Universitat Autònoma de Barcelona

Enhancing the Pronunciation of the /i:/-/ɪ/ English Vowel Contrast Through Authentic Materials: A Study on Spanish/Catalan Teenagers Using Short Films

MA in Advanced English Studies: Multilingualism and Acquisition of English

Supervisor: Dr Celia GORBA

El Wafi BADJI

July 2025

Acknowledgments

To my mum Nacira who made the dream of living, studying, and working in Barcelona a reality for me; nothing is worth more than seeing a proud look in her eyes.

To Mohamed Yalles and Yassine Bouchair, whom I am dearly missing.

I am sincerely grateful to Dr Celia Gorba for her instrumental role in the successful completion of this dissertation; her consistent guidance and incredible expertise in the field cleared every roadblock throughout bringing this work to fruition.

I would also like to express my profound appreciation to Dr Juli Cebrian and Dr Maria-Josep Solé; their enjoyable courses in Phonology and Quantitative Analyses, respectively, equipped me with the skills and knowledge required to tackle such a topic. I am likewise thankful to Meriem Serhani, my former Linguistics teacher, and to Layachi Zemmouri, my former Phonetics teacher, from both of whom I learnt so much during my time at the University of Guelma.

It goes without saying that the results of this study would not have been achieved without the collaboration of my wonderful students who agreed to take part in the training sessions (and brought me an authentic, home-made tortilla de patatas!), as well as some other students who participated as a control group. I also have to thank the Kids&Us centre in Cerdanyola del Vallès, Barcelona who supported the implementation of the experiment.

Table of Contents

Index of Figures and Tables

List of Figures

List of Tables

Abstract

This study introduces and evaluates a novel approach to High-Variability Phonetic Training (HVPT) by investigating the benefits of using authentic, audiovisual materials to improve L2 vowel perception. Given the well-documented difficulty that L1-Spanish/Catalan learners have with the English tense-lax vowel contrast /iː/-/ɪ/, the primary goal was to determine the effectiveness of an HVPT intervention that used short films as the source of linguistic input. The study employed a pre-test/post-test design with twenty teenage learners enrolled in a language school in Barcelona, Spain. An experimental group (N=10) participated in two training sessions where they viewed authentic short films and then completed identification tasks using clips extracted from the films. A control group (N=10), which did not receive the training, also completed the tests. GLMM analysis results revealed that the experimental group demonstrated a statistically significant improvement in their perceptual accuracy of the /iː/-/ɪ/ contrast following the training, while the control group showed no significant change. Contrary to expectations, this improvement was not different for either vowel, with learners showing similar gains for both /iː/ and /ɪ/. These findings suggest that an audiovisual HVPT approach incorporating authentic materials is a highly effective method for enhancing learners' ability to perceive challenging non-native vowel sounds, offering clear pedagogical implications for integrating engaging, contextualised media into pronunciation instruction and training.

Keywords: High-Variability Phonetic Training (HVPT); Identification Tasks (IDs); authentic materials; short films; L2 vowel perception; L1-Spanish/Catalan; tense-lax vowel contrast /iː/-/ɪ/.

1. Introduction

The acquisition of 'native-like' pronunciation in a second language (L2) presents a significant and often persistent challenge for many learners. Unlike grammatical or lexical errors that might be overcome with explicit instruction and practice, phonological competence involves intricate motor skills, perceptual acuity, and the restructuring of deeply ingrained first language (L1) phonetic categories (Piske, MacKay, & Flege, 2001). Difficulties can manifest in various ways, including inaccurate production of individual phonemes, incorrect stress patterns, inappropriate intonation, and a generally non-native rhythm, all of which can impede intelligibility and communicative effectiveness (Derwing & Munro, 2005). These pronunciation hurdles are not merely collateral outcomes of language learning; they can impact a learner's confidence (Zielinski, 2012), social integration (Salsabila, Nurfidha, & Lubis, 2025), and even academic or professional opportunities (Hoque, Oli, Chowdhury, & Reza, 2024).

Amongst the most common sources of L2 pronunciation difficulty are the discrepancies between the phonological inventories of a learner's L1 and the target language. When the target L2 contains phonemes or phonemic distinctions that do not exist in the L1, learners often struggle with both perceiving and producing these new sounds accurately. Instead, they may assimilate L2 sounds to the closest L1 phoneme, leading to a collapsed contrast and potential misunderstandings (Flege, 1995; Best & Tyler, 2007). This is particularly evident in vowel production, where the precise articulation and acoustic targets can vary substantially across languages (Iverson & Evans, 2007; Chodroff & Wilson, 2022).

For Spanish and Catalan learners of English, a well-documented area of difficulty lies in distinguishing and producing certain English vowel contrasts that lack direct counterparts in their native language (L1) systems (Carlet & Souza, 2018). For instance, the English high front vowels /i:/ (as in 'beet') and /ɪ/ (as in 'bit) can be particularly problematic. While both Spanish

and Catalan possess a single high front vowel /i/, neither language phonemically distinguishes vowel length, quality, and laxness in the same way English does for the /i:/-/ɪ/ contrast (Cebrian, 2006; Cebrian, 2019). This structural difference often leads to potentially substituting both with their L1 /i/ (e.g., Cebrian, 2009, regardless of experience) or even confusing /ɪ/ with the mid-front vowel /e/ (Cebrian, 2006; Fabra, 2005), thereby possibly affecting lexical clarity (e.g., "Is he going to *leave*?" vs. "Is he going to *live*?").

In light of this, the present study implemented a HVPT approach using authentic materials in the form of short film clips with the aim of improving the perception of the /ɪ/-/i:/ distinction amongst Spanish/Catalan bilingual learners of English. The study hinges on a body of literature that will be explored in the following section.

2. Theoretical Background

2.1. Spanish/Catalan vs. English Vowel Contrasts

The Spanish and Catalan vowel systems are relatively more constrained, featuring five and seven (to eight) vowel-inventory, respectively (Wheeler, 2005; Ladefoged & Disner, 2012). On the other hand, English language (General American for a case in point), comprises up to 15 distinct vowels (Disner & Ladefoged, 2012).

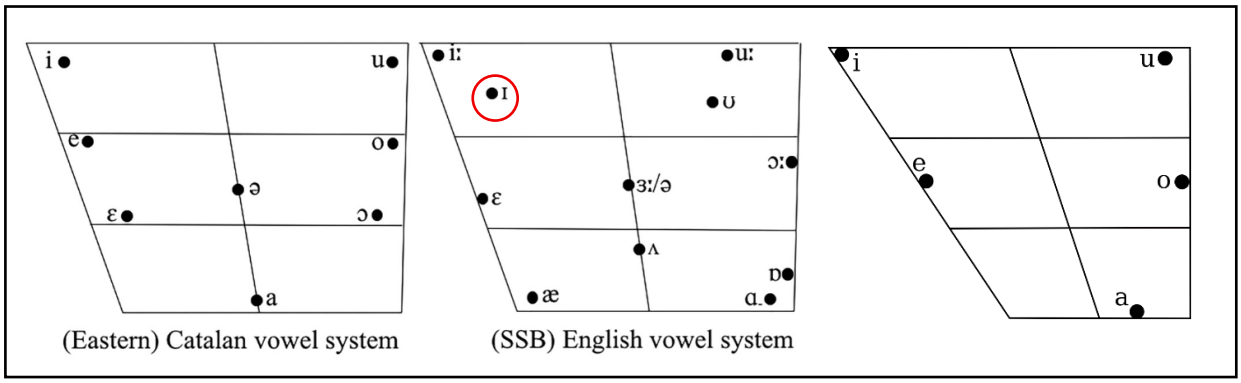

Figure 1. *Vowel systems of SSB English, Eastern Catalan (Carbonell and Llisterri, 1992, Roach, 2004, as cited in Cebrian, 2021 [emphasis on /ɪ/ added]), and Spanish [image quality enhanced by Wikimedia Commons, adapted from Ladefoged (2006)].*

The phonological implications of this structural difference are significant. As shown in **Figure 1** above, the Spanish and Catalan systems have a 'phonological empty space' where the English lax vowel /ɪ/ resides. Lacking a pre-existing category for one phoneme, learners

from these L1 backgrounds tend to find it difficult to differentiate words distinguished solely by this vowel contrast (e.g., *sheep* vs. *ship*) due to perceptual inaccuracies (Best & Tyler, 2007), thereby providing the theoretical foundation for the present investigation. The following table (**Table 1**) further describes the high front vowels vowel quality differences across English, Spanish, and Catalan:

Language	Vowel Symbol	F1 (Hz)	F2 (Hz)	Description
English	/i/	~270	~2290	Close front unrounded
English	/ɪ/	~390	~1990	Near-close near-front unrounded
Spanish	/i/	~240	~2100	Close front unrounded
Catalan	/i/	~280	~2200	Close front unrounded
Catalan	/e/	~460	~2050	Close-mid front unrounded

Table 1. *High Front Vowels in General American English, Peninsular Spanish (Castilian), and Central Catalan (Eastern dialect) by F1 and F2 formant values (Data adapted from Fabra & Romero, 2012).*

Formant data shows F1/F2 space for Catalan /i/ overlaps with English /i/, while /ɪ/ is closer to Catalan /e/. Therefore, the contrast between the English vowels /i:/ and /ɪ/ in particular poses unique challenges for Catalan-Spanish bilingual learners of English (Cebrian, 2006), primarily due to the underlined differences in the vowel systems of these languages leading to an erroneous cue weighting. That is, Catalan-speaking learners may over-rely on duration cues when attempting to differentiate English /i:/ and /ɪ/ more than native English speakers who primarily use spectral cues for high vowels (Salcedo & Romero, 2024). This overreliance on duration can occur regardless of the amount of L2 experience (e.g., Cebrian, 2006). In the same vein, a key study by Flege, Bohn, and Jang (1997) also found that Spanish learners, whose L1 also lacks vowel length contrasts, relied more on temporal cues than native speakers for the English /i/-/ɪ/ contrast. However, they did not consistently differentiate the

two vowels /i/ and /ɪ/ — neither by vowel quality (formant space) nor by length (duration), regardless of the amount of experience they had with English.

2.2. Theoretical Models of SLA Speech Acquisition

Understanding how learners acquire the sound system of a second language (L2) is crucial for effective pronunciation training. Several theoretical models have been proposed to explain the mechanisms and processes underlying L2 speech development, particularly concerning the perception and production of new phonetic contrasts. This study will focus on three prominent models: Flege's Speech Learning Model (SLM), its more recent revision of this model (SLM-r), and Best and Tyler's Perceptual Assimilation Model-L2 (PAM-L2). Additionally, the challenges faced by Spanish/Catalan learners of English due to L1-L2 phonological inventory differences will be briefly addressed.

2.2.1. Speech Learning Model (SLM) and Revised Speech Learning Model (SLM-r)

The Speech Learning Model (SLM), with its more recent updated version (SLM-r), is a comprehensive theory addressing the perception and production of second language (L2) sounds, primarily focusing on the ultimate attainment of L2 pronunciation by learners with extended experience (Flege, 1995; Flege & Bohn, 2021). A core tenet of the SLM model is that the mechanisms for speech learning remain active throughout the lifespan, and that L1 and L2 sound systems interact within a "common phonological space" (Flege, 1995, p. 239). Another central claim of the SLM is that the difficulty of learning an L2 sound varies based on its perceived phonetic relationship to L1 sounds. Therefore, L2 sounds can be categorised as "identical," "similar," or "new" relative to L1 categories. "Identical" sounds are easily acquired via transfer, while "new" sounds, which are phonetically very different from any L1 sound, are also highly learnable because learners can establish a new phonetic category for them. For instance, experienced English-speaking learners of French were found to produce the "new" French vowel /y/ authentically (Flege, 1987). However, the most problematic sounds are those

deemed "similar", which are close to an L1 sound but not identical. For these sounds, learners tend to engage in "equivalence classification", equating the L2 sound with an existing L1 category. This perceptual linkage, creating what are termed "diaphones" (Weinreich, 1957, as cited in Flege, 1995), can block the formation of a new, distinct category and result in a persistent non-native-like production, often with phonetic properties intermediate between the L1 and L2 norms (Flege, 1987; Flege, 1995).

The SLM also proposes that this cross-linguistic influence is bidirectional, affecting both "forward transfer" (L1 to L2) and "backward transfer" (L2 to L1). A classic example is the production of Voice Onset Time (VOT), where Flege (1987) found that experienced native French speakers of English produced French /t/ with longer, more English-like VOT values, while highly experienced native English speakers of French produced English /t/ with shorter, more French-like VOTs, suggesting a *mutual restructuring* within the shared phonetic space.

The original model (SLM) linked the ability to discern L1-L2 phonetic differences to the learner's age of learning (AOL) and assumed that accurate perception was a prerequisite for accurate production (Flege, 1995). On other hand, the more recently introduced Revised Speech Learning Model (SLM-r) significantly updates these principles, shifting the focus from age-related limits to the lifelong reorganisation of phonetic systems in response to L2 *input* (Flege & Bohn, 2021). Firstly, a major revision is the move away from a "perception-before-production" sequence to a "co-evolution" hypothesis, which posits that L2 perception and production skills develop in parallel without a fixed precedence, reflecting a strong bidirectional link between the two modalities. Furthermore, the original model's simple "age hypothesis" is replaced by the more nuanced "L1 category precision hypothesis". This revised tenet states that a learner's ability to discern L1-L2 phonetic differences is modulated by how precisely their own L1 phonetic categories are defined: individuals with more precise native

categories are better equipped to notice differences and thus form new L2 categories (Flege & Bohn, 2021).

