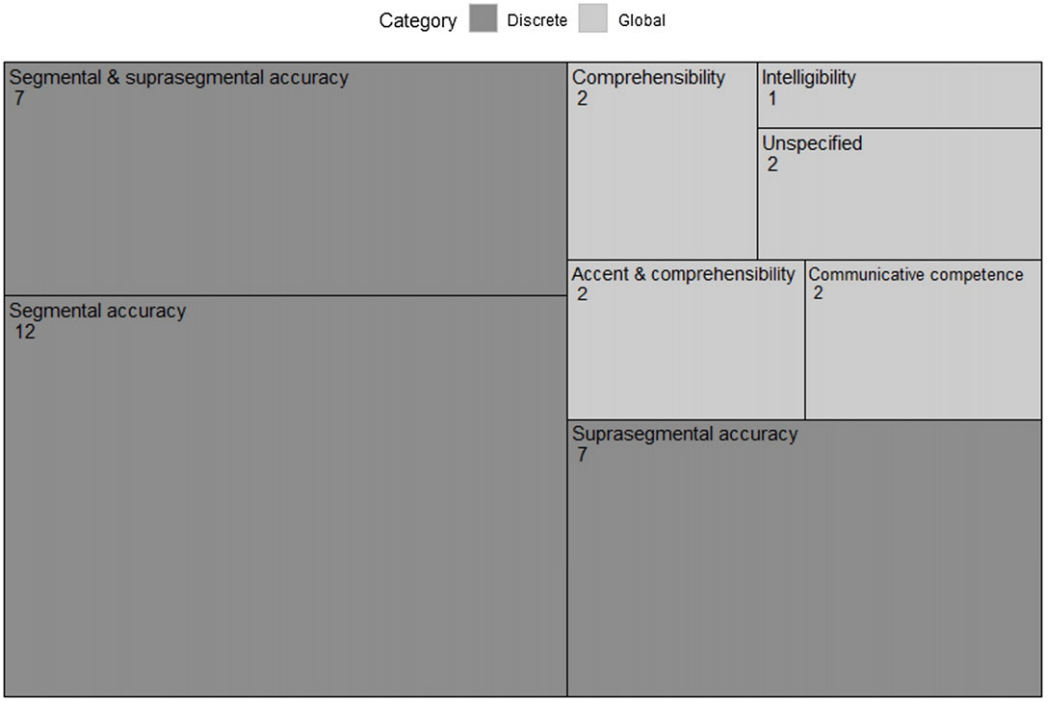


Figure 3. Discrete and global measures of pronunciation production improvement.

knowledge gap can lead to a limited applicability of CAPT research findings and missed opportunities to optimize CAPT systems that could be effective with other languages.

With the exception of a few studies (e.g. Elimat & AbuSeileek, 2014; Gao & Hanna, 2016; Neri, Mich, Gerosa & Giuliani, 2008), the review showed that little work is done with children or teenage L2 learners. This is probably because many of the studies used CAPT applications that employ advanced ASR and speech visualization systems that are designed for adult learners (Gelin *et al.*, 2021). To tackle this issue, future systems should integrate simplified user interfaces and be trained to recognize young L2 learners' speech. Alternatively, studies with young participants can engage learners in collaborative CAPT, where they can receive technical assistance from teachers or peers (e.g. Amrate, 2022; Elimat & AbuSeileek, 2014; Tsai, 2015). This would shed more light on the potential of CAPT with wider L2 populations.

Most of the studies reviewed employed the quasi-experimental control group design, while only a few studies used a comparison group design or a single group design. This demonstrates that the authors of these studies were keen to highlight the extent to which CAPT systems are responsible for pronunciation learning gains. Despite this, the generalizability of the results may be compromised due to the sampling approaches used in many studies. While the studies used various sample sizes and treatment durations (see Table 6), many did not specify the sampling approach or employed non-random sampling (see Table 5). The generalizability of the results obtained in future CAPT studies is, therefore, highly dependent on specific and detailed explanations of the sampling and group assignment approaches.

