

Review

A Systematic Review of Empirical Mobile-Assisted Pronunciation Studies through a Perception–Production Lens

Anne M. Stoughton * and Okim Kang

Department of English, Northern Arizona University, Flagstaff, AZ 86011, USA; okim.kang@nau.edu

* Correspondence: ams3282@nau.edu

Abstract: The communicative approach to language learning, a teaching method commonly used in second language (L2) classrooms, places little to no emphasis on pronunciation training. As a result, mobile-assisted pronunciation training (MAPT) platforms provide an alternative to classroom-based pronunciation training. To date, there have been several meta-analyses and systematic reviews of mobile-assisted language learning (MALL) studies, but only a few of these meta-analyses have concentrated on pronunciation. To better understand MAPT's impact on L2 learners' perceptions and production of targeted pronunciation features, this study conducted a systematic review of the MAPT literature following PRISMA 2020 guidelines. Potential mobile-assisted articles were identified through searches of the ERIC, Educational Full Text, Linguistics and Language Behavior Abstract, MLI International, and Scopus databases and specific journal searches. Criteria for article inclusion in this study included the following: the article must be a peer-reviewed empirical or quasi-empirical research study using both experimental and control groups to assess the impact of pronunciation training. Pronunciation training must have been conducted via MALL or MAPT technologies, and the studies must have been published between 2014 and 2024. A total of 232 papers were identified; however, only ten articles with a total of 524 participants met the established criteria. Data pertaining to the participants used in the study (nationality and education level), the MPAT applications and platforms used, the pronunciation features targeted, the concentration on perception and/or production of these features, and the methods used for training and assessments were collected and discussed. Effect sizes using Cohen's *d* were also calculated for each study. The findings of this review reveal that only two of the articles assessed the impact of MAPT on L2 learners' perceptions of targeted features, with results indicating that the use of MPAT did not significantly improve L2 learners' abilities to perceive segmental features. In terms of production, all ten articles assessed MPAT's impact on L2 learners' production of the targeted features. The results of these assessments varied greatly, with some studies indicating a significant and large effect of MAPT and others citing non-significant gains and negligible effect sizes. The variation in these results, in addition to differences in the types of participants, the targeted pronunciation features, and MAPT apps and platforms used, makes it difficult to conclude that MAPT has a significant impact on L2 learners' production. Furthermore, the selected studies' concentration on mostly segmental features (i.e., phoneme and word pronunciation) is likely to have had only a limited impact on participants' intelligibility. This paper provides suggestions for further MAPT research, including increased emphasis on suprasegmental features and perception assessments, to further our understanding of the effectiveness of MAPT for pronunciation training.



Citation: Stoughton, Anne M., and Okim Kang. 2024. A Systematic Review of Empirical Mobile-Assisted Pronunciation Studies through a Perception–Production Lens.

Languages 9: 251. <https://doi.org/10.3390/languages9070251>

Academic Editors: Lucrecia Rallo Fabra, Joan C. Mora and Ying Ying Tan

Received: 14 November 2023

Revised: 14 June 2024

Accepted: 3 July 2024

Published: 16 July 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Keywords: mobile-assisted pronunciation training; mobile-assisted language learning; MALL; MAPT; perception; production

1. Introduction

Pronunciation is an often overlooked, yet key component, of second language (L2) learning. L2 learners with good pronunciation skills are frequently considered to be intelligible even with deficiencies in the target language (Nair et al. 2017), while L2 learners with

strong vocabularies and grammar skills are deemed unintelligible when their pronunciation falls below a certain threshold (Hinoftis and Bailey 1980). Research suggests that L2 pronunciation can be improved through explicit pronunciation training. Studies have shown that, through such training, L2 learners have been able to improve their perception and production of L2 sounds, enhance the intelligibility of their L2 speech, and reduce their accentedness (Flege 1995, 2003; Lee et al. 2020; Pourhossein Gilakjani 2017; Sakai and Moorman 2018).

Despite the efficacy of pronunciation training, such training is often de-emphasized in language programs using the communicative approach to language learning. In these programs, little to no class time is allocated to pronunciation training (Meisarah 2020). Many language instructors also lack pronunciation teaching skills (Nair et al. 2017) or deem pronunciation teaching to be ineffective or inappropriate for their students (Pourhossein Gilakjani 2017). As a result, computer- and mobile-assisted pronunciation applications, which provide ample pronunciation input and real-time feedback, are providing an alternative, both inside and outside of the classroom, to enhance L2 learners' pronunciation skills (Arashnia and Shahrokhi 2016, p. 151; Meisarah 2020; O'Brien et al. 2018). As smartphones have become ubiquitous and are considered more flexible and affordable than computers, many L2 learners are using mobile-assisted language learning (MALL) and pronunciation training (MAPT) applications and platforms for pronunciation training. Although MALL is commonly used in the literature and includes mobile-assisted pronunciation training, this study will use MAPT to refer specifically to any mobile-assisted application and platform that provides pronunciation training; MALL will be used to refer to more encompassing mobile-assisted language learning. This paper specifically focuses on pronunciation training via MAPT to determine the technologies' effectiveness in improving L2 learners' productive and receptive skills.

With the rise of mobile-assisted applications for pronunciation and overall language learning, interest in both MALL and MAPT research has been increasing. Since 2015, numerous meta-analyses and systematic reviews have been conducted on existing MALL and MAPT research. These analyses and reviews have focused on the effectiveness of MALL apps (Burston 2015; Elaish et al. 2023; Tommerdahl et al. 2022), as well as MALL's impact on vocabulary learning, student achievement and learning performance (Cho et al. 2018; Sung et al. 2015), game-based learning (Su et al. 2021; Nitisakunwut and Hwang 2023), and speaking skills (Li 2024). To date, however, there has only been one meta-analysis (Tseng et al. 2022) and one systematic review (Metruk 2024) conducted on MAPT's impact on pronunciation training. Tseng et al. (2022) reviewed 13 empirical studies from 2009 to 2020, finding that mobile-based pronunciation training had a significant effect ($d = 0.66$) on L2 learners' pronunciation. Metruk (2024), in his review of 15 MAPT articles, found that smartphones were the most commonly used mobile device for MAPT learning and that, while MAPT applications did improve L2 learners' pronunciation, these applications were often not grounded in pedagogical theory and failed to provide sufficient feedback to their users (p. 22).

Further research is needed to better understand MAPT's impact on pronunciation. Specifically, while the studies in Tseng et al.'s (2022) meta-analysis demonstrated a positive impact of MAPT on pronunciation, there is a lack of synthesized research on MAPT's impact on L2 learners' perception and production of segmental and suprasegmental features. That is, no study to date has systematically reviewed the existing empirical and quasi-empirical studies to analyze MAPT's impact on L2 learners' abilities to perceive and/or produce consonants, vowels, word stress, sentence stress, and other pronunciation features. As perception and production interact to promote language acquisition and suprasegmental features have a greater impact on L2 learners' intelligibility and comprehensibility than segmental features, such a review would allow us to better understand how MAPT impacts these important constructs. To fill this gap, this paper follows PRISMA 2020 guidelines to identify and review ten MAPT-based pronunciation studies, focusing on the studies' impacts on L2 learners' perception and production gains. This paper also analyzes the types

of participants selected, the pronunciation features focused on, and the assessments used within each study. MAPT impacts on L2 learners' speech intelligibility, comprehensibility, and accentedness (subconstructs of production) are also discussed. This paper first defines the speech constructs of intelligibility, comprehensibility, and accentedness and then defines the concepts of perception and production. Next, a brief history of MALL and MAPT research is provided, followed by a review of the ten empirical MAPT-based pronunciation studies. Gaps within the MAPT research are identified and discussed, and future directions for MAPT-based pronunciation research are provided.

