

DEPARTMENT OF ENGLISH AND GERMAN STUDIES

**Shadowing and Phonetic Self-imitation through
Voice Manipulation for L2 English Pronunciation
Learning: a Critical Review**

Treball de Fi de Grau/ BA dissertation

Author: Júlia Fernández Cantón

Supervisor: Ingrid Mora Plaza

Departament de Filologia Anglesa i de Germanística

Facultat de Filosofia i Lletres

Grau d'Estudis d'Anglès i Francès

June 2026

Statement of Intellectual Honesty

Your name: Júlia Fernández Cantón

Title of Assignment: Shadowing and Phonetic Self-imitation through Voice Manipulation for L2

English Pronunciation Learning: a Critical Review

I declare that this TFG is an original piece of work written by me, and that all secondary sources have been correctly cited.

I understand that plagiarism—including the unacknowledged use of text generated by Artificial Intelligence (AI) tools—is a breach of academic integrity and may result in a failing grade.

I additionally declare **one of the following**:

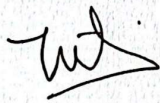
☒ **I have not used any AI tools in the preparation of this BA dissertation.**

☐ **I have used AI tools in a limited and transparent manner, with the prior written consent of my tutor, and in accordance with the course regulations.** In all such cases, I have clearly identified where and for what purpose AI was used, and I have included a brief critical reflection on how this use affected the research process and final outcome. Relevant documentation is included in an annex.

I understand that any lack of transparency regarding the use of AI will be treated as a breach of academic integrity.

Signature and date:

28th May 2026



ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to my TFG supervisor, Ingrid Mora Plaza, for her guidance, suggestions, and feedback throughout this project. Her expertise, professionalism, and dedication have been essential to the completion of this dissertation.

Furthermore, I would also like to thank my family, especially my parents and my sister, as well as my friends, for their encouragement and support, and for always being by my side.

TABLE OF CONTENTS

Abstract.....	1
Résumé du travail	2
1. Introduction	11
2. L2 pronunciation instruction techniques	15
2.1 Non-technology-oriented pronunciation techniques.....	15
2.2 Technology-oriented pronunciation techniques.....	17
3. Imitating others' voices vs. acoustically-modified self-imitation	18
3.1 Shadowing	18
3.1.1 Definition	18
3.1.2 Procedure.....	19
3.1.3 Materials.....	20
3.1.4 Effectiveness	21
3.1.5 Target audience	22
3.1.6 Pedagogical implications.....	22
3.1.7 Limitations	23
3.2 Phonetic self-imitation through voice manipulation.....	24
3.2.1 Definition	24
3.2.2. Procedure.....	25
3.2.3. Materials.....	26
3.2.4. Effectiveness	26
3.2.5. Target audience	27

3.2.6. Pedagogical implications.....	28
3.2.7. Limitations	29
3.3 Comparison of the techniques.....	30
4. Proposal for an empirical study	32
4.1 Objectives	32
4.2 Methodology	33
4.2.1 Participants	33
4.2.2 Research design.....	34
4.2.3 Stimuli creation	36
4.2.4 Tests	36
4.2.5 Analyses	37
4.3 Expected results	37
5. Conclusions, future research and pedagogical implications.....	38
References	39

INDEX OF TABLES AND FIGURES

Table 1. Participants and English levels by technique (shadowing, self-imitation through PSOLA, and self-imitation through Artificial Intelligence)..... 33

Table 2. Step-by-step procedure for each of the techniques..... 36

Figure 1. Examples distinguishing between shadowing with small delay and repetition (adapted from Hamada, 2016a) 19

Figure 2. Examples distinguishing between simultaneous shadowing and repetition (adapted from Hamada, 2019)..... 20

Abstract

L2 pronunciation instruction has received limited attention in foreign language (FL) classrooms due to time constraints and strict syllabi, and teachers' limited knowledge of the current effective pronunciation techniques. With the recent technological advances, notably Artificial Intelligence (AI), some Second Language (L2) pronunciation training techniques have emerged. One such example is the use of shadowing and phonetic self-imitation with both Pitch-Synchronous Overlap and Add (PSOLA)-manipulated and AI-synthesised speech, which has provided teachers and students with opportunities to improve L2 fluency and comprehensibility, two of the main goals of FL instruction. While both techniques have advantages as out-of-class activities, they face challenges regarding their application in classrooms and which of the three techniques is most effective for different types of learners. Given these limitations, an empirical study is proposed to examine these gaps in current L2 pronunciation instruction research.

Keywords: L2 pronunciation, shadowing, self-imitation, voice manipulation, comprehensibility.

Résumé du travail

1. Introduction

L'enseignement des langues étrangères a traditionnellement priorisé la grammaire et le vocabulaire, en faisant moins attention à la prononciation. Cette absence d'instruction est due à différentes raisons liées aux difficultés de la mettre en pratique dans les cours de langues secondes (L2).

Tout d'abord, les enseignants manquent souvent de temps pour couvrir le programme scolaire qui n'évalue pas toujours les compétences orales (Darcy, 2018 ; Tragant et al., 2014). Ensuite, à cause du grand nombre d'élèves par professeur, les enseignants pourraient affronter des difficultés pour fournir aux étudiants une rétroaction personnalisée. De surcroît, les professeurs ressentent souvent qu'ils manquent de formation pour enseigner efficacement la prononciation (Darcy, 2018).

Dans ce dernier siècle, l'enseignement de la prononciation a vécu un changement. Avant les années 1960, le principal objectif de la didactique de la prononciation était d'obtenir une prononciation native (Levis, 2005). Or, des recherches postérieures ont démontré que l'acquisition de la prononciation native était biologiquement liée aux effets de l'âge.

Par ailleurs, il serait pertinent de souligner la situation actuelle de l'anglais, considérée comme une lingua franca, qui permet la communication avec des interlocuteurs natifs et aussi avec des interlocuteurs non-natifs. Actuellement, l'enseignement de la prononciation privilégie la compréhensibilité et l'intelligibilité plutôt que l'instruction des accents natifs. La compréhensibilité pourrait être définie par

l'évaluation de la personne qui écoute quant à la facilité ou la difficulté de comprendre un énoncé (Derwing et Munro, 1995) et l'intelligibilité serait la mesure dans laquelle la personne qui écoute peut réellement comprendre le message de la personne qui parle (Derwing et Munro, 1995).

Avec cette transition dans l'instruction de la prononciation, l'objectif principal dans ce mémoire est d'examiner en détail les techniques du *shadowing* et de l'auto-imitation phonétique par la manipulation de la parole pour l'amélioration de la prononciation en L2. D'une part, le *shadowing* est une technique largement étudiée pour le développement de la compréhension orale, et dans une moindre mesure, la prononciation en L2. D'autre part, la méthode de l'auto-imitation de la voix modifiée est encore très récente comme résultat des développements de l'intelligence artificielle (IA).

2. Méthodes pour enseigner la prononciation en L2

Les étudiants et les professeurs pourraient ne pas être au courant de toutes les techniques à leur disposition pour travailler la prononciation (Wei, 2006). Comme résultat, la présentation des alternatives classées entre celles orientées ou non orientées vers la technologie serait appropriée.

En ce qui concerne les techniques non orientées vers la technologie, les paires minimales aident les apprenants à renforcer leur conscience phonologique, en distinguant les contrastes sonores qui sont souvent confus (Rakat et al., 2024). De plus, enseigner l'alphabet phonétique international facilite la connaissance de tous les sons présents dans les langues du monde. Considérant que chaque symbole représente un son (Guevara Gavilanes, 2024), cela permet aux étudiants de connaître la production correcte des sons qui ne sont pas présents dans leur langue maternelle. En outre, l'enseignement centré sur

la forme consiste à encourager les étudiants à prêter attention à la prononciation (soit implicitement soit explicitement) dans des activités communicatives orientées vers le sens (Saito, 2013). De surcroît, l'enseignement de la prononciation basé sur les tâches offre aux apprenants l'opportunité d'utiliser la prononciation correcte pour accomplir des tâches orientées vers le sens (Mora et Mora-Plaza, 2023). Finalement, l'autoévaluation et l'évaluation par les pairs favorisent la prise de conscience de la prononciation de leurs camarades, avec lesquels ils peuvent partager certaines caractéristiques phonologiques.