4.2 The nature of training systems in CAPT research

The studies reviewed used both commercial and prototype CAPT systems (see Table 7). Commercial systems were likely used due to their accessibility and availability on most devices

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

(Bajorek, 2017). Sometimes, however, commercial systems fail to deliver effective training, as they have traditionally focused on presenting technological innovations rather than following established pedagogical guidelines (Neri, Cucchiarini & Strik, 2002). Moreover, commercial systems are designed with a broad audience in mind and thus fail to meet the specific pronunciation needs of different L2 groups. Alternatively, prototypes (e.g. Neri, Mich, *et al*., 2008; Thomson, 2011) are often more pedagogically appropriate because they are designed with specific L2 groups in mind. If future commercial systems are audience-specific and adhere to sound pedagogical principles, they would have the potential to be more effective.

The studies reviewed also involved both ASR and non-ASR systems almost equally. This means that many studies utilized CAPT systems that were developed specifically for pronunciation training but lacked ASR features like instant speech-to-text representation, error detection, or immediate personalized feedback on learners' output. Instead, they employed systems that simply recorded learners' speech and provided informative spectrographic visualization. Other non-ASR-based CAPT systems were also focused on providing perceptual training through phonetic identification and discrimination activities. Despite the prevalent challenge of accurately recognizing accented L2 pronunciation (e.g. Henrichsen, 2021; Martin & Wright, 2023), the integration of well-trained ASR-based CAPT systems is still important in future studies, as they can deliver personalized error detection and feedback.

4.3 Pronunciation training in CAPT research

The review shows that CAPT studies are mostly focused on the production of segmental features, while suprasegmentals and perception are considerably understudied. This is similar to the patterns observed in conventional pronunciation instruction (Thomson & Derwing, 2015). A possible explanation for the overemphasis on segmental features could be the primary influence of these features on pronunciation intelligibility. Another factor might be that segmental features can be more easily assessed with ASR technology than suprasegmental features (Isaacs, 2013). However, this does not detract from the need for more CAPT research focusing on suprasegmental features, since these are highly correlated with pronunciation comprehensibility (Saito *et al*., 2016). Such features could also be targeted through perceptual training, which has been consistently shown to have a positive correlation with production (e.g. Nagle, 2021; O'Brien *et al*., 2018).

This review also showed an overwhelming use of controlled practice through listen-and-repeat or read-aloud activities. This is likely a side effect of the focus on segmental features in CAPT research, which can require the drilling of specific sounds in an attempt to match the models (e.g. Immonen *et al*., 2022). This drilling approach, however, does not guarantee the transfer of the learning gains to the untrained contexts or features (e.g. Qian, Chukharev-Hudilainen & Levis, 2018). The generalizability of the learning gains can also be negatively affected by the lack of input variability, as most studies relied on natural speech recordings of L1 speakers. This stems from a belief that considers natural L1 speech models as sufficient input (Thomson & Derwing, 2015). However, to emphasize the diverse uses of pronunciation features, it is essential to have a range of raw and manipulated input forms, including different kinds of voice, gender, accent, and context (e.g. Qian *et al*., 2018; Thomson, 2011).

As for feedback, most of the studies integrated systems that provide explicit feedback, including error detection and visualization. However, while important advances have been made in automated feedback, CAPT systems still struggle to accurately evaluate L2 speech (Henrichsen, 2021). This can lead to erroneous feedback on pronunciation that is perfectly comprehensible, negatively impacting the learning process. CAPT feedback is also criticized for being difficult to interpret, as it often highlights errors without further clarification. To address this, developers can train future systems on a variety of L2 corpora to minimize L2 speech recognition failures. Alternatively, practitioners can engage learners in collaborative or supervised CAPT, where peers

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

can compensate for erroneous feedback (e.g. Amrate, 2022; Elimat & AbuSeileek, 2014; Tsai, 2015).