2. Literature Review

2.1. Intelligibility, Comprehensibility, and Accentedness

As this systematic review is concentrated on mobile pronunciation applications, an understanding of the major constructs surrounding pronunciation is required. Within pronunciation research and instruction, pronunciation training is commonly targeted at global or specific aspects of pronunciation. Global aspects incorporate listeners' perceptions of speakers' overall speaking performances; specific aspects target the segmental (vowel and consonants) and suprasegmental (e.g., word stress, sentence stress, and intonation) features that impact overall speech performance (Saito and Plonsky 2019). Derwing and Munro (2015) suggest that three pronunciation constructs impact listeners' perceptions of speech including the following: the intelligibility, comprehensibility, and accentedness of the speech. Intelligibility is a listener's ability to understand a speaker's intended message (Derwing and Munro 2015). Inaccurate usage of suprasegmental features such as misplaced word stress, improper placement of pauses, and overuse of falling intonation can result in decreased intelligibility of the L2 learners' speech (Kang et al. 2020). The second construct, comprehensibility, is the effort exerted by a listener to understand a speaker's utterances (Munro and Derwing 1995). Comprehensibility issues arise when word and sentence stress is inappropriately placed, the speech rate is either too slow or too fast, and there is an overuse of pauses within speech (Saito 2021; Yang 2021). Comprehensibility is further impacted by limited or inappropriate word usage and a lack of proper grammar within speech (Derwing and Munro 2015; Trofimovich and Isaac 2012). The final construct, accentedness, is the deviance between an L2 learner's speech and a native speaker's speech (Derwing and Munro 2015). Accentedness often results from inaccurate production of vowel and consonant sounds (segmental features). While high accentedness can cause comprehensibility issues, accentedness does not necessarily impede the intelligibility of L2 learners' speech as even highly accented persons can be considered to be intelligible (Trofimovich and Isaac 2012; Kang et al. 2020). Therefore, intelligibility and comprehensibility are considered the most important constructs to target during pronunciation training (Derwing and Munro 2015). Training on suprasegmental features, which are considered to have the greatest impact on speech intelligibility and comprehensibility, is therefore recommended over segmental training (Avery and Ehrlich 1992; Morley 1991).

Historically, perception–production research (discussed in detail below) has focused on the perception and production of segmental features (i.e., consonant and vowel sounds) (Flege 2003). In their 2020 study, Lee et al. extended perception–production research to include the suprasegmental feature of word stress. Their findings revealed that perception and production of word stress improved after training (Lee et al. 2020). This shift in the focus of perception–production research is significant. As discussed above, suprasegmental features can impact the intelligibility and comprehensibility of L2 learners' speech (e.g., Kang et al. 2020; Saito 2021; Trofimovich and Isaac 2012), and an increased focus on suprasegmental features may result in the better understanding of L2 speakers by their interlocutors.

2.2. Speech Perception and Production

A primary goal of pronunciation instruction is to improve an L2 learner's perception and/or production of targeted pronunciation features. Speech perception is defined as

an individual's ability to understand the speech of others. Listeners first perceive, or distinguish, speech sounds (Hardison 2013; Mitterer and Cutler 2006) and then decode the sound elements (i.e., the frequency, tone, duration, and intensity of the sounds). Using these decoded sound elements, listeners interpret the linguistic intent of the interlocutor (Rost 2016, p. 20). In L2 learning, speech perception becomes challenging as L2 learners are often unable to effectively perceive the sounds of the target language (Nagle 2018). Several theories, including Best's Perceptual Assimilation Model (PAM-L2) (Best 1994, 1995; Best and Tyler 2007) and Flege's Speech Learning Model (SLM) (Flege 1995, 2003), have been developed to explain this phenomenon. For the purposes of this paper, the revised SLM-r model (Flege and Bohn 2021) is used as it theorizes the relationship between speech perception and production (Lee et al. 2020). According to the SLM-r model, L2 learners associate the sounds of an L2 with the closest sound in their native language (L1), even if the phonetic characteristics of the sounds differ (e.g., voice onset time or formant spacing) (Flege and Bohn 2021). This causes L2 learners to filter the sounds of the target language through their L1, making it difficult for learners to discern L2 sounds accurately. Through quality L2 speech input, it is suggested that language learners of all ages can learn to discern the phonetic characteristics of L2 sounds (Flege 2003; Flege and Bohn 2021). Hardison (2013) suggests that speech perception training is most effective when it includes multiple speakers, implicit training, and natural speech with a wide variability of sounds.

2.3. The Perception–Production Relationship

According to Flege's (1995, 2003) original SLM model, accurate production of L2 sounds could only occur if the L2 sound is first perceived by the language learner. While some research suggests that perception precedes production (e.g., Sakai and Moorman 2018; Saito and van Poeteren 2018), a meta-analysis of perception and production studies from 1988 to 2013 revealed that relationships between perception gains and production gains were not statistically significant, suggesting that the results of these studies could not confirm that improvements in perception led to improvements in production (Sakai and Moorman 2018). Furthermore, other empirical research either failed to demonstrate statistically significant correlations between perception training and speech production or showed inverse relationships between the two constructs (see Flege and Bohn 2021). Based on this research, Flege and Bohn (2021) introduced the SLM-r model suggesting that "a strong bidirectional connection exists between production and perception" (p. 30). While the perception–production relationship continues to be explored, the ability to perceive speaker messages accurately and to form accurate production of pronunciation features are considered to be fundamental requirements of communicative exchanges.

2.4. The History of MALL and MAPT Research

With the introduction of mobile devices in the 1990s, the concept of mobile-assisted learning emerged as a means to provide learner-focused instruction that was accessible anytime and anywhere (Persson and Nouri 2018). Along with the increased accessibility and portability of mobile devices over computers, mobile devices provided adaptable, personalized instruction and opportunities for individualized learning both within and outside the classroom (Persson and Nouri 2018). Mobile-assisted learning eventually expanded into L2 pronunciation training, with platforms such as YouTube videos, podcasts, social media platforms, and applications (apps) being used by L2 learners to improve their pronunciation of target languages. In particular, MAPT apps have been broadly used. These apps frequently target segmental accuracy by having users listen to and produce selected sounds and then, through automatic speech recognition (ASR), receive feedback regarding their pronunciation accuracy (O'Brien et al. 2018; Yaw 2020). Social media platforms have provided pronunciation training through collaborative learning situations in which L2 learners interact with each other, share oral content, and provide feedback to each other (Persson and Nouri 2018; Tseng et al. 2022). L2 learners have also "shadowed" or "dubbed"

the content of YouTube videos and podcasts to improve their pronunciation (Foote and McDonough 2017; Wei et al. 2022).