En ce qui concerne les méthodologies qui ont besoin de la technologie, le *high-variability phonetic training* améliore la perception et la production précise de sons où les apprenants doivent distinguer les sons présents dans la L2 (Uchihara et al., 2025). L'entraînement multimodal à la prononciation à travers des vidéos sous-titrées en L2 utilise de l'input audiovisuel pour pratiquer la parole en L2 (Wisniewska et Mora, 2020). Finalement, avec l'évolution de l'IA, des techniques de prononciation basées sur cette technologie ont émergé et sont en cours de développement. L'avantage de l'IA est qu'elle permet d'adapter les méthodes à chaque apprenant en fonction de ses nécessités et préférences individuelles. Par exemple, des méthodes telles que le *Automatic Speech Recognition*, les outils de *Text-to-Speech* et la *Computer-based Assisted Pronunciation Training* ont été démontrés d'améliorer significativement la prononciation (Sungkur et Shibdeen, 2025). De plus, les méthodologies du *shadowing* et l'auto-imitation par la manipulation de la parole travaillent sur ces nouvelles technologies.

3. L'imitation des voix externes vs. l'auto-imitation de la parole modifiée acoustiquement

3.1 *Shadowing*

Cette technique consiste à apporter un input sonore aux apprenants, tandis qu'ils le traitent cognitivement, ils doivent le répéter de la manière la plus similaire et la plus rapide possible (Luo et al., 2008) simultanément ou avec un petit décalage (Foote et McDonough, 2017).

Quelques études ont utilisé différents types d'input. Notamment, des séries télévisées populaires (Foote et McDonough, 2017), des *TED Talks* (Jin, 2023) ou encore des dialogues conversationnels quotidiens (Zafarova, 2025). Des applications et sites web offrent des matériaux additionnels pour pratiquer cette technique.

Le *shadowing* apporte des bénéfices pour perfectionner la fluence, l'intonation au niveau de la phrase, le rythme, la production des formes faibles, la compréhensibilité et la hauteur de la voix (Whitworth et Rose, 2025).

Des auteurs soutiennent l'idée qu'un certain niveau de maîtrise dans la L2 serait fondamental pour pratiquer le *shadowing* (Hamada, 2019). Cependant, les élèves de tous les niveaux (i.e., débutant, intermédiaire et avancé) peuvent pratiquer cette technique également. La seule problématique, c'est que les apprenants des niveaux A1-A2 pourraient trouver l'input auditif compliqué, ce qui peut risquer de les surcharger cognitivement.

Tandis que l'implémentation du *shadowing* dans une classe semble difficile, il paraît être une excellente alternative complémentaire qui peut se réaliser avec succès en dehors de la classe (Foote et McDonough, 2017).

En revanche, cette méthode présente quelques limitations. Les apprenants, qui considèrent le *shadowing* ennuyeux ou frustrant, vont facilement renoncer à sa pratique

(Foote et McDonough, 2017). De plus, non seulement le manque de feedback est une lacune du *shadowing*, mais aussi cette technique peut être cognitivement exigeante.

3.2 L'auto-imitation phonétique par la manipulation de la parole

La manipulation de la parole implique d'intégrer une voix synthétisée contenant les propriétés acoustiques de la prononciation native dans la parole des apprenants de L2. Alors, dans l'auto-imitation de la voix manipulée, les étudiants écoutent leur propre voix modifiée présentant les caractéristiques d'un accent natif, ils doivent la traiter cognitivement, et simultanément ou avec un petit décalage, et ils doivent répéter cet input auditif (Kusz et Pawliszko, 2026).

Il y a deux différentes stratégies pour modifier acoustiquement la propre voix pour après l'imiter : soit en imitant la voix manipulée manuellement (e.g. grâce au *Pitch-Synchronous Overlap and Add* - PSOLA), qui modifie les structures mélodiques et temporelles de la parole en fonction des paramètres consciemment choisis (Gluhareva et Angelovska, 2026), soit en reproduisant la voix générée automatiquement grâce à une IA qui modifie tous les éléments segmentaux et suprasegmentaux. Dans les deux cas, la qualité vocale est préservée. Pour créer manuellement la voix, Praat est le software qui s'utilise souvent (comme dans Skarnitzl et Bořil, 2024). Concernant la création automatique de la voix manipulée, il y a différentes plateformes qui travaillent sur l'IA, telles que le *Golden Speaker Builder* (utilisé dans Ding et al., 2019) ou des outils de *text-to-speech* comme *Revoicer* ou *Speechify* (utilisés dans Foote et McDonough, 2017).

Des études ont prouvé que l'imitation de sa propre voix transformée manuellement améliore la performance prosodique (Skarnitzl et Bořil, 2024) ainsi que la compréhensibilité et l'accent (Gluhareva et Angelovska, 2026). Selon l'étude de

Gluhareva et Angelovska (2026), la voix modifiée par le PSOLA et l'IA présentent des progrès similaires. À propos de l'auto-imitation par la voix manipulée par l'IA, il y a des résultats positifs sur la fluence et la compréhensibilité de la parole (Ding et al., 2019 ; Kusz et Pawliszko, 2026).

Cette technique peut être aussi appliquée pour tous les niveaux des apprenants. Mais deux limitations majeures se présentent dues au manque de recherche de cette technique. D'une part, les études ont analysé cette technique dans des niveaux intermédiaires et avancés (Kusz, 2022, 2023 ; Kusz et Pawliszko, 2026 ; Skarnitzl et Bořil, 2024), et beaucoup moins d'études ont examiné cette technique dans les niveaux débutants. D'autre part, ces études ont observé cette pratique dans des adultes jeunes (i.e., entre 19 et approximativement 26 ans). De plus, ces recherches ont utilisé des participants qui avaient été inscrits dans des matières de phonétique et de linguistique.

En outre, l'application de cette technique dans une classe serait très compliquée. Néanmoins, l'entraînement phonétique par l'auto-imitation vocale modifiée par le PSOLA ou l'IA pourrait être une excellente ressource comme une activité de devoirs.

Cette technique pourrait présenter certaines restrictions. Les enseignants et les apprenants pourraient rencontrer des difficultés pour obtenir la voix modifiée à cause de la nécessité des connaissances digitales pour la modification à travers le PSOLA. De plus, le degré de contrôle des propriétés acoustiques varie en fonction de la méthode utilisée. Alors, dans le PSOLA, le manipulateur a tout le contrôle des caractéristiques qui seront modifiées (Gluhareva et Angelovska, 2026). Par contre, l'IA change automatiquement toutes les caractéristiques. L'usage de l'IA pourrait susciter de la méfiance due à des problèmes associés à l'emploi non éthique de ces technologies pour commettre des

fraudes. De surcroît, cette technique offre un feedback partiel. Finalement, l'IA peut générer une voix robotique, ce qui peut impacter négativement l'efficacité des résultats.

3.3 Comparaison des techniques

Le *shadowing* semble la technique la plus réalisable grâce à l'accessibilité aux matériaux. Mais, l'auto-imitation par le PSOLA serait la technique la plus difficile à appliquer parce qu'elle requiert des compétences technologiques avancées. Concernant l'IA, elle serait plus facile à pratiquer que le PSOLA mais plus difficile que le *shadowing*.

Toutes les techniques peuvent améliorer la prononciation, en revanche, l'auto-imitation pourrait être une meilleure alternative grâce aux avantages d'utiliser sa propre voix (Kusz, 2022). De plus, l'efficacité des techniques dépendrait de la compétence linguistique des apprenants.

Mais ces techniques présentent des limitations. Le *shadowing* fournit peu de feedback et offre un contrôle limité des paramètres prosodiques. Le PSOLA présente un contrôle complet et l'IA un contrôle limité. En outre, l'IA et, en certaines occasions le PSOLA, pourraient créer une voix robotique.

4. Proposition d'une étude empirique

4.1 Objectifs

L'objectif principal de cette étude est d'aborder quelques trous qui existent à propos du *shadowing* et de l'auto-imitation phonétique modifiée par le PSOLA et l'IA dans le champ de l'apprentissage de la prononciation en L2. Notamment, cette étude est destinée à examiner les effets des trois techniques (le *shadowing* et l'auto-imitation vocale

modifiée par le PSOLA et l'IA) en ce qui concerne la compréhensibilité, la fluence et l'accent dans des apprenants d'anglais comme L2 de niveaux intermédiaire et avancé.

4.2 Méthodologie

Le nombre de participants serait d'environ 60 distribués selon le niveau de maîtrise de la langue (selon le CECR) et la technique utilisée. Par conséquent, 10 participants de niveau B1 et 10 pour le C1 pour chaque technique (i.e., *shadowing*, PSOLA et IA).