4.4 Pronunciation assessment in CAPT research

The review found that most CAPT studies were interested in employing discrete learning measures to assess pronunciation production accuracy, mostly relying on controlled speech elicitation tasks. This aligns with trends observed in conventional pronunciation instruction research (e.g. Saito & Plonsky, 2019; Thomson & Derwing, 2015). These results indicate that the studies were attentive to evaluating the outcomes of their training, which concentrated on specific segmental and suprasegmental features. However, with few studies employing global measures, it is difficult to determine the extent to which phonetic accuracy gains in CAPT translate into a comprehensible or intelligible pronunciation. Furthermore, as argued by Saito and Plonsky (2019), the predominance of controlled speech elicitation tasks in assessment can give an unclear image about the transferability of the learning gains to untrained contexts. Therefore, future research can shed light on the effectiveness of CAPT in improving L2 learners' pronunciation by employing both discrete- and global-level measures using elicitation tasks that can function as a better predictor for the transferability of the learning gains.

Pronunciation ratings in the studies reviewed mostly relied on human rating rather than acoustic measurements in assessing both discrete and global features. As noted by Saito and Plonsky (2019) and Thomson and Derwing (2015), pronunciation researchers likely lean towards human evaluations because of their interest in identifying pronunciation enhancements that are perceptible to listeners, as opposed to subtle acoustic refinements that might only be detectable through acoustic analysis. Moreover, acoustic measurements can require resources and expertise in acoustic analysis to guarantee accurate pronunciation evaluation. Nevertheless, future CAPT studies can shed more light on the correlation between learning improvements detected by acoustic measures and human rating by employing a combination of both rating methods.

5. Conclusion

The aim of the study was to conduct a pedagogically informed systematic review that maps the pronunciation training and assessment practices in empirical studies investigating the effectiveness of CAPT. Overall, the studies reviewed showed that research in this area is mostly conducted with adult intermediate learners of English. On a methodological level, the studies mostly took the form of randomized controlled trials with a pre-test/post-test design. These studies employed both ASR and non-ASR commercial as well as software prototypes equally. The scope of training was mostly focused on the controlled production of segmental features. As for the training approach, the studies were entirely reliant on the controlled practice with natural speech models and a combination of implicit and explicit feedback. Meanwhile, assessment mostly targeted the production of discrete features through the use of controlled speech elicitation. As for the learning measures, the studies mostly employed human listeners to assess discrete phonetic accuracy, with few studies addressing global pronunciation quality.

The findings underscore the need for methodologically diverse and pedagogically informed CAPT research to harness the full potential of this technology. On a methodological level, CAPT research should target wider L2 populations of different proficiency levels and ages using designs that generate generalizable results. This can be made easier if future CAPT systems are trained with larger and more diverse corpora to be optimized for such populations. As for training, future research must also target pronunciation features that directly enhance learners' comprehensibility and intelligibility instead of accent. To do this, equal attention should be given to the perception and production of segmental as well as suprasegmental features through innovative training approaches beyond drilling. When assessing learning, future studies should employ discrete as

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

well as global pronunciation measures equally with speech elicitation tasks that simulate real-life use of pronunciation and better predict the transferability of the learning gains.

Several limitations should be considered when interpreting the findings of this study. Although the review focused solely on empirical studies of CAPT systems, the field encompasses diverse research areas, from software testing to teacher and learner perceptions. Therefore, the findings may not fully represent the entire CAPT literature. Furthermore, although coder disagreements in data extraction were resolved through discussion, readers need to recognize that minor individual variations in coding are inevitable, potentially impacting the interpretation of findings. The variability in systems and training, combined with the absence of key methodological details in some of the studies reviewed, further hindered the possibility of conducting advanced statistical analyses. Despite these constraints, the review provides a critical roadmap, pinpointing gaps and setting the stage for future research to enhance CAPT efficacy and applicability, ultimately advancing L2 pronunciation teaching and learning.

Supplementary material. To view supplementary material referred to in this article, please visit https://doi.org/10.1017/S0958344024000181

Acknowledgements. We extend our sincere gratitude to the anonymous reviewers who provided valuable feedback contributing to the enhancement of the manuscript.

Ethical statement and competing interests. This study did not involve human participants. All the reviewed papers are publicly available online. The authors ensured adherence to ethical research practices in their respective countries. The authors declare no competing interests. The authors also declare no use of generative AI.

References

References marked with an asterisk indicate studies included in the systematic review.

