

---

This is the **accepted version** of the journal article:

Carlet Foresti, Angelica; Cebrian, Juli. «The roles of task, segment type, and attention in L2 perceptual training». *Applied Psycholinguistics*, Vol. 43 Núm. 2 (2022), p. 271-299. 29 pàg. DOI 10.1017/s0142716421000515

---

This version is available at <https://ddd.uab.cat/record/274876>

under the terms of the  **CC BY-NC-ND** license

## **The roles of task, segment type and attention in L2 perceptual training.**

**Angélica Carlet, Juli Cebrian**

### **Abstract**

Previous studies show that attention plays an important role in second language (L2) phonetic attainment. This study compares the effect of two high variability phonetic training methods (identification (ID) and categorical AX discrimination (DIS)) on specifically targeted sounds and on implicitly exposed but untargeted sounds. Four groups of Spanish/Catalan bilingual learners of English were trained on either a subset of English vowels (/i: ɪ æ ʌ ɜ:/) or word-initial and word-final stops. The study also examined if the potential effect of training generalized to new stimuli and persisted two months after training. Results revealed that trainees significantly outperformed the controls in their identification of targeted sounds, and improvement generalized to new stimuli and was maintained after training, showing the efficacy of both training methodologies. However, while all trainees performed similarly with initial stops, ID trainees outperformed DIS trainees in vowel perception. Interestingly, only DIS trainees showed a significant improvement in the perception of untargeted sounds, indicating that this training method (possibly due to the absence of labels and the exposure to two physically present stimuli in each trial) may be more suited to enhance learners' perception of both targeted and untargeted sounds.

Second language speech, Perceptual tasks, Attention, Phonetic training

## Introduction

This study assesses the role of three factors (namely attention, target segment and type of perceptual task) in second language (L2) phonetic learning during a high-variability phonetic training (HVPT) regime in a foreign language instructional setting. It is well known that this environment is characterized by limited native input (Muñoz, 2008; Saito, 2015) and consequently little improvement in L2 pronunciation is generally reported (Darcy et al., 2012; Muñoz, 2008). Current L2 speech models such as the Perceptual Assimilation Model (PAM-L2, Best & Tyler, 2007), the Native Language Magnet Model (Kuhl & Iverson, 1995) and the Speech Learning Model (Flege, 1995a; and its revised version, Flege and Bohn's (2021) SLM-r) emphasize the role of experience in accurate L2 speech perception and production. HVPT, which simulates the phonetic variability characteristic of real speech by means of the inclusion of variable stimuli (i.e., from multiple talkers, phonetic contexts, stimuli types), becomes an important learning tool in this context, as it offers an alternative source of specialized input and specific focus on challenging L2 sounds in order to trigger the necessary processes required for L2 category learning. Perceptual training has been shown to be efficient in improving learners' ability to perceive (Iverson & Evans, 2009; Logan et al., 1991, among others) and to a lesser extent produce directly targeted L2 sounds (e.g., Iverson et al., 2012; see Sakai & Moorman, 2017 for a review).

To date, a fair amount of HVPT studies have been carried out looking at different L1-L2 language combinations, *different target structures and methodologies*, and reporting varying degrees of success (see Bradlow (2008) and Thomson (2018) for detailed reviews of phonetic training studies). These studies provide empirical evidence that L2 learners' perception can be modified as a result of laboratory training provided highly variable stimuli and appropriate

tasks are applied. The efficacy of phonetic training is typically evaluated by assessing L2 learners' changes in perceptual and/or production abilities as a consequence of training, particularly in comparison to a group of untrained learners. Importantly, evidence of stronger or robust learning is provided when there is generalization of learning to untrained structures and retention of learning beyond the immediate training effect (Bradlow, 2008; Flege, 1995b; Logan & Pruitt, 1995; Sakai & Moorman, 2017).

According to Logan and Pruitt (1995), the choice of training and generalization tasks is critical in order to understand the processes that underlie successful perceptual learning. Although both identification (ID) and discrimination (DIS) training tasks may be useful in training different perceptual skills (Logan & Pruitt, 1995), ID training has been said to be more effective and to lead to greater generalization effects than DIS training (Jamieson & Morosan, 1986; Strange & Dittmann, 1984). A possible reason might be the fact that ID procedures force listeners to attend to relevant between-category variability, while DIS tasks focus on within-category variability (Jamieson & Morosan, 1986). However, the superiority of ID tasks is often concluded from comparisons across studies, and studies that used auditory discrimination tasks, in which "same" trials involve physically identical stimuli. Flege (1995b) directly compared a categorical AX discrimination task and a forced-choice identification task in a study involving Mandarin speakers' perception of English unreleased final /t/ and /d/. Results revealed that both types of training enhanced perception and promoted generalization of learning equally, challenging previous results and views on each methodology. Similarly, three recent studies have also shown the effectiveness of both ID and DIS tasks when training perception of tone contrasts (Wayland & Li, 2008), coda nasals (Nozawa, 2015), and the /r/-/l/ contrast (Shinohara & Iverson, 2018). Still, greater improvement with ID tasks has been reported for vowel perception (Nozawa, 2015), and for the perception of the /z/ vs. /dz/ contrast

in coda position, a difficult contrast for Japanese learners of English (Law et al., 2019). Task effectiveness may thus also depend on the type and the level of difficulty of the target structure.

The current study follows up on these investigations and expands them by looking at the effect of the two types of HVPT methods – identification (ID) and categorical discrimination (DIS) - on the perception of targeted (i.e., trained) and untargeted (i.e., present in the stimuli but untrained) sounds. In particular, this study investigates how HVPT affects the perception of a subset of Standard Southern British English (SSBE) vowels and stops by native speakers of Spanish and Catalan. Furthermore, this study addresses the issue of whether the knowledge obtained through training generalizes to new stimuli and is retained over time.

### **Training attention**

In the fields of cognitive science and L2 speech, several studies have addressed the issue of attention. However, studies that have investigated the role of attention and explicit instruction are still scarce in the L2 phonetic training literature (e.g., Alves & Luchini, 2017; Pederson & Guion-Anderson, 2010; Nozawa, 2015). Attention is the cognitive process that allows learners to select and focus on specific stimuli in the input, while ignoring others (Schmidt, 2001). In L2 acquisition, researchers evaluate what kind of information learners pay attention to in the input, as it has been reported that not all information present in the stimuli is targeted (Broadbent, 1958). Importantly, paying attention has been found to be a pre-requisite for learning to take place (Schmidt, 2001), and greater attention control may lead to improved performance in the L2 (Mora & Mora-Plaza, 2019).

Regarding the role of attention in L2 sound perception, Pederson and Guion-Anderson (2010) suggest that attention orienting, which is the “aligning of attention with a source of sensory input or an internal semantic structure stored in memory” (Posner, 1980, p.4), is crucial in order to perceive L2 sounds accurately. In the L2 speech literature, researchers have attempted to manipulate attention orienting through three different

methods: a) input enhancement (e.g., Smith, 1993), b) cue enhancement (e.g., Jamieson & Morosan, 1986), and c) instruction manipulation (e.g., Alves & Luchini, 2017; Pederson & Guion-Anderson, 2010; Nozawa, 2015). As the latter case is the most closely related to the present investigation, a few studies that orient attention through instruction will be reviewed next.

Pederson and Guion-Anderson (2010) tested if directed attention promoted learning of phonetic information. The authors controlled participants' attention by training two groups of 21 English monolingual speakers on Hindi vowels and initial stop consonants with the exact same stimuli (i.e., 27 monosyllabic words beginning with one of the target stops and containing one of the target vowels). However, while one group was told to attend to vowels present in the stimuli, the other group was instructed to attend to consonants present in the same stimuli. Both groups of learners were trained by means of identification training with immediate feedback on the targeted (i.e., trained) segment. Testing consisted of a DIS task, which assessed learners' perception of the targeted and untargeted segments, that is, vowels and consonants for both groups. Their findings showed that ID training was successful in improving learners' perceptual abilities on the targeted sounds (this was found for the consonants, but not for the vowels, which were accurately discriminated already at pretest); however, training did not promote improvement with the untargeted segments. This result was replicated in a later study involving non-native consonants and tones, where improvement was only found with targeted contrasts (Chen & Pederson, 2021). The authors concluded that directing attention to specific target sounds with explicit instruction may be effective in promoting learning of the targeted sounds. Nozawa (2015) trained four different groups of Japanese learners of English on vowels and final nasal sounds with the exact same stimuli. Testing and training stimuli involved words containing seven American English vowel sounds embedded in /bVb/, /bVd/, bVg/, /bVm/, /bVn/ and

/bVŋ/ contexts. Two groups were trained on vowel perception, one by means of an ID task and the other by means of a categorical AXB task. The other two groups were trained on nasal consonant perception, also by means of the same tasks. Testing assessed the participants' perception of both vowels and nasals, independently of the focus of their training. Interestingly, results patterned differently for vowels and consonants. The ID group improved on vowel identification, outperforming the other experimental group, and thus suggesting that ID is superior to ABX DIS for training L2 vowels. However, DIS was as effective as ID for training word-final nasals. Furthermore, the two vowel-oriented groups showed a tendency of improvement with the untargeted sounds, as the nasal identification after training was numerically enhanced, albeit non-significantly (possibly due to a small sample size,  $n = 7$  per group). This finding suggests that simple exposure to sounds present in the stimuli may result in greater sensitivity to those sounds even if they are not the training focus.

Another study assessed whether explicit instruction had an impact on the effectiveness of a HVPT procedure (Alves & Luchini, 2017). The authors tested the identification and production of English word-initial voiceless stops by 24 Argentinian Spanish speakers. Participants were divided into two experimental groups and a control group. Both experimental groups were trained to identify voiced and voiceless stops with the same set of stimuli (voiced stop stimuli had 0 VOT and voiceless stimuli had long-lag VOT). However, one group was explicitly taught that initial voiceless stops in English are aspirated, and were instructed to attend to these segments, whereas the other group was not aware of which phonological aspect they should attend to during training. Both groups showed enhanced perception after the training regime, but a significant improvement in the production of two out of the three trained segments (/p/ and /t/) was observed only with the group in which learners were told to attend to the target stops. These findings suggest that

instruction, awareness of trained sounds and attention may play an important role in HVPT studies. Moreover, Nozawa's (2015) findings point to the possibility that, so long as sufficient and balanced exposure to targeted and untargeted sounds is provided, perception in both targeted and untargeted conditions may be enhanced. The present study thus aims to explore this issue further. Note that while training studies often make use of CVC stimuli in order to train a given target phone (typically either the consonants or the vowels), hardly any studies have investigated the effect of exposure to the untargeted sounds present in the CVC training stimuli trial after trial. It is possible that exposure to non-targeted contrasts present in the training stimuli across trials may be implicitly attended to by the learners in the course of a specifically designed training regime.

With these issues in mind, this study aims to assess the effects of two different training methods (categorical DIS and ID) on the perception of targeted and untargeted segments (namely vowels and stop consonants in both initial and final position) by Spanish/Catalan learners of English. Following previous studies that relate the efficacy of a training method to evidence of learning beyond the object of training (Flege, 1995b; Logan and Pruitt, 1995), this study also aims to assess if training gains generalize to new (i.e., untrained) stimuli and new voices, and if training benefits are maintained beyond the training period. In order to investigate this, two groups of learners were trained on vowels and two groups were trained on consonants. All groups were trained using the exact same stimuli and were tested on both consonants and vowels.