Les participants devraient pratiquer leur technique assignée au moins 10 minutes par session, 4 fois par semaine et pendant 8 semaines. La procédure serait similaire pour tous les étudiants, indépendamment de la technique employée. Ceux qui s'engagent dans le *shadowing*, imiteraient des extraits des séries ou des films. Par contre, ceux qui sont dans l'auto-imitation par le PSOLA, imiteraient une version déjà modifiée de leur propre voix, dans laquelle la durée syllabique aurait été manipulée. Ceux dans l'auto-imitation par l'IA, suivront le même processus, sauf que la modification aurait été réalisée à travers une IA qui aurait modifié automatiquement toute la prosodie. Les logiciels utilisés seraient Praat (concernant le PSOLA) et *Speechify* (pour manipuler la voix à travers l'IA).

Afin d'évaluer l'état initial et les améliorations en compréhensibilité, fluence et accent, il serait pertinent que les élèves lisent un texte ou produisent un texte oral court en différents moments de l'étude (i.e., au pré-test, au post-test et au post-test différé). De plus, des parlants natifs d'anglais seront nécessaires pour déterminer l'amélioration des étudiants en utilisant une échelle de Likert en 9 points.

4.3 Résultats attendus

D'un côté, les apprenants des niveaux intermédiaires pourraient trouver l'activité difficile. En revanche, les étudiants avancés ne devraient pas rencontrer de difficultés majeures. D'autre côté, on pourrait penser que les apprenants qui vont bénéficier le plus sont les apprenants intermédiaires parce qu'ils ont beaucoup de possibilités d'amélioration. Par contre, les apprenants avancés ont déjà une maîtrise de la langue et montrent souvent une stabilisation de la progression en L2. De plus, on pourrait expecter que la technique d'auto-imitation serait plus avantageuse que le *shadowing* comme Kusz (2022) a remarqué.

5. Conclusions, recherches futures et implications pédagogiques

Le *shadowing* et l'auto-imitation par la manipulation par le PSOLA et par l'IA semblent de techniques appropriées pour perfectionner la prononciation en L2. Ces techniques permettent d'améliorer la compréhensibilité et la fluence, justement les paramètres priorisés dans l'enseignement de la prononciation des L2.

Des recherches plus approfondies seraient nécessaires pour résoudre les trous dans le champ de recherche de ces techniques. Des investigations expérimentales n'ont pas encore déterminé le niveau de la langue qui va bénéficier le plus de cette pratique, ni laquelle de ces techniques est la plus avantageuse pour améliorer la compréhensibilité, la fluence, et l'accent. Conséquemment, la proposition de l'étude empirique pourrait être un point de départ adéquat. Il serait pertinent d'analyser ces techniques pour des apprenants de différents âges et d'apprenants qui n'ont pas des connaissances préalables en linguistique.

1. Introduction

Although pronunciation has generally been considered important by both teachers and learners and should ideally be taught in connection with other language skills (Darcy, 2018), grammar and vocabulary have conventionally received more attention in second language (L2) instruction than pronunciation. Furthermore, pronunciation has often been perceived as a “supplementary activity rather than a central part of the syllabus” (Cenoz & García Lecumberri, 1999, p. 4). Several reasons may explain why pronunciation is difficult to implement in L2 classroom contexts.

Firstly, teachers often face significant time constraints when covering the already overloaded syllabus (Darcy, 2018; Tragant et al., 2014). Secondly, the lack of pronunciation and oral skills assessment in language curricula could also contribute to its lack of attention. For example, the English exam in the *Proves d'Accés a la Universitat* (PAU) (Tragant et al., 2014) consists of three parts: 30% of the final mark corresponds to oral comprehension, 30% for written comprehension, and 40% for written production (Canal Universitats, 2026). Remarkably, oral production is entirely absent in this exam. As a result, it would not be surprising that if teachers do not have the time to cover the entire syllabus and pronunciation is not evaluated, they may not focus on pronunciation in their classes, and they will instead spend more time and effort on teaching grammar and vocabulary, which do play an essential role in this official exam (Tragant et al., 2014).

Finally, other practical matters result in the marginalisation of pronunciation. For instance, providing individualised pronunciation feedback might be challenging, especially in those classroom contexts where there might be a high student-teacher ratio or where teachers might lack the necessary knowledge to teach pronunciation effectively.

Another factor may be the belief that pronunciation will be naturally acquired without the need for explicit instruction (Darcy, 2018).

In the last century, there has been a transition in the way pronunciation is taught. Before the 1960s, pronunciation instruction of a foreign language (FL) aimed at teaching native-like pronunciation in foreign languages (Levis, 2005). However, later research challenged this goal by demonstrating that nativeness was biologically related to age effects. Therefore, this made it particularly difficult for adult learners to achieve native-like pronunciation in an L2. For instance, as shown in Flege et al. (1995), “the number of subjects (...) who met the statistical criterion for "authentic" [native-like] pronunciation (...) decreased steadily as AOL [age of learning] increased” (p. 3129). That is to say, the likelihood of achieving native-like pronunciation decreased as the age of learning an L2 increased. Thus, non-native speakers of English, who began studying English after the age of 15, were reported to be foreign-accented. In contrast, early L2 learners were more likely to achieve native-like pronunciation.

For example, the Audiolingual Method was one of the old methods used in pronunciation instruction before this paradigm shift. This technique required L2 learners “to listen to and mimic native-speaker models” to acquire accurate pronunciation, which at the time was considered essential for successful L2 acquisition (Munro, 2018, p. 268).

The limitations of traditional pronunciation instruction methods, together with the emergence of the English as a Lingua Franca paradigm (i. e., a common language used as a means of communication among individuals who do not share the same first language (L1) (Baker, 2018)) contributed to this transition. In this context of using English not only with native speakers but also with non-native speakers, the use of native accents appeared

to be unnecessary. Therefore, currently, a successful conversation is one in which interlocutors are both comprehensible and intelligible, regardless of native-like features. In other words, nowadays, comprehensibility and intelligibility have become the main objectives of pronunciation instruction. Comprehensibility is understood as the listener's assessment of how easy or difficult it is to understand an utterance (Derwing & Munro, 1995), and intelligibility is defined as the degree to which a listener can actually understand a speaker's message, yet there is no widely recognised way of assessing it (Derwing & Munro, 1995). Building on this, communication can also be effective when foreign accents are strongly noticeable; that is, accent and understanding are not clearly correlated (Levis, 2005; Munro & Derwing, 1999). Hence, current pronunciation instruction should focus on those segmental or suprasegmental elements that may enhance understanding and should de-emphasise those that do not impede communication (Levis, 2005).

Given the general lack of attention to the field of L2 pronunciation in the last decades, teachers and learners may not be fully aware of the available pronunciation tools at their disposal (Wei, 2006), both inside and outside the classroom. A wide range of non-digital techniques (e.g., explicit instruction, drills, etc.) and digital ones (e.g., the Golden Speaker, the YouGlish website, etc.) are accessible for both students and instructors to help improve learners' FL pronunciation.

In light of this shift from native-like pronunciation to intelligibility and comprehensibility, there is a need to develop and encourage the use of techniques that prioritise these latter goals. Therefore, the purpose of this dissertation is to thoroughly examine the shadowing and the phonetic self-imitation through voice manipulation, two

techniques used to improve L2 pronunciation. As proven by Foote and McDonough (2017) regarding shadowing and by Kusz and Pawliszko (2026) concerning phonetic self-imitation through AI voice cloning, both techniques enhance fluency and comprehensibility rather than accentedness, which are the main aims in current pronunciation instruction. Additionally, in both cases, learners have to listen to an audio input, so not only is pronunciation practised, but also listening skills are developed in the two methodologies.

Building on this, one of the main positive effects of pronunciation instruction is the use of one's own voice. Familiar sounds (e.g., one's own voice) are easier to be cognitively processed (Holmes et al., 2018). Those learners engaging in self-imitation through voice manipulation may have an advantage over those imitating an external speaker, since self-imitation allows students to preserve vocal identity, supports L2 phonetic perception (Kusz & Pawliszko, 2026), and promotes the internalisation of L2 phonetic patterns (Adank et al., 2009).

Moreover, the analysis of both strategies would be of particular relevance for addressing the existing gap in the field of L2 pronunciation acquisition. This gap may be explained by, on the one hand, shadowing has been widely studied as a technique to promote L2 listening skills (Hamada, 2025) and, to a lesser extent, L2 pronunciation development (Hamada & Suzuki, 2024). On the other hand, phonetic self-imitation through voice manipulation is a more recent approach that has emerged as a result of Artificial Intelligence (AI) advancements. These developments have made this technique accessible to learners because it offers additional opportunities for pronunciation practice and improvement while supporting their autonomous learning.

