Research Article

Joan C. Mora*, Mireia Ortega, Ingrid Mora-Plaza and Cristina Aliaga-García

Training the pronunciation of L2 vowels under different conditions: the use of non-lexical materials and masking noise

https://doi.org/10.1515/phon-2022-2018
Published online April 15, 2022

Abstract: The current study extends traditional perceptual high-variability phonetic training (HVPT) in a foreign language learning context by implementing a comprehensive training paradigm that combines perception (discrimination and identification) and production (immediate repetition) training tasks and by exploring two potentially enhancing training conditions: the use of non-lexical training stimuli and the presence of masking noise during production training. We assessed training effects on L1-Spanish/Catalan bilingual EFL learners' production of a difficult English vowel contrast (/æ/-/ʌ/). The participants (N = 62) were randomly assigned to either non-lexical (N = 24) or lexical (N = 24) training and were further subdivided into two groups, one trained in noise (N = 12) and one in silence (N = 12). An untrained control group (N = 14) was also tested. Training gains, measured through spectral distance scores (Euclidean distances) with respect to native speakers' productions of /æ/ and /ʌ/, were assessed through delayed word and sentence repetition tasks. The results showed an advantage of non-lexical training over lexical training, detrimental effects of noise for participants trained with nonwords, but not for those trained with words, and less accurate production of vowels elicited in isolated words than in words embedded in sentences, where training gains were only observable for participants trained with nonwords.

***Corresponding author: Joan C. Mora**, Department of Modern Languages and Literatures and English Studies, Faculty of Philology and Communication, Universitat de Barcelona, Gran Via de les Corts Catalanes 585, 08007 Barcelona, Spain, E-mail: mora@ub.edu
Mireia Ortega, Ingrid Mora-Plaza and Cristina Aliaga-García, Department of Modern Languages and Literatures and English Studies, Faculty of Philology and Communication, Universitat de Barcelona, Barcelona, Spain, E-mail: m.ortega@ub.edu (M. Ortega), imoraplaza@ub.edu (I. Mora-Plaza), cristinaaliaga@ub.edu (C. Aliaga-García)

Keywords: high-variability phonetic training (HVPT); L2 vowel production; masking noise; nonword training

1 Introduction

When acquiring a second language (L2), learners are faced with the challenging task of learning the foreign language (FL) phonology, which entails the acquisition of the L2 segmental inventory, as well as the constraints that govern the distribution of L2 phonemes and the phonological processes of the FL. This is an arduous task for L2 learners, since their perceptual systems have become attuned to their L1 sound inventory (Kuhl et al. 2008), which can bias categorical perception, speech segmentation and lexical activation and retrieval processes in the L2 (Ramus et al. 2010).

According to L2 speech learning models such as PAM-L2 (Best and Tyler 2007), SLM (Flege 1995), or SLM-r (Flege and Bohn 2021), L1-based perception causes difficulties in phonetic learning, especially when phonetically similar L2 sounds are perceptually mapped onto single L1 sound categories. For example, the English phonemes /æ/-/ʌ/-/ɑː/ (*cat, cut, cart*) perceptually assimilate to the Spanish single low vowel category /a/, making it difficult for Spanish learners of English to develop distinctive L2 sound categories for these English vowels. According to these models, the likelihood of developing segmental L2 categories for L2 sounds depends, among other contextual and experience-related factors, on the degree of cross-language perceptual similarity between L2 sounds and L1 categories, such that more perceptually dissimilar L2 sounds are easier to acquire than those that are perceptually closer to the L1 sound. This has consequences for processing at both L2 phonetic and lexical levels. At the phonetic processing level, perceiving the qualitative distinction in the L2 vowels /æ/-/ʌ/ (e.g. *cat – cut*) is difficult because the Spanish learners' perceptual system maps both vowels to their native single low vowel category /a/ (Rallo Fabra and Romero 2012). In addition, given their difficulty in the perception of a quality distinction between /æ/, /ʌ/ and /ɑː/, Spanish learners might rely on the wrong perceptual cue (e.g. duration) to perceive and produce the distinction, just as they may not attend to VOT duration but would attend to the presence of closure voicing (as they would in Spanish) to distinguish voiced from voiceless stops word-initially.

At the lexical processing level, numerous priming studies (e.g. Broersma 2012) and lexical competition eye-tracking studies that used the visual world paradigm (e.g. Cutler et al. 2006; Escudero et al. 2008; Weber and Cutler 2004) have shown that L2 learners activate phonemically distinct L2 words (e.g. *cap* and *cup*) when exposed to a L2 contrast (/æ/-/ʌ/) that assimilates perceptually to one L1 sound. Difficulty in distinguishing perceptually similar L2 sounds would increase the

activation of unintended lexical items, generating "phantom" lexical competition (Broersma and Cutler 2008), making word recognition difficult and potentially leading to confusion of L2 minimal-pair words when processing L2 input. Accurate categorization of L2 contrasting sounds (e.g. /æ/-/ʌ/), however, does not guarantee accurate discrimination at the lexical level. Words containing them (e.g. *cap* and *cup*) may still be perceived to be homophonous if the target contrast has not been accurately encoded in the lexicon (Darcy et al. 2012). The implications of this for phonetic training is that training L2 learners to perceive /æ/-/ʌ/ through confusable lexical minimal-pair words that would become co-activated during training (*cap-cup*) may be less effective in making them perceive the contrast at the phonetic level than using nonword items for which they have not developed lexical representations.

The task of learning how to accurately perceive and produce sounds in the L2 is even more challenging in instructed second language acquisition (SLA) contexts (e.g. Cooke and García-Lecumberri 2018; Gomez Lacabex et al. 2008). In this setting, the opportunities for acquiring L2 sounds are more limited than in a L2 context (Tyler 2019), as L2 input is frequently restricted to a few hours of weekly instruction of teacher-centred and grammar-oriented lessons, which might be partially delivered in the learners' native language (Muñoz 2014). Some learners, however, may be exposed to out-of-classroom activities in the L2. Therefore, differences in quality of exposure range from being exposed to a single-input source (a non-native teacher) to a wider range of native speech (García-Lecumberri and García-Mayo 2003). When the teaching approach is communicative and meaning-oriented, even when difficulties at integrating pronunciation instruction in the communicative class are overcome (Sicola and Darcy 2015), learners are likely to be exposed to foreign-accented speech from their teachers and peers, who may not maintain the phonological distinctions of the L2. As a result, irrespective of amount of formal instruction received or age of onset of L2 learning (Gallardo del Puerto et al. 2006), instructed learners develop large vocabularies (mostly through written input) before being explicitly taught about L2 sound pronunciation and phonetic differences between contrasting L2 sounds. This would not only increase the number of inaccurate word form representations stored in the mental lexicon, but it might also practically hinder the discrimination of L2 sounds in perception and production (Tyler 2019). The extent to which phonetic training may be effective at correcting already established phonological word forms in instructed foreign language learning contexts remains an empirical question.

High-variability phonetic training (HVPT), which typically trains L2 learners to perceive L2 sounds in a variety of phonetic environments produced by multiple speakers, is an effective method for improving the perception and production of L2 sounds (Barriuso and Hayes-Harb 2018; Thomson 2018), leading to training gains that generalize learning to new voices and words (Aliaga-García and Mora 2009;

Carlet and Cebrian 2019; Cebrian and Carlet 2014) as well as non-target sounds (Carlet 2017; Rato and Rauber 2015). However, generalization of accuracy gains obtained for L2 sounds in words to sounds in words elicited in sentences has rarely been attested for perception (Gomez Lacabex et al. 2008; Hirata 2004), and to the best of our knowledge never for production. For example, Thomson and Derwing (2016) elicited the production of words in sentences by asking participants to create a sentence using a noun (containing the target vowel) elicited through a picture shown on the screen. Perceptual judgements revealed no training gains on vowel accuracy for words elicited in such contexts, whereas vowel intelligibility improved for words elicited in carrier phrases (i.e. They heard "*The next word is* ___," and they responded by repeating the word they had just heard in the carrier phrase, "*Now I say* ___."). Vowel productions elicited in meaning-focused sentences learners retrieve from memory are more likely to reflect vowel quality as produced in extemporaneous speech than vowels produced through a task involving the repetition of isolated words after a native speaker model, which would foster conscious attention to the phonetic form of individual items (Bradlow et al. 1999). We are unaware of any research showing generalization of phonetic training gains to words produced in extemporaneous speech or providing evidence of which training procedures might best promote such generalization effects. This type of generalization effects would indicate that the benefits of HVPT would extend to the phonological representations of words in the L2 mental lexicon, effectively leading to L2 pronunciation improvement (Darcy and Holliday 2019).

The effectiveness of HVPT in improving L2 vowel production accuracy may be enhanced by using nonwords instead of words, demonstrating that focusing on distinctive phonetic features becomes much more feasible by eliminating interference from meaning (Thomson and Derwing 2016). In addition, training the production of the target vowels in noise may also enhance HVPT outcomes in production through the elicitation of clear speech, which might enhance the full realization of articulatory targets (Hazan and Baker 2011) and lead to vowel productions of increased duration and intensity (Leung et al. 2016; Smiljanić and Bradlow 2011). The present study therefore investigates the potential benefits of using nonwords (vs. words) as training stimuli, and the use of noise (vs. silence) during production training for L2 learners' production of English /æ/ and /ʌ/ in isolated words and in sentences.

2 Acquisition of L2 sound contrasts

In order to produce L2 sounds accurately (in isolation and/or in words), the phonetic differences that serve to distinguish L2 sounds must be learned. The degree of

difficulty in L2 speech sound production has been attributed to a perceptual origin. For example, when a pair of L2 vowels fall within the perceptual space of a single L1 vowel category, learners fail to perceive the phonetic differences that distinguish them. This prevents learners from forming distinct perceptual categories for the L2 vowels, which consequently prevents them from distinguishing between them in production. This is the case of the English vowel contrast /ɛ/-/æ/ (*pen* vs. *pan*) for Dutch learners of English, who perceive these vowels as instances of their L1 vowel category /ɛ/ (Escudero et al. 2008). Similarly, Spanish learners of English have difficulties in the perception and production of the English contrast /æ/-/ʌ/ (*cap* vs. *cup*), which they perceptually map onto their low L1 vowel category /a/ (Rallo Fabra and Romero 2012).

Acquiring L2 sound contrasts is linked to vocabulary development from the very early stages of the L2 acquisition process. The acquisition of L2 words involves storing their morphological, syntactic, semantic, phonological and orthographic properties in the mental lexicon (Hayes-Harb and Masuda 2008). The phonological form in which words are stored in the mental lexicon (i.e. their phono-lexical representations) is the basis of the psycholinguistic mechanisms of lexical activation and retrieval involved in spoken word recognition processes (Broersma 2012), but it is also crucial for further phonological development. Storing English words like *cap* and *cup* with a single non-native phonological form (e.g. /kap/ with Spanish /a/) may hinder L2 learners' ability to perceive and produce the vowel quality difference between other words involving the same contrast (e.g. *pan* vs. *pun*; Darcy and Holliday 2019; Darcy and Thomas 2019; John and Cardoso 2017). At the same time, empirical evidence has accumulated suggesting that L2 learners may accurately distinguish L2 sounds phonetically while failing to encode them accurately in the lexicon (e.g. Amengual 2016; Darcy et al. 2012, 2013; Hayes-Harb and Masuda 2008; Llompart and Reinisch 2018; Simonchyk and Darcy 2017).

The strength of the link between phonetic learning and vocabulary development may differ as a function of learning context. For example, vocabulary size and phonological acquisition have been found to be positively related in immersion settings (Bundgaard-Nielsen et al. 2011, 2012), where learners' vocabularies and pronunciation develop under conditions of rich L2 exposure and use. In contrast, this relationship does not necessarily hold in foreign language learning environments, where conditions include very limited L2 exposure and use, as well as the pervasive influence of accented speech (Mairano and Santiago 2020). For example, vocabulary size has been shown to predict the lexical encoding of a difficult L2 phonological contrast in advanced learners, but not in intermediate learners for whom phonological categorization skills do so (Llompart 2021). Vocabulary growth over time may eventually lead to L2 learners' inability to accurately encode phonological contrasts in the mental lexicon, which may affect

L2 pronunciation development (Tyler 2019). For example, L1 French learners of English might develop identical phonological representations for the words *three* /θri:/ and *tree* /tri:/, both having the form /θri:/ (Trofimovich and John 2011), which will make the perception of the acoustic difference between *three* and *tree* difficult for them. Likewise, L1-Spanish learners' perceptual difficulties with the English vowel contrast /æ/-/ʌ/ (Cebrian et al. 2011; Rallo Fabra and Romero 2012) may have led them to encode it inaccurately and to develop homophonous L1-like phonological representations for pairs of words such as *cap* */kap/ and *cup* */kap/ that are minimally contrastive in English. They may have also developed "fuzzy" phonological representations for other words (e.g. *sun* as /sæn/ rather than /sʌn/) containing these vowels in the mental lexicon despite having achieved the ability to distinguish them phonetically in perception (Darcy et al. 2013). Evidence for the effectiveness of phonetic training at improving the encoding of difficult L2 sound contrasts and at updating the phonological representations containing them is scarce (but see Melnik-Leroy and Peperkamp 2021). Still, both pronunciation instruction (Lee et al. 2015; Saito and Plonsky 2019; Thomson and Derwing 2014) and phonetic training (Sakai and Moorman 2018) have demonstrated to be effective at improving the perception and production of L2 sounds, and there is evidence that learners can update inaccurate phonological representations as their pronunciation develops (Darcy and Holliday 2019).