### **Training vowels and consonants**

Previous research points to the fact that vowels and consonants are perceived differently – e.g., perceptual phoneme boundaries are more sharply defined for consonants than for vowels – (Fry et al., 1962; Strange, 2007), involve different levels of processing in order to be learned (Pisoni, 1973), activate different neuronal patterns when processed



(Carreiras & Price, 2008) and thus may require different types of L2 phonetic training (Nozawa, 2015). For instance, Aliaga-García and Mora (2009) investigated the effect of six two-hour mixed-methods HVPT sessions on the perception and production of initial English stops (/p-b/ and /t-d/) and four English vowels (/æ-ʌ/ and /i-i:/) by a group of Catalan/Spanish native speakers. The tasks included articulatory explanations, perception tasks and practice with a production-based software. Training was effective in modifying learners' perceptual categorization of English /p/ and /b/. Moreover, a significant modification towards longer VOT production for the voiceless bilabial stop /p/ and a marginally significant effect for the alveolar voiceless stop /t/ were found. As for the vowels, training resulted in significantly enhanced discrimination scores for all target sounds. However, no improvement on vowel production accuracy, as examined acoustically in terms of first and second formant changes from pre-test to post-test, was observed. Interestingly, the findings of this study suggest that the phonetic training regime applied, which combined perceptual and production training tasks, promoted pronunciation gains differently for vowels and consonants. Moreover, the authors reported that the effect of phonetic training was found to be different in perception and in production.

Studies involving perceptual training only have also revealed different results for different target phones. In a study involving Catalan/Spanish bilinguals, Cebrian and Carlet (2014) assessed the effect of a three-week HVPT regime on the perception of four English consonant sounds (/v/-/b/ and /d/-/ð/) and two vowel pairs (/i:/-/ɪ/ and /æ/-/ʌ/) by advanced learners. The results showed a significant positive effect of training for a subset of the target consonants and vowels, namely /v/, /d/, /i:/, /ʌ/ and marginally for /b/. Interestingly, a non-native sound such as /v/ was already relatively successfully identified at pre-test and improved around 10% due to training. On the other hand, the English /i:/, which has been reported to have a near-identical counterpart in Catalan and Spanish (Cebrian, 2019, 2021),

was the least successfully identified vowel at the outset of the study. The investigators suggested that different factors might have influenced the studies' outcomes: the advanced learners' phonetic and metalinguistic knowledge in the case of /v/-/b/ and /æ/-/ʌ/, word frequency differences in the case of /d/-/ð/, and the influence of vowel duration in the /i:/- /ɪ/ distinction. Specifically, the tense vowel was more successfully identified when preceding a voiced consonant than a voiceless consonant, and the opposite held true for the lax vowel /ɪ/. This over-reliance on temporal cues was also found in previous L2 perceptual studies (Aliaga-Garcia & Mora, 2009; Cebrian, 2006; Kivistö-de Souza & Carlet, 2014). Thus, the findings in Cebrian and Carlet (2014) add to the existing HVPT literature by providing empirical evidence that the learnability of different target sounds may not be affected homogeneously, and underscore the role of context-dependent variability in L2 perception. Finally, further exploring the relationship between vowel and consonant acquisition, Lee and Hwang (2016) trained a group of 12-year-old Korean learners of English by means of identification tasks. The stimuli consisted of highly variable large sets of English vowel and consonant distinctions, covering the entire vowel and consonant systems of English. Results revealed that consonants were initially better identified and obtained greater improvement than vowels.

Still, some similarities between training vowels and training consonants have been reported. For instance, Lerdpaisalwong (2015) explored if training set size had an effect on consonant training. Expanding on Nishi and Kewley-Port's (2007) findings that training on a full set of vowels is more beneficial than training on a subset of vowels, Lerdpaisalwong trained two groups of Thai EFL learners on vowels and consonants with different training set sizes. The results provided empirical evidence that full set training is also more beneficial in the case of consonants. Lerdpaisalwong suggests that while vowels and

consonants differ in many respects, as far as training set size exposure is concerned, there is a relationship between the acquisition of L2 vowels and consonants.

Further studies contrasting vowel and consonant training are needed to understand better which methods are more effective with each type of sound. For instance, the discussion about the effectiveness of training tasks (ID vs. DIS) may not be settled without considering that learning consonants and vowels does not entail the same degree of effort from L2 learners (Pereira, 2014). A few previous studies have concluded that both ID and DIS tasks can lead to significant perceptual improvement, but these studies have focused on only one type of segment (Flege, 1995b; Shinohara & Iverson, 2018; Wee et al., 2019). To date, one study has suggested that ID tasks may be more suitable for training vowel identification than DIS tasks (Nozawa, 2015; c.f. Wee et al., 2019), and that there might be an increased sensitivity to untrained segments as a result of training (Nozawa, 2015).

All in all, the studies described above report training benefits for both vowels and consonants. However, some results indicate different degrees of improvement for different types of trained sounds (Aliaga-García & Mora, 2009; Cebrian & Carlet, 2014), which may be linked to differences in the way consonants and vowels are perceived and processed (Fry et al., 1962; Pisoni, 1973; Strange, 2007). Finally, few previous studies have examined the role of attention in L2 speech acquisition (Guion & Pederson, 2007; Pederson & Guion-Anderson, 2010), and one study has found evidence of training effects on untrained segments (Nozawa, 2015). Taking all of this into consideration, the questions motivating the current study are the following:

**1. Is HVPT equally effective for training L2 vowel and L2 stop identification?**

Which type of task, ID or categorical DIS, if any, is more suitable for each type of sound?

2. Can HVPT have an effect on the identification of untargeted as well as targeted L2 segments? Which type of training task, ID or categorical DIS, is more effective?

3. Does the effect of HVPT generalize to novel stimuli and persist two months after training? Which type of training task, ID or categorical DIS, is more effective for promoting generalization and retention of learning?

These research questions are addressed in a single study in which two groups of learners are trained on vowels and two groups are trained on consonants (either with ID or DIS tasks), while all participants are tested on both vowels and consonants, and the results obtained at pre-test are compared with results obtained at post-test, a generalization test and a delayed post-test. Improvement as a result of training, generalization and retention of knowledge are predicted to occur with targeted sounds (Pederson & Guion-Anderson, 2010; Nozawa, 2015) due to the role of feedback (Logan & Pruitt, 1995) and explicit instruction (Alves & Luchini, 2017). However, improvement of untargeted sounds might also occur if an indirect but continuous exposure to untargeted structures is provided in the training regime, as tendencies found in previous work may suggest (Nozawa, 2015). Despite the predicted overall improvement, the performance with vowels and consonants may differ, as vowels and consonants may differ in the way they are perceptually categorized (Fry et al., 1962; Strange, 2007), may involve different levels of processing (Pisoni, 1973) and may benefit from different types of perceptual training (Nozawa, 2015). Following the few previous studies that contrasted the two tasks investigated here (ID and categorical DIS), we may predict that ID can be more appropriate for training vowels (Nozawa, 2015), whereas both training tasks might be equally effective for training consonants (Flege, 1995b; Nozawa, 2015, Shinohara & Iverson, 2018).

## Methods

### Participants

Eighty-nine bilingual Catalan/Spanish speakers (14 male, 75 female; Mean age = 19.9, SD:1.36) participated in the present study. All participants were second-year English majors at Universitat Autònoma de Barcelona, in Barcelona, Spain. They were enrolled in an introductory phonetics and phonology course, and received course credit for their participation. At the time of testing, participants were also enrolled in their third semester of instrumental English courses, corresponding to a level of English between upper-intermediate and advanced (approximately a B2/C1 level in the Common European Framework of Reference for Languages: Learning, Teaching, Assessment (CEFR, Council of Europe, 2001)). Participants' mean age of first exposure to English was six (ranging from 2 to 12) and the mean number of years of formal instruction was 13 (ranging from 5 to 27). Moreover, none of the participants reported having spent longer than four months in an English-speaking country or reported having any vision or hearing difficulties. Participants were distributed among five homogeneous groups based on the global perceptual scores obtained at pre-test: four experimental groups and a control group (CG). The experimental groups were trained on either consonants (C) or vowels (V), either by means of discrimination tasks (DIS\_V, DIS\_C) or identification tasks (ID\_V, ID\_C)

### Target sounds and stimuli

The target sounds were the five Standard Southern British English (SSBE) vowels /i: ɪ æ ʌ ɜ:/ and the six SSBE stop consonants /p t k b d g/, placed either word initially or finally, since these English sounds are challenging for Catalan/Spanish learners of English (Aliaga-García & Mora, 2009; Cebrian et al., 2021; Fullana & Mora, 2007).

### *Training stimuli*

The training material consisted of 72 unmodified CVC non-words produced by four different SSBE native speakers (two males and two females) adding up to a total of 288 stimuli. Recordings took place in a soundproof chamber at the speech laboratory at University College London, in London, United Kingdom. A rhyming carrier sentence was used to ensure the desired pronunciation of the nonsense words (e.g., *It rhymes with badge, dagde. I say dadge now. I say dadge again.* Recordings were carried out using Cool edit 2000 software, a Rode NT-1AX microphone, Edirol UA25 audio interface and were digitized at a 44.1 kHz sampling rate and 16 bit quantification. The use of non-words made it possible to obtain a single and balanced set of stimuli with an equal number of word pairs used to train consonant contrasts and vowel contrasts. In addition, non-words eliminate a potential effect of word familiarity, and have been found to be more efficient than real words for training non-native contrasts (Ortega et al., 2021; Thomson & Derwing, 2016). As a general rule, the first non-word produced in the carrier sentence (before a pause) was extracted, so that most stimuli contained a naturally released final stop. In a few cases when the first word was not accurately pronounced or clearly articulated, the non-word produced before a vowel (again) or before a consonant (now) were selected. Importantly, all selected stimuli were validated by three native English speakers who correctly identified and positively rated the stimuli.

Table 1 presents all the non-word stimuli used in training. The stimuli were designed to train participants explicitly on either the consonant contrasts or the vowel contrasts, while exposing all participants to both vowel and consonant contrasts across trials. To that effect every non-word contained one of the five selected English vowels /æ ʌ ɪ i: ɜ:/ (plus /e/ and /ɑ:/)<sup>i</sup>

and one of the six English stops /p t k b d g/ either initially or finally, so that the same stimuli could be used in vowel perception and consonant perception training. For instance, *vap*, *vup*, *vab* and *vub* were used to train vowel trainees on the /æ/-/ʌ/ distinction (*vap* vs. *vup*, *vab* vs. *vub*), while implicitly exposing trainees to the final consonant voicing distinction across trials (*vap* vs. *vab*, *vup* vs. *vub*). Moreover, the same quadruplet of non-words was used to train consonant trainees on the final consonant voicing distinction /p/-/b/ (*vap* vs. *vab*, *vup* vs. *vub*) while indirectly exposing trainees to the vowel /æ/-/ʌ/ distinction across trials (*vap* vs. *vup*, *vab* vs. *vub*). As can be seen in Table 1, each quadruplet contrasted one of the target vowel pairs (/æ/-/ʌ/, /ɪ/-/i:/, /ɑ:/-/ɜ:/, /e/-/ɜ:/; see footnote i), and at the same time, one of the target stops pairs (/t/-/d/, /b/-/d/, /k/-/g/) either in initial or final position.