The SLM-r also formalises the "full access" hypothesis, which maintains that learners are not constrained by their L1 feature inventory and can learn to *re-weight* any set of perceptual cues based on the statistical properties of the L2 input they receive (Flege & Bohn, 2021). This underscores the model's increased emphasis on the critical role of both the *quantity* and *quality* of L2 input, acknowledging that most L2 learners are exposed to highly variable input, including multiple dialects and other foreign-accented speech, which inevitably shapes their learning outcomes differently from that of a monolingual native speaker (Flege & Bohn, 2021). Therefore, it notes that simple "length of residence" (LOR) is often an inadequate measure of experience and advocates for more precise metrics like Full-Time Equivalent (FTE) years of input. Moreover, SLM-r simplifies its approach to phonetic similarity: whereas earlier conceptualisations of the model often relied on a three-way classification of L2 sounds as "identical," "similar," or "new", it was noted that no principled method existed for consistently applying this trichotomy. The SLM-r, hence, reframes this concept, moving away from the strict three-way division mentioned earlier; it now posits that the likelihood of forming a new L2 category is influenced by the "degree of perceived phonetic dissimilarity" between an L2 sound and the closest L1 sound. This perceived dissimilarity is viewed as a continuum and acts as one of several key factors, alongside L1 category precision and the quality of L2 input, in predicting learning outcomes.

These revisions highlight SLM-r's current focus on the individual learner, recognising that *intersubject variability* is significant and that the bilingual phonetic system is a dynamic and continuously evolving entity rather than one that reaches a fixed, monolingual-like "end-state" (Flege & Bohn, 2021).

2.2.2. Perceptual Assimilation Model-L2 (PAM-L2)

The Perceptual Assimilation Model (PAM), initially developed for native listeners (Best, 1994), was later extended to L2 learners as PAM-L2 by Best and Tyler (2007). PAM-L2, like its predecessor, is primarily a model of speech perception and posits that non-native sounds are perceived in terms of their similarity to native L1 categories through a process of "perceptual assimilation" (Best, McRoberts, & Sithole, 1988; Chang, 2019). The model is grounded in a direct realist approach to speech perception, assuming articulatory gestures as the basic phonetic unit (Best, 1994). PAM-L2 outlines several types of assimilation patterns for a pair of L2 sounds, each predicting a different level of discrimination difficulty for the L2 learner (Best & Tyler, 2007, as summarised in Chang, 2019):

- **Two-Category (TC) Assimilation:** The two L2 sounds are assimilated to two different L1 categories. This typically results in excellent discrimination.
- **Single-Category (SC) Assimilation:** Both L2 sounds are assimilated to the same L1 category with a similarly good fit. This leads to poor discrimination.
- **Category-Goodness (CG) Assimilation:** Both L2 sounds are assimilated to the same L1 category but differ in how well they fit that category (one is a better exemplar than the other). This results in fair to good discrimination, intermediate between TC and SC.
- **Uncategorised-Categorised (UC) Assimilation:** One L2 sound is assimilated to an L1 category, while the other is perceived as unlike any L1 sound (uncategorised). Discrimination is typically good.
- **Uncategorised-Uncategorised (UU) Assimilation:** Neither L2 sound is assimilated to an L1 category. Discrimination accuracy can range from poor to intermediate.
- **Non-Assimilable (NA)**: The L2 sounds are so different from any L1 sounds that they are perceived as “non-speech”. In this rare case, discrimination can be good, potentially benefiting from a non-speech auditory processing mode not thwarted by linguistic

categories (Best, McRoberts, & Sithole, 1988; Chang, 2019). An example provided is the good discrimination of Zulu click contrasts by English listeners, who have no clicks in their L1 (Best, McRoberts, & Sithole, 1988).

Finally, PAM-L2 acknowledges that for L2 learners (unlike naïve listeners), abstract phonological knowledge of the L2 can influence the L2-to-L1 mapping process, in addition to similarity at the gestural level (Chang, 2019). Also, the model emphasises how attunement to the L1 constrains the perception of non-native segments (Best, 1995).

2.3. Phonetic Training

To help learners overcome pronunciation hurdles caused by differences between learners' L1 and L2 inventories, developing effective strategies remains a compelling necessity. To this end, phonetic training aimed at enhancing the perception and production of English vowels amongst Spanish and Catalan learners has evolved through diverse methodologies that integrated theoretical knowledge with practical applications. As will be discussed in the current section, numerous studies have documented the effectiveness of these training methods, emphasising both perceptual tasks and production exercises designed to improve learners' linguistic abilities. In L2 speech perception research, identification and discrimination tasks remain two of the most fundamental experimental methods used to assess how learners process sound contrasts (Cebrian, Carlet, Gavaldà, & Gorba, 2024). Both tasks have proven fundamental in assessing and training the perception of vowels. This is mainly because they engage learners in recognising and distinguishing vowel sounds, respectively, which is crucial for improving both perceptual and production abilities (van Leussen & Escudero, 2015).

2.3.1. High-Variability Phonetic Training (HVPT)

High-variability phonetic training, henceforth HVPT, allows learners to familiarise themselves with various L2 sounds produced by different native speakers. In principle, HVPT is a perceptual training method that enhances L2 learners' ability to distinguish difficult phonemic

contrasts. Unlike traditional phonetic drills rooted in contrastive analysis (Lado, 1957) and usually focusing on repetition of minimal pairs to teach pronunciation (e.g., Woodsworth, 1967), HVPT exposes learners to multiple talkers, speaking styles, and varied phonetic contexts, which fosters better generalisation and retention of L2 sounds through enhancing learners' *phonemic* awareness (perception) and pronunciation using *varied input* (Thomson, 2018). First developed by Logan, Lively, and Pisoni (1991) for Japanese learners distinguishing English /r/ and /l/, this approach proved to be particularly effective for training learners in distinguishing and/or identifying vowel contrasts (Thomson, 2011) that are absent in their native language or usually confounded, such as English /i:/ vs. /ɪ/ for L1 Spanish-Catalan learners (Carlet & Cebrian, 2019). Studies have reinforced the significance of HVPT in both identification and discrimination tasks. Lengeris and Hazan (2010) found that high-variability exposure during identification tasks led to improved overall identification accuracy as well as production amongst Greek native speakers learning English vowels, with a substantial enhancement observed in their ability to recognise and produce unfamiliar vowel categories. Moreover, Iverson, Pinet, and Evans' (2012) study examined the effectiveness of HVPT for native French speakers learning English vowels, comparing experienced learners, who had lived in England for at least six months and used English daily, with inexperienced learners, who had studied English in school but had little exposure to native English speakers. Their interventions made participants undergo eight training sessions and were tested before and after on vowel identification, category discrimination, and vowel production. The results showed that both groups improved significantly, with inexperienced learners increasing identification accuracy from 40% to 61%, and experienced learners from 59% to 76%. However, category discrimination and production showed only minor improvements, suggesting that phonetic training enhances specific aspects of phonetic learning that natural exposure alone does not

provide. The study highlights that focused training plays a distinct role in second-language learning, benefiting learners regardless of prior language experience.

Moving back to the context of this study and to highlight the benefits of HVPT for Spanish-Catalan native speakers, a study by Aliaga-Garcia (2017) compared the effectiveness of two HPVT training methods using audiovisual stimuli—auditory identification (ID) and direct articulatory (ART) instruction—on the perception and production of English vowels by Catalan-Spanish learners. Results relevant to this study showed that the training led to significant gains in L2 vowel perception, with learners successfully generalising this improvement to new words, talkers, and contexts. Notably, the training prompted a shift in perceptual strategy, moving learners away from an over-reliance on vowel duration and towards a greater use of more native-like spectral cues. Specifically, the ART group significantly increased the spectral distance for vowel pairs like /ɪ/-/i:/ and /æ/-/ʌ/, making their productions more distinct. Additionally, the study by Cebrian, Gavaldà, Gorba, and Carlet (2024) compared the effectiveness of identification (ID) and categorical discrimination (DIS). Their findings concluded that both methods successfully improved learners' performance on both identification and discrimination tasks, with generalisation to both non-words and real word stimuli and proven retention four months later. The study also revealed an interesting asymmetrical relationship: the ID-trained group improved significantly more on the identification task than the DIS group did, but both groups improved on the discrimination task by a similar amount. The authors suggested this was because the *categorical* nature of the discrimination task engaged high-level phonological processing similar to that used for identification, leading to mutual benefits.

Nonetheless, Giannakopoulou, Brown, Clayards, and Wonnacott (2017) present contrasting evidence on the aspect of talker variability in their study comparing high- and low-variability input in children and adults. The study involved 52 Greek-speaking children and 41

adults, who were exposed to the English vowel contrast /i/–/ɪ/ through a computerised word-learning game over ten training sessions. Contrary to predictions based on previous HVPT research, both age groups showed greater perceptual gains in the (one talker) low-variability condition. Notably, children improved significantly more in vowel discrimination when trained with a single talker, even on novel, untrained items. Adults displayed a slight advantage in high-variability training, but the results were inconsistent across lexical and voice conditions. Also, no significant differences emerged in lexical learning between conditions, and the study did not support the hypothesis of "greater plasticity" in children's phonetic learning. These findings, thus, challenge assumptions about the universal benefit of talker variability and suggest that, for certain contrasts and age groups, low-variability input may be more effective for early-stage perceptual gains.

2.3.2. Authentic and Audiovisual Materials in HVPT

The rationale for employing authentic audiovisual materials in phonetic training, as proposed for the current research, is well-supported by a body of literature emphasising the unique benefits authentic resources offer for L2 pronunciation development in ESL classrooms (see Espinoza, Cárdenas, Martinez, & Saavedra, 2021). In addition to the motivational factor for the learners (Peacock, 1997), using authentic materials in phonetic training was proven to enhance learners' pronunciation (Rojas, 2022). Authentic materials can be defined as resources created for native speakers, not primarily intended for educational purposes, yet they contain a rich array of phonetic, lexical, and grammatical structures that offer considerable benefits to learners by exposing them to target language speech in real contexts (Nowacka, 2015). For this reason, integrating authentic materials in pronunciation training has been linked to increased learner motivation and engagement. For instance, materials that reflect real-life language situations help learners recognise the relevance of their training, fostering a more positive and engaging learning environment (Alisoy, 2025). This effect is especially evident amongst

learners with limited initial exposure to the target language, as they navigate the nuances of a foreign phonetic system (Gilmore, 2011). Relatedly, authentic resources offer an invaluable context for phonemic awareness and production practice, encouraging learners to actively explore sounds rather than passively receive instruction (Lago-Ferreiro, Gómez-González, & López-Ardao, 2025).

While research confirms that variability in input enhances learning and retention of L2 sounds (Uchihara, Karas, & Thomson, 2025), some research also examined the use of multimodal input (e.g., audio-visual training) incorporating technology and visual feedback to improve phonetic learning, highlighting the influence of factors such as talker's face and visibility of articulatory gestures (Hardison, 2003). Furthermore, more recent research has explored the use of audiovisual input and caption-based interventions to improve L2 pronunciation, particularly amongst L1-Spanish/Catalan learners. For instance, Galimberti, Mora, and Gilabert (2023) demonstrated that synchronised, phonemically enhanced captions in authentic videos could significantly boost Spanish-Catalan learners' pronunciation accuracy and phonological awareness of a set of target English words. Similarly, Mora and Fouz-González (2024) found that contrastive enhancement in captions, particularly colour-coded IPA transcriptions, facilitated learners' perceptual sensitivity to difficult vowel contrasts such as /æ/–/ʌ/, with eye-tracking data confirming heightened attention to the target forms. In a related study, Wisniewska and Mora (2020) showed that both captioned and uncaptioned video input improved learners' awareness of segmentation and speech processing skills, although no significant betterment was reported for phonological accuracy in perception measured through a discrimination task. Interestingly, though, accentedness ratings indicated an interaction between the type of viewing mode and the learners' task focus (form or meaning) in production. That is, pronunciation improved when learners focused on phonetic form during uncaptioned video viewing; however, this effect was not observed when captions were present.

Contrariwise, in the captioned viewing condition, pronunciation gains only occurred when learners' attention was directed towards the storyline rather than the phonetic form.

Conversely, Hutchinson and Dmitrieva (2022) found that passive exposure to French films by naïve American monolinguals for one session led to a minimal yet significant improvement in French /y/ shadowing, while they did not benefit from exposure to shadow /u/. Perception-wise, the study reports no change in their perceptual assimilation between the French vowels /y/ and /u/ and their L1 vowels. Although results seem mixed, together, these studies suggest that visual input enhancement in audiovisual contexts can aid L2 pronunciation learning and perception, particularly when designed to direct learners' attention to specific, problematic phonological contrasts and extended over a number of sessions that guarantee enough exposure to quality input from a variety of speakers. Therefore, the present study used subtitling in the short films and inside the identification clips, albeit with eliminating the target words containing the vowel contrast to curb orthographic interference. In light of this, the effectiveness of using authentic materials like film clips, instead of uncontextualised audio tracks, to provide contextual and multimodal input for L1-Spanish/Catalan learners to improve their perceptual categorisation of the challenging /i:/ - /ɪ/ English contrast remains an area of research worth further exploration.

