Studies in Second Language Acquisition (2025), **47**, 1456–1468
doi:10.1017/S0272263125101113

RESEARCH REPORT

Revisiting the link between second-language sound identification and word recognition with an eye on methodological similarity

Miquel Llompart[1,2], Celia Gorba[3] and Pilar Prieto[1,4]

[1]Universitat Pompeu Fabra, Barcelona, Catalonia, Spain; [2]Friedrich Alexander University Erlangen-Nuremberg, Erlangen, Germany; [3]Universitat Autònoma de Barcelona, Barcelona, Catalonia, Spain and [4]Institució Catalana de Recerca i Estudis Avançats, Barcelona, Catalonia, Spain
Corresponding author: Miquel Llompart; Email: miguel.llompart@upf.edu

(Received 09 December 2024; Revised 13 June 2025; Accepted 25 June 2025)

Abstract

This study revisits the relationship between second-language (L2) learners' ability to distinguish sounds in non-native phonological contrasts and to recognize spoken words when recognition depends on these sounds, while addressing the role of methodological similarity. Bilingual Catalan/Spanish learners of English were tested on the identification of two vowel contrasts (VI) of diverging difficulty, /i/-/ɪ/ (difficult) and /ɛ/-/æ/ (easy), in monosyllabic minimal pairs, and on their recognition of the same pairs in a word-picture matching task (WPM). Learners performed substantially better with /i/-/ɪ/ in VI than in WPM, and individual scores were only weakly correlated. By replicating previous findings through a more symmetrical design, we show that an account of prior work rooted in methodological dissimilarity is improbable and provide additional support for the claim that accuracy in sound identification does not guarantee improvements in word recognition. This has implications for our understanding of L2-speech acquisition and L2 pronunciation training.

Keywords: L2 phonology; second language learning; sound identification; speech perception; spoken word recognition

Introduction

Learning a second language (L2) involves encountering sounds and phonological contrasts between sounds that are not part of the native inventory. For late learners, whose perceptual system is exquisitely attuned to the phonological categories that are relevant in the native language (L1), L2 contrasts often lead to perceptual difficulties. For example, native speakers of Spanish and Catalan struggle to distinguish the vowels /i/ and /ɪ/ when learning English. Spanish and Catalan only have one high-front vowel /i/

© The Author(s), 2025. Published by Cambridge University Press. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (http://creativecommons.org/licenses/by/4.0), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

and the two English vowels are perceived as being reasonably similar, although to different degrees (see more details below), to the native vowels (Cebrian, 2019; Cebrian et al., 2021; Rallo Fabra, 2005), which makes their differentiation rather challenging (e.g., Mora & Darcy, 2023). Crucially, to be able to use these L2 vowels in real communication, it is not enough that L1-Spanish/Catalan speakers/listeners learn to identify them as two different phonetic categories, as sounds are rarely produced in isolation. They also need to be able to reliably differentiate them when recognizing spoken words in the speech stream. That is, they must learn to distinguish words (and word meanings) when the distinction depends on the sounds in question (e.g., *sheep* vs. *ship*) and also to rely on these sounds to harness lexical access processes in a broader sense (e.g., activating words like *ship*, *shin* and *shift* but not *sheep* or *sheet* when the temporary input is [ʃɪ]).

In principle, one could assume that learners' identification of L2 sounds at a pre-lexical level, be it in isolation or as part of longer units, and their ability to recognize word form-meaning pairings when recognition is dependent on these sounds (as in the *sheep-ship* example above) are two sides of the same coin (see, for example, Best & Tyler's [2007] characterization of the assimilation patterns in their Perceptual Assimilation Model-L2), or at least, that the latter directly follows from the former. In fact, it is still often the case that the pre-lexical discrimination or identification of L2 sounds alone is taken as the main index of L2 phonological mastery in perception. Nevertheless, the growing literature on the topic suggests otherwise, as evidence has been accumulating that the link between L2 pre-lexical sound identification and word (i.e., lexical) recognition appears to be looser than expected (e.g., Amengual, 2016; Llompart, 2021; Simonchyk & Darcy, 2017). There are two main findings, discussed below, that speak to this issue.