An increase in MAPT research has come with the increase in mobile-assisted usage for pronunciation training. Much of the MAPT research has focused on MAPT app reviews (Becker and Edalatishams 2019; Yaw 2020), app infrastructure and design issues, and teacher training on MAPT tools (Burston 2015). While this research can be useful, empirical research on MAPT provides greater insight into how mobile-assisted training impacts users' perception and production of various speech features. As MAPT continually becomes an integral component of pronunciation training, the following question arises: What effect does MAPT have on gains in L2 learners' perception and production? To answer this question, the remainder of this paper provides a review of ten empirical MAPT studies under a perception–production lens.

3. Methodology

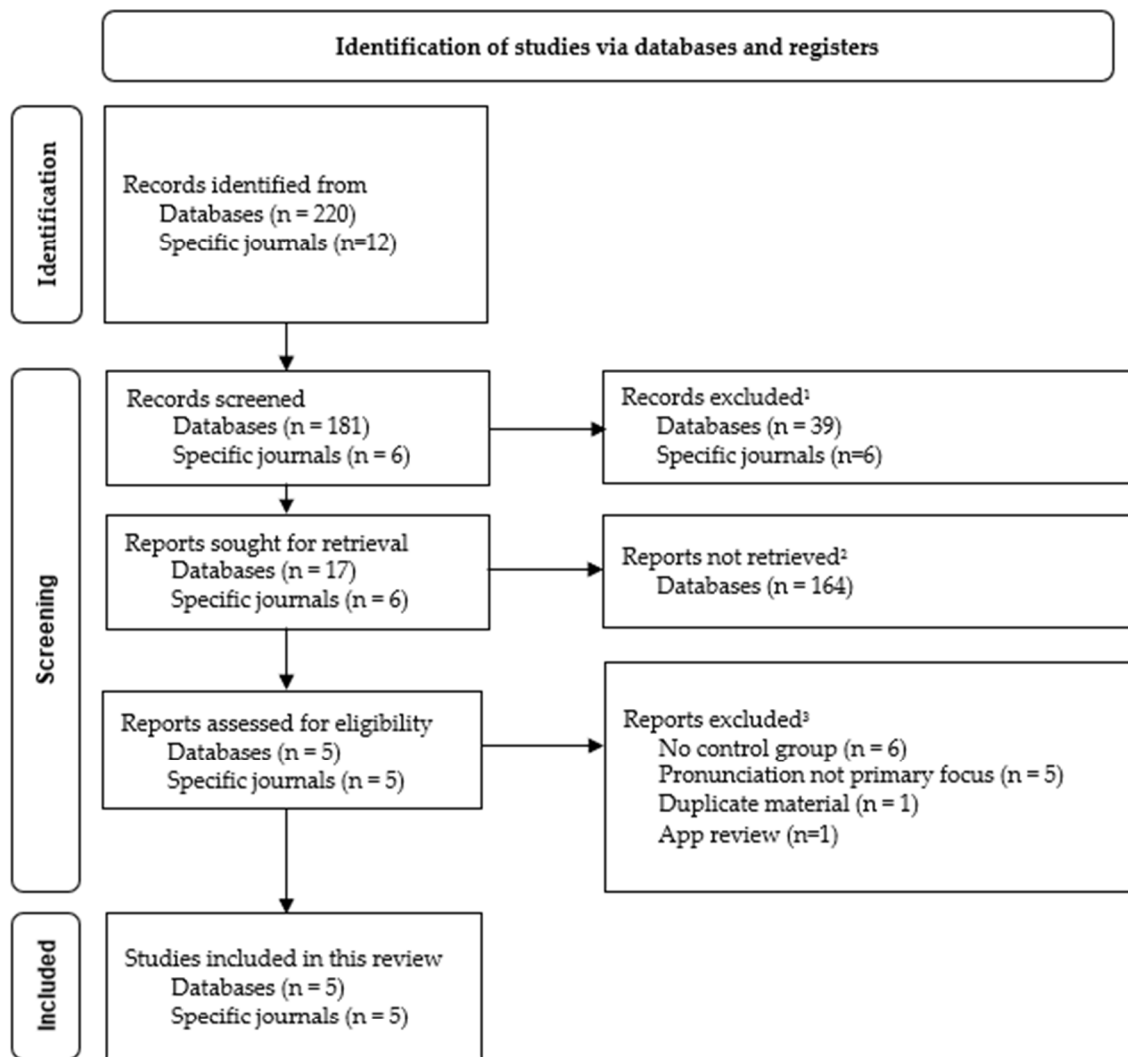
3.1. Selection of MAPT Studies

3.1.1. Identification of Articles for Inclusion

To better understand the impact MAPT has on L2 learners' perception and production of targeted features, this study conducted a systematic review of the MAPT literature following PRISMA 2020 guidelines (Page et al. 2021). Because of the fact that MAPT applications are both educational and linguistic in nature, two educational (ERIC and Educational Full Text) and two linguistic (Linguistics and Language Behavior Abstract and MLI International) databases were searched for the terms “pronunciation”, “pronunciation training”, “mobile-assisted”, “mobile-assisted pronunciation training”, and “mobile-assisted language training” for the period spanning 2014 through 2024. A general abstract and citation database (Scopus) was also searched using the same terms. These databases included several journals that intersect technology and education such as *Computers and Education*, *Journal of Educational Computing and Research*, *CALICO*, and *CALL*.

Additional journals that publish MAPT-related articles (see Sung et al. 2015; Tseng et al. 2022) were also included in this search, specifically *Language Learning and Technology*; *European Journal of Foreign Language Teaching*; *Journal of Applied Linguistics and Language Research*; *International Journal of Human-Computer Studies*; and *The New English Teacher*. These journals were searched for the terms “pronunciation”, “mobile-assisted”, and “MALL”.

A total of 232 papers were identified including 220 through the database search and 12 through the specific journal searches (see Figure 1). To narrow down the selected papers for this review, the following criteria were used to identify articles for inclusion: the articles must include training using mobile-assisted technology; pronunciation must be the primary focus of the training; an experimental or quasi-experimental research design with a control group must be used; and statistical results of the data must be provided. The articles must also be peer-reviewed and include information about the participants and the feature(s) targeted within the study. Therefore, articles including case studies, meta- or systematic-analysis, or computer-assisted training were excluded from this review, as were articles assessing MAPT's impact on non-pronunciation features (e.g., writing, vocabulary enhancement, teaching protocols) and articles assessing participants' attitudes, perceptions, or motivations towards MAPT applications. After screening the papers, the researchers found only 10 papers that met these criteria. These articles are included in this review (see Table 1). Although every effort has been made to be inclusive, it is possible that there are some MAPT-related articles that were not identified through the search terms and databases. Also, the inclusion of an additional reviewer may have resulted in the selection of different articles.



¹ Records were excluded due to duplication (n = 36) or not meeting peer-review criteria (n = 9).