**\*Insert Table 1 here\***

### ***Testing stimuli***

Testing stimuli included a total of 81 non-words produced by two novel SSBE speakers (one male and one female). Since these speakers were not heard by learners during the training phase, testing assessed generalization to novel talkers. The 81 stimuli consisted of a subset of the training words, 30 non-words to test vowels (5 target vowels x 6 words), 24 non-words (6 target consonants x 2 contexts – word initial and word final – x 2 words) to test stops, 19 non-words (12 for consonants and 7 vowels) as practice tokens, and eight non-words as used as fillers. The generalization stimuli consisted of 68 novel CVC non-words in total (34 non-words x 2 talkers). There were 24 new consonant words (6 target consonants x 2 contexts – word initial and word final – x 2 words) and 10 new vowel words (5 target vowels x 2 words) produced by two familiar talkers from training.

## **Procedure**

### ***Training Tasks***

Participants were assigned to one of the four experimental groups. Two groups were trained on L2 vowel perception, one by means of a forced choice identification task and one by means of a categorical AX discrimination task, and the other two groups received training on L2 stops by means of the same tasks. The categorical AX discrimination task consisted of 288 trials involving the vowel and consonant pairs described in Table 1 above. Each trial consisted of a contrasting pair of non-word stimuli, each non-word produced by a different talker, with an interstimulus interval (ISI) of 1.15 seconds. Importantly, the order of the two non-words and of the talkers was counterbalanced throughout the experiment. The identification training task consisted of 576 trials, including 288 stimuli (72 words x 4 speakers), with two repetitions of each stimulus. Two repetitions were included in order to assure that trainees were exposed to the same stimuli and the same number of trials in both training regimes (ID and DIS), following Flege (1995b). The procedure of the training tasks used and the response options were the same as with the testing tasks, as explained below.

Training consisted of five sessions of approximately 25-35 minutes for all experimental groups. However, whilst vowel trainees were exposed to L2 vowel sounds for the whole amount of time, the consonant training time was halved, so that consonant trainees were exposed to both initial and final stops in each session. The order of the two segments was alternated, and participants were instructed to take a quick break before starting the following training session. The CG performed transcription exercises on the online platform *The web transcription tool* (Cooke et al., 2005), so that CG received a similar amount of target language instruction as the other groups without specific perceptual training. The training (and testing) tasks were administered using the software TP 3.1 (Rauber et al., 2012), which provided



immediate feedback on the directly trained segments after each trial and global feedback at the end of each session indicating the overall accuracy score.

### ***Testing Tasks***

Perception was evaluated 3 times (pre-test, post-test and delayed post-test) by means of forced choice identification tasks. Pre and post-test were administered immediately before and after training, respectively, and the delayed post-test was administered two months after the training was over. In the case of consonants, two tests were created: one for initial and one for final stops. The response options were the letters that represent the stops consonants in both the L1 and the L2, namely “p\_,” “t\_,” “k\_,” “b\_,” “d\_,” and “g\_” for initial consonants, and “\_p,” “\_t,” “\_k,” “\_b,” “\_d,” and “\_g” for final consonants. The vowel identification test had seven response alternatives consisting of phonetic-like symbols (the software did not accept phonetic fonts) together with two real words illustrating each sound tested to make sure the intended vowel was recognized, namely /æ/ *ash, mass*, /ʌ/ *sun, thus*, /ɪ/ *fish, his*, /i:/ *cheese, leaf*, /ɜ:/ *earth, first*, /e/ *less, west* and /a:/ *arm, palm*. Recent work has shown that keywords and phonetic symbols are similarly effective as labels in identification training (Fouz-González & Mompean, 2020). The words used to illustrate each sound were carefully selected by trying to minimize the use of stops as CVC onsets or codas, so they would be as dissimilar to the stimuli as possible. The total number of trials including both consonant and vowel trials was 251, including 19 practice trials and 216 test trials. The latter consisted of 96 trials to test stops (48 for initial and 48 for final stops; 6 consonants x 2 words x 2 talkers x 2 repetitions), 120 trials testing vowels (5 vowels x 6 words x 2 talkers x 2 repetitions) and 16 fillers. The generalization test, which was administered with the post-test, assessed both trained/targeted and untrained/untargeted sounds and the test layout and response options were exactly the same as

the ones used for the post-test. The generalization test thus consisted of 136 trials, that is, 48 trials for initial consonants and 48 trials for final consonants (6 consonants x 2 words x 2 talkers x 2 repetitions) and 40 trials for vowels (5 vowels x 2 words x 2 talkers x 2 repetitions).

## **Results**

The effects of two HVPT methods (ID and categorical DIS) on targeted and untargeted segments were analysed by comparing the participants' performance at pre-test, post-test, generalization test and delayed post-test. First, gain scores (i.e., the difference between post-test and pre-test scores) were computed and a series of linear mixed models were conducted on gain scores exploring the effects of (training) group and segment type. Then, a set of generalized linear mixed models were carried out to assess generalization and retention results separately for each segment type. The measure in this case was score (correctly or incorrectly identified) at each testing time. Thus, the results for each group at pre-test, post-test and generalization test were evaluated first, exploring the effects of group and test. Similarly, the results for pre-test, post-test and delayed test were examined for the participants who completed the delayed test (recall that a subset of participants did not complete the last test). For all statistical analyses (linear mixed models and generalized linear mixed models), several models were considered, which included the independent variables in each case and their interactions as the fixed effects, and different combinations of random intercepts and slopes for participant and stimulus. In all cases, the best fitting model, based on the lowest Akaike Information Criterion (AIC) obtained, was a model that included a random intercept for participant. In fact, for every analysis, the results for the different statistical models were very similar in terms the levels of significance obtained and the results of the pairwise comparisons. IBM Corp (2017) software was used. The

results for vowel-trained groups are presented first, followed by the results for consonant-trained groups.

### **Vowel training groups**

Correct identification scores at pre-test and post-test, gain scores and generalization scores for the two groups trained on vowels (ID\_V, DIS\_V) and the control group are shown in Table 2. As can be observed, the three groups performed numerically better at post-test than at pre-test on the identification of the targeted segment, vowels. This is particularly evident in the case of the ID\_V group, with an increase of 26 percentage points, and also observed with the DIS\_V group (9.8 increase). The numerical improvement obtained by the control group (3.7 percentage points) may be due to familiarity with the test, after having completed the same test at pre-test. It may also reflect a small general improvement from exposure to English in their studies. The results for the untargeted sounds were less notable and less consistent, as only the DIS\_V group appeared to show some improvement (7.7 increase with final consonants).

**\*Insert Table 2 here \***

As the groups did not differ statistically at pre-test ( $F(4, 95) = .078, p = .614$ ), the effect of training was explored, first, by comparing the amount of gain for targeted and untargeted sounds in a series of linear mixed-effects models. The best fitting model was a linear mixed model exploring the fixed effects group (ID, DIS and CG), segment type (vowels, initial consonant and final consonant) and a group by segment type interaction, with a random intercept for participant. The analysis revealed a significant main effect of group ( $F(2, 153) = 6.32, p = .002$ ), and segment type ( $F(2, 153) = 46.83, p < .001$ ) and a

significant group by segment type interaction ( $F(4, 153) = 22.11, p < .001$ ). Pairwise comparisons with a Bonferroni correction confirmed that the two experimental groups outperformed the controls (overall increase from pre- to post-test: ID\_V: 7.1, DIS\_V: 6.3, vs. CG: 1.7;  $p < .001, d = .77$  and  $p = .01, d = .59$ , respectively)<sup>ii</sup>. This is illustrated in Figure 1. The effect of segment type can be explained by the fact that the targeted segments (vowels) obtained higher gains, namely a 13.3 increase, than the untargeted segments (initial consonants: -0.03, final consonants: 2.2). This was confirmed by Bonferroni pairwise comparisons (vowels vs initial consonants:  $p < .001, d = 1.77$ ; vowels vs final consonants.  $p < .001, d = 1.47$ ). No significant difference was found between the initial and final consonants ( $p = .433$ ). The group by segment type interaction can be explained by the identification group's much greater difference in gain between vowels and consonants in comparison to the other groups, as is illustrated in Figure 2.

**\*Insert Figure 1 Here \***

**\*Insert Figure 2 Here \***

The effect of training group on each segment type was analyzed next comparing the results for the pre-test, post-test and generalization test. Recall that the generalization test assessed the identification of new stimuli produced by familiar (training) talkers. As can be observed from Table 2, regarding vowels, both vowel-trained groups obtained generalization scores that were equal or even higher than post-test scores. Interestingly, despite a small improvement from pre-test to post-test, the control group also performed comparatively well with the new stimuli, albeit reaching lower identification scores than the trained groups. A GLMM analysis, with test (pre-test, post-test, generalization), group (CG, DIS\_V and ID\_V) and their interaction as fixed effects, and a random intercept for participant yielded a significant effect of test ( $F(2, 15111) = 168.65, p < .001$ ), group ( $F(2, 15111) = 11.02, p < .001$ ), and a

significant interaction ( $F(4, 15111) = 38.86, p < .001$ ). Bonferroni-adjusted pairwise comparisons confirmed that groups did not differ at pre-test and both training groups outperformed the controls both at post-test (ID\_V - CG:  $p < .001, d = 1.77$ ; DIS\_V - CG:  $p = .035, d = 1.96$ ) and generalization test (ID\_V - CG:  $p < .001, d = 1.09$ ; DIS\_V - CG:  $p = .044, d = .72$ ). The ID\_V group also outperformed DIS\_V at post-test ( $p < .001, d = 1.18$ ). In addition, the pre-test scores for both experimental groups (ID\_V and DIS\_V) were significantly lower than their post-test scores ( $p < .001, d = 1.25$  and  $p < .001, d = 1.26$ , respectively) and generalization scores ( $p < .001, d = 2.25$  and  $p < .001, d = 2.83$  respectively). CG's small improvement from pre- to post-test (54%-58%) proved significant too ( $p = .019, d = .64$ ). Interestingly, CG's generalization scores (68%) were also significantly higher than pre-test ( $p < .001, d = 1.45$ ) and post-test scores ( $p = .007, d = 1.23$ ). Still, CG's generalization scores (68%) were significantly lower than DIS\_V's (76%,  $p = .044, d = 0.72$ ) and ID\_V's (80%,  $p < .001, d = 1.09$ ) scores, as indicated above.

Regarding initial stops, the GLMM, with the same characteristics as the analysis of vowels, yielded no significant effects and no interaction for initial consonants (group:  $F(2, 7767) = .285, p = .752$ ; test:  $F(2, 7767) = .066, p = .937$ ; test by group:  $F(4, 7767) = 1.263, p = .282$ ). In the case of the final stops, there was a significant effect of test ( $F(2, 7767) = 8.35, p < .001$ ) and a test by group interaction ( $F(4, 7767) = 2.593, p = .035$ ), but no effect of group ( $F(2, 7767) = .66, p = .516$ ). For all three groups, generalization scores were lower than pre-test or post-test scores. DIS\_V was the only group to show improvement from pre-test to post-test (65 to 72%,  $p < .001, d = .63$ ), but this improvement was not evident with new stimuli (61%).

In brief, the two vowel-trained groups improved their identification of the targeted sounds (vowels) and showed generalization effects, with ID\_V showing the greatest improvement. The controls also show improvement and generalization with vowels but they

were always outperformed by the trainees. Regarding untargeted sounds, only DIS\_V revealed some improvement for final stops, although the improvement did not generalize to new stimuli.

### **Consonant training groups**

Table 3 shows the correct identification scores obtained at pre-test and post-test, the gain score and the results of the generalization test for the consonant training groups (ID\_C, DIS\_C) and CG. The results show that in general all groups obtained the lowest correct identification scores for vowels (53-56%), followed by the scores for final stops (65-71%) and initial stops (70-78%). The consonant training groups obtained numerically higher scores at post-test than at pre-test for the targeted segments (initial and final consonants). But only the DIS group shows a notable numerical difference with the untargeted sounds (from 55% at pre-test to 64.7% at post-test).