2.3.3. Explicit Instruction in Phonetic Training

Explicit instruction is another widely adopted methodology in phonetic training literature. The findings from Carlet and Souza (2018) indicate that such phonetic training can enhance perceptual accuracy for difficult sounds in both real and non-words yet still stress that the type of training matters. Their study, which investigated 16 Spanish/Catalan learners' ability to perceive and produce English vowels after an 8-week period of formal instructions, showed improvement in the participants' perception, but no significant changes in production. Furthermore, the authors hinge on SLM-r tenets to argue that "quality input" is best achieved

when learners engage in autonomous learning activities outside the classroom, which provides the contextualised practice needed to supplement formal instruction. Moreover, according to Añorga and Benander (2015), focusing instruction on such contrasts heightened learner awareness and led to noticeable improvements in pronunciation accuracy amongst first-year learners of Spanish. In addition, findings from Aliaga-Garcia's (2017) study reported earlier suggested as well that direct articulatory instruction may be more effective than purely perceptual training for improving the distinctiveness of L2 vowel production.

Moreover, a classroom-based study by Gordon, Darcy, and Ewert (2012) investigated how explicit instruction influences the acquisition of L2 phonological features and their contribution to comprehensible speech. Three groups of ESL learners received pronunciation instruction over three weeks, but the groups differed in the type of explicit instruction received, with two experimental groups focusing on either suprasegmental or segmental features, and a third group receiving no explicit instruction. Pre- and post-test recordings indicated that only the group trained explicitly on suprasegmentals significantly improved their comprehensibility scores. The authors conclude that explicit phonetic instruction which makes learners *notice* L2 features is beneficial for developing comprehensible speech, corroborating the important role of prosody.

2.3.4. Individual Factors' Impact on L2 Speech Training

A learner's ultimate success in acquiring L2 speech has been shown to be influenced by a wide range of general factors. A central debate revolves around age, with the *Critical Period Hypothesis* proposing a biological end to native-like acquisition around puberty, a view supported by researchers like Lenneberg (1967) and Scovel (1988). This has been challenged by other authors like Flege across SLM and SLM-r models (Flege, 1995; Flege & Bohn, 2021), who argue that learning capacity does not disappear with age and that other factors are more critical.

Piske, MacKay, and Flege (2001) also identified age of L2 learning (AOL) as the most consistent predictor of foreign accent, with learners exposed to an L2 after puberty being more likely to retain a noticeable accent. Saito and Hanzawa (2018) confirmed this with longitudinal data, showing that late starters could improve, but remained below the native-like plateau. Close to AOL, the amount of L1 use has been identified as a significant, independent factor affecting L2 pronunciation. Likewise, research by Flege, Frieda, and Nozawa (1997) indicated that L2 learners who continued to use their first language frequently tended to speak their L2 with a stronger foreign accent than those who use their L1 infrequently. This effect had been observed in both early and late bilinguals, suggesting that language use patterns have an influence regardless of a bilingual's age of L2 acquisition.

Other factors have been examined by Piske, MacKay, and Flege (2001), albeit their influence appears more contested. The length of residence (LOR) in an L2-speaking country, for instance, has yielded conflicting evidence, but more recent findings allude to a consensus on the view that learners immersed in the L2 through cultural integration show stronger oral skills than those confined to L1 environments (Munro & Derwing, 2015). On the other hand, there was no consistent effect of gender on overall L2 pronunciation accuracy. In addition to motivation, individual aptitude is believed to influence success in second language (L2) pronunciation, a view supported by both anecdotal observations and research findings (Piske, MacKay, & Flege, 2001). Additionally, studies have documented exceptional talents in adult learners (Ioup, Boustagui, El Tigi, & Moselle, 1994, as cited in Munro & Derwing, 2015) and have linked pronunciation achievement to specific cognitive abilities, such as the phonemic coding ability identified in the Modern Language Aptitude Test (MLAT) (Carroll & Sapon, 1959, as cited in Munro & Derwing, 2015). More novel research has further identified other significant individual differences that affect L2 speech learning, including musical ability (Slevc & Miyake, 2006), musical experience (Cooper, Wang, & Ashley, 2017), auditory

selective attention (Mora & Mora-Plaza, 2019), short-term phonological memory (Cerviño-Povedano & Mora, 2015), and updated analyses of phonemic coding ability (Saito, Sun, & Tierney, 2019).

As for socio-affective factors, Guiora's (1991) concept of "language ego" suggests that the ability to take on a new accent is tied to psychological flexibility and identity. Closely tied to this is Schumann's (1986) Acculturation Model, which links success to the learner's degree of social and psychological integration with the target culture, driven by either integrative or instrumental motivation. Furthermore, whilst Piske, MacKay, and Flege (2001) acknowledged motivation as one of many variables, they found limited empirical support for its independent role in reducing foreign accent. However, the more current works (Munro & Derwing, 2015) provide nuanced evidence suggesting that motivation does not directly eliminate accent, but it in fact modifies learner behaviour, leading to practices (e.g., imitation, attention to feedback) that improve intelligibility and sometimes reduce perceived accent. For example, Munro and Derwing (2015) stress that learners with integrative goals (e.g. wanting to blend into the L2-speaking community) tend to focus more consciously on pronunciation than those with only instrumental goals (e.g., passing a test). They also note that such learners often self-correct more frequently, seek feedback, and pursue authentic speaking opportunities, all of which contribute to clearer and more native-like pronunciation. Likewise, Mora (2022) attests that in foreign language classroom settings, attitudes centred on social involvement are less effective at predicting a learner's ultimate success, primarily because of the limited opportunities for interactive L2 use. Despite this, affective variables like motivation and anxiety (or lack thereof) can still have a significant impact on learning outcomes.

2.3.5. Generalisation of Perceptual Training to Production

A key question in pronunciation training is whether improvements in speech perception gained through training can generalise to improvements in speech production. While this study does

not concern itself with ultimate production results of participants, it is worth understanding the link between perception and production and the possibility of generalisation of perceptual training in L2 speech research. The relationship between L2 perception and production is complex and not as straightforward as in L1 acquisition (Chang, 2019). Flege's (1995) original SLM posited that the production of a sound eventually corresponded to the properties represented in its phonetic category (Hypothesis H7), implying that accurate perceptual targets were necessary for accurate production. This suggested that enhancing perceptual abilities was, in principle, a pre-requisite for production if it led to the formation of more accurate L2 phonetic categories. However, the new "co-evolution" hypothesis in SLM-r states that L2 perception and production develop in parallel and influence each other bidirectionally, without one necessarily needing to precede the other. This change was made to account for inconsistencies in research, including findings from studies where learners' production of a contrast was found to be more accurate than their perception of it (e.g., Darcy & Kruger, 2012, as cited in Flege & Bohn, 2021). Meanwhile, this comes in line with some findings negating the link between perception and production (e.g., Sheldon & Strange, 1982; Gorba & Cebrian, 2021).

Some studies have indeed reported positive correlations and carryover of training gains from perception to production. For instance, Wang, Jongman, and Sereno (2003) found that perceptual training on Mandarin tones led to improvements in production. More recently, the potential for perceptual training to generalise to production was investigated by Carlet and Cebrian (2019), who compared the effects of two different HVPT methods on Spanish/Catalan learners' perception and production of English vowels. The study contrasted an identification (ID) training method with a categorical discrimination (DIS) training method against a control group. The results indicated that while both training methods led to improvements in perception, only the learners who underwent ID training also showed a significant gain in

production. This finding suggests that generalisation from perception to production is not automatic and may be dependent on the nature of the perceptual training itself. Interestingly, their study points to the potential superiority of identification-based especially for vowel training. On the whole, the nature of the training, the specific phonological features targeted, individual learner differences, and the way perception and production are linked (SLM) or dissociated (SLM-r) in L2 learners all play a role in generalisation of perceptual training to production (Chang, 2019).

3. Goals and Research Question

This research seeks to investigate the effectiveness of a phonetic training with clips extracted from short films to improve the pronunciation of /i:/ and /ɪ/ amongst L1-Spanish/Catalan teenage learners of English. The primary objective of the training was to enhance learners' ability to identify the vowel sounds /i:/ as in "sheep" and /ɪ/ as in "ship," thereby improving their vowel perceptual abilities. Given the findings discussed earlier, the current study employs identification tasks in an HVPT regime enhanced by authentic audiovisual materials to target perceptual improvements for the /i:/-/ɪ/ contrast. The benefits of using authentic materials align with the principles of HVPT (see **Section 2.3.3**), which advocates for training that utilises diverse phonetic stimuli and quality input to improve learning outcomes.

According to reported studies on HVPT, the paradigm was proven to be effective for increasing perceptual accuracy in identifying vowels for L2 learners. Additionally, the focus on incorporating authentic materials into such a validated phonetic training method is intended to bridge the gap between controlled experimental setups normally used in phonetic trainings (Nagle & Baese-Berk, 2022) and the real-world linguistic input learners may encounter. Therefore, the guiding research question is as follows:

RQ: To what extent does a phonetic training programme using authentic short film clips impact the perceptual accuracy of the English vowel contrast /i:/-/ɪ/ amongst L1-Spanish/Catalan teenage learners of English?

Authentic audiovisual materials are anticipated to improve learners' ability to perceive the distinction between the vowel pair by focusing their attention on the acoustic differences (Hardison, 2003). Furthermore, ID tasks are expected to be an ideal choice for a phonetic training on vowel perception (Carlet & Cebrian, 2019). Also, the weaker perceptual mapping of /ɪ/ noted in some studies (e.g., Cebrian & Mora, 2024) suggests it would be the more difficult vowel to identify, and thus more room for improvement is presumed as a result of the targeted training intervention. Therefore, it is hypothesised that:

H1: L1-Spanish/Catalan teenage learners of English who participate in the phonetic training programme using authentic short film clips through ID tasks will demonstrate improvement in their perceptual accuracy of the /i:/ vs. /ɪ/ vowel contrast from pre-test to post-test. However, a less pronounced improvement is expected for the /i:/ vowel, as a more accurate identification is expected from the onset, based on its strong perceptual similarity to the L1 /i/ category.

4. Methodology

4.1. Participants

Twenty (11 males and 9 females) participants in total aged between 15 and 17 (mean age: 15.6, SD: 0.68) took part in the study. All the participants were Spanish-Catalan dominant bilinguals and only one participant reported speaking French as well at home. This participant was still included in the study since she was using Spanish and Catalan in her daily life, consistent with the rest of the group, and has always been studying in a Spanish/Catalan school. Also, no noticeable difference in her accent was noticed compared to her classmates[1]. Initially, 13

[1] It is assumed by her teacher that her use of French is limited to occasional interactions with one of her parents.

intermediate-level L1-Spanish/Catalan-speaking teenage learners of English were recruited for the experimental group. However, only 10 completed the full training, while 3 missed the training sessions but completed the pre-test and post-test. These 3 were therefore reassigned to the control group, bringing the final control group total to 10 participants. All the participants were enrolled in a C1 course in a specialised English language learning centre in Cerdanyola del Vallès, Barcelona, Spain and they had been learning English at the school since the ages of 1 to 8 (Mean: 3.6, SD: 1.85) following the Kids&Us teaching method (Kids&Us, 2021). It is worth noting that the learners were not engaged in any other formal phonetic training outside of this experiment, and their regular course curriculum did not feature any targeted pronunciation training of this sort.

4.2. Study Design

The study boasts a pre-test/post-test design with an intervening training phase of two sessions for the experimental group. The control group completed the pre-test and post-test but did not participate in the training sessions; however, they were still offered access to complete the experiment online at their own pace after the completion of the experiment to benefit from the training. This design allows for a comparison of perceptual gains between the two groups, isolating the effect of the phonetic training intervention, all while ensuring adherence to ethical considerations.

4.2.1. Testing Tasks (Pre-Test and Post-Test)

The pre-test and post-test were identical and consisted of a forced-choice identification task (see **Figure 1**) designed to measure participants' baseline and post-intervention perceptual accuracy of the /iː/-/ɪ/ contrast. To collect background information, all participants first completed a short questionnaire before proceeding (see **Appendix C**). Prior to the pre-test, learners were briefly given instructions on how to complete the tasks, highlighting the pronunciation differences between the two options (‘Sheep’ vs. ‘Ship’) and pointing out how

they represented the vowels present in the words they heard. To reinforce clarity, both words "sheep" and "ship" were enunciated clearly, with a focus on emphasising the phonetic distinction between the vowel sounds to eliminate potential ambiguities in the tasks. They were also encouraged to shadow the correct pronunciation of some other minimal pairs containing both vowels to enhance their awareness of these contrasts prior to the experiment.

4.2.1.1. Stimuli (Testing Tasks)

The stimuli for the testing tasks consisted of 20 audio clips containing words that featured the target vowels /i:/ and /ɪ/ (10 instances of each, containing 5 instances of male/female speakers). These clips were sourced online (Cambridge Free English Dictionary and Thesaurus, 2025; see **Appendix B** for a complete list).

The selection of stimuli for these tasks was carefully planned, beginning with the most constrained element: the trained words. These were identified first, based on their natural and clear occurrence within the two authentic short films, which limited the available pool. Following this, the 20-word stimulus set for the pre-test and post-test was compiled, comprising a majority of monosyllabic items and four disyllabic ones to ensure phonological variety. This test set was deliberately populated with 70% of the words not present in the films (see **Appendix B** for a reference to these words) in order to ensure generalisation of perceptual learning to novel items, which is a more robust measure of acquisition than the memorisation of trained examples. Still, a small subset of words from the films (30%) was also included in the test. The selection of these six specific words was critical, as they were chosen to form a balanced set, with three words representing the /ɪ/ vowel and three representing the /i:/ vowel. Finally, using stimuli recorded by both male and female speakers enhances the validity of the assessment by evaluating perceptual skill across varied voice qualities (Thomson, 2018). All these selected words were spoken by both male and female native speakers of British English

to introduce speaker variability and maintain a parallel balance with the training stimuli, which also featured actors who are British speakers of English.

