Studies in Second Language Acquisition (2025), 1–18
doi:10.1017/S0272263124000548

RESEARCH REPORT

Effects of phonetic training and cognitive aptitude on the perception and production of non-native speech contrasts

Susana Correia[1], Anabela Rato[2], Yuxin Ge[1,3], João Dinis Fernandes[1], Magdalena Kachlicka[4], Kazuya Saito[5] and Patrick Rebuschat[3,6]

[1]NOVA University Lisbon/Linguistics Research Centre, Portugal; [2]University of Toronto, Canada; [3]Lancaster University, UK; [4]Birkbeck, University of London, UK; [5]University College London, UK and [6]University of Tübingen, Germany
Corresponding author: Patrick Rebuschat; Email: p.rebuschat@lancaster.ac.uk

(Received 19 January 2024; Revised 19 July 2024; Accepted 07 August 2024)

Abstract

Research on second language (L2) speech learning suggests that incidental perception training can lead to the establishment of non-native phonological categories. The present study contributes to this line of enquiry by investigating how this training is mediated by individual differences in working memory capacity and domain-general auditory processing abilities. In our study, 130 native British English speakers without prior knowledge of Portuguese were randomly assigned to trained or untrained conditions. All participants completed a visual digit span task and an auditory processing test battery. We observed improvements from pretest to post-test in production only, but since both groups improved, these gains cannot be attributed to the incidental perception training. The analysis of the ID measures further confirms the important role played by auditory processing abilities in L2 speech learning. However, more research is needed to better understand the role of incidental perception training and the mediating role of cognitive aptitudes.

Keywords: speech learning; cognitive aptitude; incidental phonetic training

Introduction

Research on non-native (L2) language learning in adults has demonstrated that incidental exposure is often enough, though not necessarily sufficient, to lead to successful language development (Williams & Rebuschat, 2023). This has been amply documented in vocabulary and grammar learning (e.g., Monaghan, Schoetensack & Rebuschat 2019; Rebuschat et al., 2021; Yu & Smith, 2007), but there is still too little research on the incidental learning of L2 speech (Saito et al., 2022). This is a major gap in the literature for at least two reasons. First, to further improve our theories and models of L2 speech learning, we require a better understanding of what aspects of spoken language can or

cannot be acquired incidentally, i.e., without the intention to learn. Second, to further improve the teaching of L2 speech, we need to know what aspects need to be taught more overtly (e.g., via explicit pedagogical treatments) and which do not. The present study directly addresses this gap by investigating the effect of incidental training on the perception and production of L2 Portuguese.

The study also contributes to the growing literature on the role of cognitive individual differences (ID) in L2 speech learning (Mora, 2022; Saito et al., 2022, for review). Previous research suggests that incidental perception training can lead to the establishment of non-native phonological categories (e.g., Baese-Berk, 2019; Black, Rato & Rafat, 2024; Lim & Holt, 2011; Saito et al., 2022; Vlahou, Protopapas & Seitz, 2012). Here, we contribute to this line of enquiry by investigating if this type of training is mediated by ID in working memory (WM) capacity and domain-general auditory processing abilities.

Perception training and incidental learning in non-native speech

Many studies have investigated the effect of different types of perception training on L2 speech learning (Sakai & Moorman, 2017). Most of these studies have investigated the effect of intentional perception training on L2 speech development. That is, in these studies, participants are typically instructed to learn non-native sounds, they are usually provided with feedback, and they are often aware that they will be tested after training. A meta-analysis has revealed moderate effects of perception training on L2 speech perception and only small-level effects of perception training on L2 speech production (Sakai & Moorman, 2017). Research further shows that the effectiveness of perception training does not depend just on the use of feedback but also on other factors, such as instruction and task types, phonetic variability in the stimuli, and length of training (Nagle & Baese-Berk, 2022; Zhang, Cheng & Zhang, 2021). However, recent results revealed equal effectiveness in auditory categorization across different training regimes with highly variable and multidimensional input, suggesting that, in fact, the outcomes of phonetic training are much less contingent on training regimes than previously thought (Obasih, Luthra, Dick & Holt, 2023).

Very few studies investigated the impact of incidental perception training. This type of training requires participants to engage in primary tasks (e.g., discriminating pseudowords) but without direct instruction or feedback on the specific phonological contrasts that differentiate these pseudowords. A survey of the literature indicates that incidental training tasks might be effective in building robust phonological categories. For example, Lim and Holt (2011) used an incidental perception task (in a videogame) to train native Japanese speakers on a non-native contrast (/ra/-/la/). Results showed improved categorization of /ra/-/la/ contrasts from pre- to post-test. More recently, the results from a paper inspired by Lim and Holt's paradigm revealed substantial ID in the outcomes of incidental L2 speech learning (Saito et al., 2022).

Baese-Berk (2019) compared the effectiveness of incidental perception and production training. Native speakers of English were trained with synthesized speech on a continuum of speech sounds that English does not use contrastively. Baese-Berk reported that participants showed substantial learning in the trained modality, i.e., those trained by means of perception training improved their perception and those trained by means of production task improved their production abilities. However, the perception training also significantly improved the production abilities, thus showing a cross-modal advantage for perception training only. Finally, participants in a bimodal condition, who received both perception and production training, only improved their production abilities.

More recently, Black and colleagues (2024) trained first language (L1) English learners of Spanish on three L2 contrasts ([b]-[β], [d]-[ð], and [g]-[ɣ]) with an AX

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

discrimination task without feedback and reported significant improvements for all three contrasts. Impressively, these results were obtained with a very short exposure session, consisting of six 2-min blocks in a single session.

Finally, Vlahou and colleagues (2012) directly compared incidental and intentional training tasks. They trained Greek native speakers either incidentally or intentionally to perceive a non-native contrast (Hindi /t/-/ʈ/). Both groups showed learning gains, but there was an advantage for the incidentally trained group. Interestingly, when compared to intentional training, the findings appear to be mixed and conflicting, especially when natural speech is used. For instance, Lim and Holt's (2011) study showed significant improvement when using synthesized stimuli, but the training gains become only marginally significant in the context of natural speech.

Individual cognitive differences in non-native speech learning

The role of ID in non-native speech learning has attracted much attention in recent years, with many studies focusing on cognitive aptitudes, such as attention, WM capacity, and domain-general auditory processing capabilities (see Mora, 2022; Saito et al., 2022, for a review). Studies that examined the role of WM capacity suggest that it can predict L2 speech learning and phonological processing abilities (e.g., Aliaga-García, Mora & Cerviño-Povedano, 2011; Darcy, Park & Yang, 2015), though the evidence is still limited (e.g., Saito et al., 2024). In recent years, several studies have investigated the role of auditory processing abilities in non-native speech learning, focusing particularly on the question of whether these abilities are specific to language or pertain to a general cognitive domain. Prior work failed to find any significant predictive power of domain-specific auditory processing abilities for the outcomes of L2 acquisition (Hughes, Golonka, Tseng & Campbell, 2023; Linck et al. 2013), but recent results suggest a moderate-to-strong predictive value of auditory processing in L2 speech learning (Saito, 2023; Kachlicka, Saito & Tierney, 2019). Importantly, even those with poor auditory processing abilities may do well in L2 speech learning if they make the most of their WM capacity. This hierarchical relationship between auditory processing, cognition (WM, executive function), and language learning has extensively been discussed in L1 but not in L2 research yet (Snowling, Gooch, McArthur & Hulme, 2018).