**\*Insert Table 3 here\***

As can be observed the DIS\_C group's identification of initial and final consonants improved by 12.8 and 4.3 percentage points, respectively, from pre-test to post-test, whereas the ID\_C group's improvement for initial and final consonants reached 15.9 and 5.5, respectively. The amount of gain evidenced by the control group in this case was comparatively small (0.9 increase for initial consonants and 0.5 for final consonants). A linear mixed model with group (ID, DIS and CG), segment type (vowels, initial consonant and final consonant) and a group by segment type interaction as fixed effects and random intercept for participant revealed a significant effect of group ( $F(2, 144) = 8.71, p < .001$ ), and segment type ( $F(2, 144) = 7.40, p = .001$ ) and a significant interaction ( $F(4, 144) = 7.06, p < .001$ ). Bonferroni-adjusted pairwise contrasts revealed that the trained groups

significantly outperformed the controls (DIS\_C – CG:  $p < .001$ ,  $d = .85$ ; ID\_C – CG:  $p = .008$ ,  $d = .69$ ) on the overall performance (targeted and untargeted segments) and did not differ significantly from each other ( $p = .226$ ). The overall gain across segment types for each of the three groups is illustrated in Figure 3. The effect of segment type is explained by the greater improvement with initial stops than with the other two segment types, which was confirmed by Bonferroni pairwise comparisons ( $p = .002$ ,  $d = .73$  when comparing with vowels and  $p = .002$ ,  $d = .71$  when comparing with final consonants). The group by segment type interaction is explained by the fact that while the control group obtained similar results for different segment types, the trained groups, particularly the ID group, showed greater differences between segment types (see Figure 4).

**\*Insert Figure 3 Here\***

**\*Insert Figure 4 Here\***

As previously done with the vowel training groups, GLMM analyses were carried out to explore the effects of test and group for each segment type separately. Again, the best fitting model was a GLMM with group (ID\_C, DIS\_C, GG), test (pre-test, post-test, generalization) and their interaction as fixed effects and a random intercept for participant. Regarding the targeted segments (consonants), the results for initial consonants yielded a significant effect of test ( $F(2, 7287) = 42.8$ ,  $p < .001$ ) and a significant interaction ( $F(4, 7287) = 8.62$ ,  $p < .001$ ), but no main effect of group ( $F(2, 7287) = 1.44$ ,  $p = .236$ ). Bonferroni-adjusted pairwise comparisons revealed one more time that groups did not differ at pre-test, but ID\_C outperformed CG ( $p = .004$ ,  $d = .92$ ) and marginally DIS\_C ( $p = .054$ ) at post-test. In addition, the difference between post-test and pre-test was significant for both trained groups (ID\_C: 16 percentage points,  $p < .001$ ,  $d = 1.49$ ; DIS\_C: 13

percentage points,  $p < .001$ ,  $d = 1.04$ ), but not for CG (0.9 increase,  $p = .01$ ). Finally, improvement generalized to new stimuli for both trained groups (generalization scores were significantly higher than pre-test scores ( $p < .001$ ,  $d = .99$  for ID\_C and  $p < .001$ ,  $d = 1.05$  for DIS\_C) and equal to post-test scores).

Regarding final stops, the GLMM in this case yielded a significant effect of test ( $F(2, 7287) = 4.9$ ,  $p = .007$ ), but no group effect ( $F(2, 7287) = 2.41$ ,  $p = .09$ ) and no interaction ( $F(4, 7287) = 1.16$ ,  $p = .329$ ). Pairwise comparisons showed that only ID\_C's difference between pre-test and post-test (5.5 percentage points) was significant ( $p = .032$ ,  $d = .63$ ), and improvement generalized to new stimuli (post-test results and generalization results did not differ significantly). Interestingly, when looking at the untargeted segments (vowels) the pattern observed earlier with the vowel training groups is replicated as the DIS group shows the greatest improvement (9.7 percentage points). The GLMM analysis yielded a significant test effect ( $F(2, 14231) = 45.49$ ,  $p < .001$ ), a marginal effect of group ( $F(2, 14231) = 2.72$ ,  $p = .066$ ) and a significant interaction ( $F(4, 14231) = 8.39$ ,  $p < .001$ ). Bonferroni pairwise comparisons showed that DIS\_C trainees outperformed ID\_C at post-test ( $p = .032$ ,  $d = .90$ ). As was found for the vowel-trained groups, all groups obtained higher vowel identification scores in the generalization test than in the post-test, but DIS\_C's scores were significantly higher than CG's ( $p = .046$ ,  $d = .75$ ) and ID\_C's ( $p = .002$ ,  $d = 1.51$ ).

In summary, consonant-trained groups significantly improved their identification of targeted sounds (initial stops), outperforming the controls and showing generalization to new stimuli. This outcome was also found with final stops for ID\_C, albeit to a lesser extent. Similar to what was found with the vowel-trained groups, DIS trainees, but not ID trainees, showed significant improvement and generalization effects with untargeted sounds (vowels).



### **Delayed post-test results**

The delayed post-test was administered two months after the post-test to assess the long-term effects of the training. The scores obtained by each group and test can be seen in Table 4. Unfortunately, not all of the original participants returned to complete the delayed post-test. Thus, only the results of the trainees that were present at all three testing times were analysed (71% of the original participants, namely CG: 9 learners; ID\_V: 17; DIS\_V: 12; ID\_C: 12; DIS\_C: 13). Hence scores at pre- and post-test shown in Table 4 differ slightly from those shown in previous tables. As in the case of the generalization results, delayed post-test results were particularly relevant when a significant improvement from pre-test to post-test was observed (Flege, 1995b; Cebrian & Carlet, 2014).

**\*Insert Table 4 here \***

As can be seen in Table 4, for all groups the delayed post-test results were either numerically higher or similar to the post-test results. Each group's performance on each segment type was submitted to a series of GLMM analyses in the same fashion as was done for the pre-test, post-test and generalization results. Regarding the vowel-trained groups and their results for targeted sounds (vowels), a GLMM with time (pre-test, post-test and delayed post-test), group (ID\_V, DIS\_V, CG) and their interaction as the fixed effect and a random intercept for participant yielded significant effects of test ( $F(2, 13671) = 164.5, p < .001$ ), group ( $F(2, 13671) = 8.25, p < .001$ ), and a significant interaction ( $F(4, 13671) = 42.92, p < .001$ ). Bonferroni-adjusted pairwise comparisons confirmed that groups did not differ at pre-test. For ID\_V and DIS\_V, post-test results (80% and 63%, respectively) and delayed post-test results (80% and 60%) did not differ.

In addition, the pre-test scores for both experimental groups (ID\_V and DIS\_V) were significantly lower than their post-test scores ( $p < .001$ ,  $d = 3.77$  and  $p < .001$ ,  $d = 1.17$ , respectively) and delayed post-test scores ( $p < .001$ ,  $d = 3.44$  and  $p < .001$ ,  $d = .97$ ). The same pattern of results of improvement with CG observed for the whole group also emerged at this time, as CG's scores showed an improvement from pre- to post-test as well (57-62%,  $p = .019$ ,  $d = .94$ ), and their results at delayed test was also significantly higher than at pre-test ( $p < .001$ ,  $d = 1.14$ ).<sup>iii</sup> Still, the experimental groups' (ID\_V and DIS\_V) scores were significantly higher than CG's both at post-test ( $p < .001$ ,  $d = 1.79$  and  $p < .001$ ,  $d = .09$ , respectively) and delayed post-test ( $p < .001$ ,  $d = .96$  and  $p < .001$ ,  $d = .91$ ). With respect to the untargeted sounds, the GLMM yielded no significant effects and no interaction concerning initial stops (test:  $F(2, 5463) = .951$ ,  $p = .386$ ; group:  $F(2, 5463) = 1.53$ ,  $p = .734$ ; test by group:  $F(4, 5463) = 1.96$ ,  $p = .098$ ). Finally, there was a marginal effect of test in the analysis of final stops ( $F(2, 5463) = 2.882$ ,  $p = .56$ ), but no effect of group ( $F(2, 5463) = 1.119$ ,  $p = .327$ ) and no interaction ( $F(4, 5463) = 1.932$ ,  $p = .102$ ). Follow-up pairwise comparisons indicated that only in the case of DIS\_V there was a significant improvement from pre- to post-test (64-72%,  $p = .01$ ,  $d = .51$ ), which was maintained at delayed post-test (71%,  $p = .01$ ,  $d = .48$ ).

The same statistical analyses were conducted regarding the consonant-trained groups. One more time, the best model was a GLMM with test, group and their interaction as fixed effects and a random intercept for participant. With respect to the targeted sounds, the analysis of the initial stops yielded a significant effect of test ( $F(2, 4887) = 39.8$ ,  $p < .001$ ) and a test by group interaction ( $F(4, 4887) = 4.33$ ,  $p = .002$ ), but no effect of group ( $F(2, 4887) = 0.21$ ,  $p = .813$ ). Bonferroni-adjusted pairwise comparisons revealed that ID\_C and DIS\_C's results at pre-test (71% and 72%) differed significantly from post-test results (88% and 85%,  $p < .001$ ,  $d = 1.61$  and  $p < .001$ ,  $d = 1.02$ , respectively) and delayed post-

test (84% and 86%,  $p < .001$ ,  $d = .78$  and  $p < .001$ ,  $d = 1.16$ , respectively), and there was no difference between the two post-training tests. CG's scores at pre-, post- and delayed post-test did not differ. The analysis of final stops yielded no significant results (test:  $F(2, 4887) = 1.84$ ,  $p = .159$ ; group:  $F(2, 4887) = .31$ ,  $p = .734$ ; test by group:  $F(4, 4887) = 0.49$ ,  $p = .746$ ). Unlike the previous analysis involving the whole group, where ID\_C's improvement from pre- to post-test in final stop identification reached significance (70% to 75%), the numerical difference between pre- and post-test did not reach significance in the current analysis (71% to 75%). The lack of significance is probably related to the smaller group size. Regarding untargeted sounds (vowels), the GLMM indicated a significant effect of test ( $F(2, 12231) = 35.8$ ,  $p < .001$ ) and a test by group interaction ( $F(4, 12231) = 7.3$ ,  $p < .001$ ), but no effect of group ( $F(2, 12231) = 0.84$ ,  $p = .433$ ). In this case, the original result was replicated as DIS\_C were found to improve significantly from pre- to post-test (54%-64%,  $p < .001$ ,  $d = 1.37$ ) and improvement was maintained at the delayed post-test (65%,  $p < .001$ ,  $d = 1.39$ ).

To summarize, the analysis of the pre-test, post-test and delayed post-test results for the participants who completed all three tests follows very closely the results obtained for the whole group regarding the effect of training on targeted and untargeted sounds: vowel trained groups improved on vowels (to a greater extent than CG), consonant groups improved on initial stops, and DIS-trained groups showed improvement on untargeted sounds. Importantly, the fact that delayed post-test scores replicated post-test scores in all cases where there was an improvement from pre-test to post-test provides ample evidence for retention of knowledge after training, both for vowels and for consonants.

