Id: 1227599

Participatory Educational Research (PER)
Vol.10(3), pp. 72-88, May 2023
Available online at http://www.perjournal.com
ISSN: 2148-6123
http://dx.doi.org/10.17275/per.23.45.10.3

Computer-Assisted Pronunciation Teaching: An Analysis of Empirical Research

Behice Ceyda Cengiz*
Department of Foreign Languages and Cultures, Zonguldak Bulent Ecevit University, Zonguldak, Türkiye
ORCID: 0000-0002-9963-434X

Article history
Received:
01.12.2022

Received in revised form:
20.02.2023

Accepted:
02.03.2023

Key words:
Pronunciation; pronunciation teaching; computer- assisted pronunciation teaching

Computer-assisted pronunciation teaching has been shown to have favourable learning outcomes for language learners, which brought about the emergence of increasing body of research on this topic. Despite some attempted review studies, no systematic review of the line of the relevant literature has hitherto been observed. This study seeks to fill in this gap by reviewing the research on computer-assisted pronunciation teaching in L2 classes published between 2010 and 2021 May. 26 studies were scrutinized in terms of their contexts and participants, technology used, pronunciation aspect analysed, research design and focus, data collection tools and theoretical framework. The results showed that most of the studies were conducted with EFL/ESL learners at university level. Computer-assisted pronunciation training (CAPT) and Automatic Speech Recognition (ASR) programs were utilized more than other technological tools. Quantitative design was the dominant research method, which was followed by mixed method design while few studies were grounded in a theoretical framework. Segmental features of pronunciation received more attention. Pronunciation achievement, learner perceptions, relationship between some variables related to pronunciation learning and student interaction/ participation were the main topics of investigation. The findings suggested that teachers had a key role in promoting the effectiveness of computer-assisted pronunciation teaching by scaffolding, motivating, and guiding the learners as they use CAPT or technological tools. To this end teacher training that supports teachers in developing these competencies can be beneficial. Further research can focus on languages other than English, be conducted in primary and secondary school level and have longitudinal designs.

Introduction

Most second language (L2) researchers converge on the acknowledgment that pronunciation is a crucial subdomain of second language acquisition (Derwing et al., 2012) as it allows L2 learners to develop their communicative skills and impedes communication breakdowns (Bajorek, 2017; Gilakjani & Rahimy, 2019; Mahdi, & Al Khateeb, 2019; Morin, 2007). Pronunciation is also vital due to the globalization perspective, the requirements of "cross-cultural communication", "speaker identity and social integration" issues (Chun, 2019, p.1). Despite its vitality, pronunciation has long been overlooked in L2 research on account of

* Correspondency: behiceceydacengiz@gmail.com

some hindrances such as L2 learners' inability to reach native-like pronunciation (Scovel, 2000), insufficient materials (Busa, 2008), lack of classroom time for pronunciation teaching (Fouz-González, 2017), and the common belief that pronunciation teaching will not enhance learning of pronunciation. This perspective has also been affected by Communicative Language Teaching which has the underlying assumption that pronunciation will not develop through form-focused instruction but through exposure to language (Derwing & Munro, 2005). Since 2005, however, there has been subsequent corrective swing toward pronunciation teaching and research as reflected in the increasing number of studies published in peer-reviewed journals (e.g. Bueno-Alastuey, 2010; Engwall, 2012) as well as the emergence of L2 pronunciation conferences and of *Journal of Second Language Pronunciation* (Chun, 2019; Levis, 2016).

By virtue of the aforementioned challenges, computer-assisted pronunciation teaching has arisen as modus operandi for pronunciation instruction in the last decades (Evens & Chen, 2020; Fouz-González, 2015; Mahdi, & Al Khateeb, 2019). The technologies related to pronunciation learning include CAPT tools, websites, social network services, software, mobile devices and applications. CAPT tools are those which "detect and diagnose" learners' pronunciation errors and support learners in their efforts to correct these errors (Agarwal & Chakraborty, 2019, p.3731). There is speech visualization technology such as spectograms, waveform displays and pitch contours which have been shown to be effective for increasing L2 learners'awareness and production of different prounciation aspects including individual sounds, rhytm, intonation and prosody (Bliss, Abel, & Gick, 2018; Fouz- Gonzalez, 2019).Visual articulatory displays showing the articulation and producion of "sounds" and "words" by "speech organs" are among other CAPT tools (Chun, 2019, p.4). Another form of audio-visual articulatory displays is talking heads in which animated mouth movements help L2 learners to articulate L2 sounds (Peng, Chen, Wang, & Wang, 2018).

Finally, Automatic Speech Recognition (ASR) is one of the most commonly used CAPT technologies which have the function to decode and transcribe people's speech and provide them with pronunciation feedback (Evens & Chen, 2020; Levis & Suvorov, 2014). According to Neri, Cucchiarini, and Strik (2003) effective ASR programs feature the sequential implementation of "speech recognition", "scoring", "error detection", "error diagnosis" and "feedback presentation" (p.1158). Thus, by providing L2 learners with immediate and automatic feedback on their pronunciation mistakes that is customised to the individual learner, ASR presents a promising alternative to other CAPT technologies (Fouz- Gonzalez, 2019; Rogerson-Revell, 2021).

Several researchers have also emphasized the limitations of CAPT technologies. Fouz-González (2017) puts forward that technological aspects are at the forefront of these technologies while their pedagogical affordances being given only secondary importance. Another criticism is concerned with the hardship of making sense of the feedback provided by some of these technologies. A case in point is the spectrograms and waveforms, which are very difficult to interpret for L2 learners, thereby making them impractical tools to improve L2 pronunciation. The use of such tools, therefore, is shown to require explicit training both for L2 learners and teachers (Fouz- González, 2019). Due to the impractical nature of these technologies, ASR technology can be more preferable; however, they also have some shortcomings. Although they pinpoint specific pronunciation errors, they do not provide L2 learners with any or enough information on how to correct these errors (Levis, 2007; O'Brien, 2011; Fouz- González, 2019). Moreover, pronunciation feedback on spontaneous speech is not well-developed yet, which points at the need for more intense research on this topic

(Fouz- González, 2020).