3 Phonetic training techniques

A vast number of studies have shown that HVPT can increase learners' sensitivity to difficult L2 phonological contrasts, facilitating in this way L2 phonological acquisition (e.g. high-variability identification training: Iverson et al. 2012) and enhancing efficient phonological processing in the recognition and production of L2 vowels (e.g. high-variability discrimination training: Bradlow 2008). In other words, auditory HVPT helps learners improve the perception and production of difficult L2 phonological contrasts among learners with different L1s (Iverson and Evans 2009) and among learners with different L2 proficiency levels (Wong 2015). However, novice learners might only benefit from variability in training if they have high perceptual abilities, whereas low perceptual ability learners might not benefit from perceptual HVPT (Antoniou et al. 2015; Perrachione et al. 2011; Sadakata and McQueen 2013). Phonetic variability has been shown to lead to more stable realizations of L2 sounds, as shown by an increased compactness of vowel categories after production training (Kartushina and Martin 2018), and to lead to long-term perception training effects (Bradlow et al. 1999). HVPT gains have also generalized to new lexical items and speakers as well as across perception

(Hazan et al. 2005; Iverson et al. 2012; Lengeris 2008; Thomson 2011; Thomson and Derwing 2014) and production (Kartushina et al. 2015) training modalities.

HVPT studies have also concluded that gains obtained through perception-based training can be further enhanced under certain training conditions and using different types of training stimuli. For instance, Iverson et al. (2005) showed the effectiveness of input manipulation techniques to direct learners' attention to certain acoustic features, such as directing Japanese learners' attention to F3 frequencies in English (through All Enhancement, Perceptual Fading and Secondary Cue Variability techniques) when training them to distinguish English /r/ and /l/. Training gains can also be promoted by instructing learners to pay attention to specific aspects of the speech input, such as asking them to pay attention to consonants rather than vowels (Guion and Pederson 2007). Other studies have found audiovisual training to be superior to auditory-only training (Hardison 2018; Hazan et al. 2005, 2006). More recently, the use of nonword training stimuli (Thomson and Derwing 2016) and the use of noise during perceptual training (e.g. Cooke and García-Lecumberri 2018) have been empirically investigated as methods to enhance training benefits.

3.1 Training with words versus training with nonwords

One method to increase the benefits of HVPT is by modifying the properties of the training stimuli. For example, initial sensitivity to pitch patterns in a non-lexical context have been found to predict learning outcomes in the acquisition of tonal lexical contrasts (Wong and Perrachione 2007), and including non-lexical training, compared to including lexical pitch-pattern training alone, has been found to lead to enhanced sensitivity to pitch patterns and tone learning (Ingvalson et al. 2013). However, whether these findings hold for training L2 segmental contrasts remains an under-researched question, especially for learners who have failed to encode a target segmental contrast at the lexical level and do not distinguish *cap* (/kæp/) from *cup* (/kʌp/) in production, reflecting the use of L1-based representations for L2 words (e.g. *cap* and *cup* represented both as Spanish-like /kap/).

Training sensitivity to difficult segmental contrasts through non-lexical stimuli may lead to greater gains in the perception and production of segmental contrasts than using lexical stimuli. Thomson and Derwing (2016), for example, compared the use of nonword stimuli (i.e. phonetically-oriented training) with the use of word stimuli (i.e. lexically-oriented training) by perceptually training English learners with mixed L1 backgrounds to identify 10 English monophthong vowels. In this study, while some learners were trained on the vowels presented in

isolated open syllables, others were trained on the same vowels presented in real words. Both groups were tested on the same 70 words used in the training elicited in a carrier phrase through a repetition task. A subset of 20 of these words were also elicited through a picture-based sentence production task. Accuracy of production was assessed by two phonetically trained judges, who classified the target words as good or poor exemplars. Data analyses revealed that the group trained on nonword syllables (but not the lexically trained group) improved significantly in the pronunciation of English vowels. Interestingly, none of the groups improved significantly in the production of vowels elicited through the picture-based sentence production task, which might indicate that production gains for the group trained in nonwords did not generalize to trained words elicited in an extemporaneous speech production task. The authors interpreted the advantage of nonword training over lexical training as being due to the former promoting a focus on phonetic-level information induced by the nature of the training stimuli and the latter encouraging a focus on meaning. In addition, we suggest that the use of lexical materials might activate L2 lexical representations that could be misrepresented phonologically (L1-accented), which might interfere with the perception of L2-specific phonetic features characterizing the target contrast.

3.2 Training in noise versus training in silence

Natural exposure to speech often involves listening in adverse conditions (e.g. noise), which may have a negative effect on speech intelligibility, but the use of noise during phonetic training may enhance production accuracy by promoting the full realization of articulatory targets in the production of sounds to counteract the degraded perception of the signal (Cooke and Lu 2010; Hazan and Baker 2011). Training in noise has been shown to enhance the formation of robust sound categories (Cooke and García-Lecumberri 2018) by forcing adjustments in the speech perception and production system leading to L2 speech learning (Mattys et al. 2012). Previous studies on perceptual HPVT in adaptive adverse conditions (HVPT-AAC) have shown that this training regime is highly effective at improving speech perception (Burk et al. 2006) and that it is particularly useful at helping learners acquire native-like selective attention strategies in non-native perception (Leong et al. 2018). HVPT-AAC has then been reported to make it more difficult for learners to identify the spoken words so that more attention had to be used to detect the relevant acoustic cues in speech perception. However, the effectiveness of HVPT-AAC in improving speech production is still under-researched.

In addition to enhancing perceptual learning, the presence of noise modifies normal speech production, leading to an increase in learners' vocal effort (Lu and

Cooke 2008), which in turn might result in intelligibility gains (Pittman and Wiley 2001). Speakers hyper-articulate speech in adverse listening conditions in order to enhance intelligibility, a phenomenon known as 'clear speech' (Leung et al. 2016; Smiljanić and Bradlow 2011). In noisy conditions, speakers modify the acoustic-phonetic properties of vowels by increasing duration and intensity and changing first and second formant frequencies (Leung et al. 2016). Studies have also shown that listeners find clear speech more intelligible than regular, conversational speech (Bradlow and Bent 2002).

4 The current study

In light of the existing studies on phonetic training, in the present study we explore the extent to which HVPT is effective at improving the pronunciation of the target L2-English vowels /æ/ and /ʌ/ in untrained words elicited in two contexts: (a) a delayed word repetition task that allowed learners to focus their attention on the individual target words when repeating them after a native speaker; and (b) a delayed sentence repetition task that had to be processed and produced from memory. Testing learners' ability to identify /æ/ as distinct from /ʌ/ in a lexical context (i.e. testing the lexical encoding of the /æ/-/ʌ/ contrast) is beyond the scope of the present study, but eliciting /æ/ and /ʌ/ words in sentences was deemed to closely reflect learners' vowel production accuracy in extemporaneous speech, which would reflect how well the contrast is encoded in the lexicon. We were expecting vowel production to be less accurate and show HVPT gains of a smaller magnitude for words elicited in a sentence- than in a word-repetition task. We would interpret vowel production gains obtained in sentences to indicate improvement in the lexical encoding of the target vowels.

In the current study we compare the effectiveness of non-lexical (nonwords) and lexical (word) HVPT of the target English vowel contrast (/æ/ and /ʌ/) in a homogeneous L1-Spanish/Catalan population for whom this contrast is particularly difficult to acquire. We predict non-lexical training to be conducive to overall larger gains than lexical training, as it would enhance a focus on phonetic form while avoiding the activation of lexical forms for which learners might have developed inaccurate phonological representations.

Additionally, we manipulated (between subjects) the listening conditions during the production training portion of the HVPT paradigm by randomly assigning trainees (both within the non-lexical and the lexical training groups) to either a +noise or a –noise condition. In this way, we hoped to show that eliciting clear speech (by training production in masking noise) could enhance the benefits of perceptual HVPT for the production accuracy of a difficult L2 vowel contrast. We

expected the potential benefits of masking noise during production training to differ in magnitude across training conditions. Repeating words in noise during lexical production training might be more likely to elicit clear speech and enhance vowel production accuracy than repeating nonwords in noise during non-lexical training because trainees would have developed lexical representations for minimal-pair words being trained (e.g. *cap-cup*), but not for nonwords being trained (e.g. *fash-fush*). Enhancing the target vowel quality distinction through clear speech in noise might thus be easier for lexically- than for non-lexically trained learners. Should this be the case, we might recommend including masking noise during production training in HVPT paradigms only when training is lexically oriented, whereas non-lexical HVPT may be more effective than lexical HVPT only when training does not include a production component, or it includes a production component administered under regular quiet conditions.

The present study, therefore, examined the benefits of two HVPT conditions, the use of non-lexical training stimuli and the presence of masking noise during production training, on L1-Spanish/Catalan bilingual EFL learners' production of English /æ/ and /ʌ/, embedded in words elicited in isolation as well as in meaning-focused sentences. We addressed the following research questions:

RQ1: Is phonetic training effective at improving the production of /æ/ and /ʌ/? Are there differential training gains on /æ/ and /ʌ/?

RQ2: Do vowel production gains on trained items generalize to new items?

RQ3: Are there differential training gains as a function of the lexical status of the training stimuli (words vs. nonwords) in the production of /æ/ and /ʌ/ in isolated words and in words elicited in sentences?

RQ4: Does the use of masking noise during production training improve /æ/ and /ʌ/ production accuracy in isolated words and in words elicited in sentences?

RQ5: Is phonetic training equally effective at improving the production of /æ/ and /ʌ/ in isolated words and in words elicited in sentences?

We hypothesized that the training procedure implemented would be effective at improving the perception and production of /æ/ and /ʌ/ (RQ1), leading to gains in production accuracy that would generalize to untrained items (RQ2). However, based on perceptual assimilation studies (Cebrian 2019), we predicted English /ʌ/ to be produced less accurately and to present more room for improvement than /æ/. Based on previous research (Thomson and Derwing 2016), we hypothesized

non-lexical training to be more effective than lexical training (RQ3), as nonwords would avoid the activation of learners' phonological forms that may be inaccurate. In addition, production training in noise was expected to lead to larger benefits in learners' ability to distinguish /æ/ from /ʌ/ in production (RQ4), as producing speech in noise might lead to hyper-articulation and more peripheral vowel production (Bradlow and Bent 2002; Hazan and Baker 2011). However, we would expect an interaction between training conditions: noise might elicit a clear speech mode leading to production benefits only for those training items learners had a representation for (i.e. words), but not for nonword training items, for which learners had no representation. Specifically, lexical production training in noise would allow learners to activate a representation (despite the noise) to generate an articulatory plan intended to overcome the adverse listening conditions (hyper-articulation), whereas noise may be detrimental during non-lexical production training, as it would interfere with learners' ability to focus on the target phonetic forms. Finally, we hypothesized any training effects obtained under the various training conditions to lead to larger gains when assessed through words in isolation than when assessed through words elicited in sentences (RQ3, RQ4, RQ5). Eliciting words in isolation in the DWR task would likely allow learners to focus their attention more on the acoustic features that distinguish the target vowels and to achieve higher levels of articulatory control than when the target words are embedded in sentences.

5 Methods

The participants (L1 Spanish-Catalan EFL learners) were tested on their ability to produce English /æ/ and /ʌ/ before and after four 30-min phonetic training sessions that took place on two separate days per week during two consecutive weeks (see Table 1). The pre-test and post-test included an ABX discrimination task to assess training gains in perception (not reported in the current study), and delayed word (DWR) and sentence (DSR) repetition tasks to assess training gains in production. We implemented a comprehensive training paradigm that used both perception and production training within every session. The training consisted of AX discrimination (AX), identification (ID), and immediate repetition (IR) tasks and a short word learning task (WL). There are very few HVPT studies that combine perception and production training within the same training procedure. Baese-Berk and Samuel (2016) implemented a procedure whereby production of target items was required before perceptually categorizing them, which led to disruptive effects on perceptual learning. Other studies combining perception and production training, with feedback (Herd et al. 2013) or without feedback (Lu et al. 2015;

Thorin et al. 2018) on production accuracy, have found the inclusion of a production component to be neither advantageous nor detrimental compared to a perception-only training procedure. To the best of our knowledge, this is the first HVPT study to include a training procedure that combines AX discrimination, taping pre-lexical processing (acoustic-phonetic differences between sound categories), identification, taping a phonological processing level (identification of category representations for the target sounds) and immediate repetition, to allow learners to monitor and execute changes in production previously induced by perceptual learning. We expected the production component of the training procedure to enhance training gains in production, but it was beyond the scope of the current study to experimentally assess its effectiveness and contribution separately from that of perceptual training.

A language background questionnaire and a word familiarity questionnaire were administered online before the first training session. In the word familiarity questionnaire, we asked learners to indicate how familiar they were with the words presented by choosing "0" if they had never seen the word and did not know what it meant; and "7" if they were very familiar with it, could use it when speaking/writing, and knew what it meant.

In session 3, participants' L2 proficiency and vocabulary size were assessed through an elicited imitation (EI) task (Ortega et al. 2002) consisting of 30 sentences varying in length (7–19 syllables) and grammatical complexity, and a receptive vocabulary size test (X/Y Lex; Meara and Milton 2003; Meara and Miralpeix 2006).