## **Discussion and conclusions**

This study examined the roles of attention, segment type and perceptual task in L2 learning by assessing the effectiveness of two HVPT training methods (ID vs. categorical DIS) on the perception of targeted and untargeted L2 sounds by Spanish/Catalan learners. More specifically, it investigated the effect of each training method on different segment types by comparing the perception of five SSBE vowels (/i: ɪ æ ʌ ɜ:/) and the stop consonants in word initial and word final position. Two experimental groups were trained on vowels and two groups were trained on stop consonants, in a five-session training regime involving the exact same set of stimuli for all trainees. All trainees and a control group of untrained learners were tested on their identification of vowels and stops before, immediately after, and 2 months after training. The overall findings confirm the positive effect of HVPT in enhancing L2 learners' perceptual abilities, in line with previous studies (e.g., Bradlow, 2008; Thomson, 2018). This study provides strong evidence that both ID and DIS tasks are effective methods for training to perceive L2 vowels and initial consonants, although ID may be more suitable in the case of vowels. In addition, both tasks equally contribute to generalization and to retention of learning two months after the training regime ended. Moreover, the results show that the DIS method also facilitates learning of untargeted sounds. Finally, results for the untrained control group showed that CG also improved in their identification of vowels (but not of consonants). Still, the trained groups' improvement was of a significantly greater magnitude than CG's. CG's results may reflect an effect of task familiarity after having completed the pre-test, or may show an effect of continuous exposure to English at university, or metalinguistic knowledge acquired in their English phonetics course. Still, the superiority of the trained groups' results across segment types, and particularly of ID\_V on vowels, strongly support the effectiveness of HVPT. These findings are discussed below in light of the questions raised in the introduction.

## Segment types and tasks

Regarding the effect of training on the sounds that are the focus of training (targeted sounds), the results show that both training tasks are efficient. However, in the case of the vowels, the ID task resulted in significantly greater gains (by 26 percentage points) than the DIS task (9.8 percentage points), thus appearing to be better suited for training L2 vowels. This result is in line with Nozawa (2015), who found that the ID method was superior to DIS in promoting L2 vowel gains. By contrast, Wee et al. (2019) report that both types of task were equally effective for training the perception of English /i:/-/ɪ/ by Japanese speakers. Still, several differences exist between these studies. Nozawa's study and the current study involved natural stimuli and included a variety of vowels. On the other hand, the training and testing stimuli in Wee et al. were drawn from a /i:/-/ɪ/ temporal and spectral continuum, and the study's focus was on cue-weighting, thus examining a different type of perceptual skill. Further, unlike the current study, neither Nozawa (2015) nor Wee et al. (2019) included a control (untrained) group in their design, or a test of generalization or of retention of knowledge. Thus, the evidence provided in the current study is stronger and points to a greater effect of identification training for vowel perception. The possible reasons for this benefit are further discussed below, when summarizing the results for vowels and stops.

Regarding stops, both training methods promoted gains in the identification of stops in onset position to the same extent. This result replicates earlier findings showing the positive effect of both training methods on the identification of final stops by Mandarin learners (Flege, 1995b), coda nasals by Japanese learners (Nozawa, 2015) and on the identification of /r-l/ by Japanese learners of English (Shinohara & Iverson, 2018). However, little evidence of training effect was observed for targeted final stop consonants in the current study (ID\_C's improvement from 70% to 75% was significant, albeit of a

small size effect, but notably smaller than ID\_C's and DIS\_C's improvement with initial stops from 73% to 89% and from 69.5% to 82%, respectively). The greater difficulty when identifying the coda stops may be due to the fact that the final stop voicing distinction is non-existent in the learners' L1 (Flege, 1989). Moreover, the training time for consonants was quite limited (five thirty-minute sessions divided into two tasks, one for each word-position). This may suggest that training final consonant perception requires a longer amount of time and effort than training initial consonant perception. Flege (1995b) used seven training sessions and obtained improvement. It is also possible that the lack of an effect of training on the identification of final stops lies in the fact that final stop voicing contrasts are cued by the preceding vowel duration in English (among other cues), and thus are usually perceived more continuously than categorically (Pisoni, 1973). In addition, the difference between short lag and long lag VOT for initial stops may be a more robust cue than the more gradient vowel duration differences cuing final stops (Burnham, 1986). All in all, these results point to the fact that the trainability of initial voicing contrasts might be relatively easy in comparison to other distinctions (Strange & Dittmann, 1984), and can be modified through perceptual training (Collet et al., 2013).

Thus, regarding the first research question, this study has shown that HVPT is effective for training L2 vowel and L2 initial stop identification. However, the ID task proved superior for L2 vowel training, while both tasks seem to be equally effective for training initial stops. This difference may be due to the fact that consonants and vowels may involve different training procedures (Nozawa, 2015) and different amounts of training, since consonants and vowels do not entail the same degree of effort on the part of L2 learners (Aliaga-García & Mora, 2009; Pereira, 2014). Moreover, the phonetic and phonological differences between consonants and vowels may explain the different results obtained with different training tasks. For example, vowels and consonants have been found

to be perceived differently (Fry et al., 1962; Strange, 2007), require different levels of processing when learned (Pisoni, 1973), and activate different neuronal patterns when processed (Carreiras & Price, 2008). According to Liberman et al. (1967), consonants are generally more acoustically stable than vowels, which facilitates perception. Regarding the difference between the tasks, ID tasks are said to enhance between category sensitivity and involve higher levels of phonological encoding that are more relevant for L2 categorization, whereas DIS tasks tend to promote within-category sensitivity and tap into lower levels of phonological encoding (Iverson et al., 2012; Jamieson & Morosan, 1986; Logan & Pruitt, 1995). It may be the case that training through higher levels of phonological encoding is more beneficial for vowel perception, since vowels are perceived more continuously than categorically (Strange, 2007). On the other hand, initial stops benefit equally from a task which trains higher level of phonological encoding and a task which trains lower level of phonetic processing, which may be connected to the acoustic stability found in consonants (Liberman et al., 1967).<sup>iv</sup>

### **Training attention**

Another noteworthy outcome of this study was the effect of HPTV on untargeted but implicitly exposed segments, showing that L2 learning may occur even without “attention orienting” (Posner, 1980) and explicit instruction. This result goes in line with Alves and Luchini (2017), who found that while specific instruction led to some improvement in VOT production, an overall improvement in the identification of stops was found for all trained groups regardless of the presence of explicit instruction. Further, Nozawa (2015) found that the two vowel-oriented groups increased the accuracy with which they identified L2 coda nasals present in the stimuli without any specific training on these specific segments. On the other hand, Pederson and Guion-Anderson (2010), who trained one group of English speakers

on Hindi vowel contrasts and another group on Hindi stop contrasts, observed improvement in the identification of trained contrasts only. Still, Pederson and Guion-Anderson's study differs from the Alves and Luchini (2017), Nozawa (2015) and the current study in that Pederson and Guion-Anderson tested monolingual speakers on non-native contrasts (as opposed to L2 learners who already had some experience with the L2), and tested identification-trained participants by means of discrimination tasks, thus using a different type of measure from the other studies.

In any event, the pattern of results in the current paper reveals an effect of training on untargeted sounds for both vowels and final consonants when a categorical DIS task (DIS\_C on vowels, DIS\_V on final consonants) was applied. A few possible reasons for this difference between ID and DIS tasks in their ability to have an effect on untargeted sounds are considered. First, it is relevant to consider that the ID task introduces one stimulus at a time to the learner, which might force learners to directly compare the given stimulus with a previously stored mental representation of the corresponding sound category (covert task type, Bohn, 2002). By contrast, a categorical DIS task exposes learners to two physically present stimuli to be compared at every trial (overt task type, Bohn, 2002), which in turn might increase learners' awareness of the untargeted segments. A second and related explanation may be connected to the nature of the responses in each task (same/different response vs. responses that represent actual sound categories), as discussed below. A third explanation is related to the fact that DIS tasks might enhance within-category variability awareness, instead of between-category variability, as the ID task (see Logan & Pruitt, 1995). Since the perception of final consonants is partly cued by the within-category properties present in the preceding vowel sounds, the Spanish/Catalan learners might have been able to apply the enhanced awareness of the preceding vowel duration differences when identifying final consonants. In addition, the forced-choice ID method strictly directed learners' focal attention to the input which was



considered important for further processing (i.e. the target sounds). This is due to the presence of labels in this task, which forces learners to categorize each stimulus as one of the options provided (Jamieson & Morosan, 1986). On the other hand, the absence of such labels in the same/different DIS task may allow listeners to pay attention to the whole stimulus (targeted and untargeted segments). Furthermore, the difference between the two training tasks is also evident in the nature of the feedback they provide. In the ID task feedback consists of revealing the identity of the crucial sound in the stimulus. Thus ID feedback was more explicit and involved focus on the target sounds, i.e., focus on form (Long, 1991), since it involved the information of specific symbols and common spellings for each sound. According to Long (1991), focus on form consists of an occasional shift of attention to linguistic code features. Thus, it can be said that an ID task succeeded in orienting learners' focal attention solely to the target sounds while abstracting away from other cues present in the stimuli. This is in line with Pederson and Guion-Anderson (2010)'s findings, since their participants only showed gains on the directly-trained segment. By contrast, the feedback received through the DIS training was more implicit, since it strictly reflected the listeners' capacity to distinguish between the two sounds, not their ability to relate the given sound to a response category represented by a symbol or spelling. The DIS training may have allowed learners thus to focus their attention on any cues present in the stimuli.

Interestingly, DIS\_V was the training method that enhanced learners' perception of final stops the most from pre to post-test (7.7 percentage points), followed by ID\_C (5.5 percentage points). This finding provides further evidence of the interaction between consonant and vowel cues (Recasens & Mira, 2015; Steinlen, 2005). It is possible that DIS\_V trainees were able to detect the vowel duration cue to final obstruent voicing (Raphael, 1972; Roach, 2000) through the vowel training regime received, which provided exposure to obstruent voicing contrasts across trials (e.g., vab – vub and vap – vup). Recall

that previous studies have reported that Spanish/Catalan learners of English may be sensitive to phonetically conditioned vowel duration differences determined by the voicing nature of the following obstruent (Cebrian, & Carlet, 2014). Hence, DIS\_V may have been able to detect the relationship between vowel duration and final obstruent voicing by contrasting stimuli from subsequent trials, and then apply this knowledge correctly when identifying the coda stops. Since this systematic variation in vowel duration of the preceding vowel is a within-category characteristic (i.e. vowels becoming shorter when preceding a voiceless sound), it would make sense that only DIS trainees would become sensitive to such cue. Hence, the within-category differences present in vowels may not be necessary or strictly relevant for the identification of the vowel target sounds (Jamieson & Morosan, 1986), but they might be an applicable cue for the identification of the final stop that follows this given vowel, explaining the cross-training effect here observed.

The DIS\_C group also significantly improved in their identification of the untargeted segments (vowels), with a 9.7% gain from pre to post-test, despite not having received specific feedback on this segment. This may also be related to the role of preceding vowel duration in the final stop distinction. Being trained on final stops, participants may have been directed to focus their attention on the vowel duration cue (Raphael, 1972; Roach, 2000) rather than attempting to rely solely on the more variable characteristics of the final stop. In fact, this enhanced ability with vowels as an effect of consonant training has previously been reported in a different area of study: L1 phonological therapy. In an unpublished study, Stemberger (J. Stemberger, personal communication, September 8<sup>th</sup>, 2015) observed that, by training pre-schoolers with phonological impairment on consonant sounds, the infants improved their ability of producing consonant and vowel sounds.

In brief, regarding the second research question, the results suggest that DIS training is more likely than ID training to have an impact on untargeted L2 sounds (vowels and final

stops), since this type of task does not direct learners' attention exclusively to the target sounds and/or limit the learners' perception to the given labels (Polka, 1992). This is one of the key contributions of this study, since it shows that explicit instruction and "attention orienting" (Posner, 1980) are not the sole pre-requisites for L2 perception learning to take place in an HVPT approach.