This article surveys the literature on computer-assisted pronunciation teaching research including the studies published in established peer-reviewed L2 learning- focused journals between the years of 2010 and 2021. There have been some earlier review studies. Bliss, Abel and Bick (2018) reviewed the literature on "computer-assisted visual articulation feedback" (p. 129) and suggested that many of the reviewed studies did not follow such criteria as being replicable, having samples that are large enough, and measures that show the groups are equal at the beginning of the treatment. Mahdi and Al Khateeb (2019) conducted a meta-analysis study and showed the effectiveness of CAPT for learning pronunciation especially for beginner and intermediate level learners and university students. Similarly, Rogerson-Revell (2021) reported the evidence presented by many studies on the efficacy of using CAPT tools for pronunciation learning. In another review study Agarwall and Chakraborty (2019) classified CAPT systems as "visual simulation based systems, game based systems, comparative phonetics based systems and artificial neural network based systems" depending on the technology they used (p.3733). Of all these studies, the focus was solely on CAPT systems without including other technological tools that can help improve L2 pronunciation. The current study is broader in terms of its scope by also including such technological tools as software, websites, social network services, mobile devices and applications that can be used in pronunciation instruction. The past studies also did not present a systematic review of existent literature on computer-assisted pronunciation teaching. The current study is significant because it presents a systematic review of the studies pertaining to this topic published between 2010 and 2021 by analyzing their contexts and participants, technology, target aspects of pronunciation, research focus, research design, theoretical framework and data collection tools used in the research, which has not been done in earlier studies. By zooming into these features, this review will contribute to the identification of the research trends and different technology integrated pronunciation teaching practices by also discussing the main research findings to guide L2 researchers and practitioners for future research and good practice. To this end, the following research questions were addressed in the current study.

(1) Which research contexts and technologies have been under scrutiny in the reviewed studies?
(2) Which aspects of pronunciation have been addressed?
(3) Which theoretical frameworks underlie the studies?
(4) What research focus has been on the agenda of the articles and what are the main findings?
(5) Which research designs are utilized?
(6) Which data collection tools are employed?

Methodology

For the selection of articles, the researcher conducted an advanced search on Google Scholar in May 2021 by using the combination of the following search terms "pronunciation" OR "pronunciation learning" OR "pronunciation teaching" OR "pronunciation training" OR "pronunciation instruction" AND "computer-assisted" OR "computer-aided" OR "computer-based". The rationale behind choosing Google Scholar for article selection was that it was a "large-scale database" that is commonly utilized in the academic community due to its broad coverage of articles (Peterson, 2021, p. 9). Google Scholar was also utilized as the data search engine in many review studies published in respected journals (e.g. Li, 2018). Filters were applied in the advanced search option. The publication date was set between 2010 and 2021

and English was chosen as the language of publication. This search resulted in 1152 entries. Figure 1 below displays the selection process of the articles.

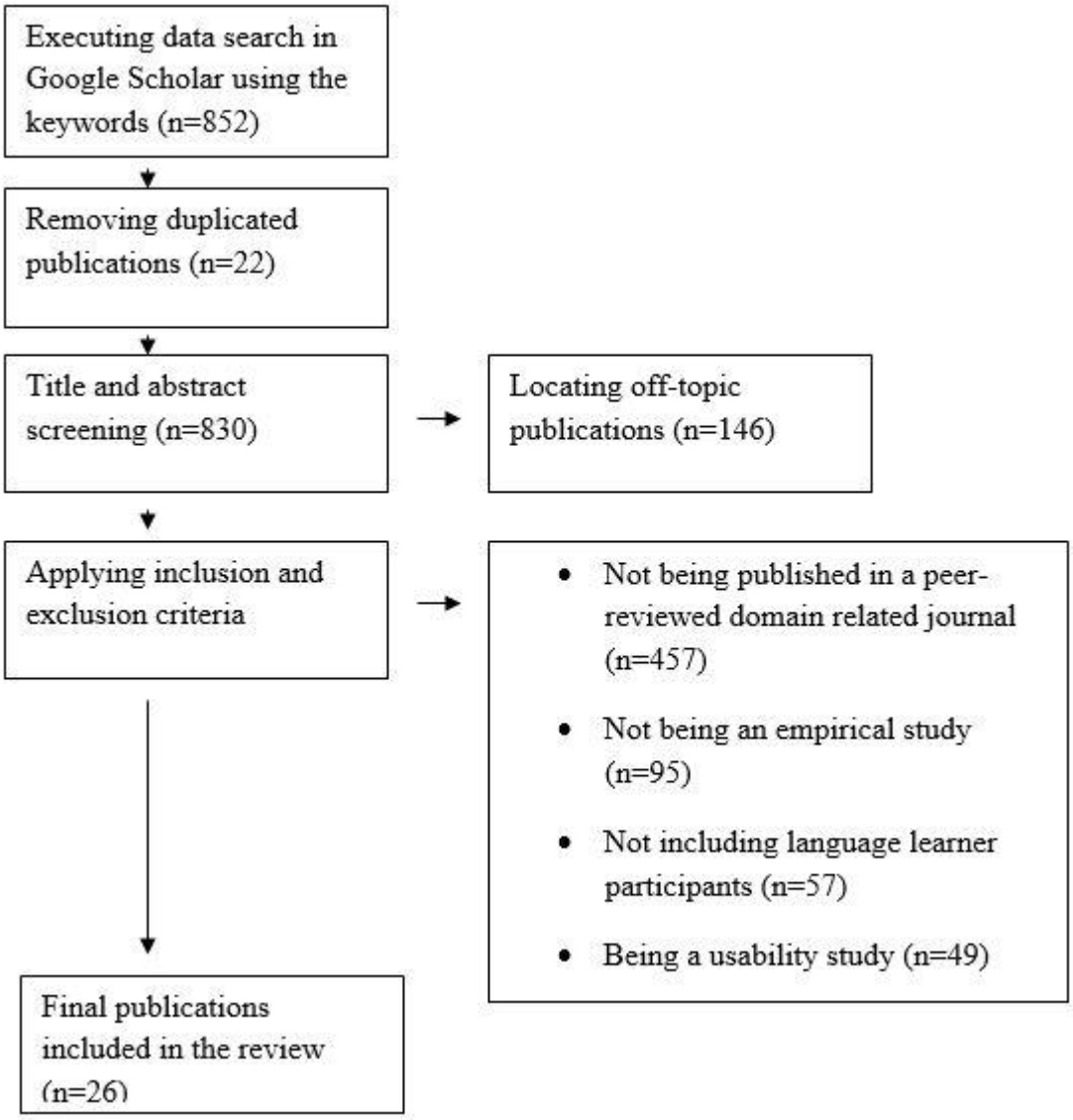

Figure 1. Article selection process.

Inclusion and exclusion criteria

After removing the duplicated publications, the researcher had a close examination of the titles and abstracts of the articles to find if they were related to computer-assisted pronunciation teaching. The publications that were related to other disciplines were excluded. Later the following inclusion and exclusion criteria were utilized to constrain the scope of the study and include high-quality publications in the review.

- The articles needed to appear in established peer-reviewed journals that centred on L2 learning research. Conference proceedings and book chapters were not included in the review. It was anticipated that reviewing the articles published in domain related and top-ranked journals would ensure the quality of the review, which is a common practise in highly cited review studies (e.g. Li, 2018)
- They needed to be an empirical research study. Review studies or conceptual papers were excluded from the current analysis.