Table 1: Research design.

	Week 1		Week 2	
	Session 1	**Session 2**	**Session 3**	**Session 4**
Pre-Test	ABX, DWR, DSR		(EI, X/Y_Lex)	
Training	WL, AX, ID, IR	WL, AX, ID, IR	WL, AX, ID, IR	WL, AX, ID, IR
Post-Test				ABX, DWR, DSR

ABX = ABX discrimination; AX = AX discrimination; DSR = delayed sentence repetition; DWR = delayed word repetition; EI = elicited imitation; ID= Identification; IR = immediate repetition; WL = word learning; X/Y_Lex = vocabulary size test.

5.1 Target vowel contrast

The vowel contrast targeted in the present study is English /æ/- /ʌ/ (*cap* vs. *cup*). Even though these two vowels differ acoustically in duration (/æ/ is longer) and F1

and F2 frequency (/æ/ is produced with a lower and more advanced tongue position than /ʌ/), and visually in degree of lip aperture (/æ/ is opener), even advanced Spanish learners of English struggle to distinguish these vowels in perception and production (Aliaga-García and Mora 2009). In addition, cross-language perception studies suggest that /æ/ and /ʌ/ differ in degree of perceived similarity with respect to the Catalan and Spanish /a/, with the English /ʌ/ being more acoustically dissimilar to the Spanish /a/ than to the English /æ/ (Cebrian 2019; Cebrian et al. 2011). Therefore, whereas we would expect L1-Spanish learners to produce English /æ/ with more target-like quality than English /ʌ/ based on its closer acoustic similarity to Spanish /a/, English /ʌ/ (acoustically more dissimilar to Spanish /a/), might be easier to improve through L2 exposure, production practice and phonetic training than English /æ/. Learning to distinguish English /æ/ from /ʌ/ in production is important for L2 learners as this will help them distinguish a large number of words based on this vowel contrast effectively and will help them enhance the comprehensibility of their speech due to its high functional load (Munro and Derwing 2006; Suzukida and Saito 2021). In addition, learning to produce this contrast accurately as L2 vocabulary develops is important because it can contribute to the establishment of contrastive lexical representations for minimal word pairs (e.g. *cap* vs. *cup*) and enhance the encoding of the /æ/-/ʌ/ phonological contrast in the lexicon (Bundgaard-Nielsen et al. 2012).

5.2 Participants

Sixty-two Spanish-Catalan bilingual learners of English participated in the study. Participants, whose ages ranged from 18 to 56 (M = 22.8, SD = 7.2), had an intermediate/upper-intermediate level of English as they were all enrolled in English studies at the time the data was collected. They were randomly assigned to experimental training groups (N = 48) or to an untrained control group (N = 14) (see Table 2). One-way ANOVAs with *Group* as the independent variable confirmed that experimental (M = 6.9, SD = 1.1) and control groups (M = 6.4, SD = 2.1) were comparable in self-estimated L2 proficiency (F[1,60] = 1.086, p = 0.302) as measured through a 9-point Likert scale where 1 meant "very poor" and 9 "native-like". Participants had learnt English mainly through formal instruction at school and reported limited weekly exposure to English. They varied in degree of Spanish and Catalan dominance (14.3% were Catalan-dominant, 35.7% were Spanish-dominant, 50% were balanced), but this was not expected to affect training outcomes, as English /æ/ and /ʌ/ are mapped onto a similar /a/ low vowel category in L1-Catalan and L1-Spanish (Cebrian et al. 2011). Participants reported no history of speech or hearing impairment and were given course credit for participation.

As shown in Table 2, experimental participants were randomly assigned to a nonword (NWD) training group (N = 24) or a word (WD) training group (N = 24). Both training groups were further sub-divided into two equal groups according to whether production training was provided in noise (N = 12) or in silence (N = 12). Experimental and control groups performed pre-test and post-test production tests, but only the experimental groups were trained in the perception and production of English /æ/ and /ʌ/. One-way ANOVAs showed that the control and training groups were comparable in terms of all demographic variables (all p values >0.05), and the 4 experimental groups were comparable in terms of L2 proficiency (F[3,44] = 1.299, p = 0.311) as measured through the EI task and vocabulary size (F[3,40] = 0.457, p = 0.714) as measured through X/Y_Lex.

Table 2: Participants' demographics.

Measure	Control group N = 14		Experimental group N = 48							
			WD training				NWD training			
			noise N = 12		silence N = 12		noise N = 12		silence N = 12	
	M	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*	*M*	*SD*
Age at testing (years)	26.7	7.4	22.6	10.7	22.3	7.7	21.1	2.4	21.3	4.7
Age of onset of L2 learning (years)	7.7	6.7	5.7	1.9	5.8	1.9	5.8	1.8	4.9	2.3
L2 instruction (years)	19.3	23.5	16.9	5.6	16.5	6.3	15.3	5.6	16.5	6.3
Spoken L2 input	19.0	22.1	13.2	7.9	15.6	9.2	11.2	3.8	10.5	5.9
L2 use[1]	9.5	11.6	7.3	7.7	7.6	4.9	6.2	3.5	4.1	3.5
Vocabulary size (0–10,000 words)[2]	–	–	6368	1012	5923	1196	6229	1121	6505	1372
L2 proficiency (0–120 points)[3]	–	–	99.7	16.8	89.6	9.5	95.5	11.5	95.5	11.5
Self-estimated proficiency (1 = very poor-9 = native-like)[4]	6.4	2.1	6.9	1.5	6.4	1.0	7.2	1.1	6.9	0.7

[1]L2 use with native and non-native speakers in hours per week. [2]Obtained through the X/Y Lex. The control group did not take this test. [3]Obtained through the Elicited Imitation task (Ortega et al. 2002). The control group did not take this test. [4]Averaged self-estimated ability to speak spontaneously, understand, read, write, and pronounce English.

5.3 Speech materials

The testing and training stimuli (see Appendix A) were elicited twice and recorded in a soundproof booth by 6 British English speakers (3 males, 3 females). Four of

the 6 speakers' voices were used in the training, the other two (1 male, 1 female) were used in the testing, so that improvement in vowel production accuracy would be indicative of generalization to new voices. Nonword and word stimuli were elicited from a randomized reading list in carrier phrases (*I say X*, *I say X again*) from which they were excised, low-pass filtered (50 Hz) and normalized for mean amplitude in *Praat* (Boersma and Weenink 2020). Sentence stimuli were recorded from a reading list and filtered and normalized in the same way as the nonword and word stimuli.

The training stimuli consisted of 16 high-variability monosyllabic CVC nonword (8) and word (8) minimal pairs produced by 4 different speakers (2 females, 2 males), with the target vowels in 8 different CVC phonetic environments (e.g. *chang* /ʧæŋ/-*chung* /ʧʌŋ/, *mad* /mæd/-*mud* /mʌd/). High stimulus variability was achieved by exposing learners to a variety of exemplars of the vowel contrast to be learned in 8 different phonetic environments and by including talker variability in the training set (4 speakers, 2 females, 2 males), in order to enhance learning of the trained stimulus set and improve generalization to novel stimuli. Testing stimuli consisted of 12 monosyllabic CVC minimal pair nonwords (6 trained, 6 untrained) and 18 monosyllabic CVC minimal pair words (6 trained, 12 untrained). In addition, 16 CVC non-minimal pair words (8 for /æ/, and 8 for /ʌ/) were elicited in isolation (e.g. *map*) in a delayed word repetition task and in context in a delayed sentence repetition task (e.g. *He looked at the map to find his way*).

5.4 Phonetic training

Training was administered in 4 30-min separate sessions with at least a day in between over a period of 2 weeks (see Table 1). Therefore, each participant attended two sessions per week, which could be either Monday and Wednesday or Tuesday and Thursday. The length of the training was consistent across participants. All training tasks started with 6 practice trials to ensure participants understood the tasks. Test trials were presented in fully randomized order within every training session. Training groups were exposed to a total of 8 minimal pairs (words or nonwords, as a function of training group). In each one of the 4 training sessions, learners were trained on 2 different word/nonword minimal pairs spoken by 4 different voices (2 females, 2 males). The same two minimal pairs were trained in all three tasks in every training session (see Appendix A).

Learners in the non-lexical training groups were trained exclusively on nonwords whereas learners in the lexical training groups were trained exclusively on words (see Appendix A-1); the procedures, interface, and feedback were identical

for both lexical and non-lexical training groups, including the type of noise the groups doing production training in noise were exposed to.

In the AX discrimination task participants decided, as fast and accurately as possible, whether two words/nonwords presented auditorily contained the same (AA, BB) or different (AB, BA) vowels. The task consisted of 96 trials × 4 sessions (384 trials). In each trial, participants heard 2 words/nonwords with a 500-ms inter-stimulus interval produced by 2 different voices as 2 response alternatives appeared on the screen ('same' or 'different') corresponding to designated labelled keys on the computer keyboard. Out of the 96 trials in every session (2 minimal pairs × 4 orders × 12 voice combinations), half of the trials started with a female voice and half with a male voice. Immediate feedback was provided in each trial so that they could monitor both accuracy and the speed with which they took decisions. In each trial, the words 'Correct!' or 'Wrong!' appeared in the middle of the screen together with the response latency in milliseconds. The aim of the AX task was to train learners' sensitivity to the primary acoustic cues qualitatively distinguishing /æ/ from /ʌ/ (first and second formant frequencies) and to improve learners' processing of such cues (first and second formant frequencies and duration).

In the identification task (ID) participants heard a single word/nonword as two response alternatives consisting of the words *cap* and *cup*, their phonetic transcription (/kæp/ and /kʌp/) and pictures representing them, appeared on the left and right sides of the screen, which corresponded to designated labelled keys on the computer keyboard. The task consisted of 32 trials (4 minimal-pair words × 2 repetitions × 4 voices) × 4 sessions (128 trials). The aim of the task was to train the identification of the category representations for /æ/ and /ʌ/, enhance generalization across contexts and talkers, and improve the categorical processing of the target vowels (Sadakata and McQueen 2013). Feedback on error and response latency was provided as in the AX task.

In the immediate repetition task (IR), participants heard a target word/nonword (e.g. /mæp/) twice and were asked to repeat it twice within 2000 ms following each of the two presentations focusing on the articulation of the vowel. The task consisted of the same 32 trials as in the ID task (4 minimal-pair words × 2 repetitions × 4 voices) × 4 sessions (128 trials). The purpose of the second presentation and repetition was to allow learners to compare their two attempts at producing the same target item and auditorily monitor accuracy of vowel production. The stimuli presentation conditions varied according to the training group: nonwords in noise, nonwords in silence, words in noise, words in silence. Particularly, two of the experimental groups were presented with the stimuli in noise, while the other two in silence. The task was the same for all groups except for the presence or absence of noise. The aim of the IR task was to train learners to

monitor their own productions and to train them to implement articulatory changes in the production of the target vowels they had just been perceptually trained on. It should be noted that the presence of background noise was only included in the production training. The AX discrimination and identification training was performed without noise. The purpose of including a condition where production was trained in noise was to explore the potential effect of noise in generating clear speech during the repetition of the training stimuli. A continuous stream of multi-talker English babble lasting the full duration of the immediate repetition task was played to participants through both ears over headphones as they performed the task, so that the background noise could be heard throughout the task, even when the participant was waiting for the next item to be presented. The background noise and the auditory stimuli for repetition were played and mixed at the same intensity level (0 dB NSR) to simulate the situation of speaking in adverse listening conditions (such as in a crowded cafeteria or a dinner party). In all cases, as confirmed in previous piloting, the background noise did not prevent learners from identifying the segmental composition of the nonwords and words for repetition and was successful in eliciting perceptually louder, hyperarticulated productions of the target items. The items in the immediate repetition task were the same as those previously trained in silence during the preceding perceptual identification task, which was deemed to facilitate recognition in noise. Full randomization of item presentation and repetition of target items twice minimized the possibility of noise having detrimental effects on specific items. Participants wore open headphones that allowed them to monitor their productions while hearing them in noise. Their productions were recorded for acoustic analysis via a voice microphone that would only very minimally capture the faint multi-talker babble noise that could escape the headphones. The authors of the current study, experienced in L2 English, carried out the acoustic analyses of the learners' productions.

In the word learning task, learners first engaged in a self-paced memorization phase with 12 unfamiliar minimal-pair words (e.g. *gash* vs. *gush*) through a presentation that showed images and sentences illustrating the meaning of the word and its orthographic form, on which they could click to hear it as many times as they wished. This was followed by a testing phase that asked learners to identify a picture illustrating the meaning of auditorily presented words. By the end of the training, all learners had learnt to identify all words in the word learning task. Familiarity with the words embedded in the task was assessed by the online Word Familiarity Questionnaire. Overall, learners reported being relatively unfamiliar (M = 2.93, SD = 1.37) with the words presented in the word learning task.

5.5 Testing

Training gains in the pronunciation /æ/ and /ʌ/ in words produced in isolation and in sentences were assessed at pre-test and post-test through a delayed word repetition (DWR) task and a delayed sentence repetition (DSR) task, respectively.