With this study, we intend to contribute to the ongoing debate about the predictive role of WM capacity and auditory processing abilities in non-native speech learning by focusing on incidental, rather than intentional, exposure conditions.

The present study

This study is part of a larger research project investigating the perception-production link and the role of cognitive factors in non-native speech learning. In this paper, we explore how short-term incidental perception training affects adult participants' ability to perceive and produce non-native speech. As highlighted above, by focusing on the effect of an incidental training regime, we directly contribute to a major gap in the L2 speech literature. Our participants are native speakers of British English without any background in the target language, Portuguese, i.e., they are naive participants. Very few studies investigated the effect of phonetic training on the perception and production of novel sounds (e.g., Baese-Berk, 2019) using ab initio learners, so comparatively little is known about the impact of phonetic training on the earliest stages of L2 speech learning. Finally, by focusing on the acquisition of L2 Portuguese, we contribute to the study of a

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

global but comparatively under-researched language. Portuguese has over 265 million native speakers and is the most widely spoken language in the southern hemisphere yet hardly features in international applied linguistics research (e.g., Plonsky, 2023). We hope that our work helps invert this trend.

The present study is set out to answer the following research questions: (a) Do perception and production abilities improve with short-term incidental perception training? (b) Is incidental training sufficient to enable phonological generalization? (d) Are cognitive factors positively correlated with perception or production gains?

We predicted that (a) a short-term incidental perception training program would result in learning gains in perception and/or production abilities; (b) the efficacy of the training would be observed in accuracy gains in perception and production abilities and in generalization to novel items; and (c) higher post-test scores would be positively correlated with WM capacity and auditory processing abilities.

Method

Participants

A total of 130 native speakers of English (62 female, 67 male, and 1 not disclosing gender) were randomly assigned to experimental (trained) and control (untrained) conditions (each $n = 65$). All adult participants (≥18 years old, with a mean age of 33.5 [standard deviation (SD) = 7.01]) were native speakers of British English and without prior experience of learning Portuguese or of having resided in a Portuguese-speaking country. A very small number of participants ($n = 12$) reported previous knowledge of typologically similar languages: Italian ($n = 1$), Spanish ($n = 4$), and/or French ($n = 9$), with nine in the experimental group and three in the control group. Two participants in the experimental group reported experience with both French and Spanish. A detailed summary of participants' language background can be found in our study's Open Science Framework (OSF) repository. Participants were recruited through Prolific and received GBP 9 per hour.

Sample size was estimated by means of Monte Carlo simulations of data (with expected power of .80). The R script for our power analysis is available in this study's Open Science Framework (OSF) here [https://osf.io/y3ufn]. The study was approved by the NOVA University Lisbon ethics review panel and conducted in accordance with the provisions of the World Medical Association Declaration of Helsinki. This study's preregistration [https://osf.io/gpu9w] can be found in the OSF registry.

Materials

The linguistic focus was on four Portuguese sound contrasts, namely two consonant contrasts, /l/-/ʎ/ (e.g., Portuguese *mala*, "suitcase", and *malha*, "mesh") and /n/-/ɲ/ (*mana*, "sister" informal, and *manha*, "ruse"), and two vowel contrasts, /e/-/ɛ/ (*sede*, "thirst", and *sede*, "head office") and /o/-/ɔ/ (*olho*, "eye", and *olho*, "I look"), which are deemed to be challenging for L1 English learners of L2 European Portuguese (EP) due to their perceived phonetic similarity (Macedo, 2015; Rato, 2019). These four contrasts were included in the perception and production pretests and post-tests, but only the contrasts /l/-/ʎ/ and /e/-/ɛ/ were the target of training. The selection of the training contrasts was based on previous findings suggesting that /l/-/ʎ/ and /e/-/ɛ/ would pose greater challenges to native English speakers than /n/-/ɲ/ and /o/-/ɔ/. For instance, an L1–L2 perceptual assimilation task (PAT) to examine cross-linguistic perceived

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

similarity between consonants in the English–Portuguese language dyad showed that the inexperienced group more consistently mapped the target L2 consonants /ʎ/ and /ɲ/ to the L1 categories /l/ (63%) and /n/ (75%) than the experienced learners (48% and 50%, respectively; Rato, 2019). Additional evidence from a rated dissimilarity task (RDT) showed that in the L1–L2 different pairs, inexperienced listeners rated perceived dissimilarity to the target contrasts /n/-/ɲ/ and /l/-/ʎ/ as 3.9 and 4.2. Beginner learners, however, show that /n/-/ɲ/ and /l/-/ʎ/ are rated as 4.4 and 4.0, respectively, for perceived dissimilarity. The results of the beginners group suggest that the contrast /l/-/ʎ/ is slightly more difficult to distinguish, as it is considered more similar than /n/-/ɲ/. A similar L1 English–L2 Portuguese PAT study that investigated the perception of EP vowels reports that the half-closed vowel /e/ is systematically mapped onto the half-open L1 Canadian English /ɛ/ (71%) and vowel /o/ is evenly assimilated to English /ɔ/ (38%) and /ʊ/ (39%) by beginner learners of Portuguese (Macedo, 2015). If we consider the fit indexes, EP vowel /e/ is considered a better fit of /ɛ/ (1.42) than vowel /o/ of /ʊ/ (0.48) and /ɔ/ (0.38), so one could predict that it would be slightly more difficult to create a new category for /e/ than for /o/. These findings seem to indicate that the target segments of this study are considered similar sounds to L1 phonemes, they may pose a challenge to native English speakers, and learning /l/-/ʎ/ and /e/-/ɛ/ contrasts, in particular, would potentially benefit from training. These contrasts further allowed us to test potential generalization in learning across contrasts /l/-/ʎ/ and /n/-/ɲ/, with coronal [+ant]/[-ant] features, and the vowel contrasts /e/-/ɛ/ and /o/-/ɔ/, involving the [-low]/[+low] features, respectively, similar to the feature generalization process found by Olson (2019).

Twelve consonants (/b, d, f, k, l, ʎ, m, n, ɲ, p, s, t/) and seven vowels (/a, e, ɛ, i, u, o, ɔ/) from the Portuguese phonemic inventory were combined to create 130 pseudowords. Each CVCV pseudoword was disyllabic and followed the phonotactics of Portuguese. The target sounds were either placed in the first syllable in the case of vowel targets (e.g., /de̲pu/ and /dɛ̲pu/) or in the onset the second syllable in the case of consonant targets (e.g., /pal̲u/ and /paʎ̲u/). Twenty-two pseudowords were only used as familiarization tokens. The stimuli used in the familiarization phase were different pseudowords that contrasted in non-target sounds. The remaining 108 pseudowords were used in the training tasks and/or tests (12 in the discrimination training task, 48 in the discrimination pre- and post-tests, and 48 in the production pre- and post-tests).

The stimuli were recorded by three native EP speakers (two female and one male) to add phonetic variability. The acoustic signal was recorded at 16 bits, with a sampling frequency of 44,100 Hz, using a Tascam DR-22WL digital recorder and an omnidirectional Monacor HSE-130/SK head-mounted microphone. The pseudowords were embedded in a carrier sentence (*Eu digo* [target] *com cuidado*, i.e., "I say [target] carefully"), which the speakers read aloud. We then extracted the target word productions and normalized the peak amplitude and duration of the stimuli.