- They needed to include language learner participants.
- The studies that evaluated the usability performances of some computer-assisted pronunciation scoring or feedback tools were not used in the review since they presented some technical knowledge and lacked a specific focus on L2 learning or teaching (e.g. Engwall, 2012; Franco, 2010; Liao, Guan, Tu,& Chen, 2014; Ouni, 2014; van Doremalen et al., 2016).

Table 1. Journal information for the reviewed articles

Journal Name	Article
Computer Assisted Language Learning	Amrate (2021); Evers & Chen, (2020); Fouz-González (2017); Hsu (2016); Liakin, Cordoso, & Liakina (2017); Luo (2016); Quintana-Lara (2014); Tsai (2019)
CALICO Journal	Barcomb & Cardoso (2020); Bueno-Alastuey (2010); Gao & Hanna (2016); Garcia, Nickolai, & Jones (2020); Liakin, Cordoso, & Liakina (2015); McCrocklin (2019a); Thomson (2011)
Language Learning & Technology	Fouz-González (2020); Martin (2020); Mompean & Fouz-González (2016); Olson (2014); Qian, Chukharev Hudilainen, & Levis (2018)
JALTCALL Journal	Elimat & AbuSeileek (2014)
Journal of Second Language Pronunciation	Foote & McDonough (2017); McCrocklin (2019b)
ReCALL	Fouz-González (2019)
System	McCrocklin (2016)
Foreign Language Annals	Martinsen, Montgomery, & Willarson (2017)

After applying these criteria, a total of 26 articles from 8 different journals were selected for the review. These journals were the following: *Computer Assisted Language Learning, CALICO Journal, Language Learning & Technology, JALTCALL Journal, Journal of Second Language Pronunciation, ReCALL, System and Foreign Language Annals.* Table 1 shows the distribution of the reviewed articles.

Data analysis

Content analysis, which is a technique widely used in social science research, was conducted to analyze the selected the studies (McMillan, 2000). For content analysis, the researcher first developed a review form, which enabled the coding of the quantitative and qualitative data from the articles. She shared the review form with a colleague holding a doctoral degree in English Language Teaching and asked her to use the form for coding 3 studies selected for the review. Two researchers, therefore, coded the data related to research contexts, technologies utilized, target pronunciation aspects, research focus, main findings, research design, theoretical framework and data collection tools as reported in the articles separately. The coding reliability was checked through percent agreement approach (Stemler, 2001). Inter-rater reliability was found to be .92, which is considered a strong agreement between raters (Cohen, 1960).

Results

The selected articles were examined in relation to the categories identified for the analysis. These categories included context and participants, technology used, pronunciation aspect(s) analyzed, research focus, research design, theoretical framework and data collection tools. Accordingly, a table was formed to make the findings more transparent. The key

information regarding the identified categories is presented in Table 2 in the Appendix while the articles are ordered according to their year of publication.

Context, participants and technology

Most of the studies were carried out at university level (n=19), while some of them were conducted at primary school (n=1) or high school context (n=3). Other studies involved adult learners from other contexts such as workplace, and language schools or information on their context was not presented. (n=3). ESL/EFL was the focus of the majority of the studies (n=20) while other studies included French as a Foreign Language (FFL) or French as a Second Language (FSL) (n=3), Spanish as a Foreign Language (SFL) (n=2), and German as a Foreign Language (GFL) (n=1). The technologies used in these studies included ASR based CAPT system (n=9) and other CAPT systems in the form of instructional software giving pronunciation training or offering speech analysis (n=5). Mobile applications (n=3), social networking service (n=2), podcast (n=1), voice recording software (n=1), LMS (n=2), videos (n=1), portable media player (n=1), presentation tool (n=1) and synchronous-voice computer-mediated communication tool (n=1) were among other technologies employed in the reviewed studies (n=1).

Research design, data collection tools and theoretical/pedagogical framework

To determine the research design of the studies, the researcher scrutinized the methodology sections and identified the research designs based on the classification of Creswell (2012). Data collection tools were categorized according to the classifications made by Peterson (2021).14 studies had quantitative designs while two studies were qualitative and 10 studies had mixed designs. Among the quantitative studies, there was one survey study while 13 studies were experimental. Depending on the state of randomization in the articles, three studies were identified as truly experimental while there were 8 quasi-experimental studies and 2 studies with repeated measures experimental design. Apart from these, two quasi experimental studies incorporated survey into its design whereas another study included both quasi experimental and correlational examination. Pre-tests and post-tests that measure the pronunciation achievement of the learners were the main data collection tools in most of the experimental studies (n=12), and these tests were also coupled with delayed post-tests in some studies (n=2). In one study (Evens & Chen, 2020), questionnaires that measured the satisfaction of learners about their pronunciation learning activities and their perceptions about the technologies' ease of use were also used as additional post-tests for a comparison between the control and experimental groups. Language learning logs, discussion board posts and surveys were also utilized as data collection tools as a complement to pre- and post- tests conducted for experimental and control groups in 2 studies (McCrocklin, 2019b; Luo, 2016).

In two qualitative studies (McCroklin, 2019a; Tsai, 2019), interviews, focus group discussions, language learning logs and a questionnaire were used as data instruments to provide a detailed qualitative account of the students' experiences with using CAPT tools. In studies with mixed method design, the general inclination was that the pre-tests and post-tests or surveys were combined with qualitative data such as questionnaire (n=4), observation (n=2), interviews (n=3), language learning logs (n=4), qualitative surveys (n=1).
It was found that few studies were grounded in a theoretical framework (n=6). Some of these frameworks centered on the interactive and socio-cultural aspects of language learning and included Interaction Hypothesis, Cooperative Learning, Collaborative Learning and Socio-Cultural Theory. Other frameworks comprised Technology Acceptance Model (n=1) and

Flege's (1995) Speech Learning and Best's (1993, 1995) Perceptual Assimilation Model (n=1).

Pronunciation aspects

Majority of the studies dealt with segmental features of pronunciation (e.g., phonemes, vowels, consonants, and so on) (n=13). Two studies focused solely on such suprasegmental features (e.g. stress, tone, intonation) while both segmental (e.g. silent letters, phoneme contrasts, and others) and suprasegmental features (e.g. pitch, intonation, linking, elision, and alike) were in the center of some studies (n=8). Three studies dealt with the learners' overall pronunciation by focusing on accuracy, accentedness, comprehensibility or fluency (n=4). In one study the target pronunciation aspect was not specified.

Research focus

The review revealed four different research strands. These were pronunciation achievement, learner perceptions, relationship between different variables affecting pronunciation learning and students' use of or interaction with the CAPT tools, which were discussed in detail below.