In the DWR task, participants were instructed to repeat a total of 76 trials (i.e. 36 pairs contrasting /æ/-/ʌ/) that included trained and untrained words and nonwords produced by untrained voices: 6 trained pairs of nonwords, 6 untrained pairs of nonwords, 6 trained pairs of words, 6 untrained pairs of words, and the 6 pairs of words used in the word learning task. Additionally, the DWR included 8 untrained pairs of words which were also used in the DSR task. We expected this delayed elicitation procedure (1500 ms silent interval followed by a 200 ms tone), previously used to investigate segmental production accuracy in L2 learners (Nagle 2021), to elicit target testing items within a phonological (rather than acoustic or phonetic) processing mode (Werker and Logan 1985), thus avoiding direct imitation from sensory memory, which ensured the elicited stimuli reflected participants' vowel representations in isolated words. All pairs of words were minimal pairs, except the 8 pairs used in the DSR task (see Appendix A).

In the DSR task, participants were instructed to read silently a sentence appearing on the screen for 4 s (e.g. *He looked at the map to find his way*) targeting an /æ/ or /ʌ/ word (e.g. *map*; see Appendix A). The sentence then disappeared from the screen and participants heard the sentence being pronounced by a native speaker. They were asked to remember it and repeat it from memory after 1500 ms upon hearing a tone signal. Sixteen sentences in 2 untrained voices (1 female, 1 male) were repeated twice. This task included the 16 untrained words previously used in the DWR test. Vowels in words elicited from memory in meaningful sentences were deemed to more closely reflect learners' vowel productions in extemporaneous speech than the vowel productions in the DWR task, as repeating sentences from memory would minimize the possibility of learners paying conscious attention to the phonetic form of the target test words.

5.6 Procedures and data analyses

All the perception and production tasks were administered in *DmDx* (Forster and Forster 2003) on laptop computers. Noise-cancelling headphones (Beyerdynamic DT 770 M) were used in the AX and ID tasks, and open headphones (Beyerdynamic DT 990 PRO) in the IR and DWR and DSR tasks. The word and sentence productions were recorded on Marantz PMD-661 solid-state digital recorders with an external Shure SM58 voice microphone at a sampling frequency of 44.1KHz. Improvement

in vowel production accuracy for /æ/ and /ʌ/ was assessed by comparing pre-test (T1) and post-test (T2) differences in quality between the vowels produced by L2 learners and those produced by the native speakers who provided the training and testing stimuli (Rato and Rauber 2015). Vowel quality was measured manually by the authors of the study in Praat. We extracted mean F0, F1 and F2 measures from a 10 ms window centred at a cursor placed in the centre of the steady state portion of the second formant of the vowel. In order to minimize age, gender and vocal tract size differences among speakers and provide speaker-independent estimates of vowel quality, we first converted frequency values in Hertz (Hz) to Bark (B) and then applied a Bark-distance normalization procedure (Syrdal and Gopal 1986), such that the difference in Bark between F1 and *f*0 (B1-B0) was used as an estimate of vowel height, whereas the difference in Bark between F2 and F1 (B2-B1) was used as an estimate of vowel frontness (Baker and Trofimovich 2005; Bohn and Flege 1990; Mora et al. 2015). Vowel quality differences between learners' and native speakers' vowel productions in the DWR and DSR test items were then estimated through normalized Bark-converted spectral distance scores (SDS, i.e. Euclidean distances).

We carried out four sets of data analyses using mixed-effects models. First, we assessed training effects by comparing vowel production accuracy between testing times for the experimental and control participant groups (Section 6.1). Then we assessed generalization effects in the DWR task by testing training effects on untrained items (relative to training effects on trained items) for the experimental participant group only (Section 6.2). We compared training effects on trained and untrained items separately by training group, as training items were different for participants trained on nonwords and those trained on words. As the production training task (immediate repetition) only involved repeating items in isolation, and in the DSR task all the target items were untrained words, a significant effect of training on the DSR items, was interpreted as learners having generalized training gains from items trained and tested in isolation to different novel untrained items tested in sentences.

Next, we tested for training condition effects, specifically, the effect of training *Group* (nonword vs. word training) and listening *Condition* during production training (in noise vs. in silence) on experimental participants (Section 6.3). These analyses were performed separately for the test items in the DWR and DSR tasks because test items in the DSR task consisted of a subsample of 16 items out of the total of 76 test items included in the DWR task. Finally, we assessed potential differential HVPT effects on vowel production accuracy between tasks (DWR, DSR) by including *Task* (DWR, DSR) as an independent variable in a mixed-effects model that was performed on the 16 test items subset common to both tasks.

Differences in spectral distance scores between testing times as a function of training group were assessed by including all testing items (76 test items, 38 /æ/-items 38 /ʌ/- items; see Appendix A), as the analysis of generalization effects showed that vowel production accuracy improved both in trained and untrained items.

6 Results

6.1 Training effects

The overall effectiveness of the training (RQ1) was assessed by fitting the spectral distance scores of test items in the DWR task to a linear mixed-effects model with *Group* (1 = Experimental: NWD+WD; 2 = Control), *Testing Time* (T1, T2), *Vowel* (/æ/, /ʌ/), and their interactions as fixed effects, and random intercepts for *Subject* and *Item* (see Table 3 for the model outcome and Appendix B for parameter estimates). This analysis yielded a significant main effect of *Vowel* (F[1, 9416] = 15.12, $p < 0.001$), as for both experimental and control groups /æ/ presented significantly smaller SDSs (i.e. more target-like productions) than /ʌ/ did (t[9568] = –5.408, $p < 0.001$ vs. t[9568] = –3.257, $p = 0.001$, respectively). The main effects of *Group* and *Testing Time* did not reach significance, but crucially, *Group* significantly interacted with *Testing Time* (F[1, 9416] = 14.75, $p < 0.001$) because for experimental learners vowel accuracy significantly improved between testing times (t[9416] = 2.64, $p = 0.008$), whereas for the untrained control group vowel accuracy significantly worsened (t[9416] = –2.93, $p = 0.003$). No other interactions reached significance.

The SDSs from the set of untrained words we used in the DSR task were also fitted to a linear mixed-effects model, as described above (see Table 3 for the model outcome and Appendix B for parameter estimates). None of the main effects reached significance, but the *Group* × *Vowel* interaction was significant (F[1, 1976] = 10.15, $p = 0.001$) because experimental participants produced /ʌ/ (but not /æ/) with significantly larger SDS (M = 0.62 Bark, SE = 0.266) than the controls did (t[1976] = 2.32, $p = 0.020$). In the DSR task, the experimental participants improved only slightly and not significantly on both vowels (0.015 Bark in /æ/ and 0.146 Bark in /ʌ/). Thus, whereas training was effective at improving vowel production accuracy in the DWR task, it did not lead to significant changes in vowel accuracy when the vowels were embedded in words elicited in a DSR task.

Table 3: Mixed-effects model outcome.

Fixed effects	DWR				DSR			
	F	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*
Group	0.182	1	9419	0.670	1.329	1	1976	0.249
Testing Time	1.759	1	9419	0.185	0.919	1	1976	0.338
Vowel	15.127	1	9419	<0.001*	0.033	1	1976	0.856
Group × *Testing Time*	14.758	1	9419	<0.001*	2.978	1	1976	0.085
Group × *Vowel*	1.492	1	9419	0.222	10.156	1	1976	0.001*
Testing Time × *Vowel*	0.007	1	9419	0.931	0.024	1	1976	0.876
Group × *Testing Time* × *Vowel*	0.043	1	9419	0.836	0.603	1	1976	0.438

Asterisks indicate significance.

6.2 Generalization to new items

Generalization effects (RQ2) were tested by comparing training effects on trained and untrained items from the DWR task only, as all the words included in the DSR task were untrained. These analyses were conducted separately for the two training groups. Linear mixed-effects models were performed with *Testing Time* (T1, T2), *Vowel* (/æ/, /ʌ/), *Trial Type* (trained nonwords, untrained nonwords, untrained words), and their interactions as fixed effects, with *Subject* and *Item* random intercepts (see Table 4 for the model outcome and Appendix B for parameter estimates). For the NWD training group, we found significant main effects of *Testing Time* (F[1, 3636] = 5.47, $p < 0.019$) and *Vowel* (F[1, 3636] = 7.38, $p < 0.007$), but neither the main effect of *Trial Type* ($p = 0.984$) nor any of the interactions reached significance. These main effects indicated shorter SDSs at post-test than at pre-test for /æ/ than for /ʌ/ for all trial types (trained as well as untrained). None of the pairwise contrasts involving trained and untrained trial types reached significance (all $p = 1.0$), indicating that training gains (i.e. shorter SDS at post-test than at pre-test) in untrained items (whether nonwords or words) were of a similar magnitude to those obtained for trained items.

For the WD training group, we found a significant main effect of *Vowel* (F[1, 3636] = 4.91, $p = 0.027$), but neither the main effect of *Testing Time* (F[1, 3636] = 0.212, $p = 0.645$) or *Trial Type* (F[1, 3636] = 0.389, $p = 0.678$), nor any of the interactions, reached significance, indicating an overall lack of improvement in vowel production accuracy. The main effect of *Vowel* was driven by SDSs being shorter for /æ/ than for /ʌ/ irrespective of testing time and trial type. Therefore, no generalization effects can be said to have occurred for this group.

Table 4: Mixed-effects model outcome.

Fixed effects	NWD training				WD training			
	F	*df1*	*df2*	*p*	*F*	*df1*	*df2*	*p*
Testing Time	5.469	1	3636	0.019*	0.212	1	3636	0.645
Vowel	7.385	1	3636	0.007*	4.908	1	3636	0.027*
Trial Type	0.016	2	3636	0.984	0.342	2	3636	0.710
Testing Time × *Vowel*	0.799	1	3636	0.372	1.198	1	3636	0.274
Testing Time × *Trial Type*	0.578	2	3636	0.561	0.389	2	3636	0.678
Vowel × *Trial Type*	0.175	2	3636	0.840	0.077	2	3636	0.926
Testing Time × *Vowel* × *Trial Type*	0.009	2	3636	0.991	0.504	2	3636	0.604

Asterisks indicate significance.

6.3 Effects of training material and listening conditions

After having established that our experimental participants (but not the control group) had improved in vowel production accuracy (Section 6.1) and that improvement took place both for trained and untrained test items (Section 6.2), we next assessed the potential differential benefits of using non-lexical (nonwords) versus lexical (words) training materials (RQ3) and using noise during production training (RQ4) on the production accuracy of the target vowels. We conducted these analyses separately for test words elicited in isolation (DWR) and in sentences (DSR) and for the non-lexical and the lexical-training groups. Because no significant differences in the magnitude of training gains had emerged between trained and untrained testing nonwords and words (see 6.2 above), all test items (trained and untrained) were included in this set of analyses.

As shown in Figure 1, participants in the WD training group obtained overall smaller SDSs than those in the NWD training group did (also at pre-test), suggesting an initial advantage of the WD training group in production accuracy of the target vowels, which could be due to differences in L2 proficiency, or in learners' ability to effectively distinguish between the target vowels in perception (Melnik-Leroy et al. 2021). To discard potential differences in proficiency between training groups we examined their L2 receptive vocabulary size and their elicited imitation scores. The groups were found to be comparable for both measures (t[42] = 0.583, p = 0.563; and t[46] = 0.230, p = 0.819, respectively). To discard potential differences in the perception of the target contrast we examined the outcome of an ABX discrimination task (120 A(/æ/)-B(/ʌ/)-X (/æ/ or /ʌ/) test trials, e.g. *cap-cup-cup* in four orders: ABA, ABB, BAB, BAA) administered at pre-test. T1 /æ/-/ʌ/ discrimination scores did not significantly predict gains in production for either group in either production task (NWD-DWR: r = 0.007, p = 0.974; NWD-DSR: r = 0.245,

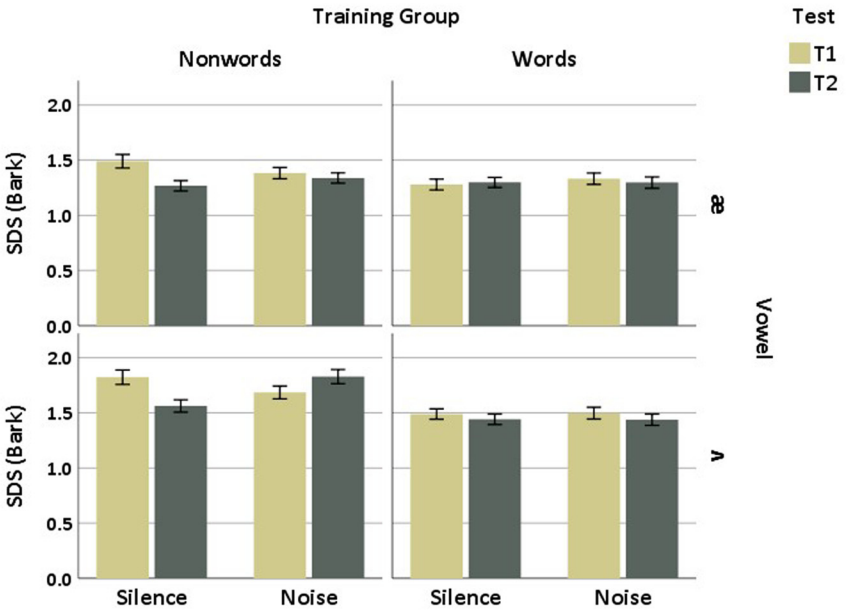

Figure 1: Spectral distance scores (SDS) as a function of testing time, vowel, training group and condition for the DWR task (error bars = ±1SE).