A complete list of the pseudowords and their respective audio files can be found in the OSF repository.

Experimental tasks

Pretest and post-test

All participants completed a pretest and a post-test involving the completion of an oddity discrimination test to assess their ability to perceptually discriminate the four contrasts (/l/-/ʎ/, /n/-/ɲ/, /e/-/ɛ/, /o/-/ɔ/) and a delayed repetition test to examine their production. We selected oddity discrimination tasks, rather than AX tasks, as the

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

former capture more robust discrimination abilities between phonetically similar—but categorically distinct—segments, akin to the contrasts examined in our experiment (Nagle & Baese-Berk, 2022).

Oddity discrimination test

Participants were presented three pseudowords in each trial (e.g., /mepu/, /mɛpu/, /mepu/), with a 1-s interstimulus interval, and were asked to indicate which, if any, of the words was different. The order of the stimuli produced by the three speakers was counterbalanced across trials. Participants had to click on one of four options: "1", "2", "3", or "SAME". There were 96 trials in the pretest (12 per contrast) and 192 trials in the post-test (24 per contrast; 96 repeated from the pretest and 96 novel items). The test began with a short practice session to familiarize participants with the task.

Delayed repetition test

Participants were shown an unusual object from the NOUN Database (Horst & Hout, 2016), while they listened to a pseudoword (e.g., /setu/), followed by a delayed beep (2,000 ms after stimulus presentation), which prompted their production of the target pseudowords. There were 48 trials (12 per contrast) in the pretest and 96 trials (24 per contrast; 48 trials repeated from the pretest and 48 novel items) in the post-test. The test began with a short practice session with familiarization items. Participants' production was coded for accuracy (target-like = 1; non–target-like = 0) by three experienced, Portuguese-native phoneticians. Coding criteria were discussed before and during the coding, and each coder covered one third of the total data (18,720 pseudowords). Target-like productions were considered equivalent to Portuguese sounds, whereas non–target-like productions were considered deviant (e.g., produced with substitutions or diphthongization in the target segments—e.g., /setu/ produced as [sɛt^{h}ʊ] or [sɪɛt^{h}ʊ], respectively). Two transcribers coded 336 trials of all production data and Cronbach's alpha was 0.82, suggesting good inter-rater agreement.

Perception training task

Participants in the training condition were exposed to the pseudowords with the trained targets (/e/-/ɛ/ and /l/-/ʎ/) by means of an oddity discrimination task, and participants in the control condition received no training. Using the same type of task for training and testing allowed us to detect learning gains more readily.

The training consisted of two 5-min sessions run on two consecutive days. The first training session occurred immediately after the pretest and the second training task occurred immediately before post-test. The decision to opt for a short training session draws on prior research indicating that adults can rapidly acquire new sounds and new words with phonological contrasts. For example, Black and colleagues (2024) report significant improvement in discrimination accuracy across six 2-min blocks. Escudero, Mulak & Vlach (2016) and Ge, Monaghan & Rebuschat (2024b), using a cross-situational word learning paradigm, showed that adult participants could learn new minimally contrasting nonwords after very short exposure periods, 3 min in the case of Escudero and colleagues (2016) and 10 min in the case of Ge and colleagues (2024b).

The task was identical to the pretest but with different pseudowords and a focus on the trained targets. There were 48 trials (24 per trained contrast) in each of the two training sessions, the presentation sequence was randomized, and no feedback was provided.

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

In the training, participants received neither feedback nor information about the learning targets. Their training consisted of repeated exposure to the target contrasts by means of a task that directed their attention to between-category differences.

Individual differences measures

Working memory capacity. Participants' WM capacity was measured by means of a visual digit span task (adapted from Saito et al., 2024). In this task, participants were presented with written digit sequences and asked to reproduce them by entering their responses on the keyboard. Sequences had to be reproduced either in the order presented (forward digit span) or in the reverse order (backward digit span).

Auditory processing abilities. Participants' auditory processing abilities were measured by means of Saito and Tierney (2022) test battery. This includes tests designed to assess participants' ability to discriminate and reproduce spectral and temporal information. In our study, participants completed two discrimination tests (formant and amplitude rise time) and two reproduction tests (melody and rhythm). The discrimination tasks required participants to listen attentively to sequences of three sounds and to decide if the second sound was similar to the first or the third. In the latter, they were asked to repeat complex rhythmic or melodic patterns by using their computer keyboard.

Procedure

The Gorilla research platform (https://app.gorilla.sc/) was used to collect data. The experiment ran over three consecutive days. Precautionary measures to ensure the quality and engagement of the participants in the study included prescreening them on Prolific's approval rate: only participants with ≥95% approval rate in previous studies were selected. The attrition rate was 29% from day 1 to day 3.

On day 1, all participants completed an eligibility check and provided informed consent. They then completed the pretest tasks (oddity discrimination and delayed repetition), which took approximately 30 min. After, the experimental group completed one 5-min session of incidental perception training, which exposed them to the trained targets /e/-/ɛ/ and /l/-/ʎ/ via an oddity discrimination task. On day 2, the experimental group repeated the same 5-min training session, though with different randomizations. Immediately afterward, they completed the post-test tasks (oddity discrimination and delayed repetition). The control group only completed the post-tests. The post-tests included novel items to assess generalization, encompassing both trained and untrained contrasts, as outlined in Table 1.

We included the novel items to test for generalization to items not included in the pretest and from the trained contrasts (/l/-/ʎ/, /e/-/ɛ/) to untrained contrasts (/n/-/ɲ/, /o/-/ɔ/) that shared the same phonological features ([ant] in the consonants, and [low] in the vowels).

Experimental participants took approximately 40 min to complete one training session and the post-test. On day 3, all participants completed the aptitude tests, which

Table 1. Item structure in the post-test

Trained contrasts (/l/-/ʎ/, /e/-/ɛ/)		Untrained contrasts (/n/-/ɲ/, /o/-/ɔ/)	
Repeated items	Novel items	Repeated items	Novel items

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

took around 30 min. Aptitude measures were run after the post-test to optimize the participants' engagement in the study, although we acknowledge the potential effect of phonetic training in the auditory processing tasks.

Data analysis

We used linear mixed-effects modelling for data analysis. Models were constructed from the null models (containing only random effects of item and participant) to the models containing fixed effects of group (experimental vs. control group), test (pre- vs. post-test) and characteristics of test items (trained vs. untrained contrasts, repeated vs. novel pseudowords) as main effects. Analysis of variance tests on log-likelihood model fit were performed to determine if adding the fixed effects contributed significantly to explaining variance.

Due to the number of fixed effects examined, we ran several mixed-effects models to test the performance on the discrimination and production tasks. First, we compared the performance of the two groups on the oddity discrimination task in the pre- and post-tests, with fixed effects of group, test and group*test interaction. Then, we looked at the performance of the experimental group on trained (/l/-/ʎ/ and /e/-/ɛ/) versus untrained contrasts (/n/-/ɲ/, /o/-/ɔ/), as well as repeated and novel pseudowords in the pre- and post-tests, with fixed effects of trained versus untrained contrasts, repeated versus novel pseudowords, pretest versus post-test, and the three-way interaction. Finally, we tested if fixed effects of ID measures (forward and backward digit span, formant discrimination, rise time discrimination, melody reproduction, and rhythm reproduction) could explain variance in perception accuracy. We also ran a similar set of analyses for the production dataset. The analyses and model constructions are explained in our preregistration.