*AbuSeileek, A. F. (2007) Computer-assisted pronunciation instruction as an effective means for teaching stress. *The JALT CALL Journal*, 3(1–2): 3–24. https://doi.org/10.29140/jaltcall.v3n1-2.33

*Amrate, M. (2022) Collaborative vs. individual computer-assisted prosody training: A mixed-method case study with Algerian EFL undergraduates. *Computer Assisted Language Learning*, 35(9): 2502–2533. https://doi.org/10.1080/09588221.2021.1882503

Bajorek, J. P. (2017) L2 pronunciation tools: The unrealized potential of prominent computer-assisted language learning software. *Issues and Trends in Learning Technologies*, 5(1): 24–51. https://doi.org/10.2458/azu_itet_v5i1_bajorek

*Bozorgian, H. & Shamsi, E. (2020) Computer-assisted pronunciation training on Iranian EFL learners' use of suprasegmental features: A case study. *CALL-EJ*, 21(2): 93–113.

*Cucchiarini, C., Neri, A. & Strik, H. (2009) Oral proficiency training in Dutch L2: The contribution of ASR-based corrective feedback. *Speech Communication*, 51(10): 853–863. https://doi.org/10.1016/j.specom.2009.03.003

Derwing, T. M. & Munro, M. J. (2015) *Pronunciation fundamentals: Evidence-based perspectives for L2 teaching and research*. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.42

*Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J. & Gutierrez-Osuna, R. (2019) Golden speaker builder – An interactive tool for pronunciation training. *Speech Communication*, 115: 51–66. https://doi.org/10.1016/J.SPECOM.2019.10.005

*Elimat, A. K. & AbuSeileek, A. F. (2014) Automatic speech recognition technology as an effective means for teaching pronunciation. *The JALT CALL Journal*, 10(1): 21–47. https://doi.org/10.29140/jaltcall.v10n1.166

Engwall, O. & Bälter, O. (2007) Pronunciation feedback from real and virtual language teachers. *Computer Assisted Language Learning*, 20(3): 235–262. https://doi.org/10.1080/09588220701489507

Fontan, L., Kim, S., De Fino, V. & Detey, S. (2022) Predicting speech fluency in children using automatic acoustic features. *Proceedings of 2022 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA ASC)*. IEEE, 1085–1090. https://doi.org/10.23919/APSIPAASC55919.2022.9979884

Fontan, L., Le Coz, M. & Detey, S. (2018) Automatically measuring L2 speech fluency without the need of ASR: A proof-of-concept study with Japanese learners of French. *Proceedings of the 19th Annual Conference of the International Speech Communication Association (INTERSPEECH 2018)*. International Speech Communication Association, 2544–2548. https://doi.org/10.21437/Interspeech.2018-1336

*Fouz-González, J. (2020) Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology*, 24(1): 62–85. https://doi.org/10125/44709