In the first place, even when learners become quite accurate at identifying sounds in challenging L2 contrasts, their ability to rely on these sounds during spoken word recognition very often lags (Darcy et al., 2013; Díaz et al., 2012; Llompart, 2021). For instance, Díaz et al. (2012) assessed L1-Dutch learners' mastery of the English vowel contrast /ɛ/-/æ/ at varying levels of processing and tested a group of native English speakers to serve as a reference level. They found that around 44% of the L2 learners in their sample were able to perceptually categorize a continuum going from /ɛ/ to /æ/ with scores that were within the L1 range. By contrast, only about 13% of them were within the L1 range in a lexical decision task in which they were presented with real words with /ɛ/ and /æ/ and nonwords created by swapping the vowels in the contrast (e.g., *l[ɛ]mp, *d[æ]sk). In this task, learners were found to accept many nonwords as real words in spite of the vowel substitutions. Similar asymmetries have been found for L1-German learners with this same contrast (Llompart, 2021; Llompart & Reinisch, 2019, 2020), as well as for an array of other L2-learning populations and non-native phonological contrasts (Amengual, 2016; Darcy et al., 2013; Darcy & Holliday, 2019; Elvin, 2016; Sebastián-Gallés & Baus, 2005).

Secondly, the few studies that have looked into how tightly related performances in pre-lexical identification/discrimination are to accuracies in lexical recognition tasks that are dependent on it have rendered mixed results (see Table 1). To illustrate, Díaz et al. (2012) report a small negative correlation for the two tasks described above, whereas Darcy and Holliday (2019) obtained a much stronger and significant correlation between similar tasks for Mandarin Chinese learners of Korean and another vowel contrast, /o/-/ʌ/. However, the most common outcome so far lies somehow in the middle, as several studies have found that, while sound identification tends to predict word recognition on the whole (Llompart, 2021; Rocca et al., 2025), correlations are

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

Table 1. Previous studies assessing the relationship between sound identification and (spoken) word recognition in the context of difficult L2 contrasts. Details are provided about the L2 learning populations and the tasks that were used. Effect sizes for the relationship of interest are reported through the correlation coefficients that were obtained

Study	L1 of learners	L2 of learners (proficiency)	Phon. Contrast	Sound-level task	Word-level task	Correlation coefficient
Darcy & Holliday (2019)	Mandarin	Korean	/o/-/ʌ/	Identification	Lexical decision	.46
Díaz et al. (2012)	Dutch	English (adv.)	/ɛ/-/æ/	Identification	Lexical decision Word identification	–.05 .42
Elvin (2016)	English Spanish	Portuguese (beg.)	Various vowel contrasts	Discrimination	Novel word recognition	.28 .50
Llompart (2021)	German	English (int. and adv.)	/ɛ/-/æ/	Identification	Lexical decision	.39
Llompart & Reinisch (2020)	German	English (int.)	/ɛ/-/æ/	Identification	Novel word recognition	.41 (Exp. 1) .38 (Exp. 2)
Simonchyk & Darcy (2017)	English	Russian (int.) Russian (adv.)	Coronal consonants	Identification	Word-picture matching	.28 .66

often rather weak, thus pointing towards a link between the two abilities, albeit a relatively loose one.

This complicated picture raises the question as to why this is the case. It seems rather intuitive to think that one's success at distinguishing [i] from [ɪ] will largely determine how well words like *sheep* and *ship* will be recognized, or how easily nonwords like *ch[ɪ]se and *w[i]nter are going to be spotted. Nonetheless, there are at least two potential nonexclusive explanations for the mismatches observed. First, spoken word recognition involves the online matching of the continuous acoustic signal to stored representations not only of phonological forms but also of meanings, which need to be very quickly retrieved. This results in high cognitive and attentional demands, especially when doing so in a second language, and means that, when recognizing words, learners have fewer resources available for pre-lexical phonetic processing. This makes their bottom-up perception of individual sounds in words less reliable than when they are simply tasked with identifying or distinguishing these same sounds (Llompart, 2021; Pajak et al., 2016). In second place, because of the difficulties posed by new non-native sounds, especially in the earliest stages of learning, the phonological forms of the words stored in the lexicon may not include a distinction between the target sounds at all (i.e., the contrast is neutralized at the lexical level), or that distinction may not be robust enough (Amengual, 2016; Cook et al., 2016; Darcy et al., 2013). In other words, the encoding of the contrast into the lexicon may be fuzzy, and that may have hindering effects on the recognition of spoken words (see Darcy et al., 2025, for a very recent overview).