In addition to the preregistered analyses, we ran correlational analyses between the ID measures and post-test results to explore to what extent IDs could account for variation in perceptual discrimination and production performance.

Results

Performance on the perception training task

We transformed raw percentage accuracy in the oddity discrimination training to A-prime measures to account for potential response biases. The A-prime scores can range from –1 to 1, with 0 indicating chance level discrimination and 1 indicating perfect discrimination.

Table 2 summarizes performance on the training task. Trained participants (but not the controls) completed two sessions of this task on days 1 and 2. There were 48 exposure trials per session, thus 96 trials in total. Each target sound (/l/, /ʎ/, /e/, /ɛ/) occurred 36 times in each session and 72 times in total.

As Table 2 shows, participants were able to discriminate the trained target contrasts well. For the consonant contrast (/l/-/ʎ/), performance did not change significantly from the first to the second session (V = 4283, $p = .599$). For the vowel contrast (/e/-/ɛ/), there was a decrease in performance from the first to the second session (V = 5283, $p = .003$). Overall, the results suggest no improvement in discrimination performance throughout training.

The trained contrasts were considerably more challenging than the untrained contrasts, as shown in the pretest (Figure 1), providing additional evidence for the selection of these targets for instruction.

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

Table 2. Performance on the oddity discrimination training task (A-prime scores)

Session		Contrasts	
		/l/-/ʎ/	/e/-/ɛ/
Day 1	M	0.64	0.62
	SD	0.25	0.23
Day 2	M	0.63	0.57
	SD	0.25	0.23

Performance on the pre- and post-tests

Discrimination tests

We transformed raw percentage accuracy in the oddity discrimination tests to A-prime measures. Figure 1 presents the performance of the two groups in the oddity discrimination pretest and post-test. Overall, participants showed accurate perception of the /o/-/ɔ/ contrast even at pretest, and the perception of the /e/-/ɛ/ contrast was relatively low. The experimental group showed a clear increase in perceptual accuracy from the pretest to the post-test only for the untrained /n/-/ɲ/ contrast. For the other contrasts, there were small drops in accuracy. The control group did not show improvement in any of the contrasts. The descriptive statistics can be found in the supplementary material section.

As described above, the post-tests contained two types of items: novel and test (i.e., repeated) items. Figure 2 shows the performance of the two groups on the repeated and novel items in the post-test. For all contrasts, performance on repeated and novel items was comparable. The performance on novel /n/-/ɲ/ items was more accurate than that of the repeated /n/-/ɲ/ items, especially for the control group (repeated items: M = 0.76, SD = 0.24; novel items: M = 0.82, SD = 0.20).

We first ran linear mixed-effects models to explore whether the two groups performed differently in the pre- and post-tests. Compared to the random effect only model, adding the fixed effect of participant group (experimental vs. control) did not significantly improve model fit ($\chi^2[1] = 0.7374$, $p = .391$), indicating that there was no significant performance difference between the two groups. Adding the effect of pre- vs. post-test ($\chi^2[1] = 0.0374$, $p = .847$) and the group*test interaction did not improve fit ($\chi^2[2] = 0.6455$, $p = .724$). The best-fitting (random effect) model is summarized in Table 3.

To further investigate the effect of perception training, we ran a separate set of models on the experimental group only. We tested the effect of trained versus untrained contrasts, repeated versus novel items, pre- versus post-test, and the three-way interaction on perceptual discrimination performance. Only adding the effect of trained/untrained contrast significantly improved model fit ($\chi^2[1] = 8.86$, $p = .003$) but not the effect of repeated/novel items ($\chi^2[1] = 0.1362$, $p = .712$), test ($\chi^2[1] = 0.636$, $p = .425$), or the interaction ($\chi^2[2] = 3.5654$, $p = .168$). Participants performed overall better on the untrained contrasts compared to the trained contrasts. Table 4 summarizes the best-fitting model.

Performance on the production tests

We analyzed participants' performance on the non-native sounds only (i.e., /ʎ/, /ɲ/, /e/, /o/). As shown in Figure 3, the production accuracy was high for the /ɲ/ sound even in the pretest, whereas the /ʎ/ and /e/ sounds were relatively low. There is a general tendency of increased production accuracy for /ʎ/ (0.20 to 0.31), /ɲ/ (0.74 to 0.79), and

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

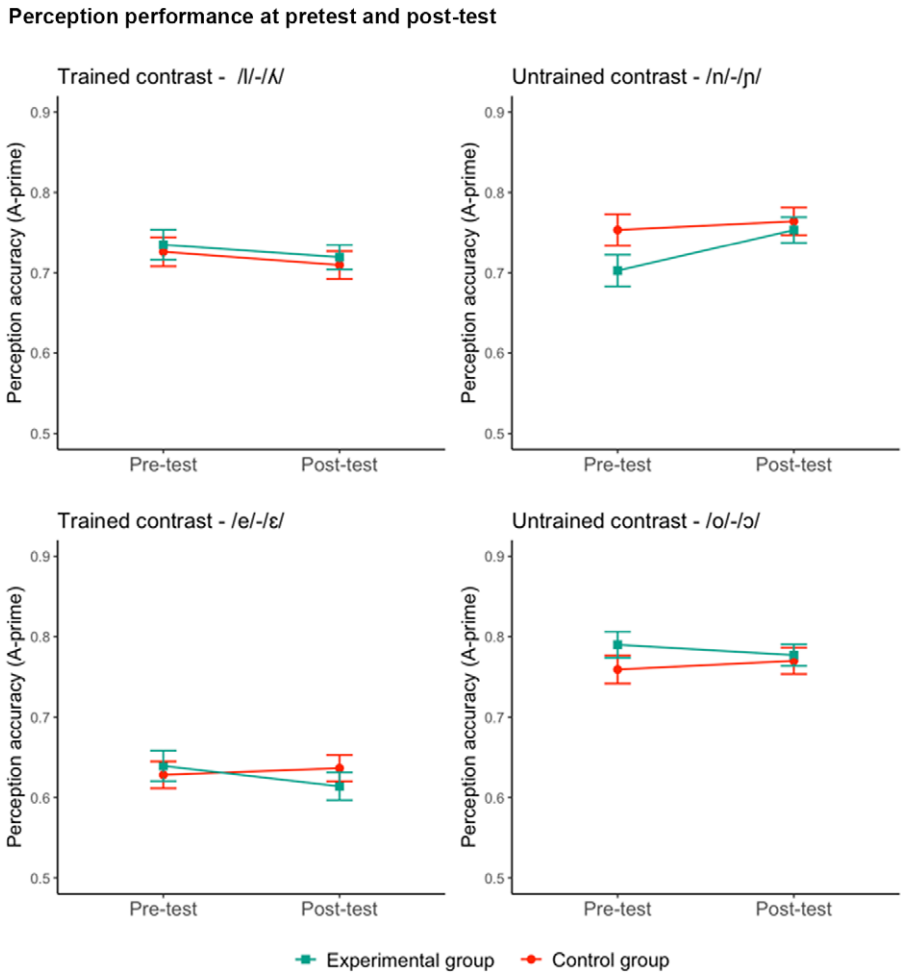

Figure 1. Performance on the trained and untrained contrasts in the oddity discrimination pretest and post-test (A-prime scores). Note: error bars represent one standard error.

/o/ (0.49 to 0.54) from pre- to post-tests across groups. The experimental and control groups performed similarly in production accuracy (full descriptives shown in the supplementary material).