$p = 0.249$; WD-DWR: $r = 0.101$, $p = 0.640$; WD-DSR: $r = 0.048$, $p = 0.823$). Thus, the training group differences observed at pre-test, which did not reach significance (see 6.3.1), could not be attributed to either L2 proficiency or their perceptual ability in distinguishing the target vowels. In fact, reduction of SDSs, indicating improvement in vowel production accuracy, only seems to have occurred for the NWD training group. Such improvement appears to be of a comparable magnitude for both vowels for those participants in this group trained in silence. However, for those in this group trained in noise no improvement was observed for /æ/, and SDS increased slightly for /ʌ/, suggesting that the presence of noise during production training was detrimental, hindering the training benefits in accuracy observed for both vowels for participants trained in silence. For the WD training group, the presence of noise during production training appeared to be neither facilitatory nor detrimental of accuracy gains. In general, SDSs were larger for /ʌ/ than for /æ/.

In order to test for these training condition effects, SDSs were submitted to a linear mixed-effects model with *Group* (NWD training, WD training), listening *Condition* (silence, noise), *Testing Time* (T1, T2), *Vowel* (/æ/, /ʌ/) and their interactions involving *Testing Time* (*Group* × *Testing Time*, *Condition* × *Testing Time*, *Vowel* × *Testing Time*, *Group* × *Vowel* × *Testing Time*, *Condition* × *Vowel* ×

Testing Time, *Group* × *Condition* × *Testing Time*) as fixed effects (with *Subject* and *Item* random intercepts). The results of these analyses (see Appendix B for parameter estimates) are presented separately for the DWR (6.3.1) and DSR (6.3.2) tasks below.

6.3.1 Training effects on words elicited in isolation (DWR)

For DWR the analyses revealed significant main effects of *Testing Time* (F[1, 7280] = 6.84, p = 0.009) and *Vowel* (F[1, 7280] = 17.38, p < 0.001), and significant *Condition* × *Testing Time* (F[1, 7280] = 7.18, p = 0.007) and *Group* × *Condition* × *Testing Time* (F[2, 7280] = 5.61, p = 0.004) and *Group* × *Vowel* × *Testing Time* (F[2, 7280] = 8.56, p < 0.001) interactions. None of the other main effects or interactions reached significance (all p > 0.17). Bonferroni-adjusted pairwise contrasts showed that the *Condition* × *Testing Time* interaction arose because overall participants trained in silence (but not those trained in noise: T1: 1.47 B; T2: 1.48 B) significantly shortened SDSs (i.e. became more accurate) between testing times (T1: 1.52 B vs. T2: 1.39 B; t[7280] = 3.74, p < 0.001). However, the significant *Group* × *Condition* × *Testing Time* interaction suggests that the advantage of production training in silence over production training in noise might be driven by the type of training received. This was in fact the case. Pairwise contrasts revealed that only learners in the NWD training group significantly shortened SDSs between testing times (T1: 1.66 B; T2: 1.41 B; t[7280] = 5.01, p < 0.001). For them noise appeared to be only very slightly (non-significantly) detrimental (T1: 1.53 B; T2: 1.58 B; t[7280] = –1.03, p = 0.304), whereas learners in the WD training group improved only very slightly and non-significantly between testing times irrespective of whether production training happened in noise (T1: 1.41 B; T2: 1.36 B; t[7280] = 0.966, p = 0.334) or in silence (T1: 1.84 B; T2: 1.69 B; t[7280] = 0.295, p = 0.768). When we further explored training condition effects in the NWD training group by vowel, we found production training in silence to improve SDSs significantly for both /æ/ (t[3639] = 3.03, p = 0.002) and /ʌ/ (t[3639] = 3.56, p < 0.001), whereas production training in noise decreased accuracy significantly (see Figure 1 above) in the case of the more difficult vowel /ʌ/ (t[3639] = –1.96, p = 0.049), but not in the case of /æ/ (t[3639] = 0.611, p = 0.542).

The NWD and WD training group learners did not significantly differ from one another at pre-test (T1) in either condition (noise: t[7280] = 0.604, p = 0.546; silence: t[7280] = 2.73, p = 0.168), suggesting that group differences at T1 do not interact with training gains. Finally, the *Group* × *Vowel* × *Testing Time* interaction arose because although both training groups improved SDSs when both vowels were taken together, the NWD training group did so significantly for /æ/ (T1: 1.43 B; T2: 1.30 B; t[7280] = 2.76, p = 0.227), but not for /ʌ/ (T1: 1.75 B; T2: 1.69 B; t[7280] = 1.21, p = 0.227), whereas the WD training group improved only very slightly

and not significantly in both /æ/ (T1: 1.31 B; T2: 1.30 B; t[7280] = 0.158, p = 0.874) and /ʌ/ (T1: 1.49 B; T2: 1.43 B; t[7280] = 1.10, p = 0.270). To summarize, the inclusion of noise during production training affected the NWD and WD training groups differentially. Whereas it was neither detrimental nor beneficial for the WD training group, for the NWD training group it was found to be detrimental to /ʌ/ production accuracy.

6.3.2 Training effects on words elicited in sentences (DSR)

Differential effects of training conditions on target untrained words elicited in sentences were similar to those found for words elicited in isolation (see Figure 2).

The statistical analysis revealed a significant main effect of *Testing Time* (F[1, 1520] = 8.27, p = 0.004) and *Vowel* (F[1, 1520] = 4.64, p = 0.031), and significant *Condition* × *Testing Time* (F[1, 1520] = 9.24, p = 0.002) and *Group* × *Condition* × *Testing Time* (F[1, 1520] = 27.79, p < 0.001) interactions. None of the other main effects or interactions reached significance. The *Condition* × *Testing Time* interaction arose because overall (both NWD- and WD-training groups together) SDSs became significantly shorter at post-test by 0.38 Bark (t[1520] = 4.18, p < 0.001) for

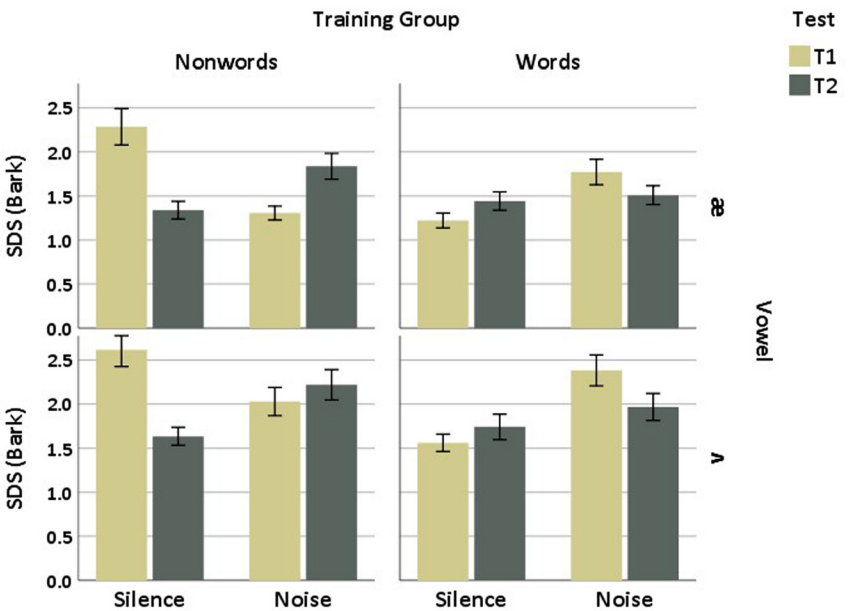

Figure 2: Spectral distance scores (SDS) as a function of testing time, vowel, training group and condition for the DSR task (error bars = ±1SE).

learners in the production training in silence condition, whereas for those trained in noise differences between pre-test and post-test SDS did not reach significance (t [1520] = −0.115, p = 0.908). However, the significant *Group* × *Condition* × *Testing Time* interaction suggests that the advantage of training production in silence, as observed for the DWR results, may depend on whether learners' training is based on non-lexical or lexical materials. Indeed, whereas we found NWD-training learners' production accuracy to increase significantly when trained in silence (T1: 2.45 B; T2: 1.48 B; t[1520] = 7.47, p < 0.001) and to decrease significantly when trained in noise (T1: 1.66 B; T2: 2.08 B; t[1520] = −2.78, p = 0.005), the WD-training learners did not significantly improve production accuracy when trained in silence (T1: 1.39 B; T2: 1.59 B; t[1520] = −1.55, p = 0.120), but obtained significant gains in production accuracy scores in the DSR task when they had trained production in noise (T1: 2.07 B; T2: 1.73 B; t[1520] = 2.62, p = 0.009). Such effects were consistent for both target vowels. These findings suggest that when training gains are assessed in words elicited in sentences (unlike what we found for words elicited in isolation) noise was detrimental for non-lexical training (i.e. for learners trained with nonwords) and beneficial for lexical training (learners trained with words), whereas training production in silence was beneficial in non-lexical training, but not in lexical training (see discussion of these effects below).

6.4 Training gains for words in isolation and in sentences

Finally, we assessed whether gains in vowel production accuracy differed as a function of elicitation task (DWR vs. DSR) by comparing the effects of training on the sub-set of test items common to both tasks (see Appendix A.2) (RQ5). As Figure 3 shows (averaged for both target vowels), SDS were larger for words elicited in sentences (DSR) than for words elicited in isolation (DWR), especially for participants in the NWD training group. Training gains for words elicited in sentences were only observable for learners trained with nonwords and for learners trained with words in noise. Thus, the noise training condition appeared to benefit learners trained with words (but not those trained with nonwords) when tested on words elicited in sentences.

SDSs were fitted to a linear mixed-effects model with *Task* (DWR, DSR), *Group* (NWD training, WD training), listening *Condition* (silence, noise) *Testing Time* (T1, T2) and the following interactions as fixed effects (with *Subject* and *Item* as random intercepts): *Task Type* × *Testing Time*, *Task Type* × *Group* × *Testing Time*, *Task Type* × *Condition* × *Testing Time*. These analyses revealed a significant main effect of *Task* (F[1, 3059] = 33.11, p < 0.001), indicating that, as predicted, overall SDS were significantly larger (less accurate vowel production) in the DSR than in the DWR

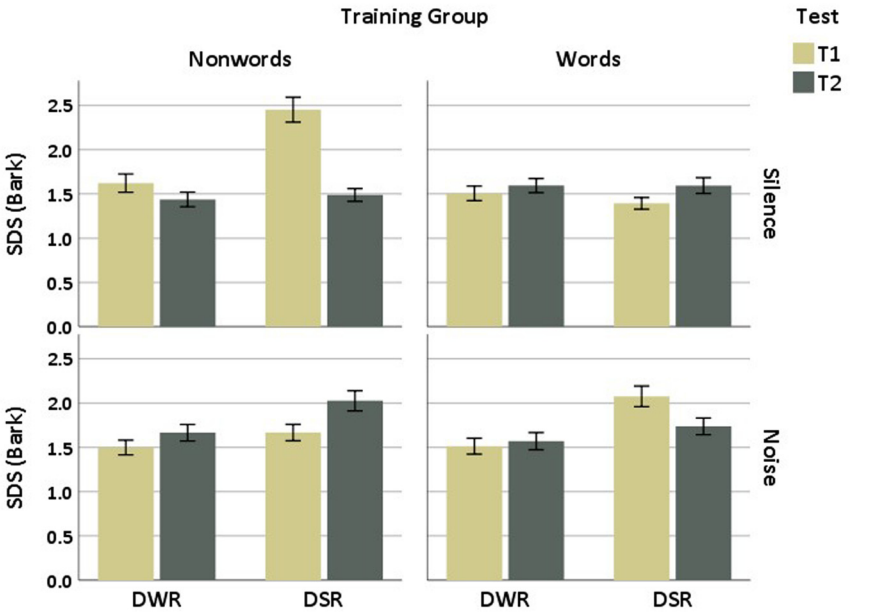

Figure 3: Spectral distance scores (SDS) as a function of testing time, listening condition, training group and elicitation task (error bars = ±1SE).

test words at both testing times (T1: 1.53 B vs. 1.89 B, $t[3060] = -3.63$, $p < 0.001$; T2: 1.56 B vs. 1.71 B, $t[3060] = -0.145$, $p = 0.020$). A significant *Task Type* × *Testing Time* interaction ($F[1, 3059] = 6.08$, $p = 0.014$) indicated that overall SDS shortened significantly between testing times in DSR words ($t[3059] = 2.97$, $p = 0.003$) but not in the DWR words ($t[3059] = -0.516$, $p = 0.606$), an effect driven by the fact that it was only for learners in the non-lexical training groups that SDS significantly shortened between testing times ($t[3059] = 3.43$, $p = 0.001$), as indicated by the significant *Task Type* × *Group* × *Testing Time* interaction ($F[3, 3059] = 3.00$, $p = 0.029$). In addition, the significant *Task Type* × *Condition* × *Testing Time* interaction ($F[1, 3059] = 4.49$, $p = 0.004$) indicated that production training in silence (but not in noise) shortened SDS between testing times ($t[3059] = 4.33$, $p < 0.001$) and did so only for the DSR words. The main outcome of these analyses confirmed that learners were overall less accurate on the production of the target vowels in words embedded in sentences (DSR) than in words elicited in isolation (DWR). Non-lexical training and production training in silence were found to favour improvement in vowel accuracy for words embedded in sentences (DSR), but not for words elicited in isolation (DWR). This is because the current analysis was carried out on a subset of 16 test items common to the DWR and DSR tasks,

which is probably why we did not observe here the training condition effects reported in 6.3.1 for the full set of 76 test items in the DWR task.