Pronunciation achievement

Majority of the studies (n=23) dealt with the effect of utilizing CAPT or other tools on the development of L2 learners' pronunciation. Bueno-Alastuey (2010) investigated how Spanish EFL learners' Skype interactions with different dyads affected their pronunciation achievement, phonetic errors, communication and phonetic breakdowns and phonetically modified output. The dyads included Spanish-Spanish EFL learners (Group 1), Spanish-Turkish EFL learners (Group 2) and Spanish EFL learners- American students (Group 3). At the end of 6 Skype interactions, it was found that the greatest improvement in pronunciation was seen in Group 2 followed by Group 1 pointing at the benefits of including non-native dyads in such voice chat implementations. Thomson (2011) looked into the efficacy of using a CAPT system using high variability phonetic training technique (HVPT). Through this technique, 22 Mandarin L1 EFL learners heard sounds from multiple voices through the computer, clicked on the perceived sound and got visual and auditory feedback on the accuracy of their answers. At the end of 8 training sessions over a three week period they displayed improvement in their vowel intelligibility. Similarly, Qian et al. (2018) investigated how a CAPT system with HPVT feature would benefit 32 Russian EFL learners' perception of some phonemes problematic for Russian speakers. The results demonstrated that the use of this prototype system that included exercises focusing on these phonemic contrasts led to improvement in the learners' perception of the target segmental fatures. Gao and Hanna (2016) compared the efficacy of teacher's pronunciation instruction, software giving pronunciation instruction and the combination of teacher instruction and pronunciation software on the pronunciation development of 60 high school students over a two week period. The software required the learners to work on segmental and suprasegmental features by listening to the pronunciation of some vowels and consonants, learning about methods of articulation, practising sounds, intonation or stress with some exercises, as well as recording themselves and comparing the visualizations of their recordings with samples. The results showed the combined teacher and software instruction worked best for improving learners' pronunciation.

Some studies centered on investigating the potential of speech analysis software for teaching

pronunciation. Olson (2014), for instance, reported a study in which the learners in the experimental group participated in three 20-minute sessions with *Praat* by self-recording themselves, receiving visual analysis of their recording, doing more practice and re-recording themselves. He reported that utilizing *Praat* as visual speech analysis software proved effective in learning segmental features. Similarly, Quintana-Lara's (2014) study lent strength to the value of *Praat* in teaching pronunciation by showing that the learners in the experimental group receiving *Praat* training and doing guided practice activities through this software in five sessions for two weeks improved their production accuracy of two vowels more than those in the control group.

The contributions of mobile applications to pronunciation achievement of L2 learners have also been scrutinized in some studies. Liakin et al. (2015), for example, researched the effect of an ASR-based mobile application called *Nuance Dragon Dictation* on the FFL learners' production and perception of a French vowel. During five 20-minute sessions one experimental group studied pronunciation activities on this application accompanied by immediate written visual feedback while another experimental group did the same actives in individual sessions with the teacher and got immediate oral feedback from the teacher. The control group, on the other hand, had individual meetings with the teacher to practice speaking skills but received no feedback on the target vowel. The findings suggested that although there was no difference in the perception improvement of the groups, the ASR group showed higher development in production compared to the other groups. In a subsequent study, Liakin et al. (2017) had a similar experimental design with the use of same treatments in experimental and control groups as in their earlier study in 2015. This time they examined the effect of using *Natural Reader* as a speech synthesis mobile application on the FFL learners' improvement in producing French liaison. The findings revealed that only the two experimental groups had pronunciation improvement over time. Fouz-González (2020) investigated the potential of using *English File Pronunciation* (EFP) app for providing training on the learners' perception and production of fossilised sound contrasts. By using this application in their own time for two weeks Spanish EFL learners were to study phonemic chart and play 10 games in the app every day. The training overall resulted in improvement in perception and production of many target sounds while it was not so for every sound and in every task.

Some studies looked at the pedagogical use of *Twitter* as a social networking service for teaching pronunciation. In their study Mompean and Fouz-González (2016) sent daily tweets about the pronunciation of problematic words for Spanish EFL learners by focusing both on segmental (e.g. silent letters, phonemes) and suprasegmental features (lexical stress) over 27 days. The tweets directed the learners' attention to form by giving short metalinguistic information on the pronunciation aspects that can be possibly mispronounced in the pre-selected words. Input enhancement was also used in these tweets by changing the typography of the salient words. The findings revealed positive effects of this type of instruction on pronunciation development of learners. One limitation of the study was the lack of a control group, which was addressed in a further study by Fouz-González's (2017) study. With the addition of an experimental group, it was shown that there was a significant improvement in the pronunciation achievement of the group receiving daily Tweets about target segmental and suprasegmental features, which was sustained over time.

Another research focus was the use of ASR based CAPT systems for pronunciation training. Elimart and AbuSeileek's (2014) study demonstrated that the 3rd grade EFL learners using an ASR based CAPT system called *Tell me More* for 8 weeks outperformed those learners

receiving traditional instruction. Additionally, the learners working individually with the system proved to improve more than those learners engaging in pair or group work among the three experimental groups. McCrockin (2019b) researched the efficacy of an ASR based Dictation program, *MyET* on the development of learners' pronunciation accuracy. The control group received three week pronunciation instruction in class, which contained information on articulation of sounds, "aural discrimination activities" as well as "focused listening" and "communicative practice" activities (p.106). The experimental group studied the instructional content online and did the production activities with the ASR dictation program. No significant difference was found between the two groups. The researcher suggested that this finding may be due to the fact that the participants in both groups were trained on the target features and did some production practice with these sounds one week before the study started. He also added that since these dictation programs do not provide explicit pronunciation feedback and guide for production practice, their use should be coupled with teacher guidance and monitoring. Likewise, Evers and Chen (2020) reported a study in which learners dictated a text to the ASR-based software, *Speechnotes* and worked on the misidentifed words on their own (in the control group) or with their peers (in the experimental group) over a 12 week time span. The results demonstrated significant difference in the learners' comprehensibility, accentedness and pronunciation in spontaneous speech in favor of the experimental group.

Garcia et al. (2020) carried out a 15 week study with 67 Spanish EFL learners in which the effect of teacher-led explicit pronunciation training was compared with that of ASR-based training. Both groups had pre-and post-tests for an examination of their development in comprehensibility, naiveness, fluency, and perceived confidence. The results indicated that explicit teacher instruction and ASR-based instruction benefited different aspects of pronunciation in the short or long term suggesting that combining both types of instruction can better facilitate learners' pronunciation improvement. In a more recent study, Amrate (2021) investigated the efficacy of using collaborative pronunciation activities through *Tell me More* in six sessions for teaching prosody and presented favourable results in terms of improvement in learners' prosodic accuracy and comprehensibility.