Considering the novel items in the post-test, the production of novel /o/ items (0.63) was more accurate than that of repeated /o/ items (0.54). A similar pattern can be seen for the /e/ items (repeated: 0.25; novel: 0.28). For the other non-native sounds, the production accuracy of the repeated and novel items was similar (Figure 4).

We ran similar analyses on both the production and perception dataset, though with generalized linear mixed effect models as production data has a binomial distribution. Compared to the random effect model, adding the fixed effect of pre- versus post-test significantly improved model fit ($\chi^2[1] = 6.6881$, $p = .010$). This shows an overall improvement from pre- to post-test across groups. Adding the effect of experimental versus control group ($\chi^2[1] = 0.0479$, $p = .827$) and the participant group*test

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

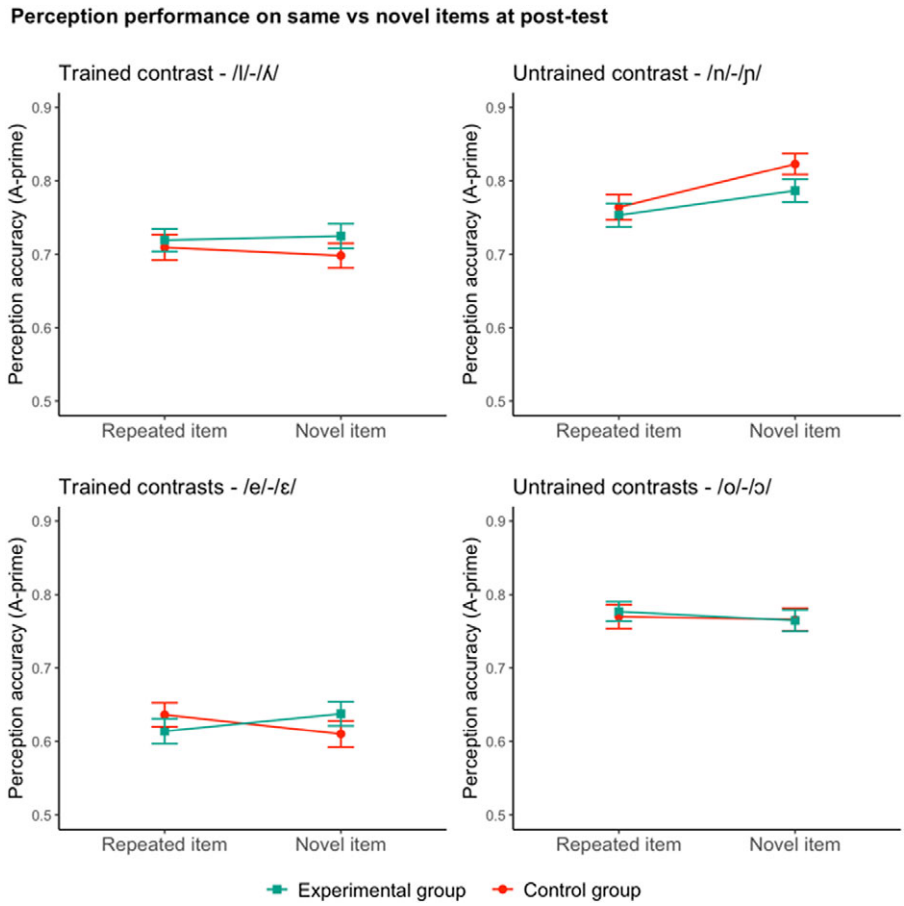

Figure 2. Performance on repeated and novel pseudowords in the oddity discrimination post-test.

Table 3. The best-fitting model for the effect of participant group (experimental/control) and test (pre/post) on discrimination

Fixed effects	Estimate	Standard error	t value	*p* value
(Intercept)	0.634	0.018	35.7	< .001***

Number of observations: 4,680; participants: 130; items, 24.
Akaike information criterion (AIC) = –661.8; Bayesian information criterion (BIC) = –623.1; log-likelihood = 336.9. *p < 0.05; **p < 0.01; ***p < 0.001.

Table 4. The best-fitting model for the effect of contrast (trained/untrained), item (repeated/novel), and test (pre/post) on perception in the experimental group

Fixed effects	Estimate	Standard error	t value	*p* value
(Intercept)	0.599	0.020	29.663	< .001***
contrastUntrained	0.087	0.027	3.202	.003**

Number of observations: 2,340; participants: 65; items, 24.
AIC = –406.3; BIC = –348.7; log-likelihood = 213.2. *p < 0.05; **p < 0.01; ***p < 0.001.

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

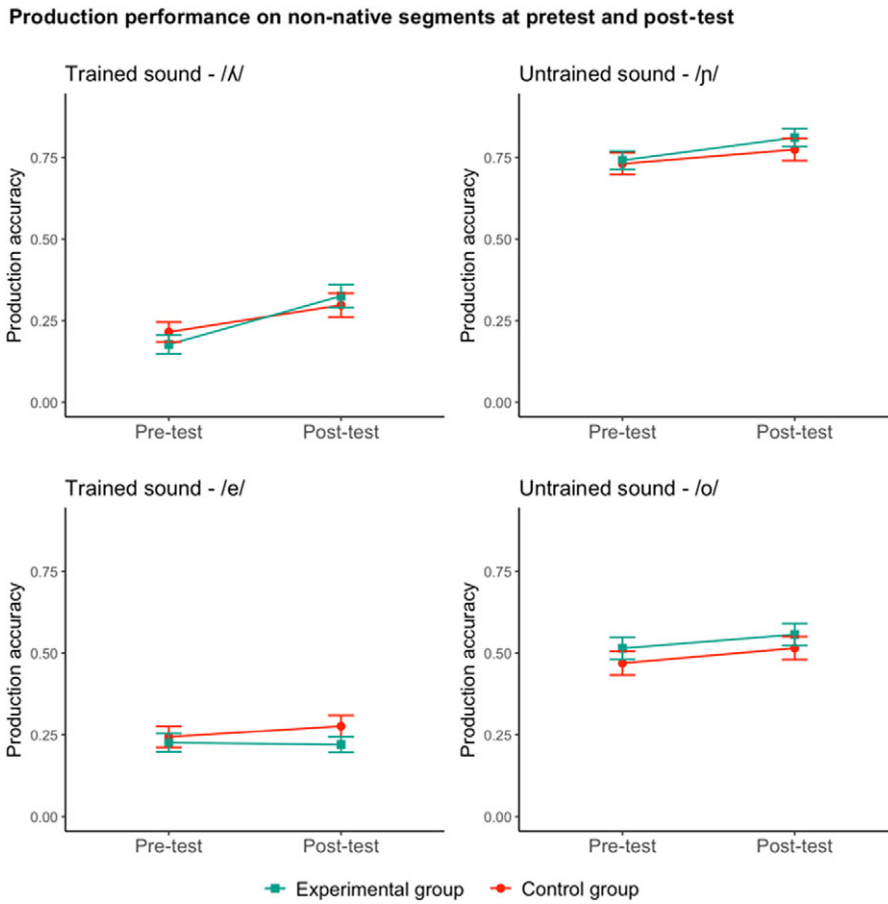

Figure 3. Performance on pseudowords with non-native target sounds in the delayed repetition pretest and post-tests (mean accuracy).

interaction did not improve fit ($\chi^2[1] = 1.9187$, $p = .166$). The best-fitting model is summarized in Table 5.