7 Discussion

The present study set out to examine the benefits of two phonetic training conditions on the production of English /æ/-/ʌ/ by 48 advanced L1-Spanish learners of English: the use of non-lexical versus lexical training stimuli (nonwords vs. words) and the inclusion of noise (or not) during production training. Training benefits were assessed in isolated and sentence-embedded words. Overall HVPT led to phonetic learning, but training gains were larger for learners trained with nonwords than for those trained with words. Production training in noise did not lead to gains, but training in silence did. In addition, the magnitude of the gains obtained were larger for /ʌ/ than for /æ/ and for words elicited in isolation than for words in sentences. Table 5 below summarizes the main findings across training conditions.

The present study reveals that the high-variability phonetic training we implemented, which combined perception and production training within every session, was effective at improving the production accuracy of /æ/ and /ʌ/ (RQ1), as indicated by generalization of gains to untrained nonwords and words (RQ2). However, training gains did not occur across the board; when they did, they differed in magnitude as a function of training conditions (lexical vs. non-lexical; production training in silence vs. in noise) and elicitation context (in isolation vs. in sentences). Training gains also differed as a function of the target vowel (/æ/ vs. /ʌ/), which in general were larger for /ʌ/ than they were for /æ/, as /æ/ was produced at higher accuracy rates than /ʌ/ due to its closer perceptual similarity to the learners' L1 low vowel /a/ (e.g. Cebrian 2019; Cebrian et al. 2011).

Table 5: Summary of main findings across conditions (√ indicates significant training gains).

		Production training			
		silence	***noise***	***silence***	***noise***
Training materials	*Nonwords*	√	–[1]	√	–[2]
	Words	–	–	–	√
Testing context		in isolation		in sentences	

/æ/ was produced more accurately than /ʌ/, but training gains were larger for /ʌ/ than for /æ/. [1]Noise was detrimental for /ʌ/; [2]noise was detrimental for /ʌ/ and /æ/.

Training gains were larger for learners trained with nonwords than for learners trained with words (RQ3), confirming previous findings (Thomson and Derwing 2016). One explanation for non-lexical training materials being superior to lexical training materials in affecting training gains is that they promote attention to phonetic form by avoiding meaning processing. In addition, the use of non-lexical forms avoids the lexical activation of word forms that might be L1-accented (Broersma and Cutler 2008), which would hinder HVPT benefits by preventing trainees to process critical quality differences between the contrasting vowels that may have been inaccurately encoded at the lexical level.

Production training conditions appeared to differentially affect the benefits of non-lexical training (RQ4). Production training in noise did not benefit vowel learning when learners were trained with nonwords, suggesting that in the absence of already established lexical representations, the presence of noise interfered with learners' ability to focus on the acoustic differences distinguishing /æ/ from /ʌ/. For learners trained with words, the presence of noise during production training was neither detrimental nor beneficial for vowels in words elicited in isolation. Namely, already established lexical representations appeared to be robust enough to be unaffected by the adverse noise condition. However, the presence of noise during lexical production training did not lead to consistent benefits in vowel production accuracy as one would expect from the elicitation of hyperarticulated speech during production training, as such positive effects were only found in sentences. This may be explained by the type of masking noise we used (multi-talker babble) not being effective at maximizing learners' attentional focus on the critical acoustic and articulatory dimensions distinguishing /æ/ from /ʌ/, as it has been shown that talkers modify their speech differently under different types of background noise (Cooke and Lu 2010), or by the noise-to-sound ratio level used not being effective at producing the expected level of spectral change in vowel production that could have led to training gains in production (Cooke and García-Lecumberri 2012; Hazan and Baker 2011). It is also worth noting that vowel production was less accurate in words elicited in sentences than in isolation and therefore there was more room for improvement in the former elicitation context. Other explanations for the effectiveness of lexical training in noise in words elicited in sentences but not in isolation include isolated words being more sensitive to perceptual degradation through noise (and therefore harder to identify) than the same words embedded in meaningful sentences, where listeners can rely on contextual cues, or hyper-articulation being more likely to occur during repetition practice over sentence-long stretches of speech than word-sized units. Further analyses would be needed to assess the validity of these tentative explanations. For example, acoustic analyses on the training data would reveal whether the background noise we used during production training was actually effective at

modifying learners' spectral distances between the contrasting vowels. In addition, further research on the use of masking noise during production training to enhance L2 vowel production accuracy in HVPT is needed to determine its effectiveness in this kind of phonetic training paradigm.

For vowels embedded in words elicited in sentences, vowel production accuracy was less accurate than for words elicited in isolation (RQ5), suggesting that meaning processing may have prevented learners from focusing on the acoustic properties of the target vowels for accurate production (Trofimovich 2008). Although vowel production was less target-like in sentences than in isolated words, training benefits did occur in this context for learners trained with nonwords in silence, suggesting that non-lexical training (unlike lexical training) succeeded at improving production accuracy both in isolated words and in words in sentences. In addition, irrespective of elicitation task, production training in noise was found to be detrimental for learners trained with nonwords, whereas for learners trained with words production training in noise led to benefits in vowel accuracy in the most demanding sentence elicitation context. Thus, when production training was in silence the NWD training group improved, but the WD training group did not, whereas noise was detrimental to learners trained with nonwords but not to those trained with words. These findings suggest that (a) training learners with words may make their lexical representations more resistant to adverse conditions than training them with non-lexical materials, and (b) improving the pronunciation of words for which learners have already established long-term phono-lexical representations is much easier if training allows learners to focus on the phonetic properties of the target vowels, such as when they are trained with non-lexical materials (nonwords) and in optimal listening conditions.

8 Conclusion

Overall, the outcome of the present study suggests that the type of comprehensive HVPT implemented in the present study, which included AX discrimination, identification and immediate repetition tasks combining both perception and production training, was effective at improving the pronunciation of English /æ/ and /ʌ/ for L1-Spanish learners. However, production accuracy was more target-like for /æ/ than it was for /ʌ/ and phonetic training gains were overall larger for the more difficult /ʌ/ than for /æ/. Future research should investigate whether and to what extent the inclusion of a production component within every training session significantly contributes to enhancing perceptual training gains or, as some research has shown, is ineffective (Herd et al. 2013; Lu et al. 2015; Thorin et al. 2018) or even

detrimental (Baese-Berk and Samuel 2016) in training L2 sound perception and production. However, training benefits in the present study appeared to be modulated by the training conditions under which training took place: training materials (non-lexical vs. lexical training) as well as listening conditions during pronunciation training (in noise vs. in silence) affected the training outcomes.

Nonword training using non-lexical materials (nonwords) was successful at helping learners focus on the phonetic features necessary for improving the production of the difficult L2 target vowels. Such training benefits in production did not hold when production training was implemented in noise, as adverse listening conditions may have prevented learners from focusing attention on the acoustic and articulatory features that would allow them to improve the pronunciation accuracy of the target vowels. The NWD training group was thus able to produce the target vowels with significantly shorter SDS at post-test than at pre-test when production training occurred under optimal listening conditions. On the other hand, although lexical training was found to be less effective at improving L2 vowel production accuracy than non-lexical training, when lexical production training was implemented in noise, it led to L2 vowel production improvement in sentences. Thus, both training stimuli type (nonwords vs. words) and production training conditions (noise vs. silence) appear to play a role in training the production of L2 vowels and may do so differentially as a function of the elicitation context in which vowels are tested. Although according to several recent research syntheses (Sakai and Moorman 2018; Thomson 2018) our sample size (12 participants per training group) is comparable to that of many other HVPT studies, it is relatively small given the number of training conditions investigated, which suggests that our results need to be interpreted with caution and further research is needed to corroborate the present findings.

In sum, the findings of the present study support advocating for the use of nonword training in silence within a HVPT paradigm because these are the conditions that appeared to be the most capable of promoting a focus on the phonetic properties of difficult L2 vowels while avoiding interference from pre-existing lexical representations, a focus on lexical meaning or the detrimental effect of noise.

To further investigate the effects of type of training stimuli on L2 phonetic learning, a follow-up of the present study could examine whether phonetic training effects on vowel production accuracy in words may vary as a function of the lexical properties of the test words. It is possible that words varying in lexical frequency, frequency of use or recency of use or acquisition might vary the extent to which their phono-lexical representations are prone to change due to phonetic training. Further avenues for future research in HVPT studies would include the role of cognitive individual differences (e.g. auditory attention control and

phonological short-term memory) and individual differences in auditory processing skills, the combination of a variety of training conditions within a single training paradigm (e.g. visual monitoring and audiovisual training combined with masking noise) or the integration of HVPT in classroom instructional settings through the development of training materials.

Acknowledgments: We are grateful to the participants that took part in the phonetic training sessions. We would like to thank Pace Bailey and Eva Cerviño for their help in data collection and acoustic analysis and Josh Frank for proofreading the manuscript. We would like to thank Phonetica's associate editor Patrick C. M. Wong and two anonymous reviewers for their very helpful and insightful comments and suggestions on a previous version of this manuscript.
Research funding: This study was supported by grant PID2019-107814GB-I00 (MCIN/AEI/10.13039/501100011033) from the Spanish Ministry of Science, Innovation and Universities.
Author contributions statements: The corresponding author Joan C. Mora and co-authors Mireia Ortega, Ingrid Mora-Plaza and Cristina Aliaga-García participated and are responsible for the conceptualization and design of the study, participant recruitment, data collection and acoustic analysis, and writing up of the manuscript. Joan C. Mora implemented the statistical analysis of the data.
Statement of ethics: The participants in the study gave their written informed consent to participate in the study. The study protocol adhered to the good practices of data collection, anonymization, processing, and storage of the Institutional Review Board of the University of Barcelona (IRB0003099).
Conflict of interest statement: The authors have no conflicts of interest to declare.

Appendix A: Training and testing items

1. Training items

Nonwords		Words	
/æ/	/ʌ/	/æ/	/ʌ/
/dæt/	/dʌt/	*cat*	*cut*
/fæʃ/	/fʌʃ/	*match*	*much*
/θæt/	/θʌt/	*mad*	*mud*
/tæz/	/tʌz/	*ban*	*bun*
/mæb/	/mʌb/	*cap*	*cup*
/tæm/	/tʌm/	*sack*	*suck*
/θæk/	/θʌk/	*bag*	*bug*
/ʧæŋ/	/ʧʌŋ/	*mag*	*mug*

2. Testing items

	DWR						DWR and DSR	
	Trained		Untrained					
Words	/æ/	/ʌ/	/æ/	/ʌ/	/æ/[1]	/ʌ/[1]	/æ/	/ʌ/
	cat	cut	hat	hut	gash	gush	map	pub
	mad	mud	back	buck	hatch	hutch	cash	gun
	cap	cup	fan	fun	mat	mutt	can	bus
	bag	bug	lack	luck	rag	rug	van	sun
	match	much	bad	bud	tab	tub	fat	fuss
	sack	suck	pan	pun	tag	tug	dad	dull
							jam	numb
							sad	nut
Nonwords	/fæʃ/	/fʌʃ/	/ʃæd/	/ʃʌd/				
	/θæt/	/θʌt/	/tædʒ/	/tʌdʒ/				
	/mæb/	/mʌb/	/θæp/	/θʌp/				
	/ʧæŋ/	/ʧʌŋ/	/gæb/	/gʌb/				
	/tæm/	/tʌm/	/sæz/	/sʌz/				
	/θæk/	/θʌk/	/væk/	/vʌk/				

[1]minimal-pair words included in the word learning task.

3. Sentences in the delayed sentence repetition task (Carlet 2017)

Target words with /æ/:
He looked at the map to find his way.
I have no cash in my pocket.
I don't want to open up that can of worms.
She bought a van to travel around the world.
He became so fat he could hardly walk.
She loves her dad more than anyone else.
She loves jam on her toast.
It's very sad that she didn't get the job.
Target words with /ʌ/:
We go to the pub on Saturdays.
Nobody saw the gun in his pocket.
I`ll get off the bus at the next stop.
He caught the sun at the beach.
Dont make such a fuss about it.
She finds it dull living in the country.
My fingers go numb in the cold.
To eat the nut first crack the shell.