4.2.1.2. Procedure (Testing Tasks)

The tasks were administered using the Gorilla Experiment Builder platform (Anwyl-Irvine, Massonnié, Flitton, Kirkham, & Evershed, 2020) through an identification task. In each trial, participants would hear one of the stimulus words and were required to identify the vowel sound by selecting one of two images presented on the screen: a picture of a "ship" to represent /ɪ/ and a picture of a "sheep" to represent /i:/ (see **Figure 2**). This use of images instead of written labels was a crucial design choice to mitigate the influence of orthographic interference, which can reinforce inaccurate phonological mappings based on spelling rather than sound (Cebrian, 2021; Hayes-Harb and Barrios, 2021). Participants received no feedback on their performance during the pre-test and post-test.

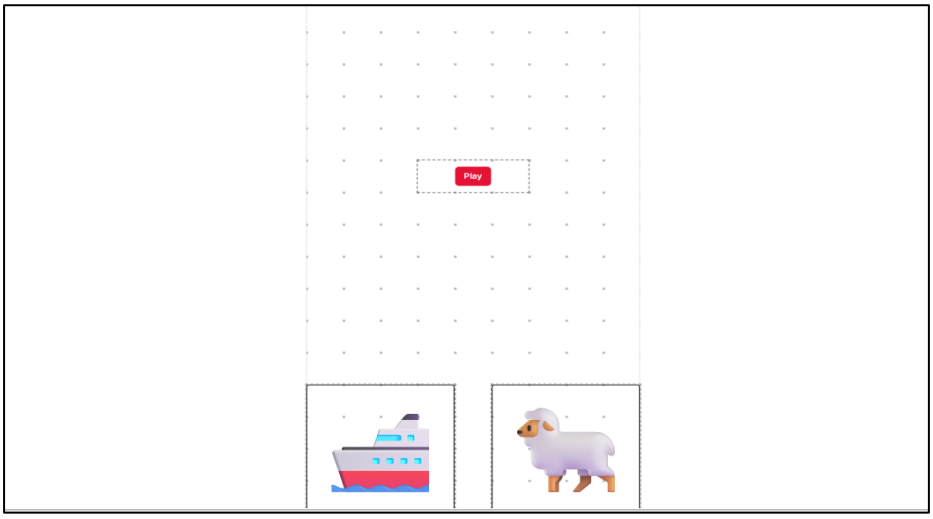

Figure 2. *Snapshot of the pre-and post-test user interface build (from Gorilla Experiment Builder)*

Before starting, a short testing sound was played to ensure that participants could hear the input clearly through their headphones **(**see **Figure 3)**.

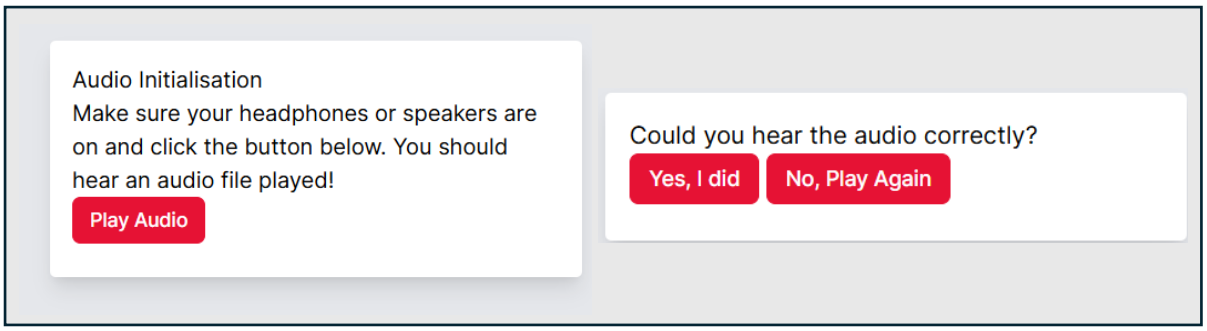

Figure 3. *Testing sound playback to ensure all the participants could listen to the input appropriately before starting the tasks (from Gorilla Experiment Builder)*

4.2.2. Training Tasks

The experimental groups participated in two training sessions, each involving watching a short film on a TV screen as a group, followed by individually completing an identification task (ID) on their phones while wearing headphones. Participants viewed the films in advance to provide contextual grounding for the subsequent ID tasks and to ensure sufficient exposure to authentic linguistic input. Moreover, the tasks were based on shorter clips to help participants concentrate specifically on the target vowel within each word.

4.2.2.1. Stimuli (Training Task)

The core training stimuli were extracted from two authentic short films sourced from YouTube (2025). In the first session, learners watched *Crossword* (NITVShorts, 2023), a romantic short film about a lonely woman who finds solace in solving daily crossword puzzles. The second session featured another short film entitled *Nadia* (Million Youth Media, 2017), a drama that follows the journey of a homeless young girl who is forced to leave home by her mother and struggles to find shelter. These films were selected after a careful review of a larger list, based on their appropriateness for a teenage audience and, crucially, for containing a sufficient number of stressed, clearly enunciated words with the target vowels /i:/ and /ɪ/ spoken by multiple characters (see **Appendix A** for a list of target words).

For each of the two training sessions, a total of 20 keywords were extracted from the authentic film clips, meticulously balanced to include 10 instances of /ɪ/ and 10 featuring /i:/.

The primary justification for using authentic materials was to provide learners with naturalistic, contextualised input from multiple speakers, which is a core tenet of HVPT. To this end, a rigorous selection process was undertaken, involving multiple viewings of the films to identify and catalogue potential keywords with each viewing. The final stimulus set of both films, which consists mainly of monosyllabic words, 10 disyllabic words, and a single three-syllable word, reflects a more natural distribution of word types found in genuine speech. While the main goal was achieving a perfect balance for the target vowel contrast in stressed positions—a goal that was successfully met—a conscious effort was also made to balance other linguistic properties. However, the inherent nature of authentic speech meant that a perfect balance across all variables, such as syllable structure, was not always possible, as the selection was ultimately constrained by the words naturally occurring in the films' dialogue. Consequently, this outcome was accepted as a necessary compromise in order to prioritise the phonetic variability and validity of the training intervention over an artificially balanced but decontextualised stimulus set.

The distribution of the vowel types in selected clips was unevenly spread across the different speakers within the films. In *Crossword*, the single male speaker produced the vast majority of the target words (7 for /ɪ/ and 9 for /i:/), while the three female speakers produced significantly fewer (a combined 3 for /ɪ/ and 1 for /i:/). A similar pattern was observed in the film *Nadia*, where the male speakers collectively accounted for most of the target vowels (7 for /ɪ/ and 8 for /i:/), compared to the female speakers (3 for /ɪ/ and 2 for /i:/). Even so, this imbalance was deemed an acceptable trade-off for stimulus clarity. This approach still aligns with HVPT principles by exposing learners to the target sounds as produced by different speakers (male and female) in natural, varied contexts. The decision to prioritise stimulus clarity over a balanced gender representation is supported by research; in Pernet, Belin, and Jones (2014),

the cognitive processing of speech sounds was shown to be more aligned with their phonetic characteristics than with the gender of the speaker.

4.2.2.2. Procedure (Training Tasks)

To prepare the training materials, English subtitles were manually embedded into the films for the initial complete viewing (see **Figures 4** and **5**). Participants watched the films in their entirety on a television screen. An immediate forced-choice identification task (see **Figure 6**) was conducted following each film, allowing playback for up to three times. The students watched the first film in the first session and came back on another day for one more session (as per their regular course schedule) to watch the second film and proceed with the subsequent ID task followed by the post-test.

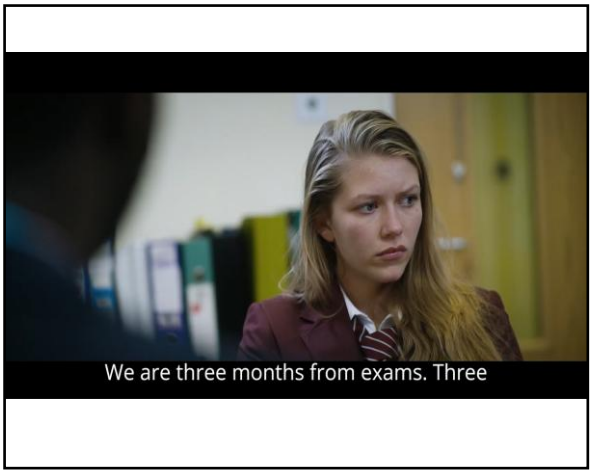

Figure 4. *A snapshot from the film Nadia (2017)*

Figure 5*. A snapshot from the film Crossword (2023)*

Clips of sentences or phrases from the film, specifically those containing the target vowels, were replayed three consecutive times. This time, the clips' subtitles were gapped, with the target words intentionally omitted. Learners were tasked with identifying whether the vowels in the missing words sounded more like /i:/ or /ɪ/ by selecting one of two options: "*sheep*" or "*ship*", again represented by corresponding images of each word to minimise orthographic interference and seek a purely phonetically driven choice.

Since the participants had already watched the films with embedded subtitles, they possessed sufficient contextual awareness to recognise the missing words, complemented by their own listening skills. This time, however, immediate feedback on the answer was provided after each selection to reinforce phonetic awareness of the contrast with each answer (see **Figures 6** and **7**).

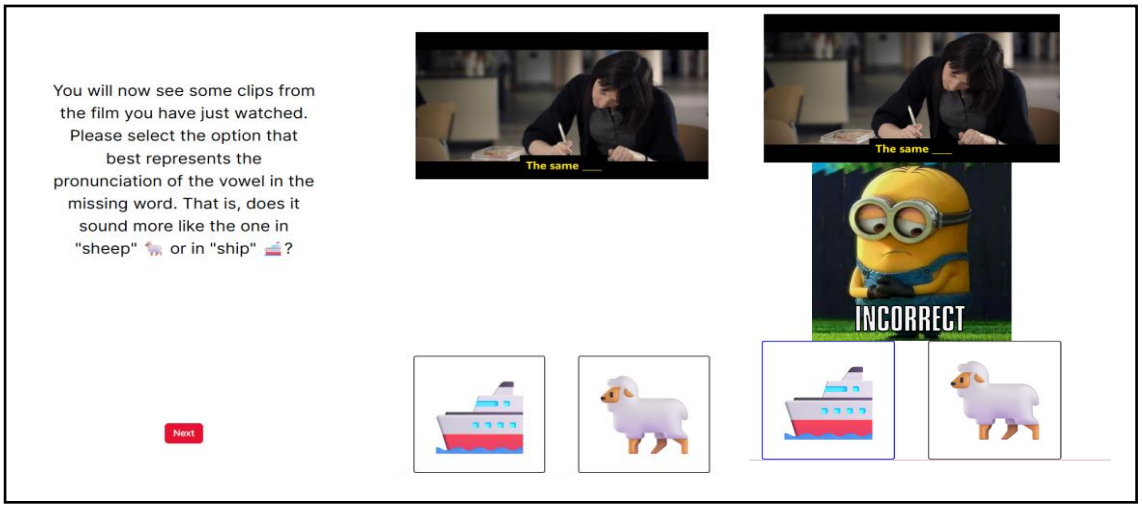

Figure 6. *A walkthrough of the first training session task, showing an example of an incorrect answer to the clip containing the word "seat".*

Figure 7. *Snapshot of correct feedback for the answer to the clip containing the word "easy"*

Since the training took place in a classroom setting and was conducted by the participants' teacher, brief group discussions occasionally emerged during the training sessions to address

any arising questions. During the ID task, for instance, learners were sometimes encouraged to pronounce the target words they heard while comparing them to the baseline items "sheep" and "ship" for reference. While it was noted that most of them showed reluctance to do so, the researcher's intervention was usually limited to re-enunciating the pronunciation of "sheep" and "ship" to clarify the acoustic cues distinguishing the vowel sounds upon the participants' request. Immediately after the second session of the training, the experimental groups were directed to the post-test task to measure improvement or lack thereof in the contrast perception.

5. Results

This chapter presents the statistical descriptive and inferential analyses of the data collected in the study to determine the effect of the phonetic training intervention on the perceptual accuracy of the English /i:/-/ɪ/ contrast. The analysis begins with a descriptive and visual exploration of the data conducted on IBM SPSS Statistics (version 26) software, followed by the main inferential statistical tests that formally address the research hypothesis.

5.1. Descriptive Statistics

The descriptive data (see **Table 2**) reveals a stark contrast between the two groups' learning trajectories. The Experimental group (N = 10) demonstrated a clear and substantial learning effect: starting from a pre-test mean of 59.5% (SD = 49.20, SE = 3.50), their performance surged to a post-test mean of 82.5% (SD = 38.10, SE = 2.70). This leap is captured by a mean gain of 23.0 percentage points (SD = 18.44, SE = 5.83). Moreover, the high median gain of 20.00%, with an interquartile range[2] (IQR) of 20.00, reflects substantial and consistently clustered gains, indicating a reliable training effect. In contrast, the Control group (N = 10) served as a stable baseline, exhibiting virtually no change between the testing sessions. That is, their mean accuracy remained relatively static, moving from 53.0% (SD = 15.12, SE

[2] The Interquartile Range (IQR) is reported alongside the median as it provides a more robust measure of spread than standard deviation, especially in the presence of outliers or skewed distributions or in small samples as in the present study (Moore, Notz, & Fligner, 2013).