For the trained group, we explored the effect of trained versus untrained contrasts and repeated versus novel items. The effect of trained/untrained contrasts led to significant improvement in model fit ($\chi^2[1] = 21.034$, $p < .001$), with higher accuracy on the untrained sound. The effect of pre- versus post-test also improved fit ($\chi^2[1] = 10.354$, $p = .001$), indicating an improvement in performance from pre- to post-test. The effect of repeated/novel items ($\chi^2[1] = 0.0925$, $p = .761$) and the three-way interaction ($\chi^2[2] = 0.207$, $p = .902$) was not significant. The best-fitting model is shown in Table 6.

Individual differences measures

We tested if the ID measures contributed to explaining variance in participants' perception and production accuracy. We first ran linear mixed-effects models for perception accuracy (A-prime scores) with forward and backward digit span, formant

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

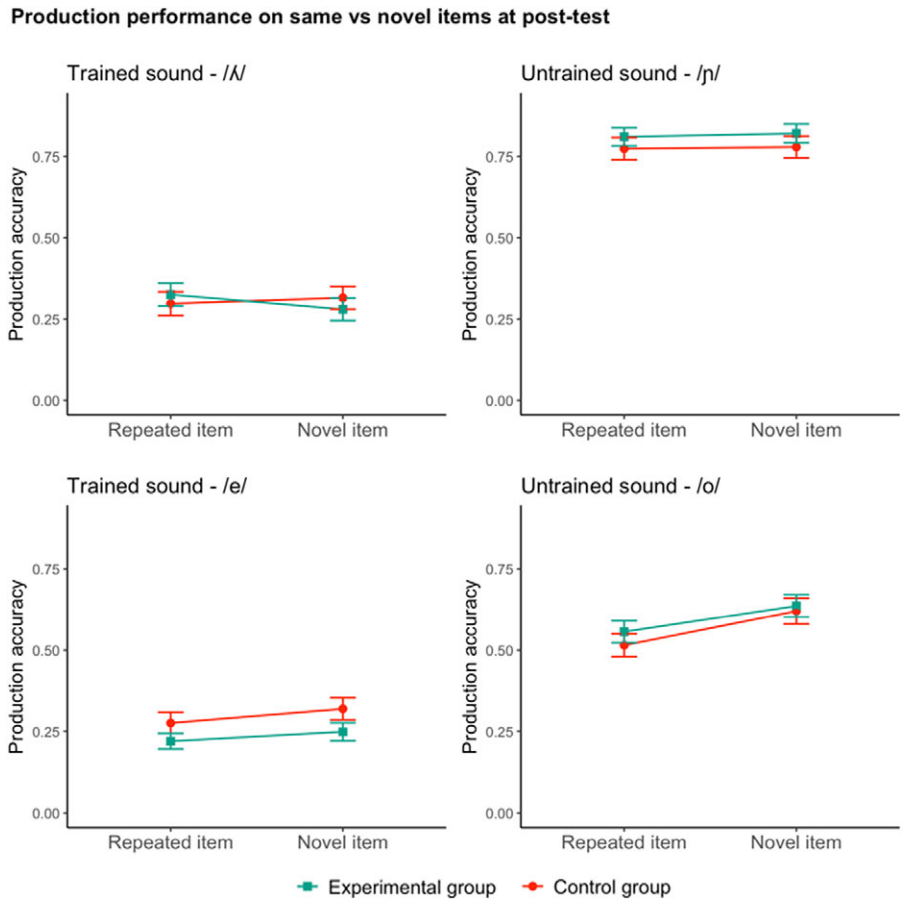

Figure 4. Performance on repeated and novel pseudowords in the delayed repetition post-test (mean accuracy).

Table 5. The best-fitting model for the effect of participant group (experimental/control) and test (pre/post) in the production test

Fixed effects	Estimate	Standard error	z value	*p* value
(Intercept)	–0.128	0.262	–0.487	0.626
Pretest	–0.316	0.105	–3.020	0.003**

Number of observations: 9,351; participants: 130; items, 24.
AIC = 9,943.7; BIC = 10,050.9; log-likelihood = –4,956.9. *p < 0.05; **p < 0.01; ***p < 0.001.

discrimination, rise time discrimination, melody reproduction and rhythm reproduction as fixed effects. We also added experimental and control groups as a fixed effect in the models to test whether there are group performance differences when IDs are considered. Only the fixed effect of forward digit span ($\chi^2[1] = 10.477$, $p = .001$) and formant discrimination ($\chi^2[1] = 8.1297$, $p = .004$) significantly improved model fit (best-fitting model summarized in Table 7). However, there was no significant effect of group or any interaction between group and ID measures.

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

Table 6. The best-fitting model for the effect of contrast (trained/untrained), item (repeated/novel), and test (pre/post) on production in the experimental group

Fixed effects	Estimate	Standard error	z value	*p* value
(Intercept)	–1.275	0.208	–6.118	< .001***
contrastUntrained	2.566	0.290	8.860	< .001***
Pretest	–0.546	0.139	–3.929	< .001***

Number of observations: 4,676; participants: 65; items, 24.
AIC = 4,931.8; BIC = 55,18.8; log-likelihood = –2374. *p < 0.05; **p < 0.01; ***p < 0.001.

Table 7. The best-fitting model for the effect of individual differences on performance in the oddity discrimination post-test

Fixed effects	Estimate	Standard error	t value	*p* value
(Intercept)	6.095e-01	4.142e-02	14.715	< .001***
groupExperimental	1.296e-02	1.806e-02	0.718	.474
Forward digit span	4.509e-03	1.760e-03	2.562	.012*
Formant	–1.562e-03	5.391e-04	–2.897	.004**

Number of observations: 3,120; participants: 130; items, 24.
AIC = –550.2; BIC = –507.9; log-likelihood = 282.1. *p < 0.05; **p < 0.01; ***p < 0.001.

Table 8. The best-fitting model for the effect of individual differences on performance in the delayed repetition post-test

Fixed effects	Estimate	Standard error	z value	*p* value
(Intercept)	1.061	0.358	2.960	.003**
groupExperimental	–0.906	0.335	–2.704	.007**
Formant	–0.008	0.006	–1.503	.133
Risetime	–0.029	0.006	–4.455	< .001***
groupExperimental:risetime	0.033	0.009	3.830	< .001***

Number of observations: 6,232; participants: 130; items, 24.
AIC = 6,679.8, BIC = 67,26.9, log-likelihood = –3,332.9. *p < 0.05; **p < 0.01; ***p < 0.001.

For the relationship between ID measures and production accuracy, we ran generalized linear mixed-effects models with the same set of fixed effects. The fixed effect of formant discrimination ($\chi^2[1] = 8.2692$, $p = .004$) and risetime discrimination ($\chi^2[1] = 5.5726$, $p = .018$) significantly improved model fit. There was also a significant interaction between group and risetime discrimination ($\chi^2[1] = 13.972$, $p < .001$), with a greater influence of risetime discrimination score on the control group. Table 8 shows the model summary.

Discussion

L2 speech research has seen conflicting results on the efficacy of phonetic training in perception and production. Weak-to-moderate effects of perception training have been observed, especially in production. Many studies investigating the effect of training in non-native speech learning highlight the superior effect of phonetic instruction and explicit feedback (Lee & Lyster, 2016; Sakai & Moorman, 2017), although evidence for the efficacy of incidental perceptual training has been also reported (Black et al., 2024; Lim & Holt, 2011; Vlahou et al., 2012).