### **Evidence of robust learning**

The third research question involved the effect of high variability perceptual training on generalization to novel items and retention of learning after two months of the training completion. First of all, improvement from pre-test to post-test already shows one type of evidence of generalization as testing stimuli were produced by different talkers from training stimuli. Thus, learners were able to apply the learning received from exposure to training stimuli to new voices. Regarding the generalization to novel non-words, the gain on the targeted segments for all experimental groups was maintained or even increased when tested on different words produced by familiar talkers. The positive generalization results on targeted segments and the large effect sizes obtained in both cases confirm that both training methods here applied (ID and categorical DIS) promoted generalization effects that went beyond the training stimuli, providing evidence of robustness of learning (Logan & Pruitt, 1995). Regarding the generalization scores of untargeted sounds, only the DIS\_C generalized the learning outcomes to novel non-word stimuli, confirming the efficacy of this training method and the strength of the effect on untargeted segments. On the other hand, the DIS\_V training group was not found to generalize the improvement on untargeted segments (final consonants) to novel non-word stimuli. It is possible that the improvement previously observed (from pre-test to post-test) was stimuli-specific, not being consolidated enough to promote generalization to novel tokens (Logan & Pruitt,

1995). In fact, all groups tended to be more successful in the generalization test than at posttest in their identification of vowels, even CG. The exception was ID\_V, whose results were equally high at post-test and generalization test, and were always higher than the other groups'. It is possible that the vowels in the generalization non-word stimuli were easier to identify than vowels in pre- and post-test due to a more careful articulation, or to a greater familiarity with the voices, as the novel stimuli were produced by the same talkers as the pre- and post-test stimuli<sup>v</sup>. Still, ID\_V, DIS\_V and DIS\_C's generalization scores were significantly higher than CG's, supporting the idea that the results were related to the training received.

Turning our attention to the retention effects, the results showed that all experimental groups were able to maintain the post-test scores at the delayed post-test phase for the targeted sounds, that is, vowels for vowel trainees and initial consonants for consonant trainees. These results revealed that training effects on targeted sounds can be retained even after the training regime is over, in line with several previous studies (Bradlow et al.,1999; Nishi & Kewley-Port, 2007;Wang & Munro, 2004). Interestingly, the categorial AX DIS methods also promoted long term effects on the untargeted sounds that had originally shown improvement with training (vowels for the DIS\_C and final consonants for the DIS\_V). This result confirms that the training effect observed with the untargeted sounds was consolidated. According to Flege (1995b), if knowledge acquired during training is retained over time, it might indicate that robust L2 categories have been established in the L2 learners' perceptual space. Moreover, it confirms the relevance of phonetic training as an L2 teaching tool and the importance of using this tool in the L2 classroom.

### **Limitations and final conclusions**

This study had some limitations, such as a possible task familiarity effect for ID training groups, since testing involved only an ID task, and, although the talkers were different, the testing words were a subset of the training words. However, ID trainees outperformed DIS trainees only in the case of vowel-trained groups, as the performance was comparable in the case of consonant-trained groups. Thus, it remains unclear whether a task familiarity effect could have influenced the outcomes of the present study. Another limitation involves the difference in duration of training between consonants and vowels. Recall that in the case of stop consonants the total amount of training time was halved in order to train initial and final stops. According to Flege (1989), learners should be exposed to a large number of trials in order for training to work, especially when the L2 feature is non-existent in the L1. Thus the shorter length of training for consonants might have limited its impact on the results, as shown by the lack of significant improvement with final stops. Future research should aim to assess cross-training effects for consonants and vowels with longer and equal amounts of training time. Furthermore, it may be the case that the improvement on untargeted vowels resulted from learners' attention to the preceding vowel as a cue to final stop voicing and thus was the consequence of final stop training only. Thus it would be more appropriate to train initial and final stops separately to fully evaluate the effect of training on both the targeted and the untargeted sounds. Finally, the current study focused on a particular population (English-major undergraduates with knowledge of phonetics). Although previous research has found training to be effective for different levels of proficiency (Iverson et al. 2012), further research is also necessary to assess if the training regimes used in the current study would be equally beneficial to learners of different levels of L2 English. The current findings about the overall effect of HVPT, and especially the effect on untargeted sounds, can be examined further by investigating generalization to

additional types of L2 sound contrasts, contexts and populations, as well as exploring its practical implications for second/foreign language pronunciation teaching.

In conclusion, the present investigation provided evidence that different training methods (ID and categorical DIS) play different roles when training L2 vowels and L2 consonants. An ID method was found to enhance L2 vowel perception to a greater extent than a categorical DIS method, whereas both methods were equally effective in modifying initial consonant perception, and ID promoted some improvement of final stop identification. In addition, the DIS method was found to enhance the perception of the untargeted but implicitly exposed sounds, showing that controlled exposure to stimuli might be able to enhance listeners' sensitivity even when their focal attention is not oriented solely towards the target sound. It suggests that the learners trained by a DIS method were able to reorient their attention to different cues present in the stimuli which contributed to the perception of both the targeted and the untargeted L2 sounds. Since the majority of training studies make use of CVC words as training stimuli, this may be an important result. This is because whilst training one type of segment, the implicit exposure to the untargeted segments present in the stimuli may contribute to learners' overall perception. In line with these results, a combination of both tasks (ID and categorical DIS) is suggested in order to enhance different perceptual abilities and maximize the effects of training (Cebrian & Carlet, 2014; Shinohara & Iverson, 2018). Perhaps DIS tasks can be used at the early stages of training (Logan & Pruitt, 1995) and ID can be introduced once the learners are more familiar with the different categories of the target sounds (Shinohara & Iverson, 2018). Nevertheless, it is relevant to note that previous studies have shown that L2 learners find an ID task more motivating and enjoyable than a DIS task, described as more demanding and tiring (Carlet, 2017; Flege, 1995b). In that regard, a greater contribution from L2 training studies may be achieved by examining not only the efficacy of the training regime

but also participants' opinions about the training tasks to assess the practical applications of phonetic training.

### Acknowledgements

This research was supported by a grant from the Spanish Ministry of Economy and Competitiveness (FFI2017-88016-P) and by a grant from the Catalan Government (2017SGR34).

---

<sup>i</sup> In the case of English /ɜ:/, it was contrasted with two potentially confusable vowels, /ɑ:/ and /e/, so that within trial contrasts (DIS task) and across trial contrasts (ID task) involved /ɑ:/- /ɜ:/ and /e/-/ɜ:/ pairs. There were 12 pairs illustrating the /æ/-/ʌ/contrast, 12 for the /ɪ/- /i:/contrast, 6 pairs for the /ɑ:/- /ɜ:/ and 6 for the /e/- /ɜ:/ contrast.

<sup>ii</sup> Cohen's *d* is given as a measure of effect size (e.g., Cohen, 1988). Following Plonsky & Oswald (2014), the benchmarks for small, medium and large effect sizes are .40, .70 and 1, respectively, for between groups comparisons, and .60, 1 and 1.40 for within groups comparisons.

<sup>iii</sup> CG was the group with the lowest number of participants who completed the delayed post-test (9 of the original 16). It is possible that the subset of participants who completed all tests were also the best performers. In fact, a comparison between these 9 participants and the remaining 7 at pre- and post-test does show that the former were more successful than the latter. This may also explain the relatively high CG scores at the delayed post-test.

<sup>iv</sup> An anonymous reviewer suggested that another explanation for the different results obtained for stops and vowels may be related to differences between these segments in the extent to which the target phones map onto the closest L1 categories. Since the current study did not specifically select target L2 sounds on the basis of differences in L2 to L1 mapping, this issue is left for future study.

<sup>v</sup> All stimuli presented vowels between obstruents, half before a final voiced obstruent, half before a final voiceless obstruent. Thus it is unlikely that the higher scores in vowel identification at the generalization test are a consequence of the phonetic context. Specifically, the novel stimuli were *dack*, *pag*, *dut*, *jud*, *vert*, *derg*, *fip*, *pid*, *geep* and *keeb*. Recall that all stimuli were similarly accurately identified by native English speakers in a stimuli validation task (see section “*Training Stimuli*”).

---

## References

- Aliaga-García, C., & Mora, J. C. (2009). Assessing the effects of phonetic training on L2 sound perception and production. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent research in second language phonetics/phonology: Perception and production*. (pp. 2-31). Cambridge Scholars Publishing.
- Alves, U. K., & Luchini, P. L. (2017). Effects of perceptual training on the identification and production of word-initial voiceless stops by Argentinean learners of English. *Ilha do Desterro* 70(3), 15-32. <https://10.5007/2175-8026.2017v70n3p15>
- Best, C., & Tyler, M. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 13-34). John Benjamins Publishing Company.
- Bohn, O-S. (2002). On phonetic similarity. In Burmeister, P., Piske, T., & Rohde, A. (Eds.), *An integrated view of language development: Papers in honor of Henning Wode* (pp. 191-216). Wissenschaftlicher Verlag.
- Bradlow, A. R. (2008). Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/ - /l/ contrast. In Hansen Edwards, J. G., & Zampini, M. L. (Eds.), *Phonology and Second Language Acquisition* (pp. 287-308). John Benjamins Publishing Company. <https://doi.org/10.1075/sibil.36.14bra>
- Bradlow, A. R., Akahane-Yamada, R., Pisoni, D. B., & Tohkura, Y. (1999). Training Japanese listeners to identify English /r/and /l/: Long-term retention of learning in perception and production. *Perception & Psychophysics*, 61(5), 977-985. <https://doi.org/10.3758/bf03206911>



- 
- Broadbent, D. E. (1958). Effects of noise on behaviour. In C.M. Harris (Ed.), *Handbook of Noise Control* (pp. 10-34). McGraw-Hi.
- Burnham, D. K. (1986). Developmental loss of speech perception: Exposure to and experience with a first language. *Applied Psycholinguistics*, 7(03), 207. <https://doi.org/10.1017/s0142716400007542>
- Carlet, A. (2017). *L2 perception and production of English consonants and vowels by Catalan speakers: The effects of attention and training task in a cross-training study*. [Unpublished doctoral dissertation]. Universitat Autònoma de Barcelona.
- Carreiras, M., & Price, C. J. (2008). Brain activation for consonants and vowels. *Cerebral Cortex*, 18(7), 1727-1735. <https://doi.org/10.1093/cercor/bhm202>
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics*, 34(3), 372-387. <https://doi.org/10.1016/j.wocn.2005.08.003>
- Cebrian, J. (2019) Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America*, 145(1), EL52-EL58. <https://doi.org/10.1121/1.5087645>.
- Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *Journal of the Acoustical Society of America*, 149(4), 2671–2685. <https://doi.org/10.1121/10.0004257>.
- Cebrian, J., & Carlet, A. (2014). Second-language learners' identification of target-language phonemes: A short-term phonetic training study. *Canadian Modern Language Review*, 70(4), 474-499. <https://doi.org/10.3138/cmlr.2318>

- 
- Cebrian, J., Gorba, C., Gavaldà, N. (2021). When the easy becomes difficult: Factors affecting the acquisition of the English /i:/-/ɪ/ contrast. *Frontiers in Communication*, 6, 1-17. <https://doi.org/10.3389/fcomm.2021.660917>.
- Chen, Y., & Pederson, E. (2021). The role of orienting attention during perceptual training in learning non-native tones and consonants. In R. Wayland (Ed.), *Second language speech learning: theoretical and empirical progress* (pp. 485–502): Cambridge University Press.
- Collet, G., Colin, C., Serniclaes, W., Hoonhorst, I., Markessis, E., Deltenre, P., & Leybaert, J. (2013). Changes in voicing perception by adult French speakers after identification training. *Applied Psycholinguistics*, 36(02), 463-483. <https://doi.org/10.1017/s0142716413000313>
- Cooke, M., García-Lecumberri, M.L., Maidment, J., & Ericsson, A. (2005). The web transcription tool. Retrieved February 18, 2017, from <http://www.wtt.org.uk/>.
- Darcy, I., Ewert, D., & Lidster, R. (2012). Bringing pronunciation instruction back into the classroom: An ESL teachers' pronunciation "toolbox". In J. Levis & K. Lavelle (Eds.), *Proceedings of the 3rd Pronunciation in Second Language Learning and Teaching Conference, Sept. 2011* (pp. 93-108). Iowa State University.
- Flege, J. E. (1989). Chinese subjects' perception of the word-final English /t/-/d/ contrast: Performance before and after training. *The Journal of the Acoustical Society of America*, 86(5), 1684-1697.
- Flege, J. (1995a). 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, J. (1995b). Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics*, 16, 425-442.