= 4.78) at pre-test to 52.5% (SD = 14.00, SE = 4.42) at post-test, resulting in a negligible negative mean gain of -0.5%. Also, the median gain for this group was only 5.00%, whereas a wide IQR of 22.50 suggests that individual scores fluctuated without a clear, systematic pattern. This lack of a consistent trend confirms that no meaningful learning occurred in the absence of the intervention.

Group (N)	Measure	Mean	SD	SE	Median	IQR
Control (10)	**Pre-test**	53.0	15.12	4.78	52.50	22.50
	Post-test	52.5	14.00	4.42	52.50	31.25
	Gain	-0.5	15.35	4.85	5.00	22.50
Experimental (10)	**Pre-test**	59.5	49.20	3.50	65.00	41.25
	Post-test	82.5	38.10	2.70	87.50	30.00
	Gain	23.0	18.44	5.83	20.00	20.00

Note. All values are expressed as percentages. SD = Standard Deviation; SE = Standard Error of the Mean; IQR = Interquartile Range (calculated as the 75th percentile minus the 25th percentile).

Table 2. *Descriptive Statistics of Pre-Test, Post-Test, and Gain Scores for the Experimental and Control Groups*

The numerical data is further illuminated by a visual inspection of the gain score distributions for both groups. The distribution of gain scores for the Control group (**Figure 8**) is centred around zero, indicating that some participants' performance slightly improved while others' declined. Likewise, the mean gain (-0.5%) being slightly lower than the median gain (5.00%) suggests a distribution with a slight negative skew (skewness = -0.63, SE = 0.69) where a few lower scores pull the mean down, reinforcing the lack of consistent improvement. On the other hand, the distribution for the Experimental group (**Figure 9**) is clearly shifted towards positive values, with a concentration of scores in the positive gain range, visually confirming the substantial improvement reported in the descriptive statistics. Correspondingly, with the mean gain (23.0%) being slightly higher than the median (20.00%), the distribution

exhibits a slight positive skew (skewness = 0.10, SE = 0.69), indicating that while most participants improved significantly, a few achieved particularly high gains.

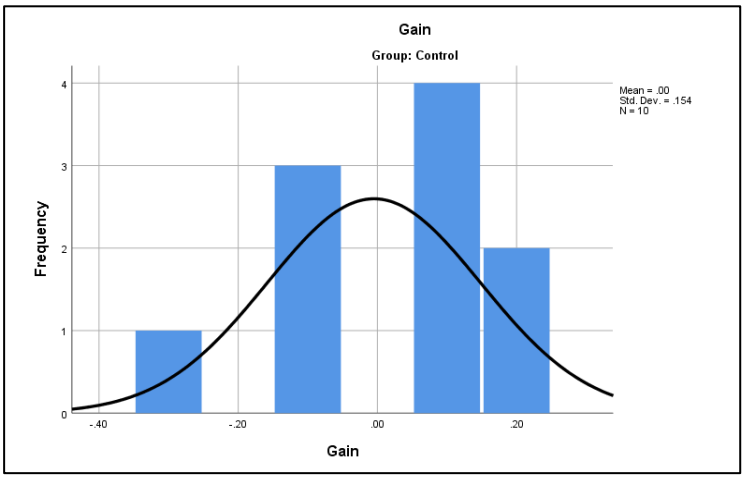

Figure 8. *Distribution of Gain Scores for the Control Group*

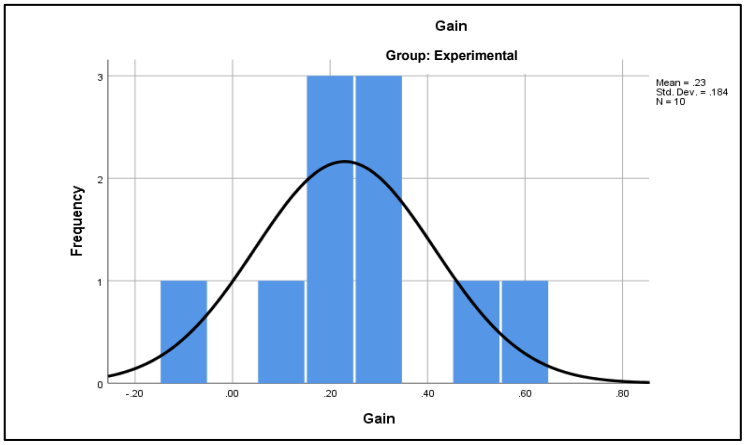

Figure 9. *Distribution of Gain Scores for the Experimental Group*

Further than that, the boxplots (**Figure 10**) provide a concise visual summary of these gain trends. The box for the Control group is centred near zero, with its interquartile range capturing both minor gains and losses. Conversely, the box for the Experimental group is located entirely in the positive range, reinforcing that the vast majority of participants benefited significantly from the training intervention. This clear visual separation between the two groups

underscores the effectiveness of the training. In summary, these descriptive statistics provide a strong preliminary indication that the training was highly effective.

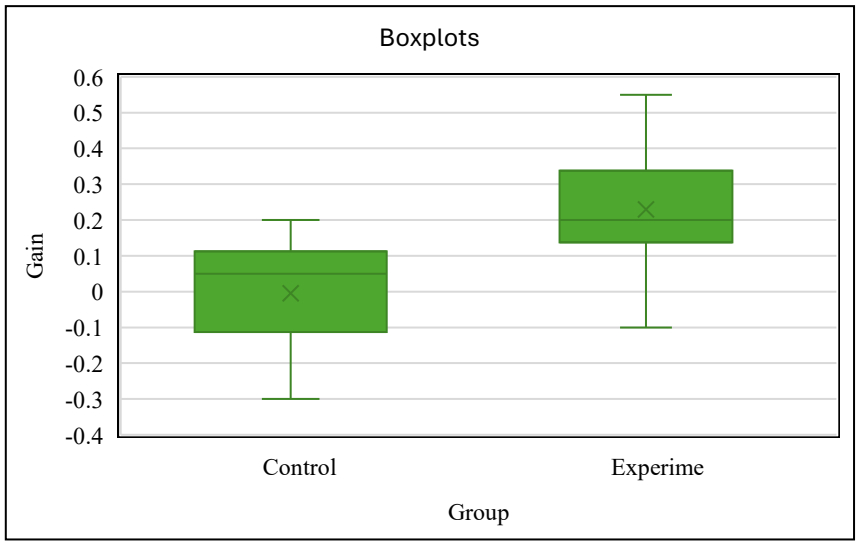

***Figure 10**. Boxplots of Gain Scores by Group*

Lastly, to examine the initial prediction regarding the difficulty of the two target vowels, a further breakdown by Vowel Type (**Table 3**) shows that at pre-test, both groups performed similarly for both vowels[3]. Following the intervention, the Experimental group showed comparable gains for both /iː/ and /ɪ/. The Control group's scores, in contrast, remained stable, with negligible changes for both vowels. Hence, both vowels may have posed the same level of difficulty to participants from the outset and did not influence their answers.

Group (N)	Measure	/iː/ Mean (SD)	/ɪ/ Mean (SD)
Control (10)	**Pre-test**	53.0% (50.2%)	53.0% (50.2%)
	Post-test	54.0% (50.1%)	51.0% (50.2%)
Experimental (10)	**Pre-test**	58.0% (49.6%)	61.0% (49.0%)
	Post-test	84.0% (36.8%)	81.0% (39.4%)

Table 3. *Mean identification accuracy scores (standard deviation in parentheses) for the /iː/ and /ɪ/ vowel contrasts, broken down by experimental group and test phase.*

[3] A separate GLMM including Vowel Type (/iː/, /ɪ/) as a fixed factor revealed no main effect of Vowel Type ($p = 0.833$), nor any significant two-way or three-way interactions involving this factor (all $p > 0.40$). This indicates that the training effect was consistent across both target vowels.

5.2. Inferential Analyses

Inferential analyses were conducted using open-source software Jamovi (The Jamovi Project, 2024), which runs on R statistical language (R Core Team, 2024). To formally test the research hypothesis, the data were modelled through a Generalised Linear Mixed Model (GLMM) with a binomial linking function, examining identification accuracy (0 = incorrect, 1 = correct) as a function of *Group* (Control, Experimental) and *Task* (Pre-test, Post-test) as fixed effects[4]. Random intercepts were included for participants (ID) and items (stimuli) to account for variability and non-independence of observations[5].

The analysis (**Table 4**) yielded a significant main effect of *Group*, such that the Experimental group performed more accurately overall than the Control group ($\beta = 1.01$; SE = 0.33; $z = 3.10$; $p = 0.002$). Also, a significant main effect of *Task* emerged, indicating that accuracy increased from pre-test to post-test ($\beta = 0.67$; SE = 0.17; $z = 4.01$; $p < 0.001$). Crucially, these main effects were qualified by a significant *Group* × *Task* interaction ($\beta = 1.38$; SE = 0.33; $z = 4.14$; $p < 0.001$). This significant interaction indicates that the pattern of accuracy change from pre-test to post-test differed significantly between the Control and Experimental groups, suggesting that the training had a positive impact on the trained group.

Parameter Estimates (Fixed Coefficients)

				95% Confidence Intervals			
Names	**Effect**	**Estimate**	**SE**	**Lower**	**Upper**	**z**	**p**
(Intercept)	(Intercept)	0.628	0.214	0.209	1.048	2.93	0.003
Group	Experimental - Control	1.012	0.326	0.373	1.652	3.10	0.002
Task	Post-test - Pre-test	0.667	0.166	0.341	0.993	4.01	<.001
Group ✻ Task	(Experimental - Control) ✻ (Post-test - Pre-test)	1.380	0.333	0.727	2.032	4.14	<.001

Table 4. *GLMM's Fixed Coefficients for Group, Task, and the Interaction Group*Task*

[4] Factor coding for these fixed effects was set to 'simple' (indicator coding), which is suitable for designs with a clear baseline (reference) condition (Hayes, 2017)

[5] Random intercepts for Participant ID (Variance = 0.388, ICC = 0.106; N=20) and individual audio files (Variance = 0.384, ICC = 0.104; N=20) revealed significant variance components, confirming clustering effects across individuals and stimuli, and thus justifying the accounting for non-independence in the data.

To further interpret this interaction, the Estimated Marginal Means (EMMs), which represent the model's predicted probability of a correct response for each condition, were examined (**Table 5**). The EMMs showed that the predicted accuracy for the Control group remained stable (Pre-test = 53.3%, Post-test = 52.8%), whereas the predicted accuracy for the Experimental group increased substantially from 61.2% at pre-test to 86.0% at post-test. This pattern of change directly reflects the observed data from the gain scores, indicating a substantial positive impact of the intervention.

Estimate Marginal Means - Group ✻ Task

				95% Confidence Intervals	
Group	**Task**	**Mean**	**SE**	**Lower**	**Upper**
Control	Pre-test	53.3%	7.08%	39.6%	66.6%
Control	Post-test	52.8%	7.09%	39.0%	66.1%
Experimental	Pre-test	61.2%	6.84%	47.3%	73.5%
Experimental	Post-test	86.0%	3.82%	76.7%	91.9%

Table 5. GLMM Estimated Marginal Means (as percentages) for the Group × Task Interaction

Figure 11 provides a visual representation of this significant interaction. The plot illustrates the estimated mean identification accuracy from the model for both groups across the pre-test and post-test phases. The distinct trajectories are immediately apparent: the Control group's performance remains stable, exhibiting a nearly flat line, while the Experimental group's performance shows a steep upward slope. Thus, this pronounced divergence between the groups' trajectories, or expected improvement, provides compelling visual evidence that the improvement over time was significantly greater for the trained participants.

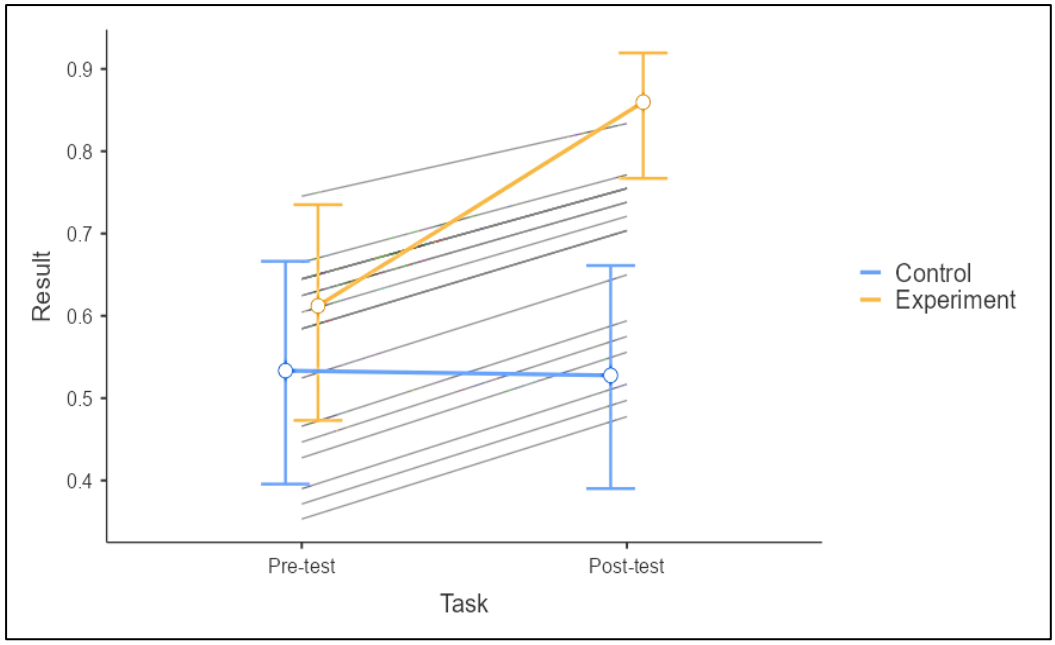

Figure 11. *GLMM Group × Task interaction plot*

Ultimately, to formally decompose the interaction, post-hoc comparisons were conducted on these estimated means. The results confirmed that while the Experimental group showed a significant improvement from pre-test to post-test ($p < 0.001$), the Control group showed no significant change ($p = 1.000$). Furthermore, while the groups did not differ significantly at pre-test ($p = 1.000$), the Experimental group significantly outperformed the Control group at post-test ($p < 0.001$). This pattern confirms that the training intervention was the source of the significant improvement noted for the experimental group. Further comparisons to follow:

Comparison								
Group	**Task**	**vs**	**Group**	**Task**	**Difference**	**SE**	**z**	**$p_{bonferroni}$**
Control	Pre-test	-	Control	Post-test	1.023	0.2187	0.107	1.000
Control	Pre-test	-	Experimental	Pre-test	0.724	0.2565	-0.910	1.000
Control	Pre-test	-	Experimental	Post-test	0.187	0.0705	-4.443	<.001
Control	Post-test	-	Experimental	Pre-test	0.708	0.2507	-0.975	1.000
Control	Post-test	-	Experimental	Post-test	0.182	0.0689	-4.503	<.001
Experimental	Pre-test	-	Experimental	Post-test	0.257	0.0657	-5.317	<.001

Table 6. Post Hoc comparison: Group ✻ Task interaction

6. Discussion

This study aimed to investigate the benefits of a HVPT regime using authentic materials on the perception of English /i:/ and /ɪ/ vowels by L1-Spanish/Catalan learners of English in a local language academy in the province of Barcelona, Spain. To this end, the study employed a pre-test/post-test design to assess the participants' identification accuracy of the contrast. The performance of an experimental group (N=10), which received the training intervention, was compared to that of a control group (N=10). The training consisted of two sessions in which the experimental group watched two short films and then completed an identification (ID) task featuring clips extracted from those films, alongside occasional interactions with the participants over the word pair pronunciation for further clarification. This section will interpret the findings, discussing their implications in relation to the guiding research question and the relevant theoretical literature.

The research question contemplated the effectiveness of a HVPT using audiovisual authentic materials in improving the perceptual accuracy of the /i:/- /ɪ/ contrast, whereby mainly it was hypothesised that the training would yield an improvement. There was also a secondary estimate that the L1-similar /i:/ vowel would benefit less from the training due to its familiarity to the learners' phonological inventory (Cebrian & Mora, 2024). The incentive behind using authentic materials instead of mainstream auditory materials was to appeal to the young participants' interests[6] and increase their motivation to complete the tasks (Alisoy, 2025). Furthermore, the decision of solely using identification tasks (ID) was motivated by their proven superior efficacy in improving perceptual awareness over discrimination tasks, especially in vowel training (Aliaga-Garcia, 2017; Cebrian, Gavaldà, Gorba, & Carlet, 2024).

[6] Elevated levels of engagement were observed amongst the learners during the first training session, in which the film *Crossword* successfully maintained their attention throughout. Positive feedback regarding the film was expressed by most of the teenage audience.

In support of the research hypothesis, the primary finding of this study was the significant Group × Task interaction ($\beta = 1.38$; SE = 0.33; $z = 4.14$; $p < 0.001$), which confirmed that the training was highly effective. Descriptively, the Experimental group achieved a substantial improvement in identification accuracy from pre-test to post-test (mean gain = 23.0%, SE = 5.83; IQR = 20.0), whereas the Control group's performance remained static (mean gain = -0.5%, SE = 4.85; IQR = 22.5). Moreover, post-hoc analyses of the interaction confirmed that the Experimental group's improvement was statistically significant ($p < 0.001$), while the Control group showed no significant change ($p = 1.000$). This indicates that the observed gains can be confidently attributed to the training intervention rather than to practice effects or other external factors. The study also predicted an asymmetrical learning pattern, where the L1-similar vowel /i:/ would be identified more accurately at pre-test than the more challenging /ɪ/ vowel. However, the descriptive results did not support this prediction; at pre-test, the experimental group's accuracy was comparable for both /i:/ (58.0%) and /ɪ/ (61.0%), suggesting a similar level of difficulty for both sounds before the intervention.

The outcome of the experiment aligns with a large body of research confirming the efficacy of HVPT in enhancing L2 learners' perception of difficult phonetic contrasts (Logan, Lively, & Pisoni, 1991; Thomson, 2018; Uchihara, Karas, & Thomson, 2025). The success of the intervention can be directly attributed to its multi-faceted approach, which effectively targeted the core of the learning problem: the structural differences between the L1 (Spanish/Catalan) and L2 (English) vowel inventories. Spanish and Catalan lack a phonemic contrast equivalent to the English /i:/-/ɪ/ distinction. This structural gap leads learners to assimilate the L2 sounds into their L1 system in complex ways; while /i:/ is typically mapped onto the native /i/, the /ɪ/ vowel is often assimilated inconsistently, sometimes to the native /i/ and other times to /e/, creating significant perceptual confusion (Fabra, 2005; Cebrian, 2006; Flege, Bohn, & Jang, 1997). By exposing learners to the target vowels /i:/ and /ɪ/ as produced by multiple speakers

in varied phonetic contexts within the short films, the training likely facilitated the development of more flexible phonemic categories for these L2 sounds. Then again, the training was not merely passive exposure. The explicit instruction provided before and during the tasks, specifically, the clear enunciation of the minimal pair "sheep" vs. "ship" by the researcher and encouragement to shadow the sounds, likely played a catalytic role. This explicit component, while limited, may have served to heighten the learners' awareness[7] of the contrast (Añorga & Benander, 2015) and directed their attention to the relevant acoustic cues (i.e., spectral quality over duration). This combination, where rich, varied input from HVPT is combined with focused, explicit guidance, appears to be a powerful mechanism for facilitating perceptual re-tuning in L2 learners (Carlet & Souza, 2018).

While the benefits of HPVT remain undeniable in pronunciation training practices (Lengeris and Hazan, 2010; Aliaga-Garcia, 2017; Cebrian, Gavaldà, Gorba, & Carlet, 2024), the challenge outlined in the research question was to incorporate authentic audiovisual materials through the use of short films and clips. These, far from being exotic in ESL classrooms, serve as a highly practical tool for exposing learners to meaningful input while simultaneously enhancing their motivation and attention span (Warden & Lin, 2000). Likewise, success was reported in studies incorporating audiovisual input in phonetic trainings (Galimberti, Mora, & Gilabert, 2023).

Furthermore, the outcome of this intervention can be interpreted through the lens of the theoretical models discussed previously. According to the Speech Learning Model-Revised (SLM-r), the formation of new L2 phonetic categories is heavily dependent on the quality and quantity of L2 input (Flege & Bohn, 2021). Specifically, the use of authentic short films

[7] Few students were noticed to discuss the differences between the two sounds and engage in a metalinguistic conversation, although it is worth noting that some of them were still referring to how *sheep* is “shorter” than *ship,* indicating an expressed conscious reliance on duration cues.

followed with ID task clips in the experiment provided learners with high-quality, contextualised, and varied input, which is crucial for noticing the subtle phonetic differences between L2 sounds and corresponding L1 categories. Besides, the SLM-r, which posits a more dynamic and interactive phonological system instead of rigid categories, also suggests that learning is influenced by the "degree of perceived phonetic dissimilarity". Accordingly, the training likely increased this perceived dissimilarity, allowing learners to establish a new category for /ɪ/. Crucially, by forcing learners to attend to the fine spectral cues distinguishing the two vowels, the training also prompted a refinement of the perceptual representation for the L1-like /i:/ (in line with Hu et al., 2016). This finding strongly supports the notion of an interactive phonological system where learning a new contrast can simultaneously sharpen the precision of existing categories. Such a concept, which was foundational to the original SLM (Flege, 1995), remains central to the dynamic, lifelong perspective of the SLM-r (Flege & Bohn, 2021). The particular relevance of this perspective is underscored by the participants' background: they have been learning English since a very young age (mean = 3.6 years) through a method that relied heavily on exposure to target language input from that early age (Kids&Us, 2021).

Nevertheless, the strength of this study's design lies not just in providing quality input, but in actively helping learners to process it. By pairing the rich L2 input from the films with explicit instruction and feedback, the training fostered a heightened metalinguistic awareness of the target contrast (Gordon, Darcy, & Ewert, 2012). Therefore, this powerful combination likely enabled learners to more effectively "re-weight" the perceptual cues they attend to, shifting from an over-reliance on duration, a common strategy for Spanish/Catalan learners (Cebrian, 2006; Salcedo & Romero, 2024), towards the more native-like spectral cues that primarily distinguish the /i:/-/ɪ/ contrast. Also, the significant improvement suggests that the intervention successfully prompted learners to perceive the phonetic dissimilarity between the

English /ɪ/ vowel and their native /i/ category, thereby preventing the "equivalence classification" that often hinders the acquisition of "similar" L2 sounds (Flege, 1995).

From the perspective of the Perceptual Assimilation Model-L2 (PAM-L2), the initial perceptual state for these learners is best described as a Category-Goodness (CG) Assimilation (Best & Tyler, 2007). In this pattern, both English /i:/ and /ɪ/ are assimilated to the single native /i/ category, but they differ in how well they fit: /i:/ is a good exemplar, while /ɪ/ is a poor one. This poor fit for /ɪ/ is consistent with findings that it is sometimes perceived as intermediate between the native /i/ and /e/ vowels (Cebrian, 2006; Fabra, 2005), leading to significant perceptual difficulty (Best & Tyler, 2007). Another possibility is a Categorised-Uncategorised (UC) Assimilation, where /i:/ is clearly mapped to the native /i/, but the acoustically intermediate /ɪ/ is not consistently mapped to any single native category (as per Cebrian, Gorba, & Gavaldà, 2021, which adopted a 70% assimilation threshold). However, the results of the present study, which show a symmetrical level of difficulty in identifying both vowels at pre-test, could also point towards a Single-Category (SC) Assimilation. In this scenario, both L2 vowels are mapped equally (and confused to a similar extent) within the single native /i/ category, without a clear goodness-of-fit difference. While these three assimilation patterns predict varying degrees of difficulty—with SC being the most problematic, followed by CG, and then UC (Best & Tyler, 2007)—they all represent a perceptual challenge that impedes accurate identification. The training, with its explicit focus on the contrast and immediate feedback (see **Figure 7**), probably helped learners to restructure this perceptual mapping. The significant improvement suggests a successful shift away from this unstable single-category mapping towards the establishment of a new, distinct category for /ɪ/, a process moving towards a Two-Category (TC) Assimilation. The use of ID tasks, which force a categorical choice, is particularly well-suited to promoting this kind of phonological reorganisation, a finding

consistent with studies that show ID training to be highly effective, sometimes more so than discrimination training (Carlet & Cebrian, 2019; Cebrian, Gavaldà, Gorba, & Carlet, 2024).

An interesting and somewhat unexpected finding was the lack of a substantial difference in the identification accuracy between the two vowels. This outcome is particularly revealing when viewed through the evolution of the Speech Learning Model. The original SLM (Flege, 1995) would predict an asymmetrical learning pattern, as the primary challenge lies with the "similar" vowel /ɪ/, which learners tend to equate with their L1 /i/, or sometimes /e/. The L1-like /i:/ vowel, being phonetically "identical" or very close due to its acoustic similarity to the native vowel (Flege, Bohn, & Jang, 1997), would be expected to be perceived accurately from the outset, leaving little room for improvement. Nevertheless, the finding that learners improved on both vowels equally challenges this prediction.

In summary, the findings of this study provide strong evidence that a focused, short-term HVPT intervention using authentic audiovisual materials and ID tasks can significantly improve the perceptual accuracy of the challenging /i:/-/ɪ/ vowel contrast for L1-Spanish/Catalan learners. The results support key tenets of the SLM-r and PAM-L2 models and underscore the pedagogical value of integrating principled phonetic training with engaging, authentic content in the L2 classroom.

7. Limitations of the Study and Recommendations for Future Research

Despite the encouraging findings regarding the effectiveness of pronunciation training through using authentic materials, this study presents several limitations that warrant consideration for future research and broader generalisation. First, methodological constraints limit the generalisability of the results. This was primarily dictated by participant availability, as the study's recruitment pool was confined to students enrolled at the language learning centre where the researcher was concurrently employed, resulting in a limited sample size. Similarly,

the duration and intensity of the intervention were constrained. Due to the demanding schedules of both the teacher and the teenage participants, it was not remotely feasible to extend beyond two total sessions without losing more participants as the sessions went on. In spite of that, gains were observed for the experimental group even with a short training. Together, these factors reduce the statistical power and prevent conclusions about the long-term retention of the perceptual gains. Future research should employ larger samples and more extended intervention periods to assess the durability of learning.

A more significant methodological limitation is the confounding of two instructional variables: HPVT and explicit instruction. While this interference appears effective, this study's design cannot disentangle their individual contributions. It remains unclear whether the gains were driven primarily by the rich, varied input of the HVPT, the focused attentional guidance of the explicit instruction by the teacher (the researcher himself), or the combination of the two. Even so, this combined approach, while not isolating each instructional variable, has high ecological validity as it reflects a real classroom environment. In such settings, rich input is often paired with teacher guidance to foster the metalinguistic awareness and 'noticing' needed to process that input effectively, a combination shown to be beneficial for L2 learners (Gordon, Darcy, & Ewert, 2012).