2. L2 pronunciation instruction techniques

L2 speakers of English have various techniques at their disposal aimed at improving their FL pronunciation. One potential classification of these strategies is into technology-oriented and non-technology-oriented pronunciation techniques. In the former, pronunciation can be practised through the use of digital tools, whereas in the latter, the use of technologies is not required.

2.1 Non-technology-oriented pronunciation techniques

Various pronunciation techniques rely on traditional approaches without the use of digital resources; one of them is minimal pair¹ training. This technique increases phonological awareness by helping learners perceive sound contrasts in minimal pair sets and enabling L2 speakers to distinguish words that are usually confusable (especially segmental sounds) (Rakat et al., 2024). As a result, this training enhances learners' mastery of pronunciation and would allow them to know how to pronounce words correctly.

Another widely used technique involves teaching the International Phonetic Alphabet (IPA). The IPA is a standardised collection of 44 symbols that systematically depict the world's language sounds, each symbol faithfully representing one different sound (Guevara Gavilanes, 2024). Moreover, phonetic notation can be used to show aspects of speech that are not visible through the simple use of orthography, such as lexical stress, intonation, and connected speech processes (Mompean & Fouz-González, 2021). Hence, the IPA allows L2 learners to master the proper pronunciation of words

¹ Minimal pairs are pairs of words that differ only in one sound (e.g., *sheep* /'ʃi:p/ vs. *ship* /'ʃɪp/).

and sounds that may not be present in their native language, enabling students to practise the perception and, consequently, the articulation of English sounds.

Another strategy used in pronunciation training is form-focused instruction. It is defined as a pedagogical approach to draw learners' attention to L2 pronunciation (explicitly or implicitly) in meaning-oriented communicative activities (Saito, 2013). This method has been proven to positively influence learners' developing L2 phonological systems while going from more controlled to more spontaneous communicative tasks.

Concerning communicative approaches, task-based pronunciation teaching (TBPT) improves learners' attention to phonetic form in educational activities that involve meaning-based interaction, and integrates the principles of task-based language teaching (Mora & Mora-Plaza, 2023). In TBPT, pronunciation targets must be used for learners in order to complete the task successfully, forcing learners to produce these sounds appropriately (Mora & Mora-Plaza, 2023). TBPT has been found to lead to improvements in the perception and production of L2 segmental and suprasegmental features.

Finally, pronunciation self- and peer-assessment also play an important role in pronunciation training. Self-assessment is the ability to evaluate one's own oral performance. Peer-assessment refers to (usually) a classmate assessing another classmate's performance. Self-assessment and peer-assessment have been shown to raise awareness of L2 pronunciation features and enhance learning. Pronunciation assessment would be more complete and fairer by including the figure of a peer-rater in the self-assessment and teacher-assessment perspectives since some sounds are difficult to self-

assess, some L2 sounds are perceived incorrectly, and learners tend to overestimate or underestimate their own performance (Edo-Marzá, 2017; Mora-Plaza & Ortega, 2026). As a result, students would be more critical and more aware of their classmates' pronunciation, who may share similar pronunciation difficulties as theirs.

2.2 Technology-oriented pronunciation techniques

Technology-oriented techniques depend on digital tools to provide students with the necessary input, feedback, and opportunities to practise their oral skills in order to improve their pronunciation inside and outside the classroom. One of the widely used approaches is high-variability phonetic training, a perceptual training that enhances perception and production accuracy, where learners need to distinguish the target L2 sounds (Uchihara et al., 2025). Not only do learners receive varied input of L2 sounds produced by multiple talkers (optimally around 6), but they also receive explicit feedback on how accurately they perceive the sounds (Uchihara et al., 2025).

Moreover, the multimodal pronunciation training through L2-captioned video uses audiovisual input (i. e., viewing videos with captions in the target language) for training L2 speech processing (Wisniewska & Mora, 2020). Students exposed to captioned viewing have shown improvements in pronunciation, speech segmentation, and speech processing skills faster and more efficiently.

With the rise of AI, AI-driven pronunciation training methodologies have emerged and are developing. AI-powered tools can provide individualised pronunciation training according to students' preferences and needs. Some methods, such as Automatic Speech Recognition, Text-to-Speech, and Computer-Assisted Pronunciation Training,

have been demonstrated to significantly improve learners' pronunciation accuracy (Sungkur & Shibdeen, 2025). Moreover, other AI-assisted tools, such as shadowing, defined as the immediate repetition of an auditory stimulus to improve L2 pronunciation, and phonetic self-imitation through voice manipulation, the repetition of one's own synthesised voice with native-like acoustic characteristics, will be further developed in section 3.

3. Imitating others' voices vs. acoustically-modified self-imitation

3.1 Shadowing

3.1.1 Definition

Shadowing is an activity where L2 learners imitate a given speech stimulus “as closely and quickly as possible” (Luo et al., 2008, p. 2807), simultaneously or with a small delay (Foote & McDonough, 2017). Hamada (2019) illustrates and defines the strategy as “a learner repeats what he/she is listening to, just as a shadow follows someone walking” (p. 387).

This technique was first used for the training of simultaneous interpreters (Lambert, 1992). Then, in the 1990s, it was established in the Japanese EFL context as a teaching technique for listening improvement (Hamada, 2019). Since then, it has spread globally (Hamada, 2019) as a tool for both L2 listening and pronunciation development. However, research has focused more on the benefits of shadowing for listening comprehension (Hamada, 2019) than for pronunciation development.

3.1.2 Procedure

During shadowing, first, learners receive a speech input, and while cognitively processing the input, they have to repeat what is heard. Shadowing requires paying attention to sounds (both segmental and suprasegmental features). Moreover, the lag between the audio input and the learner's output has to be minimal (Hamada, 2016a). If not, then the technique would not be shadowing but listen-and-repeat, in which the lag between the stimulus and the repetition is more prolonged, and the implications are different. Therefore, in simultaneous shadowing, learners direct their attention to sounds uniquely. In contrast, listen-and-repeat learners may focus on both sounds and meaning, making it more cognitively demanding (Hamada, 2016a). Figure 1 and Figure 2 illustrate examples of both shadowing and listen-and-repeat.

SHADOWING

Audio: English learning requires constant practice to improve

Learner: English learning requires constant practice to improve

REPETITION

Audio: English learning requires constant practice

Learner: English learning requires

Figure 1. Examples distinguishing between shadowing with small delay and repetition (adapted from Hamada, 2016a)

SHADOWING

Audio: English learning requires constant practice to improve

Learner: English learning requires constant practice to improve

REPETITION

Audio: English learning requires constant practice

Learner: English learning requires

Figure 2. Examples distinguishing between simultaneous shadowing and repetition (adapted from Hamada, 2019)

3.1.3 Materials

In general terms, in order to practise shadowing, there is a minimum technological requirement, that is, learners should have access to an electronic device (e.g., computer, phone, radio, etc.) to be able to receive an auditory input to afterwards mimic it.

As suggested in some studies (Foote & McDonough, 2017; Sumiyoshi & Svetanant, 2017), motivation and enjoyment are closely related to the continuity of shadowing practice. Lack of motivation and high boredom with the activity were factors attributed to the quitting of shadowing practice by some students. Therefore, the selection of materials is of essential importance. What should be prioritised is the use of material that seems interesting to learners, that is to say, students' individual preferences ought to be considered in order to make shadowing an appealing task for them, and to try to avoid them giving up their language learning practice.

Some studies have used different types of input. For example, popular television sitcoms such as *Friends*, *The Big Bang Theory*, or *How I Met Your Mother*, among others (Foote & McDonough, 2017), TED Talks (Jin, 2023), or even everyday conversational

dialogues (Zafarova, 2025) were used as input. Furthermore, currently, mobile apps (e.g., ELSA Speak, Tubeshad, etc.) and internet videos and websites (e.g., the *parrot.app* website and *parrot shadowing* on YouTube) include materials for shadowing practice.

3.1.4 Effectiveness

In spite of the strategy facing criticism for being closely associated with repetition and for its resemblance to audio-lingual drills (Hamada, 2021), often compared to a “meaningless parrot-like practice” (Bovee & Stewart, 2009, p. 20), research has demonstrated that it is a useful technique for L2 pronunciation enhancement (Sari & Jaya, 2025).

Studies generally agree on the benefits that shadowing offers. On the one hand, shadowing is an effective tool for enhancing listening comprehension because it actively supports students in processing and understanding spoken language (Bapir & FageAbdulla, 2025). In addition to that, training enables learners to recognise connected speech processes, phonetic reductions, and prosodic features (Thao & LY, 2025). Another benefit is learners’ improvement in phoneme perception, which allows learners to recognise words and vocabulary in rapid speech (Hamada, 2016b; Thao & LY, 2025). As for what shadowing for speaking and pronunciation concerns, Whitworth and Rose (2025)’s systematic review shows that shadowing improves fluency, sentence-level intonation, rhythm, the production of weak forms, comprehensibility, and shows some mixed results regarding pitch enhancement.