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

*Gao, Y. & Hanna, B. E. (2016) Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in China. *CALICO Journal*, 33(2): 201–230. https://doi.org/10.1558/cj.v33i2.26054
*García, C., Nickolai, D. & Jones, L. (2020) Traditional versus ASR-based pronunciation instruction: An empirical study. *CALICO Journal*, 37(3): 213–232. https://doi.org/10.1558/cj.40379
Gelin, L., Pellegrini, T., Pinquier, J. & Daniel, M. (2021) Simulating reading mistakes for child speech transformer-based phone recognition. *Proceedings of the 22nd Annual Conference of the International Speech Communication Association (INTERSPEECH 2021)*. International Speech Communication Association, 3860–3864. https://doi.org/10.21437/Interspeech.2021-2202
Henrichsen, L. E. (2021) An illustrated taxonomy of online CAPT resources: *RELC Journal*, 52(1): 179–188. https://doi.org/10.1177/0033688220954560
*Hew, S.-H. & Ohki, M. (2004) Effect of animated graphic annotations and immediate visual feedback in aiding Japanese pronunciation learning: A comparative study. *CALICO Journal*, 21(2): 397–419. https://doi.org/10.1558/cj.v21i2.397-419
*Hincks, R. (2003) Speech technologies for pronunciation feedback and evaluation. *ReCALL*, 15(1): 3–20. https://doi.org/10.1017/S0958344003000211
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
Isaacs, T. (2013) Assessing pronunciation. In Kunnan, A. J. (ed.), *The companion to language assessment*. Chichester: John Wiley & Sons, 140–155. https://doi.org/10.1002/9781118411360.WBCLA012
Isaacs, T. & Thomson, R. I. (2013) Rater experience, rating scale length, and judgments of L2 pronunciation: Revisiting research conventions. *Language Assessment Quarterly*, 10(2): 135–159. https://doi.org/10.1080/15434303.2013.769545
Kang, O. (2010) Relative salience of suprasegmental features on judgments of L2 comprehensibility and accentedness. *System*, 38(2): 301–315. https://doi.org/10.1016/j.system.2010.01.005
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
*Kawai, G. & Hirose, K. (2000) Teaching the pronunciation of Japanese double-mora phonemes using speech recognition technology. *Speech Communication*, 30(2–3): 131–143. https://doi.org/10.1016/S0167-6393(99)00041-2
*Lai, Y.-S., Tsai, H.-H. & Yu, P.-T. (2009) A multimedia English learning system using HMMs to improve phonemic awareness for English learning. *Educational Technology and Society*, 12(3): 266–281.
*Lan, E.-M. (2022) A comparative study of computer and mobile-assisted pronunciation training: The case of university students in Taiwan. *Education and Information Technologies*, 27(2): 1559–1583. https://doi.org/10.1007/s10639-021-10647-4
Lee, B., Plonsky, L. & Saito, K. (2020) The effects of perception- vs. production-based pronunciation instruction. *System*, 88: Article 102185. https://doi.org/10.1016/j.system.2019.102185
Levis, J. (2007) Computer technology in teaching and researching pronunciation. *Annual Review of Applied Linguistics*, 27: 184–202. https://doi.org/10.1017/S0267190508070098
*Liao, F.-H. (2010) A new perspective of CALL software for English perceptual training in pronunciation instruction. *The JALT CALL Journal*, 6(2): 85–102. https://doi.org/10.29140/jaltcall.v6n2.94
Lightbown, P. & Spada, N. M. (1999) *How languages are learned* (2nd ed.). Oxford: Oxford University Press.
*Liu, X., Wu, D., Ye, Y., Xu, M., Jiao, J. & Lin, W. (2020) Improving accuracy in imitating and reading aloud via speech visualization technology. *International Journal of Emerging Technologies in Learning*, 15(8): 144–160. https://doi.org/10.3991/IJET.V15I08.11475
Mahdi, H. S. & Al Khateeb, A. A. (2019) The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education*, 7(3): 733–753. https://doi.org/10.1002/rev3.3165
Martin, J. L. & Wright, K. E. (2023) Bias in automatic speech recognition: The case of African American language. *Applied Linguistics*, 44(4): 613–630. https://doi.org/10.1093/applin/amac066
*Mehrpour, S., Shoushtari, S. A. & Shirazi, P. H. N. (2016) Computer-assisted pronunciation training: The effect of integrating accent reduction software on Iranian EFL learners' pronunciation. *CALL-EJ*, 17(1): 97–112.
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. (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? *Radboud Repository of the Radboud University Nijmegen*. http://hdl.handle.net/2066/76209
*Neri, A., Cucchiarini, C. & Strik, H. (2008) The effectiveness of computer-based speech corrective feedback for improving segmental quality in L2 Dutch. *ReCALL*, 20(2): 225–243. https://doi.org/10.1017/S0958344008000724
Neri, A., Cucchiarini, C., Strik, H. & Boves, L. (2002) The pedagogy-technology interface in computer assisted pronunciation training. *Computer Assisted Language Learning*, 15(5): 441–467. https://doi.org/10.1076/call.15.5.441.13473

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press

*Neri, A., Mich, O., Gerosa, M. & Giuliani, D. (2008) The effectiveness of computer assisted pronunciation training for foreign language learning by children. *Computer Assisted Language Learning*, 21(5): 393–408. https://doi.org/10.1080/09588220802447651

Ngo, T. T.-N., Chen, H. H.-J. & Lai, K. K.-W. (2024) The effectiveness of automatic speech recognition in ESL/EFL pronunciation: A meta-analysis. *ReCALL*, 36(1): 4–21. https://doi.org/10.1017/S0958344023000113

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

*Qian, M., Chukharev-Hudilainen, E. & Levis, J. (2018) A system for adaptive high-variability segmental perceptual training: Implementation, effectiveness, transfer. *Language Learning & Technology*, 22(1): 69–96.