Furthermore, practical challenges during intervention delivery may have influenced the overall training experience; managing the training sessions effectively with teenagers sometimes proved tedious due to the inherent demands of behaviour management. Additionally, the necessity for the researcher to occasionally provide spare headphones, and even phones, to students who did not bring theirs could have introduced minor inconsistencies or disruptions to the smooth flow of the sessions.

Lastly, the scope of the study was narrowly focused on the perception of the English /i:/ and /ɪ/ vowel contrast, relying exclusively on identification tasks. While this approach effectively addressed the research question, it did not include an assessment of participants' production of these sounds, meaning it cannot speak to whether the observed perceptual gains generalised to improved pronunciation. This is a particularly important omission in the context of the SLM-r's "co-evolution" hypothesis, which posits a parallel, bidirectional link between perception and production (Flege & Bohn, 2021). By measuring only one modality, the study provides an incomplete picture of the learning process, especially since it is an ESL context where communicative achievement is key to learners.

These limitations point to several avenues for future research. A crucial next step would be to design studies that isolate the core instructional components. For instance, a three-group design (HVPT only; explicit instruction only; combined HVPT and explicit instruction) could determine the unique contribution of each element. Similarly, to specifically test the benefit of using rich, contextualised media, another study could directly compare an HVPT intervention using authentic audiovisual materials against one using only decontextualised audio stimuli. Furthermore, future studies must incorporate production tasks to test for perception-production generalisation and provide a more comprehensive test of L2 speech learning models. Finally, to move beyond the study's group-level analysis and verify the inferred motivational benefits of authentic materials in HVPT, future endeavours should use questionnaires and aptitude measures to explore the role of individual differences. This would allow for an investigation into how factors such as learner motivation (Mora, 2022) or phonological memory (Cerviño-Povedano & Mora, 2015) might interact with the training intervention to predict learning outcomes. Such work would build upon these promising results and contribute to a more informed understanding of how to best facilitate L2 phonological acquisition in the classroom.

8. Conclusions

This study investigated the effectiveness of a novel audiovisual HVPT intervention on the perception of the challenging English /i:/-/ɪ/ vowel contrast by Spanish/Catalan teenage learners. The results provide clear and compelling evidence that the training was highly successful. Participants in the experimental group demonstrated a significant improvement in their perceptual accuracy, while the control group showed no such improvement, confirming that the gains were a direct result of the intervention. The findings also revealed that, contrary to some theoretical predictions, learners improved their perception of both the L1-similar /i:/ and the more challenging /ɪ/ vowel to a similar degree, with neither of the vowels in particular proving more challenging.

These results carry significant theoretical and pedagogical implications. From a theoretical standpoint, the findings support contemporary models of second language speech acquisition, such as the SLM-r (Flege & Bohn, 2021) and PAM-L2 (Best & Tyler, 2007), which emphasise the importance of high-quality input and the dynamic nature of perceptual re-tuning. Pedagogically, this study demonstrates that integrating authentic, engaging materials like short films into a principled pronunciation training regime is a highly effective and practical approach for the language classroom (e.g., Galimberti, Mora, & Gilabert, 2023). The powerful combination of providing rich, contextualised input while also fostering metalinguistic awareness through explicit guidance appears to be a key driver of perceptual learning (Carlet & Souza, 2018). Finally, while the study is limited by its small sample size and lack of production data, it successfully establishes this audiovisual training method as a promising ground for future research and a valuable tool for teachers seeking to improve their students' pronunciation skills.

References

Aliaga-Garcia, Cristina (2017). The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English [Doctoral dissertation, Universitat de Barcelona]. TDX (Theses and Dissertations Online).

Alisoy, Hasan (2025). Exploring new teaching paradigms: A systematic analysis of university classroom practices. *Rivus Scientiarum Applicatarum*, *1*(1), 4–14.

Anwyl-Irvine, Alexander L., Massonnié, Jessica, Flitton, Adam, Kirkham, Natasha, & Evershed, Jo K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior research methods*, *52*, 388–407.

Añorga, Angel, & Benander, Ruth (2015). Creating a pronunciation profile of First-Year Spanish students. *Foreign Language Annals*, *48*(3), 434–446.

Best, Catherine T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. In J. C. Goodman & H. C. Nusbaum (Eds.), *The development of speech perception: The transition from speech sounds to spoken words* (pp. 167–224). MIT Press.

Best, Catherine T. (1995). A direct realist perspective on cross-language speech perception. In W. Strange (Ed.), Speech perception and linguistic experience: Theoretical and methodological issues in cross-language speech research (pp. 167–200). York Press.

A direct realist perspective on cross-language speech perception. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues in Cross-language Speech Research*, edited by W. Strange (York, Timonium, MD), pp. 167–200.

Best, Catherine T., & Tyler, Michael D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O. S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege* (pp. 13–34). John Benjamins Publishing.

Best, Catherine T., McRoberts, Gerald W., & Sithole, Nomathemba M. (1988). Examination of perceptual reorganization for nonnative speech contrasts: Zulu click discrimination by English-speaking adults and infants. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 345–360.

Cambridge Dictionary. (2025). Cambridge University Press & Assessment. Retrieved April 1, 2025, from https://dictionary.cambridge.org/

Carlet, Angelica, & Cebrian, Juli. (2019). Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In J. Levis, C. Nagle, & E. Todey (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 123-138). Iowa State University Digital Press.

Carlet, Angelica, & Souza, Hanna K. (2018). Improving L2 pronunciation inside and outside the classroom. *Ilha Do Desterro a Journal of English Language, Literatures in English and Cultural Studies*, *71*(3), 99–124.

Cebrian, Juli (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, *34*(3), 372–387.

Cebrian, Juli (2009). Interaction of L2 phonotactics and L1 syllable structure in L2 vowel production. In A. M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production* (1st ed., pp. 234–243). Cambridge Scholars Publishing.

Cebrian, Juli. (2019). Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *The Journal of the Acoustical Society of America*, *145*(1), EL52–EL58.

Cebrian, Juli. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *The Journal of the Acoustical Society of America*, *149*(4), 2671–2685.

Cebrian, Juli., & Mora, Joan C. (2024). Online assessment of cross-linguistic similarity as a measure of L2 perceptual categorization accuracy. *Languages*, *9*(5), 152.

Cebrian, Juli, Gorba, Celia, & Gavaldà, Núria (2021). When the easy becomes difficult: Factors affecting the acquisition of the English/iː/-/ɪ/contrast. Frontiers in Communication, 6, 660917.

Cebrian, Juli, Gavaldà, Núria, Gorba, Celia, & Carlet, Angelica. (2024). Differential effects of identification and discrimination training tasks on L2 vowel identification and discrimination. *Studies in Second Language Acquisition*, 1–25.

Cerviño-Povedano, Eva, & Mora, Joan C. (2015). Spanish EFL learners' categorization of /iː-ɪ/ and phonological short-term memory. *Procedia-Social and Behavioral Sciences*, *173*, 18–23.

Chang, Charles B. (2019). The phonetics of second language learning and bilingualism. In W. F. Katz & P. F. Assmann (Eds.), *The Routledge handbook of phonetics* (pp. 427–447). Routledge.

Chodroff, Eleanor, & Wilson, Colin. (2022). Uniformity in phonetic realization: Evidence from sibilant place of articulation in American English. *Language*, *98*(2), 250–289.

Cooper, Angela, Wang, Yue, & Ashley, Richard. (2017). Thai rate-varied vowel length perception and the impact of musical experience. *Language and Speech*, *60*(1), 65–84.

Derwing, Tracey M., & Munro, Murray J. (2005). Second language accent and pronunciation teaching: A research-based approach. *TESOL Quarterly*, *39*(3), 379–397.

Espinoza, Maria G. T., Cárdenas, Yola I. C., Martinez, Claudia V. P., & Saavedra, Francisco I. B. (2021). The Use of Audiovisual Materials to Teach Pronunciation in the ESL/EFL

Classroom: El uso de materiales audiovisuales para enseñar la pronunciación en el aula de ESL/EFL. *South Florida Journal of Development*, *2*(5), 7345–7358.

Flege, James E. (1987). The production of "new" and "similar" phones in a foreign language: Evidence for the effect of equivalence classification. *Journal of Phonetics*, *15*(1), 47–65.

Flege, James E. (1995). Second language speech learning: Theory, findings, and problems. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 233–277). York Press.

Flege, James E., & Bohn, Ocke-Schwen (2021). The revised speech learning model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: Theoretical and empirical progress* (pp. 84–118). Cambridge University Press.

Flege, James E., Bohn, O.-S., & Jang, Sunyoung. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, *25*(4), 437–470.

Flege, J. E., Frieda, Elaina M., & Nozawa, Takeshi. (1997). Amount of native-language (L1) use affects the pronunciation of an L2. *Journal of Phonetics*, *25*(2), 169–186.

Galimberti, Valeria, Mora, Joan C., & Gilabert, Roger. (2023). Audio-synchronized textual enhancement in foreign language pronunciation learning from videos. *System*, *116*, 103078.

Georgiou, Georgios P. (2022). The impact of auditory perceptual training on the perception and production of English vowels by Cypriot Greek children and adults. *Language Learning and Development*, *18*(4), 379–392.

Giannakopoulou, Anastasia, Brown, Helen, Clayards, Megan, & Wonnacott, Elizabeth. (2017). High or low? Comparing high and low-variability phonetic training in adult and child second language learners. *PeerJ*, *5*, e3209.

Gilmore, Alex. (2011). “I prefer not text”: Developing Japanese learners’ communicative competence with authentic materials. *Language Learning*, *61*(3), 786–819.

Gorba, Celia, & Cebrian, Juli. (2021). The role of L2 experience in L1 and L2 perception and production of voiceless stops by English learners of Spanish. *Journal of Phonetics*, *88*, 101094.

Gordon, Joshua, Isabelle Darcy, & Doreen Ewert. (2012). Pronunciation teaching and learning: Effects of explicit phonetic instruction in the L2 classroom. *Pronunciation in Second Language Learning and Teaching Proceedings*, *4*(1), 1–15.

Guiora, Alexander Z. (1991). The two faces of language ego. *Toegepaste Taalwetenschap in Artikelen*, *41*(1), 5–14.

Hardison, Debra M. (2003). Acquisition of second-language speech: Effects of visual cues, context, and talker variability. *Applied Psycholinguistics*, *24*(4), 495–522.

Hayes, Andrew F. (2017). Introduction to mediation, moderation, and conditional process analysis: A regression-based approach. Guilford publications.

Hayes-Harb, Rachel, & Barrios, Shannon. (2021). The influence of orthography in second language phonological acquisition. *Language Teaching*, *54*(3), 297–326.

Hoque, Mohammed Shamsul, Oli, Lamesa, Thamarai Selvi, M. D., Chowdhury, Abu Sayed Md Mahmudul Haque, & Reza, MD Mohoshin (2024). Untangling the tongues: Exploring pronunciation development strategies for learners of English at high schools in Oromia region of Ethiopia. *World Journal of English Language*, *14*(5), 257.

Hu, Wei, Mi, Lin, Yang, Zhen, Tao, Sha, Li, Mingshuang, Wang, Wenjing, Dong, Qi, & Liu, Chang (2016). Shifting perceptual weights in L2 vowel identification after training. *PLoS ONE*, *11*(9). https://doi.org/10.1371/journal.pone.0162876.

Hutchinson, Amy E., & Dmitrieva, Olga (2022). Exposure to speech via foreign film and its effects on non-native vowel production and perception. *Journal of Phonetics*, *95*, 101189.

Iverson, Paul, & Evans, Bronwen G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *The Journal of the Acoustical Society of America*, *122*(5), 2842–2854. https://doi.org/10.1121/1.2783198

Iverson, Paul, Pinet, Melanie, & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, *33*(1), 145–160.

The jamovi project. (2024). *jamovi* (Version 2.6) [Computer Software]. Retrieved from https://www.jamovi.org.

Kids&Us. (2021). *El método para aprender inglés*. https://www.kidsandus.es/es/metodo-aprender-ingles/

Ladefoged, Peter (2006). *A course in phonetics* (5th ed.). Thomson Wadsworth.

Ladefoged, Peter, & Disner, Sandra F. (2012). *Vowels and consonants*. John Wiley & Sons.

Lado, Robert (1957). *Linguistics across cultures: Applied linguistics for language teachers*. University of Michigan Press.

Lago-Ferreiro, Alfonso, Gómez-González, María Ángeles, & López-Ardao, José Carlos (2025). A new serious game (e-SoundWay) for learning English phonetics. *Multimodal Technologies and Interaction*, *9*(6), 54.

Lengeris, Angelos, & Hazan, Valerie (2010). The effect of native vowel processing ability and frequency discrimination acuity on the phonetic training of English vowels for native speakers of Greek. *The Journal of the Acoustical Society of America*, *128*(6), 3757–3768.

Lenneberg, Eric H. (1967). The biological foundations of language. *Hospital Practice*, *2*(12), 59–67.

Logan, John S., Lively, Scott E., & Pisoni, David B. (1991). Training Japanese listeners to identify English /r/ and /l/: A first report. *The Journal of the Acoustical Society of America*, *89*(2), 874–886.

Million Youth Media. (2017, July 6). *NADIA | Drama Short Film | MYM* [Video]. YouTube. https://www.youtube.com/watch?v=ffqp6f0_rzw

Moore, David S., Notz, William I., & Flinger, Michael A. (2013). *The basic practice of statistics* (6th ed.). W. H. Freeman and Company.