Another bunch of studies examined the integration of a learning management system (LMS), videos, ipods, podcasts or other computer programs into pronunciation instruction. In Luo's (2016) experimental study in-class pronunciation instruction was combined with some pronunciation activities in *Blackboard* as an LMS. These activities comprised reading a text out loud, recording oneself, comparing one's own recording with that of a native speaker and finally posting the recording to *Blackboard*. The students later provided peer feedback to the recordings. The findings demonstrated that this implementation in the experimental group resulted in better pronunciation improvement than the only in-class instruction. Similarly, Barcomb and Cardoso (2020) had a two-week gamified course giving pronunciation training in *Moodle*. The target sounds were English /r/ and /l/ segments problematic for Japanese EFL learners. The course included such gamifying elements as earning points and budgets with the completion of certain missions and leader boards. The missions contained watching the videos that gave explicit instruction on the pronunciation of the target features, doing production activities and a minimal-pair listening quiz. This one-group quasi-experimental study revealed that the gamified course resulted in increased awareness about the target segments and improvement in pronunciation. In Martin's (2020) study the focus was on examining the effect of using innovative cued pronunciation readings (ICPR) as a CAPT method in a distance language course over a 10-week period on the learners' perception and production of target segmental and suprasegmental features. This method also incorporating

HVPT technique presented information on these features and included perception and production exercises. The results suggested that the experimental group receiving ICPR-based pronunciation training had significantly higher improvement in pronunciation than the control group.

The possible pedagogical affordances of shadowing was investigated in some studies. Foote and McDonough (2017), for example, implemented shadowing activities with the use of *iPod* over a 8 week period. 16 ESL learners had pronunciation practice by shadowing short dialogues for a minimum of 10 minutes every week and were assessed on the comprehensibility, accentedness and fluency of their speech as well as on their ability to imitate the speech model at the beginning, middle and end of the study. It was shown that learners had improvement in all these aspects except for accentedness. There was no control group in this study. Therefore, although the researcher acknowledged the value of shadowing activities for developing L2 pronunciation, he also exercised caution not to consider technology enhanced shadowing activities as a substitute for in-class pronunciation instruction. Similar to the afore-mentioned study, Martinsen et al. (2017) demonstrated that utilizing subtitled videos in French for shadowing and tracking activities both in-class and in a computer lab with 19 FFL learners led to pronunciation development. However, the lack of a control group was also a limitation in this study.

Utilizing podcasts as a complement to in-class pronunciation activities received attention in Fouz-González's (2019) study in which 47 Spanish EFL learners acted as both control and experimental groups. Pronunciation training activities included the following: (1) receiving in-class explicit instruction on some phonemic contrasts (2) listening to podcasts and doing such podcast-based activities as finding the words including target sounds and classifying the target sounds and (3) recording a text containing these sounds and uploading the recordings to Edmodo for peer feedback. Overall this implementation, which lasted for three weeks, resulted in significant improvements in the learners' perception and production of target segmental features. However, the training did not have the same effect on "every sound in every task" (p.1), which was attributed to the link between the target features and the participants' native language.

Learner perceptions

The reviewed studies overall reported positive learner perceptions about having computer-assisted pronunciation instruction (e.g. Foote & McDonough, 2017; Martinsen et al., 2017). It was shown that learners generally liked working with CAPT or other technological tools for improving their pronunciation (Evers & Chen, 2020; Tsai, 2019). They also believed that doing pronunciation practice with these tools was effective for the development of their pronunciation (Foote & McDonough, 2017; Fouz-González, 2017; Gao & Hanna, 2016; Martin, 2020) and even for improving their speaking skills (Barcomb & Cardoso, 2020).

It was stated by many learners that they benefited greatly from doing extensive pronunciation practice with these tools and developed an increased awareness about different pronunciation features (McCrocklin, 2019a; Tsai, 2019). They highly appreciated getting feedback on their pronunciation (Mompean & Fouz-González, 2016) especially in the absence of other people, which they found quite relaxing (McCrocklin, 2016). Pronunciation tips and metalinguistic explanations provided by these tools or programs were also found useful (Martin, 2020; Tsai, 2019). Although some learners favored working with these tools in a self-paced and individualized manner (Martinsen et al., 2017), the collaborative work integrated into the pronunciation tasks in some studies were also valued by some learners. For example, learners

in Luo's (2016) study indicated that they found it helpful to listen to their peers' recordings and compare them to their own recordings. Evers and Chen (2020) showed that the learners engaged in ASR-based pronunciation activities coupled with peer feedback had higher levels of satisfaction than those learners doing these activities alone without any peer feedback. In a similar vein, Tsai (2019) also suggested that peer scaffolding, guidance and support can be beneficial when learners face technological difficulties and need human assistance while using CAPT or other technological tools.

When videos were integrated into pronunciation instruction, the learners cared about the authenticity of the videos (Martinsen et al., 2017) and enjoyed having the option to pause the videos or manipulating their speed (Barcomb & Cardoso, 2020; Martinsen et al., 2017; McCorcklin 2016). In Barcomb and Cardoso's (2020) study, the gamification elements integrated into a distance language course were also enjoyed by the learners (Barcomb & Cardoso, 2020)

Negative learner perceptions were mostly pertinent to some challenges learners faced while receiving computer-assisted pronunciation instruction. Recognition failures of the CAPT or ASR programs were commonly reported problems, which resulted in learners' frustration with these programs (McCrocklin, 2016, 2019a). Accordingly, some learners also had doubts about the reliability of the scoring system of these programs and were unsatisfied with these programs due to the "invariable feedback" they provided (Tsai, 2019, p.732). In some studies, the learners seemed to develop some strategies to deal with low recognition rates of these programs. McCrocklin (2016, 2019a), for example, indicated that some learners used the target words in phrases or sentences in order for the ASR program to better recognize these words. Other learners employed e-dictionaries for checking the pronunciation of words before recording themselves for a better recognition rate.

Many studies reported that learner perceptions of the activities and the efficiency of computer-assisted pronunciation activities (Foote & McDonough, 2017; Martinsen et al., 2017; McCrocklin, 2019a) became more positive over time. It was shown that some learners initially found it difficult or frustrating to use the CAPT or ASR programs due to some technological problems (McCrocklin, 2019a). However, as they got more familiar with these programs for pronunciation practice, they were more satisfied with using them at the end of the study (Martinsen et al., 2017). Some learners pointed out that it was not convenient for them to use the computer for pronunciation practice all the time and suggested that mobile tools would be more practical and preferable (McCrocklin, 2019a). The lack of teacher feedback was also noted as a problem in some studies (Martin, 2020). Although some learners stated that they benefited from peer feedback, others indicated their need for teacher feedback since peer feedback was not always reliable or helpful for them (Luo, 2016; Martin, 2020). Regarding this issue, Tsai (2019) put forward that CAPT software and peer interaction can be incorporated into the different stages of pronunciation instruction. That is, the learners can work with the CAPT program individually for some time and later do some collaborative work and learn some strategies from their peers, which might better prepare them for independent practice. Peers can also be given guidance on how to scaffold each other and collaborate more efficiently. In this way teacher can have more time to deal with individual pronunciation needs of the learners.