Saito, K. & Lyster, R. (2012) Effects of form-focused instruction and corrective feedback on L2 pronunciation development of /ɹ/ by Japanese learners of English. *Language Learning*, 62(2): 595–633. https://doi.org/10.1111/j.1467-9922.2011.00639.x

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

Saito, K., Suzukida, Y. & Sun, H. (2019) Aptitude, experience, and second language pronunciation proficiency development in classroom settings: A longitudinal study. *Studies in Second Language Acquisition*, 41(1): 201–225. https://doi.org/10.1017/S0272263117000432

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

*Seferoğlu, G. (2005) Improving students' pronunciation through accent reduction software. *British Journal of Educational Technology*, 36(2): 303–316. https://doi.org/10.1111/j.1467-8535.2005.00459.x

*Teeranon, P. (2020) Chinese learners learning Thai language with an application: Evidence from an acoustic study and a perception test. *Asian Journal of Education and Training*, 6(2): 330–340. https://doi.org/10.20448/journal.522.2020.62.330.340

*Tejedor-García, C., Escudero-Mancebo, D., Cámara-Arenas, E., González-Ferreras, C. & Cardeñoso-Payo, V. (2020) Assessing pronunciation improvement in students of English using a controlled computer-assisted pronunciation tool. *IEEE Transactions on Learning Technologies*, 13(2): 269–282. https://doi.org/10.1109/TLT.2020.2980261

*Thomson, R. I. (2011) Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *CALICO Journal*, 28(3): 744–765. https://doi.org/10.11139/cj.28.3.744-765

*Thomson, R. I. (2012) Improving L2 listeners' perception of English vowels: A computer-mediated approach. *Language Learning*, 62(4): 1231–1258. https://doi.org/10.1111/j.1467-9922.2012.00724.x

Thomson, R. I. & Derwing, T. M. (2015) The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics*, 36(3): 326–344. https://doi.org/10.1093/applin/amu076

*Tsai, P. (2015) Computer-assisted pronunciation learning in a collaborative context: A case study in Taiwan. *The Turkish Online Journal of Educational Technology*, 14(4): 1–13.

*Walker, N. R., Trofimovich, P., Cedergren, H. & Gatbonton, E. (2011) Using ASR technology in language training for specific purposes: A perspective from Quebec, Canada. *CALICO Journal*, 28(3): 721–743. https://doi.org/10.11139/cj.28.3.721-743

*Wang, M. & Chen, H. C. (2009) Pedagogical practice and students' perceived effectiveness of web-based automated speech evaluation. *The Journal of Asia TEFL*, 6(4): 217–243.

*Wang, X. & Munro, M. J. (2004) Computer-based training for learning English vowel contrasts. *System*, 32(4): 539–552. https://doi.org/10.1016/j.system.2004.09.011

*Yenkimaleki, M. & van Heuven, V. J. (2019) The relative contribution of computer assisted prosody training vs. instructor based prosody teaching in developing speaking skills by interpreter trainees: An experimental study. *Speech Communication*, 107: 48–57. https://doi.org/10.1016/j.specom.2019.01.006

About the authors

Moustafa Amrate is a lecturer in applied linguistics and TEFL at the Department of English, University of Biskra. He holds an MA in applied linguistics from the University of Biskra and a PhD in education from the University of York, United Kingdom. His research interests revolve around second language speech and computer-assisted pronunciation training.

Pi-hua Tsai, MA in linguistics and PhD in TESOL, brings over 34 years of English teaching experience. Her interdisciplinary research spans computer-assisted pronunciation training, discourse analysis, and medical humanities. Published in esteemed journals like *Computer Assisted Language Learning*, her work reflects her passion for educational technology and linguistic inquiry.

https://doi.org/10.1017/S0958344024000181 Published online by Cambridge University Press