3.1.5 Target audience

Numerous studies (Dang & Tran, 2025; Foote & McDonough, 2017) have analysed the impact of shadowing for pronunciation improvement in university contexts where the vast majority of the participants had a certain degree of knowledge of English, some of them were even English major students.

Different authors have defended the position that “beginner level learners should start from standard shadowing for listening. Once learners achieve the upper-intermediate or advanced level, they are ready for shadowing for speaking” (Hamada, 2019, p. 391). That is to say, in order to engage in shadowing, a certain high degree of proficiency in the language is required.

Nevertheless, shadowing can be practised by beginner (Oanh & Tri, 2019), intermediate (Jin, 2023), and advanced (Sumiyoshi & Svetanant, 2017) L2 learners. The difference between each language level could be that learners may find the audio difficult, and shadowing may appear to be an overloading task. Despite this, when it comes to L2 acquisition, the quantity of input matters, and shadowing may offer learners the possibility to, on the one hand, receive auditory input and, on the other hand, practise and be aware of pronunciation.

3.1.6 Pedagogical implications

Its implementation inside the classroom could be challenging. On the one hand, it is difficult to make all students individually practise shadowing because there is little time in classes and, as already mentioned in section 1, other aspects of language learning are usually prioritised before pronunciation; therefore, it would be very time-consuming. On

the other hand, the alternative of making everyone practise shadowing at the same time (i.e., choral shadowing) is not an ideal solution since no individualised feedback may be provided, the classroom may not meet the ideal conditions to attentively listen to an audio input due to the presence of background noise, for instance, that may interfere with the process of performing the task correctly, or even students' anxiety about practising oral pronunciation in front of classmates may affect the effectiveness of shadowing training.

Nonetheless, shadowing appears to be an excellent tool for those learners who do not have access to formal pronunciation instruction, as they do not have a teacher or a partner with whom to practise pronunciation skills. Moreover, it is a low-cost activity that can be successfully conducted outside of the classroom (Foote & McDonough, 2017). Nevertheless, it should be noted that shadowing itself should not be used to replace pronunciation instruction inside the classroom (Foote & McDonough, 2017), but rather as a complement for those who desire extra training.

3.1.7 Limitations

Motivation plays an essential role in the continuity of shadowing practice (Foote & McDonough, 2017). Therefore, those learners who find the activity boring and frustrating may easily decide to give up practising, as suggested in Foote and McDonough (2017). One possible solution to this problem is to try to make the activity as enjoyable as possible by using appealing materials, to encourage learners to continue training despite not noticing their own pronunciation improvement, etc.

One of the limitations of shadowing training is the lack of feedback. No feedback regarding how accurately the sounds are produced is provided. This lack of assessment

may lead to the fossilisation of mistakes, as the students may not be aware of their own mistakes, and they will internalise the incorrect production of these sounds, believing it is the correct way. For example, L1-Spanish speakers who are L2 English learners might not notice contrasts in minimal pairs, believing that *beat* /'bi:t/ and *bit* /'bɪt/ are pronounced exactly the same because both /i:/ and /ɪ/ are assimilated to /i/ in Spanish. In some studies (Foote & McDonough, 2017), learners practising shadowing were asked to self-record themselves in order to, once the repetition task had finished, listen to themselves to notice the mistakes they made. This self-recording is not a replacement for a teachers' correction, but at least, it can give learners hints to know the correctness of their pronunciation.

Furthermore, shadowing can be a cognitively overloading task, particularly for beginner learners who may not yet be accustomed to this type of practice. In this respect, shadowing training requires constant practice. It may be beneficial to engage in short but frequent practice sessions at different points throughout the week. Without regular practice, improvements in pronunciation may be less noticeable and more difficult to attain. Likewise, progress in pronunciation imitation may occur at a slower rate, potentially delaying overall pronunciation development. For instance, in the study by Foote & McDonough (2017), learners were asked to practise shadowing at least four times per week for ten minutes each time, and improvements in pronunciation were observed.

3.2 Phonetic self-imitation through voice manipulation

3.2.1 Definition

Voice manipulation involves modifying the acoustic characteristics of L2 speech. In self-imitation of a manipulated voice, L2 speakers listen to their own modified voice with native-accented features and mimic this self-auditory input (Kusz & Pawliszko, 2026). This approach is also called the “golden speaker” and it is used as a pronunciation teaching technique.

There are two different strategies to self-imitate an acoustically modified version of one’s voice: imitating a manually manipulated voice (e.g., through Pitch-Synchronous Overlap and Add - PSOLA), which modifies melodic and temporal patterns of speech (e.g., pitch, intonation, duration, etc.) according to the parameters adjusted (Gluhareva & Angelovska, 2026), or imitating an AI-generated voice (e.g., the Golden Speaker Builder), which automatically changes all segmental and suprasegmental speech features. In both voice manipulation through PSOLA or through AI, voice quality is preserved.

3.2.2. Procedure

Before engaging in self-imitation, learners need to obtain their modified voice in the form of an audio input. So, first, they need to record themselves through their phone or Praat software doing a specific task. To manually manipulate their speech through PSOLA, students may need the aid of an expert. In contrast, to modify the speech through an AI, it can be done autonomously. Second, in order to engage in the practice, they must necessarily pay attention to this manipulated speech input, process it cognitively, and, with a small delay or simultaneously, mimic the model audio.

3.2.3. Materials

As previously mentioned, motivation has a strong impact on learners' perseverance in repetitive tasks. Therefore, once again, learners' own topics of interest may be relevant in the creation of materials used for self-imitation. Consequently, students would not perceive the task as monotonous, and they may engage more actively in the activity.

In order to create the manually manipulated voice which improves the target L2 speech features through PSOLA, Praat is the software used (as in Skarnitzl & Bořil, 2024)². Concerning automatic manipulated speech, there are different platforms that support AI voice cloning, although pronunciation instruction is not the primary aim of these platforms, with the exception of the *Golden Speaker Builder*. Examples of such platforms include text-to-speech resources such as *Revoicer* and *Speechify* (used in Kusz & Pawliszko, 2026), the *Golden Speaker Builder* (used in Ding et al., 2019), or even other resources that have not been used in research, such as *ElevenLabs* or *Voices AI: Voice Changer*, among others. Once the manipulated speech has been obtained, self-imitation can be implemented in the foreign language classroom through the use of intelligent technological devices (e.g., computers, smartphones, tablets).

3.2.4. Effectiveness

Research on self-imitation through manually manipulated speech reported enhancements in prosodic performance (Skarnitzl & Bořil, 2024) and gains in comprehensibility and accentedness (Gluhareva & Angelovska, 2026). According to a recent study by Gluhareva

² A step-by-step tutorial so that users can familiarise themselves with PSOLA's procedure through Praat. Available online: <https://fonetika.ff.cuni.cz/en/research/from-our-research/psola-modification/>

and Angelovska (2026), both PSOLA and AI-modified speech have been found to result in similar gains in terms of comprehensibility and accentedness.

Additionally, regarding AI-driven phonetic self-imitation, it has also had positive outcomes in L2 pronunciation improvement, particularly concerning fluency and comprehensibility (Ding et al., 2019; Kusz & Pawliszko, 2026). After 3 weeks of training in Ding et al. (2019) or after 8 weeks of practising three times a week for 15 minutes in Kusz and Pawliszko (2026), learners' L2 speech became significantly more comprehensible and fluent. However, researchers reported no improvements regarding accentedness, which does not seem to be very concerning as comprehensibility itself is considered key for successful communication.

3.2.5. Target audience

Phonetic self-imitation through voice manipulation can be practised by all types of learners. Nonetheless, two major problems arise from the lack of research in this pronunciation technique. On the one hand, while the majority of studies have focused on intermediate and advanced learners of English (B1, B2+, and C1) (Kusz, 2022, 2023; Kusz & Pawliszko, 2026; Skarnitzl & Bořil, 2024), far less research has explored the effectiveness of self-imitation in beginner learners. On the other hand, these studies have investigated young adult learners' (i.e., participants' age ranged from 19 to approximately 26 years old). Moreover, the aforementioned studies targeted students of Applied Linguistics at university or students enrolled in English Phonetics and Phonology courses, except for Skarnitzl & Bořil (2024) where students were not registered for phonetics courses. Therefore, further studies could be carried out to investigate the efficacy of self-

imitation through voice manipulation among other age groups and among other populations who do not have any previous knowledge of phonetics or linguistics.