² Records were not retrieved because the title indicated that the article did not meet search criteria.

³ Reports were primarily excluded from database search. Only one specific journal report was excluded due to no control group.

Figure 1. Identification of MAPT articles for review.

Table 1. Summary of MAPT articles included in this review.

Study	Participants ²	MPAT Tool	Target Feature	Assessment
Perception and Production				
Fouz-González (2020)	Spanish College	English File Pron App	Phonemes	Perception tasks produced words, sentences, etc.
Liakin et al. (2015) ¹	Canadian College	Nuance Dragon (ASR)	Phoneme	Perception tasks, produced words, etc.

Table 1. Cont.

Study	Participants ²	MPAT Tool	Target Feature	Assessment
Production				
Dillon and Wells (2023)	Korean College	Google Doc (ASR Feedback)	Phonemes	Produced passages
Liakin et al. (2017) ¹	Canadian College	NaturalReader TTS	Phonemic	Produced sentences
Sufi and Shalmani (2018)	Iranian College	Engl-to-Engl TFlat App	Phonemes	Produced words
Arashnia and Shahrokhi (2016)	Iranian MS	English File Pron App	Words	Produced words
Cerezo et al. (2019)	Spanish PreK	HolograFX ³ App	Words	Produced words
Di (2018)	Chinese College	Speak English More App + ASR	Word, sentence stress	Produced words, sentences, etc.
Sun et al. (2017)	Chinese ES	Papa App (SNS)	Fluency; pronunciation	Produced descriptions
Wei et al. (2022)	Chinese College	English Fun Dubbing	Intelligibility; comprehen.	Produced passages

Notes: ¹ MAPT studies targeted French phonemes (Liakin et al. 2015) and liaisons (Liakin et al. 2017). ² ES = elementary school, MS = middle school, and HS = high school. ³ App was created by the authors for the study.

3.1.2. Data Collection and Coding

For this study, the researcher collected data from the selected articles to identify the participants used in the studies (e.g., nationality and education level), the MPAT applications and platforms used, the pronunciation features targeted, and the methods used for assessment. Nearly all the data, with the exception of some perception and production coding, was readily available from within the articles. If not stated, the coding of the assessment tasks as either perception or production was based on the nature of the task. Tasks asking participants to read sentences or paragraphs were deemed as production tasks; tasks asking participants to listen to recordings and identify targeted features were labeled as perception tasks.

4. Results

The reviewed studies targeted participants from various countries worldwide including Canada, China, Iran, Korea, and Spain (see Table 1). The majority of the participants were college students (seven studies). Over half of the studies (six studies) assessed the impact of MAPT apps; the other studies analyzed the impact of speech recognition software, text-to-speech (TTS) applications, ASR feedback, and a dubbing program. Five of the studies targeted phonemic (i.e., segmental) features, one targeted word and sentence stress (suprasegmental features), and two studies targeted overall speech performance (i.e., intelligibility and comprehensibility; fluency and overall pronunciation). While all the studies assessed participants' production of either words, sentences, or passages, only two assessed participants' perceptions of targeted features.

The studies selected for review occurred between 2015 and 2024 and included an average of 52.4 participants per study (see Table 2). Nearly half of the studies (four studies) conducted MAPT training sessions over four- to six-week periods. Except for the Fouz-González (2020) study, which utilized the correlation r to determine effect sizes for perception tasks, effect sizes were determined using Cohen's d . Seven of the studies showed large effect sizes ($d > 0.8$); the other studies showed a mixture of small, medium, and large effect sizes.

Table 2. Participant data, study duration, and effect sizes for included studies.

Study	Participants ¹	Study Duration	Effect Size <i>d</i> or <i>r</i>
Fouz-González (2020)	54 EG1 = 2; EG2 = 27; CG = 27	2 weeks	Perception—Stimuli (Familiar): EG1: /æ/ <i>r</i> = 0.24 /ʌ/ <i>r</i> = 0.44 /ɑ:/ <i>r</i> = 0.55 /ə/ <i>r</i> = 0.38 /z/ <i>r</i> = 0.41 EG2: /æ/ <i>r</i> = 0.83 /ʌ/ <i>r</i> = 0.79 /ɑ:/ <i>d</i> = 1.39 /ə/ <i>r</i> = 0.69 /z/ <i>r</i> = 0.77 Perception—Stimuli (Novel): EG1: /æ/ <i>r</i> = 0.09 /ʌ/ <i>r</i> = 0.31 /ɑ:/ <i>r</i> = 0.10 /ə/ <i>r</i> = 0.15 /z/ <i>r</i> = 0.19 EG2: /æ/ <i>d</i> = 0.29 /ʌ/ <i>r</i> = 0.60 /ɑ:/ <i>r</i> = 0.19 /ə/ <i>d</i> = 0.68 * /z/ <i>r</i> = 0.62 * Production—Imitation Task: EG1: /æ/ <i>d</i> = 0.57 * /ʌ/ <i>d</i> = 0.86 /ɑ:/ <i>d</i> = 0.60 /ə/ <i>d</i> = 0.08 /z/ <i>d</i> = 0.33 EG2: /æ/ <i>d</i> = 0.09 /ʌ/ <i>d</i> = 0.35 /ɑ:/ <i>d</i> = 0.69 /ə/ <i>d</i> = 0.16 /z/ <i>d</i> = 0.28 Production—Sentence Task (Familiar): EG1: /æ/ <i>d</i> = 0.72 /ʌ/ <i>d</i> = 0.44 /ɑ:/ <i>d</i> = 0.70 * /ə/ <i>d</i> = 0.24 /z/ <i>d</i> = 0.23 * EG2: /æ/ <i>d</i> = 0.19 /ʌ/ <i>d</i> = 0.28 /ɑ:/ <i>d</i> = 0.35 /ə/ <i>d</i> = 0.20 /z/ <i>d</i> = 0.00 Production—Sentence Task (Novel): EG1: /æ/ <i>d</i> = 0.17 /ʌ/ <i>d</i> = 0.63 * /ɑ:/ <i>d</i> = 0.76 /ə/ <i>d</i> = 0.60 * /z/ <i>d</i> = 0.18 EG2: /æ/ <i>d</i> = 0.40 /ʌ/ <i>d</i> = 0.19 /ɑ:/ <i>d</i> = 0.34 /ə/ <i>d</i> = -0.04 /z/ <i>d</i> = 0.30 Production—Timed-picture Description: EG1: /æ/ <i>d</i> = 0.43 /ʌ/ <i>d</i> = 0.60 * /ɑ:/ <i>d</i> = 0.52 /ə/ <i>d</i> = 0.74 * /z/ <i>d</i> = 0.17 EG2: /æ/ <i>d</i> = 0.31 /ʌ/ <i>d</i> = 0.11 /ɑ:/ <i>d</i> = 0.25 /ə/ <i>d</i> = 0.13 /z/ <i>d</i> = 0.21
Liakin et al. (2015)	42 ASR = 14; NonASR = 14; CG = 14	5 weeks	Perception: ASR: <i>d</i> = 0.57 NonASR: <i>d</i> = 0.17 Production: ASR: <i>d</i> = 0.74 * NonASR: <i>d</i> = 0.52
Dillon and Wells (2023)	59 EG = 34; CG = 25	4 weeks	Overall error reduction: <i>d</i> = -0.28 * Vowel error reduction: <i>d</i> = -0.31 /l/—/r/ error reduction: <i>d</i> = -0.21 Epenthesis error reduction: <i>d</i> = -0.18 /bpfv/ error reduction: <i>d</i> = 0.02
Liakin et al. (2017)	27 TTS = 9; NonTTS = 9; CG = 9	5 weeks	TTS: <i>d</i> = 1.51 Non-TTS: <i>d</i> = 0.98
Sufi and Shalmani (2018)	30 EG = 15; CG = 15	6 weeks	<i>d</i> = 3.47 *
Arashnia and Shahrokhi (2016)	60 EG = 30; CG = 30	3 months	<i>d</i> = 2.46 *
Cerezo et al. (2019)	70 EG1 = 23; EG2 = 23; CG = 24	3 weeks	EG1: <i>d</i> = 2.79 * EG2: <i>d</i> = 3.64 *
Di (2018)	60 EG = 30; CG = 30	8 lessons	<i>d</i> = 3.88
Sun et al. (2017)	72 EG = 37; CG = 35	13 weeks	Fluency: <i>d</i> = 1.00 * Pronunciation: <i>d</i> = 0.96
Wei et al. (2022)	50 EG = 24; CG = 26	8 weeks	Intelligibility: <i>d</i> = 1.57 Comprehensibility: <i>d</i> = -0.09 *