<https://doi.org/10.1017/S0142716400066029>

Flege, J. E., & Bohn, O. S. (2021). The Revised Speech Learning Model (SLM-r). In R. Wayland (Ed.), *Second language speech learning: theoretical and empirical progress* (pp. 3–83): Cambridge University Press.

Fouz-González, J., & A Mompean, J.A. (2020). Exploring the potential of phonetic symbols and keywords as labels for perceptual training. *Studies in Second Language Acquisition*, 1-32. <https://doi.org/10.1017/S0272263120000455>

Fry, D. B., Abramson, A. S., Eimas, P. D., & Liberman, A. M. (1962). The identification and discrimination of synthetic vowels. *Language and Speech*, 5(4), 171-189. <https://doi.org/10.1177/002383096200500401>

Fullana, N., & Mora, J. C. (2009) Production and perception of voicing contrasts in English word-final obstruents: Assessing the effects of experience and starting age. In M. A. Watkins, A. S. Rauber, & B. O. Baptista (Eds.), *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. (pp. 97-117). Cambridge Scholars Publishing.

Guion, S. G., & Pederson, E. (2007). Investigating the role of attention in phonetic learning. *Language Experience in Second Language Speech Learning Language Learning & Language Teaching*, 57-77. <https://doi.org/10.1075/llt.17.09gui>

IBM Corp. (Released 2015). IBM SPSS Statistics for Windows, Version 23.0. IBM Corp.

- 
- Iverson, P., & Evans, B. G. (2009). Learning English vowels with different first-language vowel systems II: Auditory training for native Spanish and German speakers. *The Journal of the Acoustical Society of America*, 126(2), 866-877. <https://doi.org/10.1121/1.3148196>
- Iverson, P., Pinet, M., & Evans, B. G. (2012). Auditory training for experienced and inexperienced second-language learners: Native French speakers learning English vowels. *Applied Psycholinguistics*, 33(01), 145-160. <https://doi.org/10.1017/s0142716411000300>
- Jamieson, D. G., & Morosan, D. E. (1986). Training non-native speech contrasts in adults: Acquisition of the English /ð/-/θ/ contrast by francophones. *Perception & Psychophysics*, 40(4), 205-215. <https://doi.org/10.3758/bf03211500>
- Kivistö-de Souza, H., & Carlet, A. (2014). Vowel inventory size and the use of temporal cues in non-native vowel perception by Catalan and Danish EFL learners. In *Proceedings of the International Symposium on the Acquisition of Second Language Speech* (Vol. 5, pp. 322-336). Concordia Working Papers in Applied Linguistics (COPAL).
- Kuhl, P., & Iverson, P. (1995). Linguistic experience and the “Perceptual Magnet Effect”. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 121-154). York Press.
- Law, I. L. G., Grenon, I., Sheppard, C., & Archibald, J. (2019). Which is better: Identification or discrimination training for the acquisition of an English coda contrast. In Calhoun, S., Escudero, P., Tabain, M., & Warren, P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS), Melbourne*,

---

*Australia 2019* (pp. 939-943). Australasian Speech Science and Technology Association Inc. 2019.

Lee, H. Y., & Hwang, H. (2016). Gradient of learnability in teaching English pronunciation to Korean learners. *The Journal of the Acoustical Society of America*, 139(4), 1859-1872.

Lerdpaisalwong, S. (2015). *Perception training of Thai learners: American English consonants and vowels*. [Unpublished doctoral dissertation]. The University Of Wisconsin-Milwaukee.

Liberman, A. M., Cooper, F. S., Shankweiler, D. P., & Studdert-Kennedy, M. (1967). Perception of the speech code. *Psychological Review*, 74(6), 431-461.

Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *Speech Perception and Linguistic Experience: Issues in Cross Language Research* (pp. 351-378). Timonium, MD:York Press.

Logan, J. S., Lively, S. E., & Pisoni, D. 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. <https://doi.org/10.1121/1.1894649>

Long, M. (1991). Focus on Form. *Studies in Bilingualism Foreign Language Research in Cross-Cultural Perspective*, 39-52. <https://doi.org/10.1075/sibil.2.07lon>

Mora, J. C., & Mora-Plaza, I. (2019) Contributions of Cognitive Attention Control to L2 Speech Learning. In Nyvad, A. M., Hejná, M., Højen, A., Bothe Jespersen, A., & Hjortshøj Sørensen, M. (Eds.): *A Sound Approach to Language Matters – In Honor of Ocke-Schwen Bohn*. Dept. of English, School of Communication & Culture, Aarhus University, pp. 477-499.

- 
- Muñoz, C. (2008). Symmetries and asymmetries of age effects in naturalistic and instructed L2 learning. *Applied Linguistics*, 29(4), 578-596. <https://doi.org/10.1093/applin/amm056>
- Nishi, K., & Kewley-Port, D. (2007). Training Japanese listeners to perceive American English vowels: Influence of training sets. *Journal of Speech Language and Hearing Research*, 50(6), 1496-1509. [https://doi.org/10.1044/1092-4388\(2007/103\)](https://doi.org/10.1044/1092-4388(2007/103))
- Nozawa, T. (2015). Effects of training methods and attention on the identification and discrimination of American English coda nasals by native Japanese listeners. In the Scottish Consortium for ICPHS 2015 (Ed.), *Proceedings of the 18th International Congress of Phonetic Sciences*. The University of Glasgow. ISBN 978-0-85261-941-4
- Ortega, M., Mora-Plaza, I., & Mora, J. C. (2021). Differential effects of lexical and non-lexical high-variability phonetic training on the production of L2 vowels. In Kirkova-Naskova, A., Henderson, A., & Fouz-González, J. (Eds.), *English pronunciation instruction: Research-based insights*. John Benjamins.
- Pederson, E., & Guion-Anderson, S. (2010). Orienting attention during phonetic training facilitates learning. *The Journal of the Acoustical Society of America*, 127(2), EL54-EL59. <https://doi.org/10.1121/1.3292286>
- Pereira, Y. I. (2014). *Perception and production of English vowels by Chilean learners of English: Effect of auditory and visual modalities on phonetic training* [Unpublished doctoral dissertation]. University College London.

- 
- Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & Psychophysics*, 13(2), 253-260.  
<https://doi.org/10.3758/bf03214136>
- Polka, L. (1992). Characterizing the influence of native language experience on adult speech perception. *Perception & Psychophysics*, 52(1), 37-52.  
<https://doi.org/10.3758/bf03206758>
- Posner, M. I. (1980). Orienting of attention. *Quarterly journal of experimental psychology*, 32(1), 3-25. <https://doi.org/10.1093/acprof:oso/9780199791217.003.0003>
- Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *The Journal of the Acoustical Society of America*, 51(4B), 1296-1303.  
<https://doi.org/10.1121/1.1912974>
- Rauber, A., Rato, A., Kluge, D., & Santos, G. (2012). TP (Version 3.1).[Software]. Brazil: Worken. [[http://www.worken.com.br/tp\\_regfree.php?l=i](http://www.worken.com.br/tp_regfree.php?l=i)].
- Recasens, D., & Mira, M. (2015). Place and manner assimilation in Catalan consonant clusters. *Journal of the International Phonetic Association*, 45(02), 115-147.  
<https://doi.org/10.1017/s0025100315000080>
- Roach, P. (2000). *Phonetics and phonology: A Practical course*. Cambridge University Press.
- Saito, K. (2015). Variables affecting the effects of recasts on L2 pronunciation development. *Language Teaching Research*, 19(3), 276-300.  
<https://doi.org/10.1177/1362168814541753>

- 
- Sakai, M., & Moorman, C. (2017). Can perception training improve the production of second language phonemes? A meta-analytic review of twenty-five years of perception training research. *Applied Psycholinguistics*, 1-36.
- Schmidt, R. (2001). "Attention". In P. Robinson (Ed.), *Cognition and second language instruction* (pp. 3-32). Cambridge University Press.
- Shinohara, Y. & Iverson, P. (2018). High variability identification and discrimination training for Japanese speakers learning English /r-/l/. *Journal of Phonetics* 66, 242-251.
- Smith, M. S. (1993). Input enhancement in instructed SLA. *Studies in Second Language Acquisition*, 15(02), 165-179. <https://doi.org/10.1017/s0272263100011943>
- Steinlen, A. K. (2005). *The influence of consonants on native and non-native vowel production: A cross-linguistic study*. Gunter Narr Verlag.
- Strange, W. (2007). Cross-language phonetic similarity of vowels: Theoretical and methodological issues. In O.-S. Bohn, & M. Munro (Eds.), *Language Experience in Second Language Speech Learning – In honor of James Emil Flege* (pp. 35-55). John Benjamins Publishing Company.
- Strange, W., & Dittmann, S. (1984). Effects of discrimination training on the perception of /r-/l/ by Japanese adults learning English. *Perception & Psychophysics*, 36(2), 131-145. <https://doi.org/10.3758/bf03202673>
- Thomson, R. 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.



---

Thomson, R. I., & Derwing, T. M. (2016). Is phonemic training using nonsense or real words more effective? In J. Levis, H. Le., I. Lucic, E. Simpson, & S. Vo (Eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference* (pp. 88-97). Ames, IA: Iowa State University.

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

Wayland, R. P., & Li, B. (2008). Effects of two training procedures in cross-language perception of tones. *Journal of Phonetics*, 36(2), 250-267. <https://doi.org/10.1016/j.wocn.2007.06.004>

Wee, D. T., Grenon, I., Sheppard, C., & Archibald, J. (2019). Identification and discrimination training yield comparable results for contrasting vowels. In Calhoun, S., Escudero, P., Tabain, M., & Warren, P. (Eds.), *Proceedings of the 19th International Congress of Phonetic Sciences (ICPhS), Melbourne, Australia 2019* (pp. 939-943). Australasian Speech Science and Technology Association Inc. 2019.