It can be hypothesised that upper-intermediate levels may benefit more from the technique than beginner learners, as they already have a good basis of the language. Furthermore, certain audiences, such as elderly learners who may not be very familiar with new technologies, may encounter difficulties both in generating their own modified version of their voice and in using such tools effectively. A potential solution could be to provide students with an explanation inside the classroom on how to obtain and practise imitating their own-modified voice.

3.2.6. Pedagogical implications

Phonetic self-imitation appears to be difficult to implement inside a classroom context since the classroom may not meet the ideal conditions to carry out this type of practice. Firstly, it is tailored training, so no choral activities could be conducted because that would be a shadowing activity rather than a self-imitation one. Moreover, another major problem would be that, if all learners were practising their own golden speech at the same time, there would be excessive noise and learners would not be able to concentrate.

Despite its potentially challenging implementation in the classroom, both manually and automatically manipulated phonetic self-imitation might be an outstanding tool as a homework activity for those who seek extra pronunciation practice. Nonetheless, teachers can still explain during class time how pronunciation may be practised through the two techniques, as well as directing students' attention to specific segmental and

suprasegmental features of speech that can already be trained inside the classroom (e.g., stress, intonation, or difficult vowel and consonant sounds).

3.2.7. Limitations

Although imitation of manually-manipulated and AI-manipulated speech have proven to be successful tools to improve L2 pronunciation, some limitations may arise concerning the process of obtaining the manipulated speech, the degree of control over the speech features, data protection and privacy concerns, partial lack of feedback, and the generated robotic-sounding voice.

One of the main problems that learners may encounter is the issue of obtaining one's own synthesised voice. Both students and teachers may not possess the necessary digital skills to be able to create this artificial speech. In this respect, due to its automatic nature, AI may be more practical for such users (Gluhareva & Angelovska, 2026).

Regarding the use of AI tools, as they automatically transform all the segmental and suprasegmental features, users do not have the control over the parameters that have been adapted. On the contrary, concerning PSOLA, the speech is modified according to the parameters that users have manually adjusted, thus, in this case, the modification of the speech features (e.g., pitch, intonation, etc.) is more controlled (Gluhareva & Angelovska, 2026).

Although AI voice cloning can be effectively used as an L2 pronunciation enhancement tool, it could also be used unethically, raising privacy and security concerns. For example, voice-cloning technology may be misused to generate artificial audio recordings for fraudulent purposes. Therefore, it would not be strange that students may

prefer to use and rely on other non-AI-based alternatives to improve their L2 pronunciation.

Furthermore, phonetic self-imitation through manipulated speech is defined as personalised pronunciation feedback (Kusz & Pawliszko, 2026). Learners can notice how their pronunciation should sound by applying native speakers' characteristic acoustic features. Nevertheless, after students repeat this synthesised voice, they do not receive further feedback regarding how well they are producing these sounds. So, if they do not perceive and produce accurately after this artificial speech, no one would correct the mistakes they continue to make, which may lead to fossilisation.

Finally, the use of an AI-generated robot-like sounding synthetic voice may affect the efficiency of the outcomes. Still, AI tools such as text-to-speech technologies have not yet achieved an optimal development regarding prosody, in addition to being considered less natural and less accurate-sounding than human speech (Bione & Cardoso, 2020).

3.3 Comparison of the techniques

In terms of the feasibility of implementation, shadowing appears to be the most feasible technique in both inside and outside the classroom context. It appears to be the easiest technique for homework activities and for in-class pronunciation practice because numerous materials have already been created and are accessible for shadowing. In contrast, PSOLA seems to be the most difficult technique to apply because it requires advanced phonetic and phonological knowledge as well as technological skills to manually manipulate speech. In the case of AI-modified speech, it looks easier to be

implemented as a homework activity than PSOLA, but still harder than shadowing. Therefore, this shows that tailored training is not the most suitable technique for in-class activities.

With respect to the effectiveness of the three techniques, all of them have been proven to improve L2 learners' pronunciation. Comparing imitation of others' speech vs. self-imitation, Kusz (2022) reported that self-imitation may be more advantageous than imitation practice regarding at least speech rate. Moreover, self-imitation uses a familiar voice, which has been shown to be beneficial for cognitive processing and L2 phonetic perception and pronunciation retention (Holmes et al., 2018; Kusz & Pawliszko, 2026). However, further studies should be carried out to thoroughly examine the effectiveness of the three techniques. Along the same lines, the efficacy of the techniques could depend on learners' L2 proficiency level.

Both imitation and self-imitation techniques face some limitations. While shadowing seems to fail at providing students with feedback and does not control the acoustic features that learners would improve, self-imitation through both PSOLA and AI speech modifications automatically supplies learners with partial feedback on their own pronunciation. In this instance, the level of control differs between the two self-imitation techniques. While self-imitation through PSOLA presents total control because acoustic features are consciously modified, AI also lacks control because all speech features are automatically changed. Finally, one of the main drawbacks that AI manipulation might emphasise is the production of an artificial, robotic-sounding voice. PSOLA manipulation could also create a robotic voice, but as it only changes some features and not all of them, it is less likely to happen. In contrast, shadowing uses authentic voices. Therefore, despite

having a personalised technique which uses someone's own voice (i.e., PSOLA and AI self-imitation), which seems to be more beneficial, this does not ensure that these techniques might present less limitations.

Once the three techniques have been compared, there is a gap regarding the type of learners who would benefit the most from these techniques. Additionally, there is limited information related to the comparison of shadowing and self-imitation through PSOLA and AI manipulation. Therefore, there is very little information about which technique may be more effective at enhancing L2 pronunciation, and which type of learners would benefit the most.

4. Proposal for an empirical study

4.1 Objectives

The main aim of this empirical study proposal is to address some of the existing gaps in the current literature regarding two technology-oriented pronunciation training techniques: (1) shadowing of others' speech, and (2) imitation of one's own manipulated voice. As it has been found through the realisation of this dissertation, the two techniques still remain under-researched (most notably self-imitation through speech modification). In addition, there is no general consensus regarding the type of learners that will benefit the most from these two techniques; therefore, it would be interesting to examine L2 English learners at different levels to understand which group of learners would benefit most from each technique.

As a result, the study would aim at investigating the effects of three different L2 imitation techniques (i.e., shadowing, self-imitation through PSOLA voice manipulation,

and AI speech modification) on global dimensions of L2 speech (i.e., comprehensibility, fluency, and accentedness) across L2 intermediate and advanced L2 learners of English.

4.2 Methodology

4.2.1 Participants

The participants of the experimental study would be approximately 60 L1-Catalan/Spanish learners of L2 English. That is, a minimum of 10 participants for each of the following levels should be studied in order to analyse the effectiveness of the techniques per FL level (according to the CEFR framework): 10 B1 learners for the intermediate level and 10 C1 students for the advanced level of English. So, there will be 20 participants for each technique. Table 1 summarises the division of the participants by pronunciation training technique (shadowing, PSOLA, and AI) and level (B1 and C1).

Group 1: shadowing	Group 2: PSOLA	Group 3: AI
10 beginner learners (B1)	10 beginner learners (B1)	10 beginner learners (B1)
10 advanced learners (C1)	10 advanced learners (C1)	10 advanced learners (C1)

Table 1. Participants and English levels by technique (shadowing, self-imitation through PSOLA, and self-imitation through Artificial Intelligence)

The gender of participants would be balanced, and their age, socio-economic status, and extra-curricular exposure to L2 English would be similar in order to obtain comparable results across groups. One possibility could be conducting this study with adult learners in an EOI (*Escola Oficial d'Idiomes*) as long as access is permitted and if

collecting data from secondary school learners was difficult. Nevertheless, participants with hearing or speaking impairments would not be able to participate in this study.

4.2.2 Research design

Ideally, learners should practise their assigned technique for a minimum of 10 minutes per session, four times a week, for eight weeks, following the implementation procedure by Foote & McDonough (2017).

The procedure would be similar for all participants regardless of the method they use. Firstly, those participants who are working with the shadowing technique would receive some short extracts from a B1-B2 series or film. They would not receive the transcripts, as they are not necessary for shadowing. Once they have the material, they would mimic the audio immediately after hearing it.