Appendix B: Parameter estimates of linear mixed-effects models

Training effects in the DWR task (Section 6.1)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.572	0.126	12.457	<0.001	1.325	1.820
Group	−0.005	0.137	−0.039	0.969	−0.274	0.263
Testing Time	−0.127	0.062	−2.028	0.043	−0.250	−0.004
Vowel	−0.196	0.082	−2.379	0.017	−0.357	−0.034
Group × Testing Time	0.183	0.071	2.570	0.010	0.043	0.322
Group × Vowel	−0.072	0.071	−1.010	0.313	−0.211	0.068
Testing Time × Vowel	−0.006	0.088	−0.068	0.945	−0.108	0.167
Group × Testing Time × Vowel	0.021	0.101	0.206	0.836	−0.176	0.218

Training effects in the DSR task (Section 6.1)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.58	0.322	4.910	<0.001	0.949	2.211
Group	0.354	0.305	1.160	0.246	−0.244	0.952
Testing Time	−0.381	0.286	−1.334	0.182	−0.942	0.180
Vowel	0.199	0.343	0.578	0.563	−0.475	0.872
Group × Testing Time	0.528	0.298	1.769	0.077	−0.057	1.113
Group × Vowel	−0.508	0.298	−1.704	0.088	−1.094	0.077
Testing Time × Vowel	0.197	0.404	0.486	0.627	−0.597	0.990
Group × Testing Time × Vowel	−0.328	0.422	−0.776	0.438	−1.155	0.500

(c) Generalization effects for the NWD training group (Section 6.2)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.721	0.131	13.18	<0.001	1.465	1.977
Testing Time	0.051	0.063	0.816	0.415	−0.072	0.174
Vowel	−0.427	0.120	−3.549	<0.001	−0.662	−0.191
Trial Type 1 (Trained Nonwords)	−0.109	0.196	−0.555	0.579	−0.494	0.276
Trial Type 2 (Untrained Nonwords)	−0.055	0.196	−0.281	0.779	−0.440	0.330
Testing Time × Vowel	0.069	0.089	0.773	0.439	−0.105	0.243
Testing Time × Trial Type 1	0.086	0.145	0.590	0.555	−0.199	0.370
Testing Time × Trial Type 2	−0.040	0.145	−0.274	0.784	−0.324	0.245
Vowel × Trial Type 1	0.117	0.278	0.420	0.674	−0.428	0.661
Vowel × Trial Type 2	0.096	0.278	0.346	0.729	−0.448	0.640
Testing Time × Vowel × Trial Type 1	0.025	0.205	0.120	0.905	−0.378	0.427
Testing Time × Vowel × Trial Type 2	0.016	0.205	0.079	0.937	−0.386	0.418

(d) Generalization effects for the WD training group (Section 6.2)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.380	0.127	10.884	<0.001	1.132	1.629
Testing Time	0.056	0.079	0.711	0.477	−0.099	0.212
Vowel	−0.098	0.136	−0.725	0.468	−0.364	0.168
Trial Type 1 (Trained Words)	0.101	0.166	0.659	0.510	−0.216	0.435
Trial Type 2 (Untrained Words)	0.079	0.121	0.650	0.516	−0.159	0.317
Testing Time × Vowel	−0.072	0.112	−0.638	0.523	−0.291	0.148
Testing Time × Trial Type 1	0.003	0.137	0.022	0.982	−0.266	0.272
Testing Time × Trial Type 2	−0.007	0.100	−0.067	0.946	−0.203	0.190
Vowel × Trial Type 1	−0.036	0.235	−0.152	0.879	−0.496	0.425
Vowel × Trial Type 2	−0.072	0.172	−0.420	0.675	−0.409	0.264
Testing Time × Vowel × Trial Type 1	−0.096	0.194	−0.495	0.620	−0.477	0.285
Testing Time × Vowel × Trial Type 2	0.078	0.142	0.550	0.582	−0.200	0.356

(e) Training condition effects on words in isolation (DWR) (Section 6.3.1)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.428	34154	<0.001	1	−66950	66953
Group (NWD)	0.134	34154	<0.001	1	−66952	66952
Group (WD)	0.013	34154	<0.001	1	−66952	66952
Condition	−0.004	0.216	−0.019	0.985	−0.428	0.419
Testing Time	0.047	0.068	0.691	0.490	−0.087	0.181
Vowel	−0.143	0.092	−1.566	0.117	−0.323	0.036
Group × Condition	0.271	0.306	0.887	0.375	−0.328	0.870
Group × Testing Time	0.214	0.097	2.212	0.027	0.024	0.404
Group × Vowel	−0.151	0.097	−1.560	0.119	−0.341	0.039
Condition × Testing Time	0.012	0.097	0.125	0.900	−0.178	0.202
Condition × Vowel	0.003	0.097	0.032	0.975	−0.187	0.193
Testing Time × Vowel	−0.066	0.097	−0.682	0.495	−0.256	0.124
Group × Condition × Testing Time	−0.418	0.137	−3.052	0.002	−0.686	−0.149
Group × Condition × Vowel	−0.200	0.137	−1.464	0.143	−0.469	0.068
Group × Testing Time × Vowel	0.027	0.137	0.199	0.842	−0.241	0.295
Condition × Testing Time × Vowel	0.041	0.137	0.297	0.766	−0.228	0.309
Group × Condition × Testing Time × Vowel	0.187	0.194	0.967	0.334	−0.192	0.566

(f) Training condition effects on words in sentences (DSR) (Section 6.3.2)

	β	*SE*	*t*	*p*	*95% CI*	
					Lower	*Upper*
Intercept	1.742	0.226	7.697	<0.001	1.298	2.186
Group	−0.107	0.258	−0.414	0.679	−0.613	0.400
Condition	0.225	0.258	0.873	0.383	−0.281	0.732
Testing Time	−0.180	0.183	−0.985	0.325	−0.538	0.178
Vowel	−0.299	0.263	−1.137	0.256	−0.814	0.217
Group × Testing Time	1.162	0.258	4.505	0.000	0.656	1.669
Condition × Testing Time	0.596	0.258	2.309	0.021	0.090	1.102
Vowel × Testing Time	−0.042	0.258	−0.162	0.872	−0.548	0.464

(continued)

	β	SE	t	p	95% CI	
					Lower	Upper
Group × Condition × Testing Time (T1)	−1.410	0.365	−3.863	0.000	−2.126	−0.694
Group × Condition × Testing Time (T2)	0.359	0.365	0.983	0.326	−0.357	1.075
Group × Vowel × Testing Time (T1)	0.007	0.258	0.029	0.977	−0.499	0.514
Group × Vowel × Testing Time (T2)	0.003	0.258	0.012	0.990	−0.503	0.509
Condition × Vowel × Testing Time (T1)	−0.272	0.258	−1.056	0.291	−0.778	0.234
Condition × Vowel × Testing Time (T2)	−0.159	0.258	−0.618	0.537	−0.666	0.347
Group × Condition × Vowel × Testing Time (T1)	−0.116	0.365	−0.319	0.750	−0.832	0.599
Group × Condition × Vowel × Testing Time (T2)	0.072	0.365	0.196	0.845	−0.644	0.787

(g) Training effects on words in isolation and in sentences (Section 6.4)

	β	SE	t	p	95% CI	
					Lower	Upper
Intercept	1.895	88552	<0.001	1.000	−173625	173629
Task	0.036	0.108	0.333	0.739	−0.176	0.248
Group (NWD)	−0.309	88552	<0.001	1.000	−173627	173627
Group (WD)	−0.401	88552	<0.001	1.000	−173627	173626
Condition	0.343	0.149	2.309	0.021	0.052	0.635
Testing Time	0.265	0.108	2.452	0.014	0.053	0.477
Task × Testing Time	−0.258	0.153	−1.686	0.092	−0.557	0.042
Task × Group × Testing Time (T1)	−0.042	0.125	−0.333	0.739	−0.286	0.203
Task × Group × Testing Time (T2)	−0.123	0.125	−0.983	0.326	−0.367	0.122
Task (DSR) × Group × Testing Time	0.234	0.125	1.872	0.061	−0.011	0.478
Task × Condition × Testing Time (T1)	−0.399	0.125	−3.201	0.001	−0.644	−0.155
Task × Condition × Testing Time (T2)	−0.239	0.125	−1.918	0.055	−0.484	0.005
Task (DSR) × Condition × Testing Time	−0.393	0.125	−3.147	0.002	−0.637	−0.148

References

Aliaga-García, Cristina & Joan C. Mora. 2009. Assessing the effects of phonetic training on L2 sound perception and production. In Michael A. Watkins, Andreia S. Rauber & Barbara O. Baptista (eds.), *Recent research in second language phonetics/phonology: Perception and production*, 2–31. Newcastle upon Tyne, UK: Cambridge Scholars Publishing.

Amengual, Mark. 2016. The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics* 37. 1221–1251.

Antoniou, Mark, Eric Liang, Marc Ettlinger & Patrick C. M. Wong. 2015. The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition* 18(4). 683–695.

Baese-Berk, Melissa M. & Arthur G. Samuel. 2016. Listeners beware: Speech production may be bad for learning speech sounds. *Journal of Memory and Language* 89. 23–36.

Baker, Wendy & Pavel Trofimovich. 2005. Interaction of native- and second-language vowel system(s) in early and late bilinguals. *Language and Speech* 48(1). 1–27.

Barriuso, Anne & Rachel Hayes-Harb. 2018. High variability phonetic training as a bridge from research to practice. *The CATESOL Journal* 30. 177–194.

Best, Catherine & Michael D. Tyler. 2007. Nonnative and second-language speech perception: Commonalities and complementarities. In Murray J. Munro & Ocke-Schwen Bohn (eds.), *Second language speech learning: The role of language experience in speech perception and production*. Amsterdam: John Benjamins.

Boersma, Paul & David Weenink. 2020. Praat: doing phonetics by computer (Version 6.1.09) [Computer program]. http://www.praat.org/ (accessed 1 January 2020).

Bohn, Ocke-Schwen & James Emil Flege. 1990. Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics* 11. 303–328.

Bradlow, Ann R. & Tessa Bent. 2002. The clear speech effect for non-native listeners. *The Journal of the Acoustical Society of America* 112(1). 272–284.

Bradlow, Ann R. 2008. Training non-native language sound patterns: Lessons from training Japanese adults on the English /r/-/l/ contrast. In Jette Hansen Edwards & Mary L. Zampini (eds.), *Phonology and second language acquisition*, 287–308. John Benjamins Publishing Company.

Bradlow, Ann, Lynne Nygaard & David Pisoni. 1999. Effects of talker, rate, and amplitude variation on recognition memory for spoken words. *Perception & Psychophysics* 61(2). 206–219.

Broersma, Mirjam & Anne Cutler. 2008. Phantom word activation in L2. *System* 36(1). 22–34.

Broersma, Mirjam. 2012. Increased lexical activation and reduced competition in second-language listening. *Language and Cognitive Processes* 27(7–8). 1205–1224.

Bundgaard-Nielsen, Rikke L., Catherine T. Best & Michael D. Tyler. 2011. Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics* 32(1). 51–67.

Bundgaard-Nielsen, Rikke, Catherine Best, Christian Kroos & Michael Tyler. 2012. Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics* 33(3). 643–664.

Burk, Matthew H., Larry E. Humes, Nathan E. Amor & Lauren E. Strauser. 2006. Effect of training on word-recognition performance in noise for young normal-hearing and older hearing-impaired listeners. *Ear and Hearing* 27(3). 263–278.

Carlet, Angélica. 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*. Barcelona, Spain: Universitat Autònoma de Barcelona. (Unpublished PhD Thesis).

Carlet, Angelica & Juli Cebrian. 2019. Assessing the effect of perceptual training on L2 vowel identification, generalization and long-term effects. In Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen & Mette Hjortshøj Sørensen (eds.), *A sound approach to language matters – In Honor of Ocke-Schwen Bohn*, 91–119. Dept. of English, School of Communication and Culture, Aarhus University.

Cebrian, Juli & Angelica Carlet. 2014. Second language learners' identification of target language phonemes: A short-term phonetic training study. *Canadian Modern Language Review* 70(4). 474–499.

Cebrian, Juli, Joan C. Mora & Cristina Aliaga-García. 2011. Assessing crosslinguistic similarity by means of rated discrimination and perceptual assimilation tasks. In Magdalena Wrembel, Malgorzata Kul & Kararzyna Dziubalska-Koaczyk (eds.), *Achievements and perspectives in the acquisition of second language speech: New sounds 2010*, vol. I, 41–52. Frankfurt am Main: Peter Lang.

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

Cooke, Martin & Maria Luisa García-Lecumberri. 2012. The intelligibility of Lombard speech for non-native listeners. *The Journal of the Acoustical Society of America* 132(2). 1120–1129.

Cooke, Martin & Maria Luisa García-Lecumberri. 2018. Effects of exposure to noise during perceptual training of non-native language sounds. *The Journal of the Acoustical Society of America* 143. 2602–2610.

Cooke, Martin & Youyi Lu. 2010. Spectral and temporal changes to speech produced in the presence of energetic and informational maskers. *The Journal of the Acoustical Society of America* 128(4). 2059–2069.

Cutler, Anne, Andrea Weber & Takashi Otake. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34(2). 269–284.

Darcy, Isabelle & Jeffrey Holliday. 2019. Teaching an old word new tricks: Phonological updates in the L2 lexicon. In John Levis, Charles Nagle & Erin Todey (eds.), *Proceedings of the 10th pronunciation in second language learning and teaching conference*, ISSN 2380-9566, Ames, IA, September 2018, 10–26. Ames, IA: Iowa State University.

Darcy, Isabelle & Trisha Thomas. 2019. When blue is a disyllabic word: Perceptual epenthesis in the mental lexicon of second language learners. *Bilingualism: Language and Cognition* 22(5). 1141–1159.

Darcy, Isabelle, Danielle Daidone & Chisato Kojima. 2013. Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon* 8(3). 372–420.

Darcy, Isabelle, Laurent Dekydtspotter, Rex A. Sprouse, Justin Glover, Christiane Kaden, Michael McGuire & John H. Scott. 2012. Direct mapping of acoustics to phonology: On the lexical encoding of front rounded vowels in L1 English–L2 French acquisition. *Second Language Research* 28(1). 5–40.

Escudero, Paola, Rachel Hayes-Harb & Holger Mitterer. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36(2). 345–360.

Flege, James Emil & Ocke-Schwen Bohn. 2021. The revised Speech Learning Model (SLM-r). In Ratree Wayland (ed.), *Second language speech learning: Theoretical and empirical progress*, 3–83. Cambridge: Cambridge University Press.