¹ EG = experimental group; CG = control group. * Statistically significant at (*p* < 0.05).

5. Discussion

5.1. MAPT Impact on Perception

The only MAPT studies that included an assessment of perception gains were [Fouz-González \(2020\)](#) and [Liakin et al. \(2015\)](#). The [Fouz-González \(2020\)](#) study utilized the English File Pronunciation (EFP) app to improve Spanish college students' perception and production of four English vowel sounds (/æ a: ʌ ə/) and the /s—z/ contrast. Two experimental groups were used, EG1 and EG2. EG2 originally served as a control group for EG1 and then received subsequent training. Both groups reviewed targeted sounds on a phonemic chart and practiced these and other sounds using the EFP app for a two-week period. Perception gains were assessed using the following familiar and novel stimuli for two tasks: (1) a sound identification task and (2) a sound differentiation task. EG1, the only group compared to the control group, showed no significant difference in sound perceptions from the control group for either the familiar or novel stimuli. The results showed that the training for the EG1 group had a small to medium effect ($r = 0.24$ to 0.55) on perception gains for familiar tasks, while the training for the EG2 group had a medium to large effect on perception ($r = 0.69$ to $d = 1.39$). For both groups, the training had primarily small effects on perception gains when novel stimuli were introduced. As a result, even though the participants showed improved abilities to perceive targeted sounds in familiar stimuli, they were unable to transfer these perception skills to novel stimuli.

[Liakin et al. \(2015\)](#) studied the impact of French /y/ phoneme training on native and near-native English speakers from a Canadian university. Two groups of students were trained on the French /y/ phoneme while a third (control) group received no training. The first group used Nuance Dragon Dictation, a commercially available ASR application, to receive feedback on the pronunciation of selected words (ASR group); the second group studied the selected words in class with a teacher (NonASR group); and the third (control) group received only conversation practice. The results from perception tasks (listening to French pseudowords with the targeted sound and distractors) yielded a medium effect size for the ASR group ($d = 0.57$) and a small effect size for the NonASR group ($d = 0.17$). The study, however, did not find the perception gains to be statistically different among the three groups.

Based on this review, perception gains were only assessed in two out of the ten reviewed studies. When perception was assessed, perception gains were not significant ([Fouz-González 2020](#); [Liakin et al. 2015](#)) or transferable to novel input ([Fouz-González 2020](#)). These results imply that MAPT training may not significantly improve L2 learners' abilities to perceive sounds. However, as these results were only found in two studies, more research is required to understand MAPT's impact on perception.

5.2. MAPT Impact on Production

All the studies included in this review assessed the production of the targeted features. Five of these studies targeted phonemes or segmental production, two targeted word production, and one targeted word and sentence stress production. The final two studies targeted the global features of fluency and production and intelligibility and comprehensibility.

5.2.1. Segmental Production: Vowels and Consonants

The [Fouz-González \(2020\)](#) study described above utilized three tasks to assess participants' production including an imitation task, a sentence reading task, and a picture description task. The impact of the training varied based on the production task type and the targeted vowel sounds. For example, the EG1 group realized greater effect sizes than the EG2 group in /æ/ production in the imitation and familiar sentence tasks ($d = 0.57$ and 0.72 , respectively) and the picture description task ($d = 0.43$). Conversely, the EG2 group realized higher effect sizes in /æ/ production during the novel sentence task ($d = 0.40$). Overall, the training tended to have small to medium effect sizes on production (EG1: $d = 0.08$ to 0.76 ; EG2: $d = 0.00$ to 0.69). Production gains were significant for only the /æ/

sound in the imitation task, the /ɑ:/ and /z/ sounds for the familiar sentence task, and the /ʌ/ and /ə/ sounds for the novel sentence task and the picture description task. Within the Liakin et al. (2015) study (also described above), production tasks included participants reading aloud words and phrases. Production gains for both experimental groups showed medium effect sizes of the training ($d = 0.74$ and 0.52 , respectively), though only the gains from the ASR group were considered to be statistically significant within the study.

In the Dillon and Wells (2023) study, Korean students received pronunciation training on the vowel and consonant sounds that differ between English and Korean. The groups also practiced reading a paragraph about rainbows. The students were divided into an experimental and a control group, with only the experimental group receiving training on, and practicing pronunciation with, the ASR feedback feature in Google Documents. Results from a production task (reading the rainbow paragraph) showed a significant reduction in overall pronunciation errors between the experimental and control groups, although the training had only a small effect ($d = -0.28$) on the experimental group's error reduction. Reduction in errors for individual segmental features showed no significant difference between the experimental and control groups, although the training did result in small effect sizes for reduction in vowel sound errors ($d = -0.31$), /l/-/r/ production errors ($d = -0.21$), and epenthesis errors ($d = -0.18$).

In the Liakin et al. (2017) study, two groups of students at a Canadian university were trained on French liaisons; a third (control) group received no training. The first group uploaded word phrases provided by instructors into NaturalReader, a text-to-speech (TTS) tool, and utilized the uploaded phrases to practice French liaisons. The second group practiced French liaison phrases with a teacher (NonTTS), while the third (control) group received only conversation practice. Results from the production tasks (reading sentences aloud) showed large effect sizes of the training in both the TTS and NonTTS groups ($d = 1.51$ and 0.98 , respectively). However, the authors cited no significant differences in production gains among the three groups.

In the Sufi and Shalmani (2018) study, students from an Iranian university were trained on vowels, consonants, and diphthongs in a classroom setting followed by pronunciation practice using the English-to-English TFlat App. The control group learned and practiced the sounds in a classroom setting. Results from a production task (reading words aloud) showed that only the group using the TFlat app realized significant production gains for the targeted sounds with a large effect size for the training ($d = 3.47$).