---

**Table 1.**

*Training stimuli organized by vowel contrast (columns) and consonant contrast (rows)*

	<b>æ - ʌ</b>	<b>i - i:</b>	<b>ɜ: - ʌ: / ɜ: - e</b>	
<b>t/d_#</b>	dadge-tadge dudge-tudge	deedge-teedge didge-tidge	darge-targe derge-terge*	dedge-tedge derge-terge*
<b>p/b_#</b>	pav - bav puv - buv	peedge-beedge pidge-bidge	parsh-barsh persh-bersh	perff-berf peff-beff
<b>k/g_#</b>	kak-gak kuk-guk	keedge-geedge kidge-gidge	karch-garch kerch-gerch*	ketch-getch kerch-gerch*
<b>#_t/d</b>	zat -zad zut - zud	jeet-jeed jit-jid	zart-zard zert-zerd	chert-cherd chet-ched
<b>#_p/b</b>	vap -vab vup - vub	veep-veeb vip-vib	jarp- jarb jerp-jerb*	jep-jeb jerp-jerb*
<b>#_k/g</b>	vak -vag vuk -vug	veek-veeg vik-vig	vark-varg verk-verg*	vek-veg verk-verg*

*Note.* \_# - initial; #\_ - final. Rows show consonant contrasts (dadge-tadge) and columns illustrate vowel contrasts (dadge-dudge). \* Items that appear twice on the list. Some /ɜ:/ items appear twice as /ɜ:/ was contrasted with /ʌ:/ in half the trials and with /e/ in the other half.

**Table 2.**

*Pre-test, post-test, gain scores and generalization to new stimuli for targeted and untargeted sounds obtained by the vowel groups and CG*

	CG		DIS_V		ID_V	
Vowels	%	SD	%	SD	%	SD
Pre-test	54.1	9.9	55.5	6.5	52.9	9.5
Post-test	57.8	10.2	65.3	9.7	79.1	13.3
Gain	3.7	4.2	9.8	7.6	26.3	7.6
Generalization new words	68.4	12.4	75.9	8.3	80.4	9.8
Initial consonants	%	SD	%	SD	%	SD
Pre-test	78.1	12.4	74.3	9.7	77.6	11.9
Post-test	79	13.4	76.7	11.4	74.2	11.9
Gain	0.9	5.6	2.4	10.3	-3.4	9.4
Generalization new words	79.9	8.7	78.8	7.4	75.8	11.0
Final consonants	%	SD	%	SD	%	SD
Pre-test	69.1	7.8	64.7	14.5	71	6.4
Post-test	69.6	6.29	72.4	6.9	69.4	10
Gain	0.5	4.5	7.7	13.2	-1.6	8.2
Generalization new words	64.9	11.4	60.8	9.2	64.2	8.2

**Table 3.**

*Pre-test, post-test, gain scores and generalization to new stimuli for targeted and untargeted sounds obtained by consonant trained groups and CG*

	CG		DIS_C		ID_C	
Initial consonants	%	SD	%	SD	%	SD
Pre-test	78.1	12.4	69.5	9.6	72.8	9.5
Post-test	79	13.4	82.3	9.8	88.7	7.1
Gain	0.9	5.6	12.8	11	15.9	10.6
Generalization new words	79.9	11.5	81.5	8.7	85.3	7.8
Final consonants	%	SD	%	SD	%	SD
Pre-test	69.1	7.8	70	16.8	69.9	8.8
Post-test	69.6	6.2	74.3	7.3	75.4	7.0
Gain	0.5	4.5	4.3	13.9	5.5	8.9
Generalization new words	65.0	7.9	70.2	6.8	72.9	6.5
Vowels	%	SD	%	SD	%	SD
Pre-test	54.1	9.9	55	7.5	56.2	10.1
Post-test	57.8	10.2	64.7	8.4	55.3	12.2
Gain	3.7	4.2	9.7	7.4	-0.9	7.9
Generalization new words	68.4	11.3	75.8	8.2	62.9	8.9

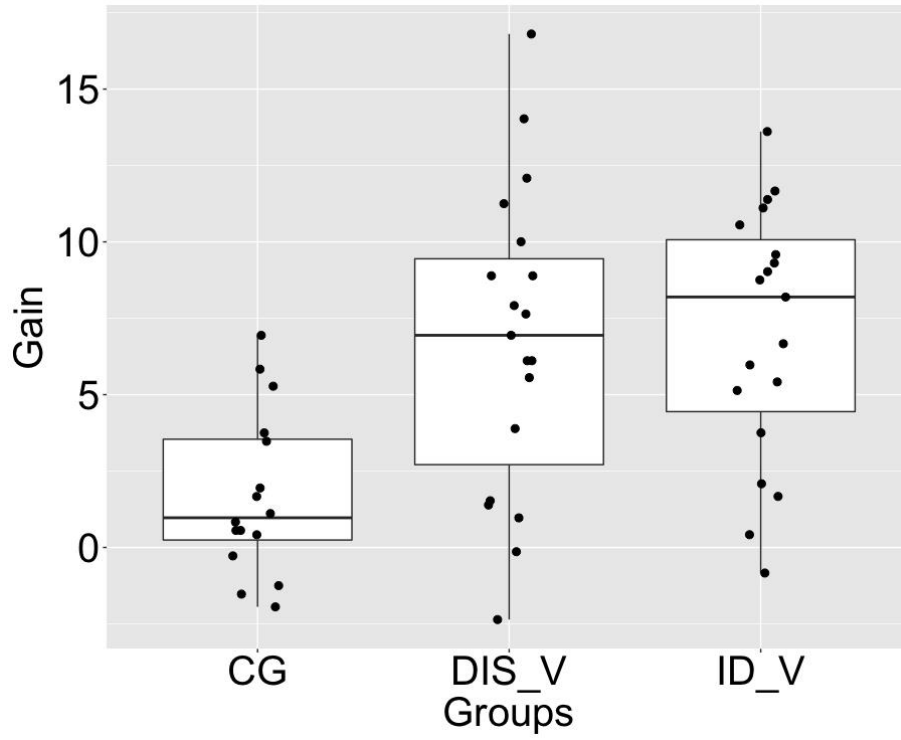
**Table 4.**

*Pre-test, post-test and delayed post-test scores for targeted and untargeted sounds (data from participants who completed all tests).*

	CONTROL		DIS_V		ID_V		DIS_C		ID_C	
Vowels	%	SD	%	SD	%	SD	%	SD	%	SD
PRE	56.7	11.3	53	4.2	51.8	9.7	53.8	7.6	57.4	10.5
POST	61.9	11.1	62.8	9.4	79.7	9.3	64	9.3	57.3	10.6
DELAYED POST	70.0	14	60.4	8.2	80.1	8.3	65.2	17.5	60	8.3
Initial consonants	%	SD	%	SD	%	SD	%	SD	%	SD
PRE	79.2	15	72.9	9.5	80	11.1	71.8	9.4	71	10.9
POST	82.6	14.7	74.3	11.7	75.4	12.2	84.6	9.3	88.4	6.8
DELAYED POST	80.2	13.9	76.7	10.5	78.3	12.4	86.4	9.7	84.3	14.4
Final consonants	%	SD	%	SD	%	SD	%	SD	%	SD
PRE	72.9	5.9	63.7	17.9	71.7	6.3	73.2	9	71.2	8.6
POST	71.5	6.4	71.7	6.6	70.6	9	75.6	6.3	75.2	6
DELAYED POST	74.1	7.6	70.8	6.9	74.1	6.5	76.3	7.4	75.2	7.5

Figure 1.

*Overall gain obtained by vowel trained groups and control group (CG) across segments*



**Figure 2.**

*Gain score for targeted and untargeted sounds obtained by vowel trained groups and CG*

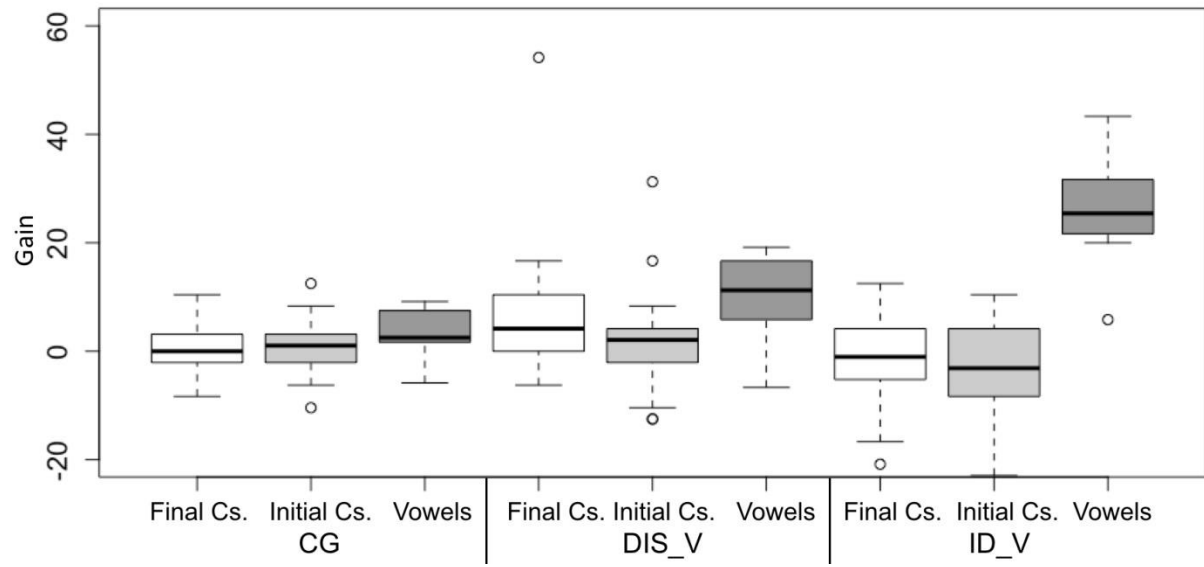


Figure 3.

*Overall gain obtained by consonant trained groups and control group (CG)*

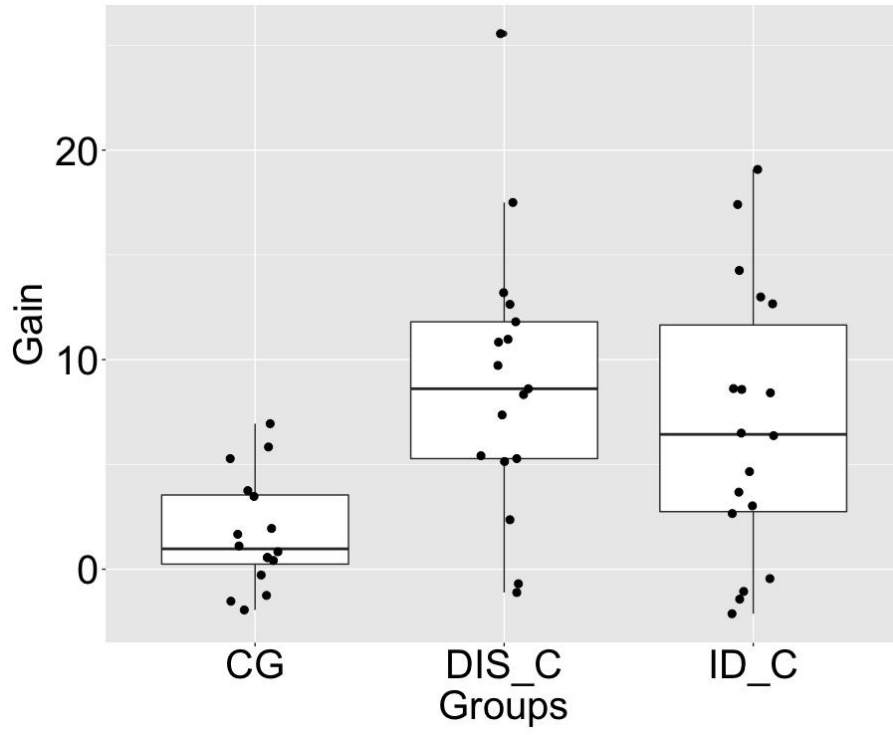




Figure 4.

Group by segment gain on targeted and untargeted sounds obtained by consonant training groups and CG