Relationship between some variables related to pronunciation learning

In some studies which integrated CAPT or other technological tools into pronunciation instruction one research focus centered on examining the relationship between different variables affecting pronunciation learning. In Elimat and AbuSileek's (2014) study, in which the researchers tested out the effectiveness of an ASR-based CAPT system, they also investigated if a correlation existed between the level of utterance ("word, sentence, and real-life dialogue") and pronunciation performance of the experimental group in the post-test but found no significant correlation (p.21). Gao and Hanna (2020) employed a CAPT system with 60 high school students with a focus on the improvement of suprasegmental features. They aimed to find out the relationship between learners' attitudes towards pronunciation learning and their pronunciation scores, which was confirmed in their study. Hsu (2016) reported using a virtual teacher as an ASR-based CAPT system for teaching pronunciation and investigated if there was a relationship between the learners' perceptual learning styles and levels of acceptance towards this system. No significant relationship was found between perceptual learning style and perceived usefulness while there was a significant relationship between perceived usefulness and learners' attitudes towards using the CAPT system. Of the learning styles, visual and kinaesthetic styles were significantly related with perceived ease of use, which in turn resulted in perceived usefulness. Finally, Fouz Gonzalez (2017) delved into the relationship between the learners' engagement in Twitter- based pronunciation instruction and their pronunciation gains and found a positive correlation, albeit a weak one. Although learners with high improvement scores generally had high engagement scores, there were also some learners with low improvement scores and high engagement scores. He also showed that learners' punctuality in reading the tweets also did not affect their improvement scores significantly.

Student interaction and/or participation in CAPT activities

Another body of research was related to an examination of the interaction taking place during learners' use of CAPT tools. Tsai (2019), for instance, conducted a qualitative study in which *MyET*, an ASR-based CAPT system was used over a period of 10 weeks with two groups of learners, namely the collaborative CAPT group (with peers) and the group consisting of learners working individually with the CAPT system (without peers). After spending 20 minutes with the system, the learners were supposed to write down their reflections on their pronunciation learning. Both groups reported more positive than negative comments about using this system. They were glad of receiving pronunciation feedback and doing more practice. When the learner interactions in the collaborative CAPT group were analyzed, two collaborative patterns were identified, which were "peer as a teacher" and "peer with equal status" (p.726). The former pattern arose when some students acted as a teacher by reading the given utterance, asking their peers to repeat after them a few times, directing their attention to their errors and providing corrections. The latter was observed when the peers acted as equal partners by pointing out one another's errors and providing feedback on these errors for pronunciation improvement. These interaction patterns had different psychological effects on the learners. While the first pattern led to feelings of inferiority on part of some learners acting as students, the second pattern tended to motivate peers better for the collaborative work.

Amrate (2021) had a similar study in which *Tell me More*, an ASR-based CAPT system was utilized either collaboratively or individually by two group of learners. The amount and type of support required in these groups was investigated. The video recordings of the CAPT sessions suggested that the collaborative CAPT group were able to work more independently

and in a more fun way with less help and guidance needed from the teacher. However, the type of support needed by the learners varied depending on the type of challenges. Peer help worked well for technical problems while it was the teacher providing this kind of support in the group working with CAPT individually. For technical problems and support both groups needed teacher's help. When the challenges were non-technical, the peers served to clarify the activities' instructions and provided their suggestions on each other's productions. Management of "the practice setup" was an area in which the learners in both groups asked for teacher support but this was more evident in the individual CAPT group (p.22).

Discussion and Conclusion

This study aimed at reviewing the empirical research on computer-assisted pronunciation teaching and portraying the research trends on this topic. In this review computer-assisted pronunciation teaching was found to be effective in improving L2 learners' pronunciation as reported in various experimental studies. However, it appears that there are some important considerations for their optimal potential to be realized in L2 classes.

One important consideration is pertinent to the key role of the teacher in the process of integrating technology into pronunciation instruction. As the research revealed, some CAPT programs have some technical limitations and lead to difficulties on part of the learners, which result in lowering learner engagement and motivation. These limitations include but not limited to these programs' low recognition rates and the lack of explicit pronunciation they provide for L2 learners (McCrocklin, 2019a; McCrocklin, 2019b), which shows that there is still room for development of these software. Learners' lack of familiarity with these programs and lack of trust felt towards their reliability also act as impediments to their effective use. To this end, as a common theme arising in many studies, it is significant for L2 teachers to provide scaffolding and guidance for their learners when incorporating these programs into their classes. They also need to motivate their learners by informing them of these programs' benefits for pronunciation development and training them for some time to develop a sense of familiarity of these programs. Use of the ASR programs or other CAPT programs that provide L2 learners with feedback on their pronunciation can be combined with teacher feedback to better facilitate pronunciation development and to compensate for the weaknesses of these programs (Garcia et al., 2020; Qian et al., 2018).

To fulfill these teacher roles outlined above, L2 teachers also need some training to develop the competence to critically evaluate CAPT programs (Thomson, 2011). In this way they can develop a better understanding of the strengths and weaknesses of these programs and find creative and ideal ways of using them with their students. A collaboration between the teachers, researchers and programmers can also prove useful for designing CAPT programs ideal for learners (Thomson, 2011).

Many experimental studies also demonstrated that learners' collaborative use of CAPT programs worked well for pronunciation improvement. Hence, it can be argued that L2 teachers can integrate collaborative work while integrating these programs into their instruction. However, as discussed earlier, the peers also need to be provided guidance on how to use these programs and collaborate effectively. Learners can still seek for teacher feedback while working with their peers; therefore, the teacher should monitor their progress and provide support when needed (Amrate, 2021). To this end, it can be ideal to design a pedagogical framework that provide learners with instructions on applying collaboration strategies for effective use of CAPT programs with peers (Tsai, 2019).

An analysis of the studies suggested that university setting was the dominant context of research and EFL/ESL was the most commonly researched target language. This finding indicated the need for more research to be conducted in other contexts (e.g. primary and secondary school context) and on different languages other than English. It was also displayed that the majority of the studies were quantitative as followed by the studies with mixed method design while there were few qualitative studies. In many studies there was a lack of focus on to what extent improvements in the perception or production aspects of pronunciation were generalized "beyond the level of isolated words" (Qian et al., 2018, p. 87), which requires more attention in future studies.