It is worth mentioning, though, that a perhaps less interesting explanation is also plausible given the data available, and this is that methodological differences across tasks have artificially widened the perceived gap between sound identification and (identification-dependent) word recognition. Previous studies vary substantially not only in the L2-learning populations tested and the paradigms used (see Table 1), but, crucially, they also display important differences between the tasks in terms of their demands, speech materials, and elicited target responses. For example, as described above, Díaz et al. (2012) assessed Dutch learners' perception of /ɛ/ and /æ/ by means of the identification of a synthesized continuum of isolated vowels, while the use that learners make of these vowels in spoken word recognition was investigated by means of a lexical decision task in which they were substituted by each other to create nonword stimuli. Llompart (2021), with L1-German learners, also used a similar categorization task, this time with the vowel embedded in a lexical minimal pair (*bet-bat*), and the same type of lexical decision task. When comparing the tasks, it is easy to notice that the stimuli are very different (many repetitions of isolated vowels/a monosyllabic minimal pair in identification vs. larger sets of naturally-produced non–minimal–pair words in lexical decision), and so are the responses to be produced. In the categorization task, there are two alternatives from a closed set of possible responses (i.e., the vowel is either /ɛ/ or /æ/), whereas the lexical decision task relies on yes-no answers to the question "is this a word (i.e., any word) in English?". Hence, all words in the L2 lexicon could potentially matter when providing an answer. This is also thought to make the task rather challenging, especially considering the additional demands that L2 processing imposes on lexical access (Broersma, 2005).

In the light of this, and adding to recent work highlighting the importance of reflecting on methodological choices when addressing foundational questions in L2-speech research (Llompart, 2024; Nagle & Baese-Berk, 2022; Saito & Plonsky, 2019), the present study revisits the relationship between non-native sound identification and identification-dependent L2 spoken word recognition while prioritizing methodological

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

similarity across tasks. We test bilingual Catalan/Spanish learners of English on the identification of the vowels of two English contrasts of diverging expected difficulty, /i/-/ɪ/ ("difficult") and /ɛ/-/æ/ ("easy") when embedded in monosyllabic minimal pairs (e.g., *peak-pick, bet-bat*), and on their recognition of these same minimal pairs in a word-picture matching task aimed at testing the form-meaning associations for them (e.g., *peak* with the upper part of a mountain). As mentioned above, prior work has shown that this learner population has difficulties distinguishing between /i/ and /ɪ/ in perception (e.g., Mora & Darcy, 2023). In contrast, /ɛ/ and /æ/ appear to be differentiated fairly easily (e.g., Aliaga-Garcia, 2017). These patterns are thought to stem from how these L2 vowels assimilate to the learners' L1 categories. Cross-linguistic assimilation tasks have shown that /ɛ/ and /æ/ are mapped onto distinct L1 categories: /ɛ/ is perceived as Catalan /ɛ/ or Spanish /e/ over 90% of the time, and /æ/ is assimilated to both Catalan and Spanish /a/ almost unequivocally (Cebrian, 2019; 2021; Rallo Fabra & Tyler, 2021), which explains their robust discrimination. On the other hand, /i/-/ɪ/ presents partially overlapping cross-linguistic assimilations, since /i/ is unequivocally mapped onto Spanish and Catalan /i/, but /ɪ/'s L1 categorization is split between /i/ (like the other English vowel) and /e/ (Cebrian, 2019; 2021; Cebrian et al., 2021; Rallo Fabra & Tyler, 2021). This overlap is therefore thought to be central to the difficulties in the acquisition of the contrast.

By using the same set of stimuli across tasks and a word recognition paradigm that prompts participants to consider particular form-meaning combinations (i.e., is the meaning of the word you heard X?)—just as the identification task requires them to make particular word-vowel associations—we aim to shed light on the extent to which prior findings of a somewhat loose relationship may have been due to the undue influence of methodological deviations that were not accounted for.

Method

Participants

A total of 97 participants (84 females; 12 males; 1 nonbinary) of between 18 and 30 years of age (*M*: 20; *SD* = 3.2) completed the experiment. They were first and second year undergraduate students majoring in linguistics or translation sciences at Universitat Pompeu Fabra and first year students majoring in English studies at Universitat Autònoma de Barcelona. Participants were L1 Catalan and Spanish speakers who had learned English as an L2 and had an upper-intermediate to advanced level (a B2 level according to the Common European Framework of Reference for Languages [CEFR] is required to access their programs of study). They had very limited knowledge of phonetics and phonology, as they had not yet received any training in this subject matter. Moreover, they had not lived in an English-speaking country for long periods of time (only visited). All participants gave their informed consent to participate. Ethics clearance was obtained from the Institutional Committee for Ethical Review of Projects of Universitat Pompeu Fabra (Application ID: 283).