Overall, the results of the MAPT studies focusing on the production of segmental features are mixed. Significant production gains for experimental groups over control groups were shown in the studies by Liakin et al. (2015) and Sufi and Shalmani (2018). Dillon and Wells (2023) reported significant error reduction overall; however, error reductions for specific segmental features were not significant. The Fouz-González (2020) study demonstrated production gains for only certain vowel features but not others. Effect sizes of the training were large in the studies by Liakin et al. (2017) and Sufi and Shalmani (2018) and medium-sized in those by Liakin et al. (2015) and Dillon and Wells (2023). Differences in the parameters of the studies (e.g., nationality of the participants, targeted features, types of MAPT platforms used) make it difficult to assess why the variances in results occurred. Additional research on MAPT studies targeting segmental production could provide greater insight into how MAPT training impacts segmental production.

5.2.2. Word Production

Two of the reviewed MAPT studies targeted the production of words specifically. In the Arashnia and Shahrokhi (2016) study, the researchers analyzed the impact of a MAPT app on Iranian middle school students' production of selected words. The study's experimental group learned word pronunciation in a classroom setting with continued learning of the word and practice via the EFP app (this app was also used by Fouz-González 2020). The control group received only in-class instruction and practice. In a production task of reading aloud selected words, the experimental group realized significantly higher

production gains than the control group. The training was also effective, resulting in a large effect size ($d = 2.46$).

In the [Cerezo et al. \(2019\)](#) study, preschool students in Spain practiced the pronunciation of previously learned vocabulary using an app developed by the authors. One version of the app included holographic images of the vocabulary words and a virtual “teacher” named Arturito. Students were divided into three groups as follows: the first group used the app without the holographic game (EG1), the second group used the app without the holographic game (EG2), and the third (control) group received in-class training only. Results from a word production task showed statistically higher gains for both experimental groups over the control group, with the group using the holographic game app realizing statistically higher production gains than the other experimental group. The training was also effective, resulting in large effect sizes for both experimental groups (EG1: $d = 2.79$; EG2: $d = 3.64$).

Overall, the results of these reviewed studies suggest that using MAPT to improve word production can be effective. The experimental groups in both MAPT studies in this category realized significant gains over the control groups. The participants in these studies were preschool and middle school students, which may suggest that these apps are effective for younger participants.

5.2.3. Suprasegmental Production: Word and Sentence Stress

The [Di \(2018\)](#) study is the only study to target suprasegmental features. The study targeted Chinese college students’ improvement in word and sentence stress patterns through the use of song lyrics embedded into the Speak English More App with ASR feedback. The experimental group listened to the lyrics via the app, practiced producing the lyrics, and received ASR feedback on their pronunciations. The control group only practiced the lyrics in class. Results from production tasks (reading of words, phrases, sentences, and a paragraph) showed a statistical improvement in word and sentence stress for the students using the app. The training had a large effect size ($d = 3.88$). Therefore, while only one study targeted suprasegmental features, the results suggest that MAPT may be effective at enhancing L2 learners’ production of suprasegmental features. However, more research on the use of MAPT for suprasegmental improvement is needed to support this assertion.

5.2.4. Overall Speech Performance

The final two studies reviewed did not target a specific segmental or suprasegmental feature. Instead, [Sun et al. \(2017\)](#) targeted fluency and overall participant pronunciation gains, and [Wei et al. \(2022\)](#) targeted improvements in the intelligibility and comprehensibility of participants’ speech. During the [Sun et al. \(2017\)](#) study, Chinese elementary school students received training on selected words and sentence structures and were given oral homework tasks to practice these items. Students in the experimental group were asked to record responses to these tasks using Papa, a social network (SNS) app, with parental assistance. The control group did not record their responses. Speech samples were elicited through picture description tasks, and the results of the study indicated that the experimental group realized statistically significant gains in fluency only; production gains for overall pronunciation were not significant. However, the training had a large effect on both fluency ($d = 1.00$) and overall pronunciation ($d = 0.96$).

In the [Wei et al. \(2022\)](#) study, an experimental group of Chinese college students learning English dubbed two 60-minute video clips during eight pronunciation training sessions. Both the experimental and control groups received English instruction during this period, but the control group did not dub videos. To measure performance, the participants were asked to read a selected passage. At the end of the eight sessions, the training was shown to have a large effect on intelligibility gains for the experimental group ($d = 1.57$), but these gains were not found to be significantly different than those of the control group. In terms of comprehensibility, the experimental group’s reading of the passage was considered

to be statistically more comprehensible than the control group's reading. However, the overall impact of the training was negligible ($d = 0.09$).

In summary, the results of the studies on overall speech performance are again mixed. Only fluency (Sun et al. 2017) and comprehensibility (Wei et al. 2022) were significantly different between the performance of the experimental and control groups. Differences in intelligibility (Wei et al. 2022) and overall pronunciation production (Sun et al. 2017) were not statistically significant. The MAPT training, however, did have a large effect on fluency, pronunciation, and intelligibility. Based on the results of these studies, MAPT training appears to be effective, but this training does not appear to have significantly higher results than in-class pronunciation training.

5.3. MAPT Impact

In the ten studies listed above, MAPT was not used as an alternative for in-class pronunciation training; instead, MAPT platforms were used to supplement and/or provide additional practice for in-class pronunciation instruction. Only Fouz-González (2020) used MAPT as the primary form of pronunciation training. Overall, there was very little consistency in the type of MAPT platforms used during these studies, with some studies assessing the impact of established MAPT apps (Arashnia and Shahrokhi 2016; Fouz-González 2020; Sufi and Shalmani 2018) and others assessing apps they developed themselves (Cerezo et al. 2019). ASR from various sources was used in three studies (Dillon and Wells 2023; Di 2018; Liakin et al. 2015), and a social network site was used in one study (Sun et al. 2017). Dubbing apps (Wei et al. 2022) and text-to-speech apps (Liakin et al. 2017) were also used. This degree of variability makes it difficult to assess MAPT's impact on pronunciation.

The results from assessments of MAPT's impact on perception indicate that MAPT provides no significant gains over in-class teaching. However, these findings are based on the results of only two studies. More research is needed to determine if these results are consistent in other studies. The results of MAPT's impact on production are more promising, with seven studies showing statistically significant results from using MAPT platforms. However, the effect of training in Dillon and Wells (2023) had only a small effect on overall error reduction, and pronunciation gains were not significant for individual features. Furthermore, while differences in comprehensibility between experimental and control groups were considered to be statistically significant in Wei et al. (2022), the participants actually demonstrated slightly less comprehensible speech after training (pre-training $M = 80.72$ versus post-training $M = 80.36$). The remaining five studies demonstrated large effect sizes of the training ($d = 0.74$ to 3.64). Therefore, only half of the studies can be considered to demonstrate a significant impact of MAPT on L2 learners' production of targeted pronunciation features.