The effect of task type and intervention duration on pronunciation learning can also be examined in future studies. In many experimental studies, the lack of delayed post-tests and the lack of control group were identified as shortcomings by the researchers, which needs to be taken into consideration in future studies. The short duration of the treatments can also be seen as a weakness of the research studies. In order to observe the long term effects of the treatment on the learners, longitudinal studies can be conducted in the future. It was also seen that the majority of the studies were not grounded in any theoretical framework, which is again a food for thought for researchers. It is suggested that linking computer-assisted pronunciation teaching research with theories of second language acquisition or theories of learning can better help to build up the theoretical foundation of this field for future research.
The study had some limitations. Due to space limits, the current study only reviewed 26 studies published in peer-reviewed journal articles. Inclusion of book chapters, and conference proceedings as well conducting the search in other databases (e.g., ERIC, Scopus) in future studies would help to provide a more comprehensive review. It is also considered that it would be better not to limit the language of publications to English but to include other languages.

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. https://doi.org/10.1007/s10639-019-09955-7

*Amrate, M. (2021). Collaborative vs. individual computer-assisted prosody training: a mixed-method case study with Algerian EFL undergraduates. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2021.1882503

Bajorek J.P (2017). L2 pronunciation in CALL: The unrealized potential of Rosetta Stone, Duolingo, Babbel, and Mango Languages. *Issues and Trends in Educational Technology* *5*(1), 24–51. Retrieved from https://journals.librarypublishing.arizona.edu/itlt/article/id/1497/

*Barcomb, M., & Cardoso, W. (2020). Rock or lock? Gamifying an online course management system for pronunciation instruction: Focus on English /r/ and /l/. *CALICO Journal*, *37*(2), 127-147. https://doi.org/10.1558/cj.36996

Bliss, H., Abel, J. & Gick, B. (2018). Computer-assisted visual articulation feedback in L2 pronunciation instruction: a review. *Journal of Second Language Pronunciation 4*, 129-153. https://doi.org/10.1075/jslp.00006.bli

*Bueno-Alastuey, C. (2010). Synchronous-voice computer-mediated communication: effects on Ppronunciation. *CALICO Journal, 28*(1), 1-20. https://doi.org/ 10.11139/cj.28.1.1-20

Busa, M. G. (2008). New perspectives in teaching pronunciation. In A. Baldry,

M. Pavesi, C. Taylor Torsello and C. Taylor (eds), *From DIDACTAS to ECOLINGUA. An ongoing research project on translation and corpus linguistics* (pp. 165–82). Trieste, Italy: Universita degli Studi di Trieste.

Chun, D. (2019). Computer-assisted pronunciation teaching.In Carol A. Chapelle (Eds.), *The Encyclopedia of Applied Linguistics (*pp.1-11). John Wiley & Sons.

Cohen, J. (1960). A coefficient of agreement for nominal scales. *Educational and Psychological Measurement*, *20*, 37-46. https://doi.org/10.1177/001316446002000104

Creswell, J. W. (2012). *Educational research: Planning, conducting, and evaluating quantitative and qualitative research* (4th ed.). Boston, MA: Pearson.

Derwing, T. M., & Munro, M. J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*, 379–397. https://doi.org/10.2307/3588486

Derwing, T.M., Diepenbroek, L. G., & Foote, J. A. (2012). How well do general-skills ESL textbooks address pronunciation? *TESL Canada Journal*, *30*(1), 22–44. https://doi.org/10.18806/tesl.v30i1.1124.

*Elimat, A. K. and AbuSeileek, A. F. (2014). Automatic speech recognition technology as an effective means for teaching pronunciation. *JALT CALL Journal*, 10(1), 21–47. Retrieved from https://eric.ed.gov/?id=EJ1107929

Engwall, O. (2012). Analysis of and feedback on phonetic features in pronunciation training with a virtual teacher. *Computer Assisted Language Learning*, *25*(1), 37-64, https://doi.org/10.1080/09588221.2011.582845

*Evers, K.,& Chen, S. (2020): Effects of an automatic speech recognition system with peer feedback on pronunciation instruction for adults. *Computer Assisted Language Learning*. https://doi.org/10.1080/09588221.2020.1839504

*Foote, J.A., & McDonough, K. (2017). Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation*, *3*(1) 34–56. https://doi.org/10.1075/jslp.3.1.02foo

Fouz-González, J. A. (2015). Trends and directions in computer-assisted pronunciation training. In M. A. Mompean & J. Fouz-Gonza´lez, J. (Eds.), *Investigating English pronunciation: Trends and directions* (pp. 314–42). Basingstoke, England: Palgrave Macmillan.

*Fouz-González, J. (2017). Pronunciation instruction through Twitter: The case of commonly mispronounced words. *Computer Assisted Language Learning 30*(7), 631–663. https://doi.org/10.1080/09588221.2017.1340309

*Fouz-González, J. (2019). Podcast-based pronunciation training: Enhancing FL learners' perception and production of fossilised segmental features. *ReCALL*, *31*(2), 150-169, https://doi.org/10.1017/S0958344018000174

*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

Franco, H., Bratt, H., Rossier, R., Rao Gadde, V., Shriberg, E., Abrash, V., & Precoda, K. (2010). EduSpeak®: A speech recognition and pronunciation scoring toolkit for computer-aided language learning applications. *Language Testing*, *27*(3), 401–418. https://doi.org/10.1177/0265532210364408

*Gao, Y., & Hanna, B. E. (2016). Exploring optimal pronunciation teaching: Integrating instructional software into intermediate-level EFL classes in China. *CALICO Journal*, *33*(2), 201–30. https://doi.org/10.1558/cj.v33i2.26054

*Garcia, C.,Nickolai,D., & Jones, L. (2020). Traditional versus ASR-based pronunciation instruction : An empirical study. *Computer Assisted Language Learning*, *37*(3), 213–232. https://doi.org/10.1558/cj.40379

Gilakjani, A., & Rahimy, R. (2019). Using computer-assisted pronunciation teaching (CAPT) in English pronunciation instruction: A study on the impact and the teacher's role.

Education and Information Technologies, *2*5(2), 1129–1159. https://doi.org /101007/s10639-019-10009-1

*Hsu, L. (2016) An empirical examination of EFL learners' perceptual learning styles and acceptance of ASR-based computer-assisted pronunciation training. *Computer Assisted Language Learning, 29*(5), 881-900. https://doi.org/ 10.1080/09588221.2015.1069747.