Flege, James Emil. 1995. Two procedures for training a novel second language phonetic contrast. *Applied Psycholinguistics* 16. 425–442.

Forster, Kenneth I. & Jonathan C. Forster. 2003. DMDX: A Windows display program with millisecond accuracy. *Behavior Research Methods* 35(1). 116–124.

Gallardo del Puerto, Francisco, Maria Luisa García-Lecumberri & Jasone Cenoz. 2006. Age and native language influence on the perception of English vowels. In Barbara O. Baptista & Michael Alan Watkins (eds.), *English with a Latin beat: Studies in Portuguese/Spanish English interphonology*, 57–79. Amsterdam: John Benjamins.

García-Lecumberri, Maria Luisa & Maria del Pilar García-Mayo. 2003. English FL sounds in school learners of different ages. In Maria Luisa García-Lecumberri & Maria del Pilar García-Mayo (eds.), *Age and the acquisition of English as a foreign language*, 115–135. UK: Multilingual Matters.

Gomez Lacabex, Esther, Maria Luisa García-Lecumberri & Martin Cooke. 2008. Identification of the contrast full vowel-schwa: Training effects and generalization to a new perceptual context. *Ilha do Desterro: A Journal of English Language, Literatures in English and Cultural Studies* 55. 173–196.

Guion, Susan G. & Eric Pederson. 2007. Investigating the role of attention in phonetic learning. In Ocke-Schwen Bohn & Murray J. Munro (eds.), *Language experience in second language learning*, 57–77. Amsterdam: John Benjamins.

Hardison, Debrah M. 2018. Effects of contextual and visual cues on spoken language processing: Enhancing L2 perceptual salience through focused training. In Susan M. Gass, Patti Spinner & Jennifer Behney (eds.), *Salience in second language acquisition*, 201–220. New York: Routledge.

Hayes-Harb, Rachel & Kyoko Masuda. 2008. Development of the ability to lexically encode novel second language phonemic contrasts. *Second Language Research* 24(1). 5–33.

Hazan, Valerie & Rachel Baker. 2011. Acoustic-phonetic characteristics of speech produced with communicative intent to counter adverse listening conditions. *The Journal of the Acoustic Society of America* 130. 2139–2152.

Hazan, Valerie, Anke Sennema, Andrew Faulkner, Marta Ortega-Llebaria, Midori Iba & Hyunsong Chung. 2006. The use of visual cues in the perception of non-native consonant contrasts. *The Journal of the Acoustical Society of America* 119(3). 1740–1751.

Hazan, Valerie, Anke Sennema, Midori Iba & Andrew Faulkner. 2005. Effect of audiovisual perceptual training on the perception and production of consonants by Japanese learners of English. *Speech Communication* 47. 360–378.

Herd, Wendy, Allard Jongman & Joan Sereno. 2013. Perceptual and production training of intervocalic/d, r, r/in American English learners of Spanish. *The Journal of the Acoustical Society of America* 133(6). 4247–4255.

Hirata, Yukari. 2004. Computer assisted pronunciation training for native English speakers learning Japanese pitch and durational contrasts. *Computer Assisted Language Learning* 17(3–4). 357–376.

Ingvalson, Erin M., Allison M. Barr & Patrick C. M. Wong. 2013. Poorer phonetic perceivers show greater benefit in phonetic-phonological speech learning. *Journal of Speech, Language, and Hearing Research* 56. 1045–1050.

Iverson, Paul & Brownen G. Evans. 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.

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

Iverson, Paul, Valerie Hazan & Kerry Bannister. 2005. Phonetic training with acoustic cue manipulations: A comparison of methods for teaching English/r/-/l/to Japanese adults. *The Journal of the Acoustical Society of America* 118(5). 3267–3278.

John, Paul & Walcir Cardoso. 2017. Are word-final consonants codas? Evidence from Brazilian Portuguese ESL/EFL learners. In Jan Volin & Radek Skarnitzl (eds.), *Pronunciation of English*

by speakers of other languages, 117–138. Newcastle upon Tyne: Cambridge Scholars Publishing.

Kartushina, Natalia & Clara Martin. 2018. Talker and acoustic variability in learning to produce nonnative sounds: Evidence from articulatory training. *Language Learning* 69(1). 71–105.

Kartushina, Natalia, Alexis Hervais-Adelman, Ulrich Hans Frauenfelder & Narli Golestani. 2015. The effect of phonetic production training with visual feedback on the perception and production of foreign speech sounds. *The Journal of the Acoustical Society of America* 138(2). 817–832.

Kuhl, Patricia, Barbara Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola & Tobey Nelson. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society* B 363. 979–1000.

Lee, Junkyu, Juhyun Jang & Luke Plonsky. 2015. The effectiveness of second language pronunciation instruction: A meta-analysis. *Applied Linguistics* 36. 345–366.

Lengeris, Angelos. 2008. The effectiveness of auditory phonetic training on Greek native speakers' perception and production of Southern British English vowels. In *Proceedings of the 2nd ISCA workshop on experimental linguistics, ExLing 2008*, 25–27 August 2008, Athens, Greece, 133–136.

Leong, Christine Xiang Ru, Jessica M. Price, Nicola J. Pitchford & Walter J. B. van Heuven. 2018. High variability phonetic training in adaptive adverse conditions is rapid, effective, and sustained. *PLoS One* 13(10). e0204888.

Leung, Keith King, Allard Jongman, Yue Wang & Joan A. Sereno. 2016. Acoustic characteristics of clearly spoken English tense and lax vowels. *The Journal of the Acoustical Society of America* 140(1). 45–58.

Llompart, Miguel & Eva Reinisch. 2018. Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition* 22(5). 1085–1100.

Llompart, Miquel. 2021. Phonetic categorization ability and vocabulary size contribute to the encoding of difficult second-language phonological contrasts into the lexicon. *Bilingualism: Language and Cognition* 24(3). 481–496.

Lu, Shuang, Ratree Wayland & Edith Kaan. 2015. Effects of production training and perception training on lexical tone perception – A behavioral and ERP study. *Brain Research* 1624. 28–44.

Lu, Youyi & Martin Cooke. 2008. Speech production modifications produced by competing talkers, babble and stationary noise. *Journal of the Acoustical Society of America* 124. 3261–3275.

Mairano, Paolo & Fabian Santiago. 2020. What vocabulary size tells us about pronunciation skills: Issues in assessing L2 learners. *Journal of French Language Studies* 30(2). 141–160.

Mattys, Sven, Matthew Davis, Ann Bradlow & Sophie Scott. 2012. Speech recognition in adverse conditions: A review. *Language and Cognitive Processes* 27(7–8). 953–978.

Meara, Paul & Imma Miralpeix. 2006. *Y_Lex: The Swansea advanced vocabulary levels test. V2. 05*. Swansea, UK: Lognostics.

Meara, Paul & James Milton. 2003. *X_Lex.* Swansea, UK: Lognostics.

Melnik-Leroy, Gerda Ana & Sharon Peperkamp. 2021. High-variability phonetic training enhances second language lexical processing: Evidence from online training of French learners of English. *Bilingualism: Language and Cognition* 24(3). 497–506.

Melnik-Leroy, Gerda Ana, Rory Turnbull & Sharon Peperkamp. 2021. On the relationship between perception and production of L2 sounds: Evidence from Anglophones' processing of the French/u/–/y/contrast. *Second Language Research*. 1–25.

Mora, Joan C., James L. Keidel & James Emil Flege. 2015. Effects of Spanish use on the production of Catalan vowels by early Spanish-Catalan bilinguals. In Joaquín Romero & María Riera (eds.), *The phonetics-phonology interface: Representations and methodologies*, 33–53. Amsterdam: John Benjamins.

Munro, Murray & Tracey Derwing. 2006. The functional load principle in ESL pronunciation instruction: An exploratory study. *System* 34. 520–531.

Muñoz, Carmen. 2014. Contrasting effects of starting age and input on the oral performance of foreign language learners. *Applied Linguistics* 35(4). 463–482.

Nagle, Charles, L. 2021. Revisiting perception–production relationships: Exploring a new approach to investigate perception as a time-varying predictor. *Language Learning* 71(1). 243–279.

Ortega, Lourdes, Noriko Iwashita, John Norris & S Rabie. 2002. An investigation of elicited imitation tasks in crosslinguistic SLA research. In *Second Language Research Forum*, Toronto.

Perrachione, Tyler, Jiyeon Lee, Louisa Ha & Patrick Wong. 2011. Learning a novel phonological contrast depends on interactions between individual differences and training paradigm design. *The Journal of the Acoustical Society of America* 130(1). 461–472.

Pittman, Andrea & Terry Wiley. 2001. Recognition of speech produced in noise. *Journal of Speech, Language, and Hearing Research* 44. 487–496.

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

Ramus, Franck, Sharon Peperkamp, Anne Christophe, Charlotte Jacquemot, Sid Kouider & Emmanuel Dupoux. 2010. A psycholinguistic perspective on the acquisition of phonology. In C. Fougeron, B. Kühnert, M. d'Imperio & N. Vallée (eds.), *Laboratory phonology 10: Variation, phonetic detail and phonological representation*, 311–340. Berlin: Mouton de Gruyter.

Rato, Anabela & Andreia Rauber. 2015. The effects of perceptual training on the production of English vowel contrasts by Portuguese learners. In the Scottish Consortium for ICPhS 2015 (ed.), *Proceedings of the 18th international congress of phonetic sciences. Paper number 656*. Glasgow, UK: Glasgow University.

Sadakata, Makiko & James McQueen. 2013. High stimulus variability in nonnative speech learning supports formation of abstract categories: Evidence from Japanese geminates. *The Journal of the Acoustical Society of America* 134(2). 1324–1335.

Saito, Kazuya & Luke Plonsky. 2019. Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning* 69(3). 652–708.

Sakai, Mari & Coleen Moorman. 2018. Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics* 39(1). 187–224.

Sicola, Laura & Isabelle Darcy. 2015. Integrating pronunciation into the language classroom. In Marnie Reed & John Levis (eds.), *The handbook of English pronunciation*, 471–487. Hoboken, NJ: John Wiley and Sons.

Simonchyk, Ala & Isabelle Darcy. 2017. Lexical encoding and perception of palatalized consonants in L2 Russian. In Mary O'Brien & John Levis (eds.), *Proceedings of the 8th Pronunciation in 34*

Second Language Learning and Teaching Conference, 121–132. ISSN 2380-9566. Calgary, AB, August 2016. Ames, IA: Iowa State University.

Smiljanić, Rajka & Ann R. Bradlow. 2011. Bidirectional clear speech perception benefit for native and high proficiency non-native talkers and listeners: Intelligibility and accentedness. *The Journal of the Acoustical Society of America* 130(6). 4020–4031.

Suzukida, Yui & Kazuya Saito. 2021. Which segmental features matter for successful L2 comprehensibility? Revisiting and generalizing the pedagogical value of the functional load principle. *Language Teaching Research* 25(3). 431–450.

Syrdal, Ann & H. S. Gopal. 1986. A perceptual model of vowel recognition based on the auditory representation of American English vowels. *Journal of the Acoustical Society of America* 79. 1086–1100.

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

Thomson, Ron I. & Tracey Derwing. 2014. The effectiveness of L2 pronunciation instruction: A narrative review. *Applied Linguistics* 36(3). 326–344.

Thomson, Ron I. & Tracey Derwing. 2016. Is phonemic training using nonsense or real words more effective? In John Levis, Huong Le, Ivana Lucic, Evan Simpson & Sonca Vo (eds.), *Proceedings of the 7th Pronunciation in Second Language Learning and Teaching Conference*, 88–97. Ames, IA: Iowa State University.

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

Thorin, Jana, Makiko Sadakatac, Peter Desain & James M. McQueen. 2018. Perception and production in interaction during non-native speech category learning. *The Journal of the Acoustical Society of America* 144(1). 92–103.

Trofimovich, Pavel. 2008. What do second language listeners know about spoken words? Effects of experience and attention in spoken word processing. *Journal of Psycholinguistic Research* 37. 309–329.

Trofimovich, Pavel & Paul John. 2011. When three equals tree: Examining the nature of phonological entries in L2 lexicons of Quebec speakers of English. In Pavel Trofimovich & Kim McDonough (eds.), *Applying priming methods to L2 learning, teaching and research: Insights from psycholinguistics*, 105–129. Amsterdam: John Benjamins.

Tyler, Michael D. 2019. PAM-L2 and phonological category assimilation in the foreign language classroom. In Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anne Bothe Jespersen & Mette HjortshøjSørensen (eds.), *A Sound approach to language matters – In honor of Ocke-Schwen Bohn*, 607–630. Denmark: Dept. of English, School of Communication and Culture, Aarhus University.

Weber, Andrea & Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50(1). 1–25.

Werker, Janet F. & John S. Logan. 1985. Cross-language evidence for three factors in speech perception. *Perception & Psychophysics* 37(1). 35–44.

Wong, Janice. 2015. The impact of L2 proficiency on vowel training. In Jose A. Mompean & Jonás Fouz-González (eds.), *Investigating English pronunciation*, 219–239. London: Palgrave Macmillan.

Wong, Patrick C.M. & Tyler K. Perrachione. 2007. Learning pitch patterns in lexical identification by native English-speaking adults. *Applied Psycholinguistics* 28(4). 565–585.