6. Implications and Future Directions

Although the findings of this review show mixed results, the studies do demonstrate that MAPT can be used to target a variety of segmental and suprasegmental features across a range of participants, both in terms of L1 background and age. As demonstrated, MAPT can be used to enhance in-class pronunciation training, increasing L2 learners' exposure to targeted pronunciation features and providing immediate feedback via ASR. Although not stated, L2 learners may also benefit from un-stressful learning environments and personalized instructions, which are key components of MAPT platforms. Therefore, this research suggests that MAPT platforms may provide viable options for practicing pronunciation within L2 classrooms. However, because of the small number of studies included in this review, more research is needed to better understand the impact of MAPT on L2 learners' perception and production. In addition, research is needed to assess the viability of MAPT as a stand-alone form of pronunciation training within L2 classrooms.

Also, these studies focused only on participants in school settings. It is likely that MAPT can be used for pronunciation learning outside of school settings and by people in

different domains of life. For example, students living in remote areas with limited access to L2 learning and adults with mobility issues, childcare concerns, or work responsibilities may find it useful to access the pronunciation tools available via smartphones. There are, therefore, opportunities to expand MAPT research to include MAPT usage outside of classroom settings and to include non-traditional L2 learners.

Moreover, research is needed to understand MAPT's impact on overall speech performance. In this review, only one study researched the impact of MAPT on speech intelligibility and comprehensibility. As these two constructs impact the understandability of L2 learners' speech, more emphasis should be placed on determining how MAPT impacts these constructs. In addition, more researchers should conduct research on how MAPT can be used to improve the perception and production of suprasegmental features, which are considered to impact L2 learners' intelligibility and comprehensibility.

7. Conclusions

While MAPT provides a range of benefits to L2 instructors, students, and other L2 learners, the ten studies included in this review did provide sufficient support for using MAPT for perception and production purposes. Perception gains were rarely assessed within these studies, and only half of the studies demonstrated a significant impact of MAPT on production gains. More robust, empirical studies are needed for MAPT research with a focus on both segmental and suprasegmental features and an assessment of both perception and production gains. To measure intelligibility and comprehensibility, gains should also be included to assess the impact of MAPT on overall speech performance. Research should also expand beyond classroom settings to better understand how MAPT impacts the pronunciation gains of L2 learners not affiliated with an educational program. Such changes could build greater confidence among L2 instructors, L2 learners, and researchers in the effectiveness of MAPT applications for pronunciation training.

Funding: This research received no external funding.

Data Availability Statement: All data used in this study is included in the following databases: EBSCO Information Services, Education Full Text; EBSCO Information Services, ERIC (Education Resources Information Center); EBSCO Information Services, MLA International Bibliography with Full Text; Scopus; ProQuest 2024. Linguistics and Language Behavior Abstracts (LLBA).

Conflicts of Interest: The authors declare no conflicts of interest.

References

- Arashnia, Maryam, and Mohsen Shahrokhi. 2016. Mobile assisted language learning: English pronunciation among Iranian pre-intermediate EFL learners. *Journal of Applied Linguistics and Language Research* 3: 149–62.
- Avery, Peter, and Susan Ehrlich. 1992. *Teaching American English Pronunciation*. Oxford: Oxford University Press.
- Becker, Kimberly, and IdéeI Edalatihams. 2019. ELSA Speak—Accent Reduction. In *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference, and IA Ames, Ames, IA, USA, September 6–8, 2018*. Edited by John Levis, Charles Nagle and Erin Todey. Ames: Iowa State University, pp. 434–38.
- Best, Catherine T. 1994. The emergence of native-language phonological influence in infants: A perceptual assimilation hypothesis. In *The Development of Speech Perception: The Transition from Speech Sounds to Spoken Words*. Edited by Judith C. Goodman and Howard C. Nusbaum. Cambridge: MIT Press, pp. 1676–224.
- Best, Catherine T. 1995. A direct-realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Speech Research*. Edited by Winifred Strange. York: Timonium, pp. 171–206.
- Best, Catherine T., and Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In *Language Experience in Second Language Speech Perception*. Edited by Ocki-Schwen Bohn. Amsterdam: John Benjamins, pp. 13–34.
- Burston, Jack. 2015. Twenty years of MALL project implementation: A meta-analysis of learning outcomes. *ReCALL* 27: 4–20. [\[CrossRef\]](#)
- Cerezo, Rebeca, Vicente Calderón, and Cristóbal Romero. 2019. A holographic mobile-based application for practicing pronunciation of basic English vocabulary for Spanish speaking children. *International Journal of Human-Computer Studies* 124: 13–25. [\[CrossRef\]](#)
- Cho, Kyunghwa, Sungwoong Lee, Min-Ho Joo, and Betsy Jane Becker. 2018. The effects of using mobile devices on student achievement in language learning: A meta-analysis. *Education Sciences* 8: 105. [\[CrossRef\]](#)