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

Levis, J. (2016). The interaction of research and pedagogy. *Journal of Second Language Pronunciation, 2*(1), 1-7. [editorial]. https://doi.org/10.1075/jslp.2.1.001lev

Levis, J., & Suvorov, R. (2014). Automated speech recognition. In C. Chapelle (Ed.*), The encyclopedia of applied linguistics*. http://onlinelibrary.wiley.com/.

Li, M. (2018). Computer-mediated collaborative writing in L2 contexts: an analysis of empirical research. *Computer Assisted Language Learning, 31*(8), 882–904. https://doi.org/ 10.1080/09588221.2018.1465981

*Liakin, D., Cardoso, W. and Liakina, N. (2015). Learning L2 pronunciation with a mobile speech recognizer: French /y/. *CALICO Journal*, 32(1), 1–25. https://doi.org/10.1558/cj.v32i1.25962

*Liakin, D., Cardoso, W. & Liakina, N. (2017) The pedagogical use of mobile speech synthesis (TTS): focus on French liaison. *Computer Assisted Language Learning, 30*(3-4), 325-342. https://doi.org/10.1080/09588221.2017.1312463

Liao, H-C., Guan, Y-H., Tu, J-J., & Chen, J-C. (2014). A prototype of an adaptive Chinese pronunciation training. *System. 45*(1), 52-66. http://dx.doi.org/10.1016/j.system.2014.04.006

*Luo, B. (2016). Evaluating a computer-assisted pronunciation training (CAPT) technique for efficient classroom instruction, *Computer Assisted Language Learning*, *29*(3), 451–476. https://doi.org/10.1080/09588221.2014.963123.

Mahdi, H.S., & Al Khateeb, A.A. (2019). Context and Implications Document for: The effectiveness of computer-assisted pronunciation training: A meta-analysis. *Review of Education, 7*(3), 754–756. https://doi.org/10.1002/rev3.3165

*Martin, I.A. (2020). Pronunciation development and instruction in distance language learning. *Language Learning & Technology*, *24*(1), 86-106. https://doi.org/10125/44711

*Martinsen,R., Montgomery, C.,, & Willardson, V. (2017). The effectiveness of video-based shadowing and tracking pronunciation exercises for foreign language learners. *Foreign Language Annals, 50*(4), 661–680. https://doi.org/10.1111/flan.12306

*McCrocklin, S. (2016). Pronunciation learner autonomy: The potential of Automatic Speech Recognition. *System, 57*, 25-42. http://dx.doi.org/10.1016/j.system.2015.12.013

*McCrocklin, S. (2019a). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation, 5*(1), 98–118. https://doi.org/10.1075/jslp.16034.mcc

*McCrocklin, S. (2019b). ASR-based dictation practice for second language pronunciation improvement. *Journal of Second Language Pronunciation 5*(1), 98–118. https://doi.org/10.1075/jslp.16034.mcc

McMillan, J. H. (2000). *Educational research: Fundamentals for the consumer* (3rd ed.). New York, NY: Harper Collins.

*Mompean, J. A, 6, Fouz-Gonzalez, J. (2016). Twitter-based EFL pronunciation instruction. *Language Learning and Technology*, *20*(1), 166-190. Retrieved from http://llt.msu.edu/issues/february2016/mompeanfouzgonzalez.pdf

Morin, R. (2007). A neglected aspect of the standards: Preparing foreign language Spanish teachers to teach Ppronunciation. *Foreign Language Annals, 40* (2), 342-260. Retrieved from https://www.actfl.org/i4a/pages/index.cfm?pageid=3320

Neri, A. Cucchiarini, C., & Strik, H. (2003). Automatic Speech Recognition for second language learning: How and why it actually works. In *Proceedings of 15th International Congress of Phonetic Sciences*, pp. 1157–1160.

O'Brien, M. G. (2011). Teaching and assessing pronunciation with computer technology. In N. Arnold & L. Ducate (Eds.), *Present and Future Promises of CALL: From Theory and Research to New Directions in Language Teaching* (2nd ed.) (pp. 375–406). San Marcos, Texas: CALICO Monograph Series, 375-406.

*Olson, D. J. (2014). Benefits of visual feedback on segmental production in the L2 classroom. *Language Learning & Technology*, *18*(3), 173–92. Retrieved from http://llt.msu.edu/issues/october2014/olson.pdf

Ouni, S. (2014) Tongue control and its implication in pronunciation training. *Computer Assisted Language Learning*, *27*(5), 439-453. https://doi.org/10.1080/09588221.2012.761637

Peng, X., Chen, H., Wang, L., & Wang, H. (2018). Evaluating a 3-D virtual talking head on pronunciation learning. *International Journal of Human-Computer Studies, 109*, 26–40. https://doi.org/10.1016/j.ijhcs.2017.08.001

Peterson, M. (2021). Digital simulation games in CALL: a research review. *Computer Assisted Language Learning.* https://doi.org/10.1080/09588221.2021.1954954

*Qian, M., Chukharev-Hudalainen, E., & Levis, J. (2018). A system for adaptive high-variability segmental-perceptual training: Implementation, effectiveness, and transfer. *Language Learning and Technology*, *22*, 69–96. https://doi.org/10125/44582

*Quintana-Lara, M. (2014) Effect of Acoustic Spectrographic Instruction on production of English /i/ and /I/ by Spanish pre-service English teachers, *Computer Assisted Language Learning*, *27*(3), 207-227. https://doi.org/10.1080/09588221.2012.724424

Rogerson-Revell, P. (2021). Computer-assisted pronunciation training (CAPT): Current Issues and Future Directions. *RELC Journal*, *52*(1), 189-205. https://doi.org/10.1177/003368822097740

Scovel, T. (2000). A critical review of the critical period research. *Annual Review of Applied Linguistics*, *20*, 213-223. https://doi.org/10.1017/S0267190500200135

Stemler, S. (2001). An overview of content analysis. *Practical Assessment, Research and Evaluation*, *7*(17), 137-146. https://doi.org/10.7275/z6fm-2e34

*Thomson, R.I. (2011). Computer- assisted pronunciation training: Target-ing second language vowel perception improves pronunciation. *CALICO Journal*, *28*(3), 744-765. Retrieved from https://www.jstor.org/stable/calicojournal.28.3.744

*Tsai, P. (2019). Beyond self-directed computer-assisted pronunciation learning: A qualitative investigation of a collaborative approach. *Computer Assisted Language Learning 32*(7), 713-744. https://doi.org/10.1080/09588221.2019. 1614069

van Doremalen, J., Boves, L., Colpaert, J., Cucchiarini, C., & Strik, H. (2016). Evaluating automatic speech recognition-based language learning systems: A case study. *Computer Assisted Language Learning*, *29*(4), 833–851. https://doi.org/10.1080/09588221.2016.1167090