Secondly, learners engaging in self-imitation through PSOLA voice manipulation would receive the extracts' transcripts of the exact series or films used by group 1, so everyone would be mimicking the same content but through different means. In order to create their manually manipulated voice, they would have to read the transcripts aloud while their L2 speech is being recorded. Then, their audio would be modified in Praat by the researcher on syllable duration to manipulate rhythm, and make it more stressed-timed, as English is. Once they have their manipulated speech, they would have to shadow it.

Thirdly, L2 speakers using self-imitation through the AI voice cloning technique would also receive the transcripts of the same extracts used by groups 1 and 2. With the intention of creating their AI-generated speech, learners should record themselves reading

those scripts, and then, the researcher would clone learners' speech through an APP (e.g., Eleven Labs, Speechify, etc.). The program would generate this artificial voice matching native-like features in the speech given. After obtaining their modified speech, learners would have to mimic it. Table 2 summarises the step-by-step procedure for each of the techniques.

Finally, all participants would be required to record themselves while imitating the audio materials. Upon completing each practice session, participants would be instructed to submit these recordings to the researchers in order to verify task completion and ensure consistent pronunciation practice throughout the study.

	Group 1: shadowing	Group 2: PSOLA	Group 3: AI
Before imitation		1. They will receive the transcripts of the extracts.	1. They will receive the extracts' transcripts.
		2. Read the transcripts aloud while recording their voice.	2. Write the transcripts in Speechify
		3. Audio modification through Praat by the researchers.	3. Clone the voice through Speechify by the researchers.
During	1. They receive extracts and they will	4. Once they have their modified voice, they	4. Once they have their modified voice, they

imitation	have to shadow them immediately after and record their speech.	shadow it immediately after and record their speech.	shadow it immediately after and record their speech.
After imitation	2. They should send the recorded speech samples to the researcher.	5. They should send the recorded speech samples to the researcher.	5. They should send the recorded speech samples to the researcher.

Table 2. Step-by-step procedure for each of the techniques

4.2.3 Stimuli creation

In order to create the stimuli, Praat would be used to manually manipulate suprasegmental features in the speech such as syllable duration, and Speechify, which is an AI text-to-speech tool, would allow to automatically create the modified speech, and YouTube could be helpful for obtaining the extracts that participants engaging in a shadowing practice would use.

4.2.4 Tests

In order to evaluate students' initial and improvement levels in comprehensibility, fluency, and accentedness, they could be asked to read a text or to produce a short oral text from a specific prompt before and after the pedagogical intervention. Therefore, students should do these tasks at different moments during the study. Optimally, students would do the tests at home and record themselves before (pre-test) and after the

intervention (post-test). If possible, the same tasks would be repeated two months after in a delayed post-test to observe any retention effects.

4.2.5 Analyses

Native-English speaking raters would assess 30 seconds of the speech produced by the learners at pre-test and post-test on a 9-Point Likert scale on comprehensibility, fluency, and accentedness, in which 1 would stand for “very difficult to understand”, “not fluent at all”, and “strong foreign accented”; and 9 would correspond to “very easy to understand”, “very fluent”, and “no foreign accent at all”. Additionally, before starting the experiment, a training session would be offered to participants to show them instructions and essential information needed to complete the study.

4.3 Expected results

On the one hand, it could be hypothesised that intermediate learners might find shadowing or the self-imitation task difficult if they are not used to this practice. On the other hand, advanced learners should not find major problems regarding the difficulty of the activity. Moreover, it might be speculated that the group of learners that might benefit the most from all the techniques would be intermediate learners, as they have much room for improvement. On the contrary, proficient learners might already have a basis of knowledge of the pronunciation of numerous words and often show plateauing (i.e., a stabilisation of L2 learning in which learners slow down progressing). Nevertheless, imitation and self-imitation can still be useful techniques for practising those pronunciation aspects still unacquired.

Finally, it could be expected that self-imitation through voice manipulation will be more advantageous for L2 pronunciation improvement than shadowing (at least for speech rate improvement), as has been reported by Kusz (2022). Another reason to predict that self-imitation might be better is the use of a familiar voice, which has been considered to be beneficial for facilitating cognitive processing effortlessly and supporting L2 phonetic perception and pronunciation retention (Holmes et al., 2018; Kusz & Pawliszko, 2026). It is difficult to assume which of the two self-imitation techniques (i.e., through PSOLA or AI) will produce better outcomes. As noted by Gluhareva & Angelovska (2026), both PSOLA and AI approaches produce similar gains in terms of comprehensibility and accentedness. However, few information is available on the comparison of the two approaches, so the results might be different for the dimension of L2 fluency. Then, the results should be considered exploratory. If in this case, there are differences between the two approaches, it could be predicted that PSOLA provides more benefits since it is a personalised method in which the L2 parameters (e.g., syllable duration) to be enhanced are manually controlled, and not all parameters (i.e., pitch, syllable duration, stress, rhythm, intonation, etc.) are automatically changed, as AI platforms such as Speechify do.

5. Conclusions, future research and pedagogical implications

This dissertation has argued that both shadowing and phonetic self-imitation through PSOLA's and AI's voice manipulation may be useful techniques for L2 pronunciation enhancement. Despite their hard implementation inside classrooms and supplementary limitations (i.e., lack of feedback, cognitively demanding nature, technical issues to

obtain the synthesised voice, etc.), they are convenient alternatives for those who seek extra pronunciation practice outside classrooms because they allow learners to improve intonation, rhythm, but most notably fluency and comprehensibility, matching with the current prioritisation in FL instruction.

Further research is needed to account for the existing deficiencies of shadowing, but especially, for the self-imitation through voice manipulation techniques. Empirical investigations have not yet determined the level of proficiency that would benefit the most from this practice, nor which of the three methodologies is more beneficial to improve comprehensibility, fluency, and accentedness. Consequently, the empirical study proposed in this project could be an adequate starting point. Finally, further investigation could examine how to adapt the technique to learners of different ages, and learners who may not have any knowledge of linguistics or phonetics, given that numerous studies have applied the technique to advanced students with prior specialised knowledge.

References

- Adank, P., Evans, B. G., Stuart-Smith, J., & Scott, S. K. (2009). Comprehension of familiar and unfamiliar native accents under adverse listening conditions. *Journal of Experimental Psychology: Human Perception and Performance*, 35(2), 520–529. <https://psycnet.apa.org/doi/10.1037/a0013552>
- Anglès. Canal Universitats. (2026). Retrieved March 3, 2026, from <https://universitats.gencat.cat/ca/pau/materies-pau/angles/index.html#estructura-de-l-examen>
- Baker, W., Dewey, M., & Jenkins, J. (Eds.). (2018). *The Routledge handbook of English as a Lingua Franca* (1st ed.) Routledge. <https://doi.org/https://doi.org/10.4324/9781315717173>
- Bapir, N. S., & FageAbdulla, B. I. (2025). The effects of shadowing technique on boosting first year students' listening comprehension. *Koya University Journal of Humanities and Social Sciences*, 8(2), 534–544. <https://doi.org/10.14500/kujhss.v8n2y2025.pp534-544>

- Bione, T., & Cardoso, W. (2020). Synthetic voices in the foreign language context. *Language Learning & Technology*, 24(1), 169–186. <https://doi.org/https://doi.org/10.64152/10125/44715>
- Bovee, N., & Stewart, J. (2009). The utility of shadowing. In A. Stoke (Ed.), *JALT 2008 Conference Proceedings* (pp. 888–900). Japan Association of Language Teaching.
- Cenoz, J., & García Lecumberri, M. L. (1999). The acquisition of English pronunciation: learners' views. *International Journal of Applied Linguistics*, 9(1), 3–17. <https://doi.org/10.1111/J.1473-4192.1999.TB00157.X>
- Dang, V. M. T., & Tran, T. D. M. (2025). The effects of shadowing technique on students' pronunciation of -ed ending sounds: the case of Saigon university. *ICTE Conference Proceedings*, 9, 68–85. https://doi.org/10.54855/979-8-9870112-9-4_5
- Darcy, I. (2018). Powerful and effective pronunciation instruction: how can we achieve it? *The CATESOL Journal*, 30(1), 13–45. <https://doi.org/10.5070/B5.35963>
- Ding, S., Liberatore, C., Sonsaat, S., Lučić, I., Silpachai, A., Zhao, G., Chukharev-Hudilainen, E., Levis, J., & Gutierrez-Osuna, R. (2019). Golden speaker builder – An interactive tool for pronunciation training. *Speech Communication*, 115, 51–66. <https://doi.org/10.1016/J.SPECOM.2019.10.005>
- Edo-Marzá, N. (2017). The self, the peer and the teacher in the EFL pronunciation class: a comparative study on assessment, perceptions and systematicity. *International Journal of Applied Linguistics & English Literature*, 6(7), 198–219. <https://doi.org/10.7575/aiac.ijalel.v.6n.7p.198>
- Flege, J. E., Munro, M. J., & Mackay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97(5, Pt 1), 3125–3134. <https://doi.org/10.1121/1.413041>
- 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>
- Gluhareva, D., & Angelovska, T. (2026, April 14–15). *Self-imitation vs AI models in CAPT: A within-subjects study of L2 English comprehensibility and accentedness gains*. [Conference presentation]. EPIP9 English pronunciation issues & practices, Manchester, UK.
- Guevara Gavilanes, D. F. (2024). *The use of the international phonetic alphabet IPA as a tool for developing English pronunciation skills* [La Libertad: Universidad Estatal Península de Santa Elena, 2024.]. <https://repositorio.upse.edu.ec/handle/46000/10970>
- Hamada, Y. (2016a). *Teaching EFL learners shadowing for listening: developing learners' bottom-up skills* (1st ed.). Routledge. <https://doi.org/https://doi.org/10.4324/9781315677118>