In this report, we described the effects of an incidental perception training study as well as the role of individual cognitive differences in the development of non-native speech perception and production. We tested 130 native British English speakers with

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

Portuguese pseudowords that included four non-native target contrasts (/l/-/ʎ/, /e/-/ɛ/, /n/-/ɲ/, /o/-/ɔ/). The cohort was naive to Portuguese, with 65 participants undergoing a short incidental training without feedback and 65 not completing any training. Generalization of learning to new items with trained (/l/-/ʎ/, /e/-/ɛ/) and untrained (/n/-/ɲ/, /o/-/ɔ/) contrasts was also tested.

Our objectives were to investigate the impact of incidental perception training on both speech modalities and the predictive role of individual cognitive measures in non-native speech learning. In addition, we intended to test whether learning could be generalized to novel and untrained contrasts.

The results showed no perception gains between pre- and post-test, indicating, therefore, no effect of incidental perception training on the L1 English speakers' perception abilities and, thus, no learning. In production, post-test improvements can arguably be attributed to production practice, since the experimental (trained) group gains did not differ from those observed in the control (untrained) group.

The performance with novel items in the post-test did not differ significantly from that of the repeated items. It is worth noticing, however, that novel items did not show any detrimental learning effects, and their overall performance was similar to the one observed in the repeated items, which suggests that novel items did not negatively impact the participants' performance.

The results further suggest that learning non-native speech sounds is considerably moderated by the nature of perception training. A short-term (<3.5 h) incidental perception training seems to be insufficient to learn new speech sounds (Sakai & Moorman, 2017). In addition, task complexity and demands in the training (an oddity discrimination task with multiple speakers and no feedback) appear to hinder learning of non-native speech sounds. Crucially, results with native British English speakers learning the same contrasts in an oddity or AX discrimination training task with feedback showed significant post-test improvements with the AX task only (Ge et al., 2024a). Thus, the type of incidental training provided in the current study, without feedback, may have not been adequate for the target contrasts, particularly for the non-native segments /ʎ/, /ɲ/, /e/, and /o/, which may have been perceived as similar sounds to L1 English categories. Furthermore, training and testing the target contrasts with different tasks would have likely produced different—eventually more positive—results. This appears as a limitation of our study.

The results of our study also showed a difference between trained and untrained contrasts, with participants performing overall better on the latter (i.e., /n/-/ɲ/, and /o/-/ɔ/), even at baseline. These findings confirm the intrinsic phonetic similarity of the contrasts /l/-/ʎ/ and /e/-/ɛ/ and the increased discrimination difficulties.

Little is known about British English listeners' perception of the EP sound system. To date, only two studies have examined the perceived cross-linguistic phonetic similarity between EP and Canadian English (CE) sounds (Macedo, 2015; Rato, 2019). Macedo's (2015) findings on the perception of EP vowels by CE L2 learners predicted that the /e/-/ɛ/ contrast could be more difficult to acquire than the /o/-/ɔ/ pair as both front vowels are most frequently assimilated as a single native CE category (/ɛ/) (L2 /e/ > L1 /ɛ/: 71%; L2 /ɛ/ > L1 /ɛ/: 76%), whereas the vowels of the back contrast are most often categorized as two English vowels, /ʊ/ (40%) and /ɔ/ (62%), respectively. This could tentatively explain the higher accuracy observed in our study, from the onset of training, for the untrained contrast /o/-/ɔ/ than for the trained /e/-/ɛ/ pair.

As for the two consonant contrasts, the results of an RDT reported by Rato (2019) suggest that the L2 contrast more difficult to distinguish is /ʎ/-/l/ as both laterals are perceived to be more similar than the pair /ɲ/-/n/. However, the results

https://doi.org/10.1017/S0272263124000548 Published online by Cambridge University Press

of a cross-linguistic PAT suggest that in each pair, the novel consonants (/ʎ/ and /ɲ/) are equally categorized as their English alveolar counterparts /l/ (48%), and /n/ (50%) with predictions of equal discrimination accuracy for both pairs. Taking all these results into consideration, the prediction was that both consonant contrasts would present similar difficulty, with the nasal pair showing a slight advantage, which was the pattern observed in our study. However, we need to interpret these results with caution, as these predictions were based on a different L1 English variety (CE), and perception may be differential cross-dialectally (e.g., Escudero & Boersma, 2004).

As aforementioned, another possible limitation of the study may have been the type of perception training task. Logan and Pruitt (1995) suggest that identification tasks may be more effective in focusing learners' attention on within-category variability, thus, in establishing new phonetic categories, and in promoting generalization to novel stimuli not presented during training, than discrimination tasks. Carlet and Cebrian (2015) examined the effects of different training tasks and reported that an identification task promoted more improvement than an AX discrimination task in the perception of English vowels by Spanish/Catalan speakers. However, in our study, a tendency toward improvement with the untrained segments was observed for the learners who completed discrimination training only. In addition, the two 5-min training sessions seem to not have provided sufficient input for naive L1 English speakers to learn to perceive and produce two non-native phonemic contrasts.

The impact of the length and type of training task and complexity (e.g., AX vs. oddity or AXB discrimination, single vs. multiple talkers) in a sample of participants learning new sounds is, thus, a topic for further research.

The analysis on the individual cognitive measures showed that WM, namely, forward digit span, has a small-to-moderate positive correlation with perception abilities. Auditory processing also predicts perception abilities, with formant discrimination positively correlating with speech contrast discrimination. For production, the results showed a positive correlation with formant and risetime discrimination, and melody reproduction, suggesting that spectral and audio-motor integration abilities played a role in non-native speech production.

Despite no correlation with learning gains between pre- and post-test, cognitive factors correlated with better discrimination and production abilities. Therefore, our results suggest that learners with better auditory-motor skills also perform better at discriminating and producing non-native speech sounds and that domain-general auditory abilities seem to be working in non-native speech processing.

Being the first experiment of an ongoing research project, this study has provided us with much insight for future avenues. Considering the target sample of the project—native English speakers learning a non-native language, Portuguese—in subsequent empirical studies, a different approach will be recommended, with fewer learning targets, increased training, and a systematic comparison between incidental and intentional exposure conditions.

Supplementary material. The supplementary material for this article can be found at http://doi.org/10.1017/S0272263124000548.

Author note and acknowledgments. Our power analysis, materials, anonymized data, and R scripts are available on our project site on the OSF platform. Our preregistration can be accessed here: https://osf.io/gpu9w. We wish to thank Joan C. Mora and Ron Thomson for their expert guidance on the ProPerL2 project. Their advice was essential to the completion of the project. Additionally, we would like to thank the anonymous reviewers for their thoughtful evaluations of the manuscript. Their suggestions and insights have significantly improved the clarity and quality of this work.

We gratefully acknowledge the financial support provided by the Foundation for Science and Technology (grant reference [2022.04013.PTDC]), the Linguistics Research Centre of NOVA University Lisbon (UIDB/LIN/03213/2020 and UIDP/LIN/03213/2020 funding program), and Lancaster University's Camões Institute Cátedra for Multilingualism and Diversity.

Competing interest. The authors declare none.

References

Aliaga-García, C., Mora, J. C., & Cerviño-Povedano, E. (2011). L2 speech learning in adulthood and phonological short-term memory. *Poznań Studies in Contemporary Linguistics*, *47*.