- Derwing, Tracey M., and Murray J. Munro. 2015. *Pronunciation Fundamentals: Evidence-Based Perspectives for L2 Teaching and Research*. Amsterdam: John Benjamins Publishing Company.
- Di, Wu. 2018. Teaching English stress: Can song-lyric reading combined with mobile learning be beneficial to non-English majors? *The New English Teacher* 12: 91–91.
- Dillon, Thomas, and Donald Wells. 2023. Effects of pronunciation training using automatic speech recognition on pronunciation accuracy of Korean English language learners. *English Teaching* 78: 3–23. [CrossRef]
- Elaish, Monther M., Mahmood H. Hussein, and Gwo-Jen Hwang. 2023. Critical research trends of mobile technology-supported English language learning: A review of the top 100 highly cited articles. *Education and Information Technologies* 28: 4849–74. [CrossRef] [PubMed]
- Flege, James Emil. 1995. Second language speech learning: Theory, findings, problems. In *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Edited by Winifred Strange. Timonium: York Press, pp. 233–77.
- Flege, James Emil. 2003. Assessing constraints on second-language segmental production and perception. In *Phonetics and Phonology in Language Comprehension and Production: Differences and Similarities*. Edited by Niels O. Schiller and Antje S. Meyer. New York: Mouton de Gruyeter, pp. 319–55.
- Flege, James Emil, and Ocki-Schwen Bohn. 2021. The revised speech learning model (SLM-r). In *Second Language Special Learning: Theoretical and Empirical Progress*. Edited by Ratrete Wayland. Cambridge: Cambridge University Press, pp. 3–83.
- Foote, Jennifer A., and Kim McDonough. 2017. Using shadowing with mobile technology to improve L2 pronunciation. *Journal of Second Language Pronunciation* 3: 34–56. [CrossRef]
- Fouz-González, Jonás. 2020. Using apps for pronunciation training: An empirical evaluation of the English File Pronunciation app. *Language Learning & Technology* 24: 62–85.
- Hardison, Debra M. 2013. Second language speech perception: A cross-disciplinary perspective on challenges and accomplishments. In *The Routledge Handbook of Second Language Acquisition*. Edited by Susan M. Gass and Alison Mackey. London: Routledge, pp. 349–63.
- Hinofotis, Frances, and Kathleen Bailey. 1980. American undergraduates' reactions to the communication skills of foreign teaching assistants. *On TESOL* 80: 120–33.
- Kang, Okim, Ron I. Thomson, and Meghan Moran. 2020. Which features of accent affect understanding? Exploring the intelligibility threshold of diverse accent varieties. *Applied Linguistics* 41: 453–80. [CrossRef]
- Lee, Bradford, Luke Plonsky, and Kazuya Saito. 2020. The effects of perception- vs. production-based pronunciation instruction. *System* 88: 102185. [CrossRef]
- Li, Rui. 2024. Effects of mobile-assisted language learning on foreign language learners' speaking skill development. *Language Learning & Technology* 28: 1–26.
- Liakin, Denis, Walcir Cardoso, and Natallia Liakina. 2015. Learning L2 pronunciation with a mobile speech recognizer: French /y/. *Calico Journal* 32: 1–25. [CrossRef]
- Liakin, Denis, Walcir Cardoso, and Natallia Liakina. 2017. The pedagogical use of mobile speech synthesis (TTS): Focus on French liaison. *Computer Assisted Language Learning* 30: 325–42. [CrossRef]
- Meisarah, Fitria. 2020. Mobile-assisted pronunciation training: The Google Play pronunciation and phonetics application. *Script Journal* 5: 487. [CrossRef]
- Metruk, Rastislav. 2024. Mobile-assisted language learning and pronunciation instruction: A systematic literature review. *Education and Information Technologies*. [CrossRef]
- Mitterer, Holger, and Anne Cutler. 2006. Speech perception. In *Encyclopedia of Language & Linguistics*, 2nd ed. Edited by Keith Brown. Amsterdam: Elsevier.
- Morley, Joan. 1991. The pronunciation component in teaching English to speakers of other languages. *TESOL Quarterly* 25: 481–520. [CrossRef]
- Munro, Murray J., and Tracey M. Derwing. 1995. Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning* 45: 73–97. [CrossRef]
- Nagle, Charles L. 2018. Perception, production, and perception-production: Research findings and implications for language pedagogy. *Contact* 44: 5–12.
- Nair, Ramesh, Rajasegaran Krishnasamy, and Geraldine De Mello. 2017. Rethinking the teaching of pronunciation in the ESL classroom. *The English Teacher* 35: 27–40.
- Nitisakunwut, Panicha, and Gwo-Jen Hwang. 2023. Effects and core design parameters of digital game-based language learning in the mobile era: A meta-analysis and systematic review. *International Journal of Mobile Learning and Organisation* 17: 470–98. [CrossRef]
- O'Brien, Mary Grantham, Tracey M. Derwing, Catia Cucchiari, Debra M. Hardison, Hansjörg Mixdorff, Ron I. Thomson, Helmer Strick, John M. Levis, Murray J. Munro, Jennifer A. Foote, and et al. 2018. Directions for the future of technology in pronunciation research and teaching. *Journal of Second Language Pronunciation* 42: 182–207. [CrossRef]
- Page, Matthew J., Joanne E. McKenzie, Patrick M. Bossuyt, Isabelle Boutron, Tammy C. Hoffmann, and Cynthia D. Mulrow. 2021. The PRISMA 2020 statement: An updated guideline for reporting systematic reviews. *BMJ* 372: n71. [CrossRef] [PubMed]
- Persson, Veronica, and Jalal Nouri. 2018. A systematic review of second language learning with mobile technologies. *International Journal of Emerging Technologies in Learning* 13: 188–210. [CrossRef]

- Pourhossein Gilakjani, Abbas. 2017. English pronunciation instruction: Views and recommendations. *Journal of Language Teaching and Research* 8: 1249–55. [\[CrossRef\]](#)
- Rost, Michael. 2016. Linguistic procession. In *Teaching and Researching Listening*, 3rd ed. London: Routledge, pp. 19–48.
- Saito, Kazuya. 2021. What characterizes comprehensible and native-like pronunciation among English as a second language speakers? Meta-analyses of phonological, rater, and instructional factors. *TESOL Quarterly* 55: 866–900. [\[CrossRef\]](#)
- Saito, Kazuya, and Luke Plonsky. 2019. Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning* 69: 652–708. [\[CrossRef\]](#)
- Saito, Kazuya, and Kim van Poeteren. 2018. The perception–production link revisited: The case of Japanese learners’ English /ɹ/ performance. *International Journal of Applied Linguistics* 28: 3–17. [\[CrossRef\]](#)
- Sakai, Mari, and Colleen Moorman. 2018. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics* 39: 187–224. [\[CrossRef\]](#)
- Su, Fan, Di Zou, Haoran Xie, and Fu Lee Wang. 2021. A comparative review of mobile and non-mobile games for language learning. *SAGE Open* 11: 21582440211067247. [\[CrossRef\]](#)
- Sufi, Effat, and Hamed Babaie Shalmani. 2018. The effects of Tflat pronunciation training in Mall on the pronunciation ability of Iranian EFL learners. *European Journal of Foreign Language Teaching* 3: 1245173. [\[CrossRef\]](#)
- Sun, Zhong, Chin-Hsi Lin, Jiaxin You, Hai Jiao Shen, Song Qi, and Liming Luo. 2017. Improving the English-speaking skills of young learners through mobile social networking. *Computer Assisted Language Learning* 30: 304–24. [\[CrossRef\]](#)
- Sung, Yao-Ting, Kuo-En Chang, and Je-Ming Yang. 2015. How effective are mobile devices for language learning? A meta-analysis. *Educational Research Review* 16: 68–84. [\[CrossRef\]](#)
- Tommerdahl, Jodi M., Chrystal S. Dragonflame, and Amanda A. Olsen. 2022. A systematic review examining the efficacy of commercially available foreign language learning mobile apps. *Computer Assisted Language Learning* 37: 333–62. [\[CrossRef\]](#)
- Trofimovich, Pavel, and Talia Isaac. 2012. Disentangling accent from comprehensibility. *Bilingualism: Language and Cognition*. 15: 1–12. [\[CrossRef\]](#)
- Tseng, Wen-Ta, Sufen Chen, Shih-Peng Wang, Hsing-Fu Cheng, Pei-Shan Yang, and Xuesong A. Gao. 2022. The effects of MALL on L2 pronunciation learning: A meta-analysis. *Journal of Educational Computing Research* 60: 1220–52. [\[CrossRef\]](#)
- Wei, Jing, Haibo Yang, and Jing Duan. 2022. Investigating the effects of online English film dubbing activities on the intelligibility and comprehensibility of Chinese students’ English pronunciation. *Theory and Practice in Language Studies* 12: 1911–20. [\[CrossRef\]](#)
- Yang, In Young. 2021. Differential contribution of English suprasegmentals to L2 foreign-accentedness and speech comprehensibility: Implications for teaching EFL pronunciation, speaking, and listening. *Korean Journal of English Language and Linguistics* 21: 818–36. [\[CrossRef\]](#)
- Yaw, Katherine. 2020. Technology Review: Blue Canoe. Paper presented at the 11th Pronunciation in Second Language Learning and Teaching Conference, Flagstaff, AZ, USA, September 12–14, 2019.

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.