- Hamada, Y. (2016b). Shadowing: who benefits and how? Uncovering a booming EFL teaching technique for listening comprehension. *Language Teaching Research*, 20(1), 35–52. <https://doi.org/10.1177/1362168815597504>
- Hamada, Y. (2019). Shadowing: what is it? How to use it. Where will it go? *RELC Journal*, 50(3), 386–393. <https://doi.org/10.1177/0033688218771380>
- Hamada, Y. (2021). Shadowing procedures in teaching and their future. *The Language Teacher*, 45(6), 32–36. <https://doi.org/10.37546/jaltlt45.6-3>
- Hamada, Y. (2025). Developing a new shadowing procedure for listening vocabulary: selective shadowing. *Innovation in Language Learning and Teaching*, 19(2), 122–135. <https://doi.org/10.1080/17501229.2024.2340567>
- Hamada, Y., & Suzuki, Y. (2024). Situating shadowing in the framework of deliberate practice: a guide to using 16 techniques. *RELC Journal*, 55(1), 219–227. <https://doi.org/10.1177/00336882221087508>
- Holmes, E., Domingo, Y., & Johnsrude, I. S. (2018). Familiar voices are more intelligible, even if they are not recognized as familiar. *Psychological Science*, 29(10), 1575–1583. <https://doi.org/10.1177/0956797618779083>
- Jin, S. (2023). Shadowing practice using TED Talks on the metaverse platform, gather: Affective factors and oral proficiency development *STEM Journal of English Teaching through Movies and Media*, 24(2), 31–46. <https://doi.org/10.16875/stem.2023.24.2.31>
- Kusz, E. (2022). Effects of imitation and self-imitation practice on L2 pronunciation progress. *Topics in Linguistics*, 23(2), 1–17. <https://doi.org/10.2478/topling-2022-0008>
- Kusz, E. (2023). Effects of self-imitation practice on L2 pronunciation with the use of a Golden Speaker Builder—an interactive tool for pronunciation training. *Pronunciation in Second Language Learning and Teaching Proceedings*, 13(1). <https://doi.org/10.31274/psllt.15721>
- Kusz, E., & Pawliszko, J. (2026). Efficacy of Artificial Intelligence (AI) voice cloning in phonetic self-imitation for L2 pronunciation training. *International Journal of Applied Linguistics*, 36(2), 1678–1695. <https://doi.org/10.1111/IJAL.70014>
- Lambert, S. (1992). Shadowing. *Meta*, 37(2), 263–273. <https://doi.org/10.7202/003378AR>
- Levis, J. M. (2005). Changing contexts and shifting paradigms in pronunciation teaching. *TESOL Quarterly*, 39(3), 369–377. <https://doi.org/10.2307/3588485>
- Luo, D., Shimomura, N., Minematsu, N., Yamauchi, Y., & Hirose, K. (2008). Automatic pronunciation evaluation of language learners' utterances generated through shadowing. *Proceedings of the Annual Conference of the International Speech Communication Association, INTERSPEECH*, 2807–2810. <https://doi.org/10.21437/interspeech.2008-476>

- Mompean, J. A., & Fouz-González, J. (2021). Phonetic symbols in contemporary pronunciation instruction. *RELC Journal*, 52(1), 155–168. <https://doi.org/10.1177/0033688220943431>
- Mora, J. C., & Mora-Plaza, I. (2023). From research in the lab to pedagogical practices in the EFL classroom: the case of task-based pronunciation teaching. *Education Sciences*, 13(10), 1042. <https://doi.org/10.3390/educsci13101042>
- Mora-Plaza, I. & Ortega, M. (2026). Listeners' self-assessment of second language speech. In M. Reed & J. Levis (Eds.), *The Handbook of Second Language Listening* (pp. 199-212). Wiley-Blackwell. <https://doi.org/10.1002/9781394312375.ch15>
- Munro, M. J. (2018). How well can we predict second language learners' pronunciation difficulties? *The CATESOL Journal*, 30(1), 267–281. <https://doi.org/10.5070/b5.35975>
- Munro, M. J., & Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97. <https://doi.org/10.1111/j.1467-1770.1995.tb00963.x>
- Munro, M. J., & Derwing, T. (1999). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 49(Supp. 1), 285–310. <https://doi.org/10.1111/0023-8333.49.s1.8>
- Oanh, N. H. & Tri, N. M. (2019). Applying shadowing technique and authentic materials to promote phonological awareness among young learners of English. *Proceedings of ELT upgrades 2019: a focus on methodology* (pp. 13–23). ISBN: 978 604 67 1450 7.
- Rakat, B., Handayani, I., & Miranty, D. (2024). The effectiveness of using minimal pairs towards students' pronunciation mastery. *Journal of English Language Teaching and Cultural Studies*, 7(1), 1–13. <https://dx.doi.org/10.48181/jelts.v7i1.22262>
- Saito, K. (2013). Reexamining effects of Form-Focused Instruction on L2 pronunciation development: the role of explicit phonetic information. *Studies in Second Language Acquisition*, 35(1), 1–29. <https://doi.org/10.1017/S0272263112000666>
- Sari, D. K. & Jaya, A. (2025). The effect of shadowing technique on students' speaking fluency and pronunciation accuracy. *English Community Journal*, 9(1), 23–31.
- Skarnitzl, R., & Bořil, T. (2024). Training of English prosody with acoustically modified voices. *Journal of Second Language Pronunciation*, 10(3), 375–403. <https://doi.org/10.1075/jslp.24041.ska>
- Sumiyoshi, H., & Svetanant, C. (2017). Motivation and attitude towards shadowing: learners perspectives in Japanese as a foreign language. *Asian-Pacific Journal of Second and Foreign Language Education*, 2(1), 1–21. <https://doi.org/10.1186/S40862-017-0039-6>

- Sungkur, R. K., & Shibdeen, N. (2025). Pronunciation trainer for second language learning using generative AI. *International Journal of Educational Technology in Higher Education*, 22(1), 64. <https://doi.org/10.1186/s41239-025-00561-x>
- Thao, D. T., & LY, N. H. (2025). A theoretical overview of the shadowing technique in developing listening and speaking skills among students. *International Journal of Educational Research and Development*, 7(3), 1–5.
- Tragant, E., Miralpeix, I., Serrano, R., Pahissa, I., Navés, T., Gilabert, R., & Serra, N. (2014). Cómo se enseña inglés en un grupo de institutos donde se obtienen resultados destacables en la prueba de lengua inglesa en las PAU. *Revista de Educación*, 363, 60–82. <https://doi.org/10.4438/1988-592X-RE-2012-363-170>
- Uchihara, T., Karas, M., & Thomson, R. I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*, 47(3), 794–827. <https://doi.org/10.1017/S0272263125100879>
- Wei, M. (2006). A literature review on strategies for teaching pronunciation. (ERIC Document Reproduction Service No. ED 491566). Retrieved from <http://www.eric.ed.gov/PDFS/ED491566.pdf>
- Whitworth, B., & Rose, H. (2025). A systematic review of research on the use of shadowing for second language pronunciation teaching. *Research Synthesis in Applied Linguistics*, 1(2), 239–269. <https://doi.org/10.1080/29984475.2025.2546827>
- Wisniewska, N., & Mora, J. C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*. 42(3), 599–624. <https://doi.org/10.1017/S0272263120000029>
- Zafarova, M. (2025). Shadowing as an effective method in language learning: theory, practice, and implementation. *Academic Journal of Science, Technology and Education*, 1(3), 7–11. <https://integrouppublication.org/index.php/ajste/article/view/31>