Mora, Joan C., & Fouz-González, Jonás (2024). Contrastive input enhancement in captioned video for L2 pronunciation learning. In *Audiovisual Input and Second Language Learning* (pp. 150–175). John Benjamins Publishing Company.

Mora, Joan C., & Mora-Plaza, I. (2019). Contributions of cognitive attention control to L2 speech learning. In J. Levis, C. Nagle, & E. Todey (Eds.), *A sound approach to language matters: In honor of Ocke-Schwen Bohn* (pp. 477–499). Iowa State University Digital Press.

Mora, Raúl Alberto. (2022). Toward more equitable language learning and teaching frameworks for our ELT community: Moving from EFL to ECL to CE. *Gist: Education and Learning Research Journal*, (24), 25–42.

Munro, Murray J., & Derwing, Tracey M. (2015). Intelligibility in research and practice: Teaching priorities. In M. Reed & J. M. Levis (Eds.), *The handbook of English pronunciation* (pp. 375–396). Wiley-Blackwell.

Nagle, Charles L., & Baese-Berk, Melissa M. (2022). Advancing the state of the art in L2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, *44*(2), 580–605.

NITVShorts. (2023, February 21). *"Crossword" | Solve life's big mystery in this romantic short film* [Video]. YouTube. https://www.youtube.com/watch?v=auKKbPBshmQ

Nowacka, Marta (2015). Phonetic inspirations in authentic materials: Stimulating students' phonetic awareness. In *Proceedings of the Pronunciation in Second Language Learning and Teaching Conference* (pp. 107–118).

Peacock, Matthew (1997). The effect of authentic materials on the motivation of EFL learners. *ELT Journal*, *51*(2), 144–156.

Pernet, Cyril R., Belin, Pascal, & Jones, Anna (2014). Behavioral evidence of a dissociation between voice gender categorization and phoneme categorization using auditory morphed stimuli. *Frontiers in Psychology*, *4*, 1018.

Piske, Thorsten, MacKay, Ian R. A., & Flege, James E. (2001). Factors affecting degree of foreign accent in an L2: A review. *Journal of Phonetics*, *29*(2), 191–215.

R Core Team. (2024). *R: A Language and environment for statistical computing* (Version 4.4) [Computer software].

Fabra, Lucrecia Rallo (2005). Predicting ease of acquisition of L2 speech sounds. A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics*, (2), 75–92.

Fabra, L., & Romero, Joaquín (2012). Native Catalan learners' perception and production of English vowels. *Journal of Phonetics*, *40*(3), 491–508.

Rojas, Carlos Andrés (2022). *Computerized pronunciation training of Spanish rhotics* [Unpublished doctoral dissertation]. The University of Wisconsin-Madison.

Saito, Kazuya, & Hanzawa, Keiko (2018). The role of input in second language oral ability development in foreign language classrooms: A longitudinal study. *Language Teaching Research*, *22*(4), 398–417.

Saito, Kazuya, Sun, Hui, & Tierney, Adam (2019). Explicit and implicit aptitude effects on second language speech learning: Scrutinizing segmental and suprasegmental sensitivity and performance via behavioural and neurophysiological measures. *Bilingualism: Language and Cognition*, *22*(5), 1123–1140.

Salcedo, Adrián, & Romero, Joaquín (2024). Examining L2 vowel contrast perception in the EFL classroom–A study on Spanish/Catalan speaker acquisition of English tense and lax vowels. In *Proceedings of the International Symposium on Applied Phonetics (ISAPh 2024)* (pp. 88–93).

Salsabila, Salwa, Nurfidha, Shinta, & Lubis, Yani (2025). The importance of English pronunciation in our life. *Morfologi: Jurnal Ilmu Pendidikan, Bahasa, Sastra dan Budaya*, *3*(2), 74–79.

Schumann, John H. (1986). Research on the acculturation model for second language acquisition. *Journal of Multilingual & Multicultural Development*, *7*(5), 379–392.

Scovel, Thomas (1988). The power of English prepositions. *Cross Currents*, *14*(2), 1–13.

Sheldon, Amy, & Strange, Winifred (1982). The acquisition of /r/ and /l/ by Japanese learners of English: Evidence that speech production can precede speech perception. *Applied Psycholinguistics*, *3*(3), 243–261.

Slevc, L. Robert, & Miyake, Akira (2006). Individual differences in second-language proficiency: Does musical ability matter? *Psychological Science*, *17*(8), 675–681.

Thomson, Ron I. (2011). Computer assisted pronunciation training: Targeting second language vowel perception improves pronunciation. *Calico Journal*, *28*(3), 744–765.

Thomson, Ron I. (2018). High variability [pronunciation] training (HVPT) A proven technique about which every language teacher and learner ought to know. *Journal of Second Language Pronunciation*, *4*(2), 208–231.

Uchihara, Takumi, Karas, Michael, & Thomson, Ron I. (2025). High variability phonetic training (HVPT): A meta-analysis of L2 perceptual training studies. *Studies in Second Language Acquisition*, 1–34.

van Leussen, Jan-Willem, & Escudero, Paola (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 1000.

Wang, Yue, Jongman, Allard, & Sereno, Joan A. (2003). Acoustic and perceptual evaluation of Mandarin tone productions before and after perceptual training. *The Journal of the Acoustical Society of America*, *113*(2), 1033–1043.

Warden, Clyde A., & Lin, Hsiu Ju (2000). Existence of integrative motivation in an Asian EFL setting. *Foreign Language Annals*, *33*(5), 535–545.

Wheeler, Max W. (2005). *The phonology of Catalan*. OUP Oxford.

Wisniewska, Natalia, & Mora, Joan C. (2020). Can captioned video benefit second language pronunciation? *Studies in Second Language Acquisition*, *42*(3), 599–624.

Woodsworth, John Alan (1967). *An analysis of the audio-lingual approach as applied to methods of teaching Russian* [Unpublished master's thesis]. University of Alberta.

Zielinski, Beth (2012). The social impact of pronunciation difficulties: Confidence and willingness to speak. *Pronunciation in Second Language Learning and Teaching Proceedings*, *3*(1), 1-15.

Appendices

Appendix A

Film	Speaker ID	Word	Number of syllables	Vowel Type	Gender
Crossword	Speaker2	*Dish*	1	/ɪ/	M
Crossword	Speaker2	*Easy*	1	/i:/	M
Crossword	Speaker2	*Eating*	2	/i:/	M
Crossword	Speaker2	*Films*	1	/ɪ/	M
Crossword	Speaker2	*Give*	1	/ɪ/	M
Crossword	Speaker3	*Hit*	1	/ɪ/	F
Crossword	Speaker2	*Keep*	1	/i:/	M
Crossword	Speaker2	*Kill*	1	/ɪ/	M
Crossword	Speaker2	*Lives*	1	/ɪ/	M
Crossword	Speaker4	*Miss*	1	/ɪ/	F
Crossword	Speaker1	*Office*	2	/ɪ/	F
Crossword	Speaker2	*Reason*	2	/i:/	M
Crossword	Speaker2	*Seat*	1	/i:/	M
Crossword	Speaker2	*Seem*	1	/i:/	M
Crossword	Speaker2	*Sit*	1	/ɪ/	M
Crossword	Speaker2	*Speak*	1	/i:/	M
Crossword	Speaker2	*Team*	1	/i:/	M
Crossword	Speaker3	*These*	1	/i:/	F
Crossword	Speaker2	*This*	1	/ɪ/	M
Crossword	Speaker2	*Week*	1	/i:/	M
Nadia	Speaker1	*Decrease*	2	/i:/	M
Nadia	Speaker3	*Difficult*	3	/ɪ/	M

Nadia	Speaker2	*Dinner*	2	/ɪ/	F
Nadia	Speaker3	*Eighteen*	2	/i:/	M
Nadia	Speaker1	*Give*	1	/ɪ/	M
Nadia	Speaker3	*Leave*	1	/i:/	M
Nadia	Speaker3	*Listen*	2	/ɪ/	M
Nadia	Speaker1	*Live*	1	/ɪ/	M
Nadia	Speaker3	*People*	2	/i:/	M
Nadia	Speaker4	*Please*	1	/i:/	F
Nadia	Speaker3	*Reason*	2	/i:/	M
Nadia	Speaker5	*Seen*	1	/i:/	M
Nadia	Speaker1	*Sit*	1	/ɪ/	M
Nadia	Speaker4	*Teeth*	1	/i:/	F
Nadia	Speaker2	*Think*	1	/ɪ/	F
Nadia	Speaker2	*This*	1	/ɪ/	F
Nadia	Speaker3	*Three*	1	/i:/	M
Nadia	Speaker3	*Victims*	2	/ɪ/	M
Nadia	Speaker5	*Weak*	1	/i:/	M
Nadia	Speaker3	*Which*	1	/ɪ/	M

Table 7. *Words included in ID tasks*

Appendix B

Word	Vowel type	Number of syllables	Gender	Appears in trained clips
Bin	/ɪ/	1	M	No
Bit	/ɪ/	1	M	No
Visit	/ɪ/	2	F	No
Grid	/ɪ/	1	F	No
Listen	/ɪ/	2	F	Yes
Grin	/ɪ/	1	F	No
Pitch	/ɪ/	1	M	No
Sin	/ɪ/	1	M	No
Sit	/ɪ/	1	M	Yes
This	/ɪ/	1	F	Yes
Bean	/i:/	1	M	No
Beat	/i:/	1	M	No
Beach	/i:/	2	M	No
Colleague	/i:/	2	F	No
Greed	/i:/	1	F	No
Green	/i:/	1	F	No
Reason	/i:/	2	M	Yes
Seat	/i:/	1	M	Yes
These	/i:/	1	F	Yes
Wheel	/i:/	1	F	No

Table 8. Words included in the pre- and post-test

Appendix C

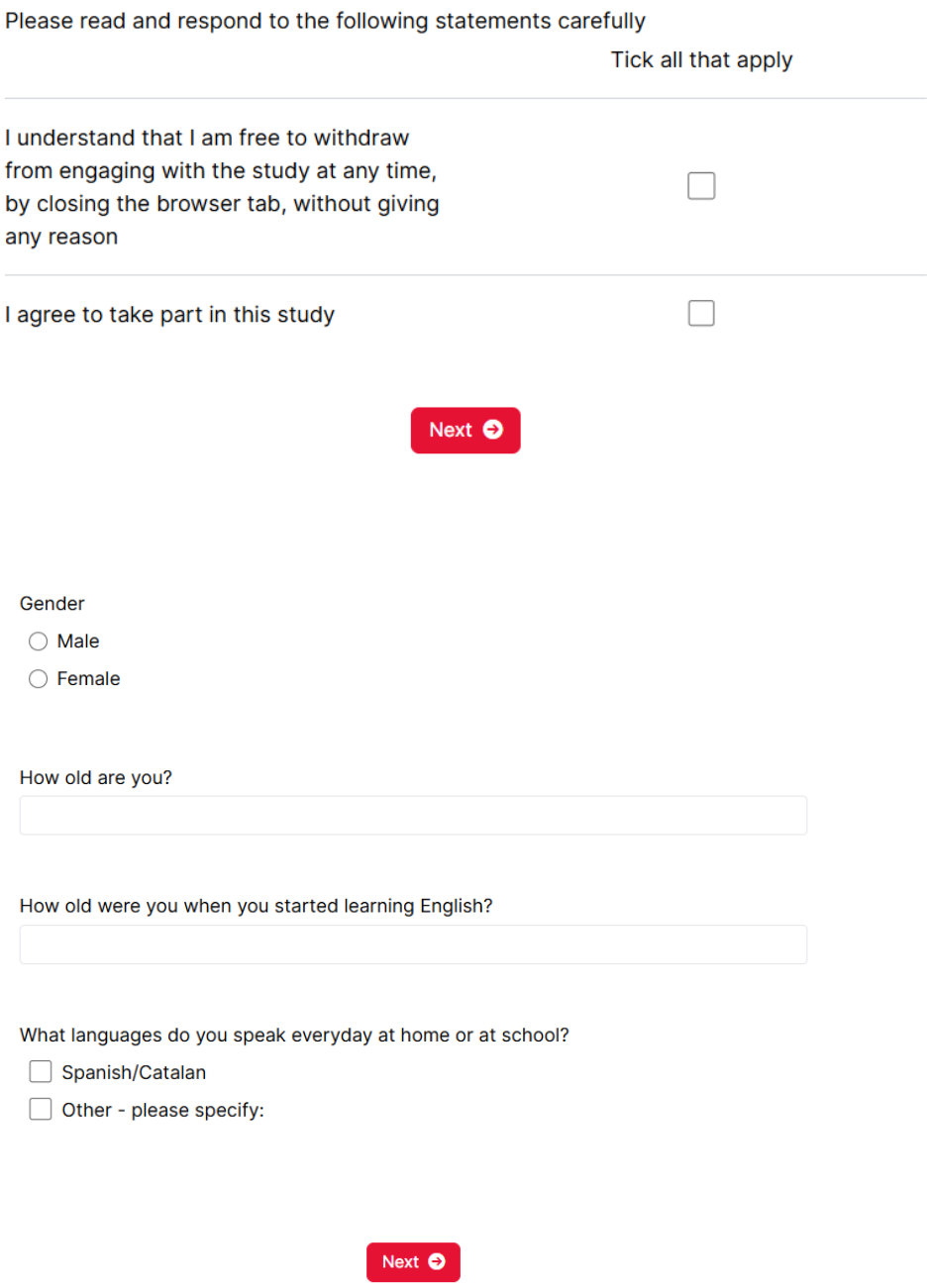

Figure 12. *Background information questionnaire*