Baese-Berk, M. (2019). Interactions between speech perception and production during learning of novel phonemic categories. *Attention, Perception, & Psychophysics*, *81*, 981–1005.

Black, M., Rato, A., & Rafat, Y. (2024). Effect of perceptual training without feedback on bilingual speech perception: Evidence from approximant-stop discrimination in L1 Spanish and L1 English late bilinguals. *Journal of Monolingual and Bilingual Speech*, *6*, 127–150.

Carlet, A., & Cebrian, J. (2015). Identification vs. discrimination training: Learning effects for trained and untrained sounds. *Proceedings of the 18th International Congress of Phonetic Sciences* (ICPhS) pp. 1–9.

Darcy, I., Park, H., & Yang, C.-L. (2015). Individual differences in L2 acquisition of English phonology: The relation between cognitive abilities and phonological processing. *Learning and Individual Differences*, *40*, 63–72.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, *26*, 551–585.

Escudero, P., Mulak, K., & Vlach, H. (2016). Cross-situational learning of minimal word pairs. *Cognitive Science*, *40*, 455–465.

Ge, Y., Correia, S., Fernandes, J. D., Hanson, K., Rato, A., & Rebuschat, P. (2024a). Does phonetic training benefit word learning? https://doi.org/10.31234/osf.io/5zspu [Preprint].

Ge, Y., Monaghan, P., & Rebuschat, P. (2024b). The role of phonology in non-native word learning: Evidence from cross-situational statistical learning. *Bilingualism: Language and Cognition*, 1–16.

Horst, J. S., & Hout, M. C. (2016). The Novel Object and Unusual Name (NOUN) database: A collection of novel images for use in experimental research. *Behavior Research Methods*, *48*, 1393–1409.

Hughes, M., Golonka, E., Tseng, A., & Campbell, S. (2023). The Hi-Level Language Aptitude Battery (Hi-LAB). Development, validation and use. In Z. Wen, P. Skehan, & R. Sparks (Eds.), *Language aptitude. theory and practice* (pp. 73–93). Cambridge University Press.

Kachlicka, M., Saito, K., & Tieney, A. (2019). Successful second language learning is tied to robust domain-general auditory processing and stable neural representations of sound. *Brain & Language*, *192*, 15–24.

Lee, A. H. & Lyster, R. (2016). The effects of corrective feedback on instructed L2 speech perception. *Studies in Second Language Acquisition*, *38*, 35–64.

Lim, S.-J., & Holt, L. (2011). Learning foreign sounds in an alien world: Videogame training improves non-native speech categorization. *Cognitive Science*, *35*, 1390–1405.

Linck, J., Hughes, M., Campbell, S., Silbert, N., Tare, M., Jackson, S., Smith, B., Bunting, M., & Doughty, C. (2013). Hi-LAB: A new measure of aptitude for high-level language proficiency. *Language Learning*, *63*, 530–655.

Logan, J., & Pruitt, J. (1995). Methodological issues in training listeners to perceive non-native phonemes. In W. Strange (Ed.), *speech perception and linguistic experience: issues in cross language research* (pp. 351–378). York Press.

Macedo, A. (2015). *Estudo da perceção de vogais e ditongos orais de alunos de PLNM, falantes de Inglês L1* [Unpublished master's thesis]. University of Minho.

Monaghan, P., Schoetensack, C., & Rebuschat, P. (2019). A single paradigm for implicit and statistical learning. *Topics in Cognitive Science*, *11*, 536–554.

Mora, J. C. (2022). Aptitude and individual differences. In Tracey M. Derwing, Murray J. Munro, & Ron I. Thomson (Eds.), *The Routledge handbook of second language acquisition and speaking* (pp. 68–82). Routledge.

Nagle, C., & Baese-Berk, M. M. (2022). Advancing the state of the art in l2 speech perception-production research: Revisiting theoretical assumptions and methodological practices. *Studies in Second Language Acquisition*, *44*, 580–605.

Obasih, C., Luthra, S., Dick, F., & Holt, L. (2023). Auditory category learning is robust across training regimes. *Cognition*, *237*, 105467.

Olson, D. (2019). Feature acquisition in second language phonetic development: Evidence from phonetic training. *Language Learning*, *69*, 366–404.

Plonsky, L. (2023). Sampling and generalizability in Lx research: A second-order synthesis. *Languages*, *8*, 75.

Rato, A. (2019). *The predictive role of cross-language phonetic similarity in L2 consonant learning* [Conference presentation]. International Symposium on the Acquisition of Second Language Speech – New Sounds 2019, Waseda University, Japan, August 30–September 1.

Rebuschat, P., Monaghan, P., & Schoetensack, C. (2021). Learning vocabulary and grammar from cross-situational statistics. *Cognition*, *206*.

Saito, K., Hanzawa, K., Petrova, K., Kachlicka, M., Suzukida, Y., & Tierney, A. (2022). Incidental and multimodal high variability phonetic training: potential, limits, and future directions. *Language Learning*, *72*: 1049–1091.

Saito, K. & Tierney, A. (2022). Domain-general auditory processing as a conceptual and measurement framework for second language speech learning aptitude: A test-retest reliability study. *Studies in Second Language Acquisition*, 1–25.

Saito, K., Kachlicka, M., Suzukida, Y., Mora-Plaza, I., Ruan, Y., & Tierney, A. (2024). Auditory processing as perceptual, cognitive, and motoric abilities underlying successful second language acquisition: Interaction model. *Journal of Experimental Psychology: Human Perception and Performance*, *50*, 119–138.

Saito, K. (2023). How does having a good ear promote successful second language speech acquisition in adulthood? Introducing Auditory Precision Hypothesis-L2. *Language Teaching*, *56*, 522–538.

Sakai, M., & Moorman, C. (2017). Can perception training improve the production of second language phonemes? A meta-analytic review of 25 years of perception training research. *Applied Psycholinguistics*, *39*, 187–224.

Snowling, M. J., Gooch, D., McArthur, G., & Hulme, C. (2018). Language skills, but not frequency discrimination, predict reading skills in children at risk of dyslexia. *Psychological Science*, *2*, 1270–1282.

Vlahou, E. L., Protopapas, A., & Seitz, A. R. (2012). Implicit training of nonnative speech stimuli. Journal of Experimental Psychology: *General*, *141*: 1–19.

Williams, J. N., & Rebuschat, P. (2023). Implicit learning and SLA: A cognitive psychology perspective. In A. Godfroid and H. Hopp (Eds.), *The Routledge Handbook of Second Language Acquisition and Psycholinguistics* (pp. 281–293). Routledge.

Yu, C., & Smith, L. B. (2007). Rapid word learning under uncertainty via cross-situational statistics. *Psychological Science*, *1*, 414–420.

Zhang, X., Cheng, B., & Zhang, Y. (2021). The role of talker variability in nonnative phonetic learning: A systematic review and meta-analysis. *Journal of Speech, Language, and Hearing Research*, *64*, 4802–4825.

Cite this article: Correia, S., Rato, A., Ge, Y., Fernandes, J. D., Kachlicka, M., Saito, K., & Rebuschat, P. (2025). Effects of phonetic training and cognitive aptitude on the perception and production of non-native speech contrasts. *Studies in Second Language Acquisition,* 1–18. https://doi.org/10.1017/S0272263124000548