Materials

This study reports on findings from a subset of the tasks used in a larger project aimed at assessing the effects of different types of perceptual training on the perception and production of L2 sounds (Gorba et al., in preparation). Specifically, here we limit our focus to two of the pretest tasks of that study: a vowel identification task with real-word stimuli (henceforth VI task), and a word-picture matching task (henceforth WPM

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

Table 2. Minimal-pair stimuli used in the VI and the WPM task. The target /i/-/ɪ/ and /ɛ/-/æ/ stimuli are presented in the first two columns

Minimal-pair stimuli by vowel contrast					
/i/-/ɪ/	/ɛ/-/æ/	/ɪ /-/ɛ/	/i/-/æ/	/i/-/ɛ/	/ɪ/-/æ/
seat - sit peak - pick cheap - chip	bet - bat beg - bag dead - dad	bill - bell lift - left lid - led	seed - sad heat - hat	read - red league - leg	sick - sack fit - fat

task). Likewise, as introduced above, here we are exclusively concerned with how participants perceive two English vowel contrasts, /i/-/ɪ/ ("difficult") and /ɛ/-/æ/ ("easy"), and how they recognize words based on their ability to distinguish the sounds in these contrasts.

The VI and WPM tasks used the same audiovisual stimuli[1]. These corresponded to 30 English words which were monosyllabic with CVC, CCVC, or CVCC structures and contained one of the four target vowels (/i, ɪ, ɛ, æ/) as the only vowel. Within the set of 30, all stimuli had a minimal pair (15 pairs total). Three minimal pairs contrasted /i/ and /ɪ/, and three contrasted /ɛ/ and /æ/. There were also three pairs for /ɪ/-/ɛ/, and 2 pairs for /i/-/æ/, /ɪ/-/æ/, and /iː/-/ɛ/, which can all be considered fillers for the purposes of this study. The list of stimuli can be found in Table 2. Two speakers of General American English (1 female, 1 male) were videorecorded producing the target words. They were instructed to read the target words, which were presented individually to them on a screen using a PowerPoint presentation. Each stimulus was presented twice, and if one of the items was not pronounced clearly, speakers were instructed to repeat it. Once the recordings had been obtained, they were checked for clarity and accuracy of production of the target vowel, and the items with the best quality were selected (i.e., one token per word per speaker).

The WPM task additionally made use of a set of 30 pictures portraying the visual referents of the target words. Unlicensed images were obtained through regular Google Images searches. Even though an effort had been made beforehand to make use of real English words that were picturable, some words were more easily picturable (e.g., *bell*, *bag*) than others (e.g., *led*, *fit*), and, similarly, some pairs may be more visually distinct by default (e.g., *bill-bell*) than others (e.g., *seat-sit*). Because of this, the authors entered an iterative process in which they rated the image for each word on a 1 (very poor match) to 5 (very good match) scale and selected new images for any items that scored below a 4 out of 5, only stopping once all items had images that reached that threshold. The set of images used in the WPM task is available at https://osf.io/awj4f (Open Science Framework), together with the dataset analyzed below and the code to reproduce all analyses reported.

Procedure

Participants completed the two tasks in a soundproof room in a linguistics laboratory at their university using a laptop and noise-cancelling headsets. The tasks were hosted and run online using the *Alchemer* presentation software (alchemer.com). The WPM task

[1]Audiovisual stimuli were used because one of the conditions of the larger training study involved presenting manual gestures and these could not be presented without video. Still, note that the talkers made no gestures in the stimuli presented in the two tasks in this study.

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

was completed first, immediately followed by the VI task[2]. Participants completed other perception and production tasks as part of the same pretest session of the larger training project mentioned above (i.e., Gorba et al., in preparation). Note that participants had not received any perceptual training when they were tested, as the pretest was designed to determine their baseline skills before any intervention.

Vowel identification task (VI)

The VI task was a four-alternative forced-choice (4AFC) vowel identification task. In each trial, participants saw a brief video clip of one of the two talkers producing one of the target words and had to select the vowel that was produced in the stimulus. The four response options, presented right below the videoclip, corresponded to the phonetic symbols of the four possible target phones (i.e., /i, ɪ, ɛ, æ/). Participants received a short training right before completing the VI task in which they were instructed on the sound–to–IPA–symbol correspondences. They were presented with the symbols one by one and, for each of them, they were provided with two auditory words that were not part of the task to listen to, plus their orthographic representations, as examples of that category. In the main task, each item was presented twice, once produced by each of the two speakers, for a total of 60 trials. The order of trials was randomized. The VI task took approximately 6 minutes to complete.

Word-picture matching task (WPM)

In the WPM task, participants were presented with the same audiovisual stimuli as in the VI task. However, on each trial, they first saw the videoclip of a talker producing a word and, immediately afterwards, a new screen was presented on which a picture appeared. They were then instructed to determine whether the word they heard matched the visual referent of the picture by clicking on one of two response options: YES and NO. Crucially, there were two types of trials, and these were determined by the relationship between the word and the picture presented. The picture could either match the word (e.g., *seat*-SEAT; Match condition) or depict the referent of its minimal pair (e.g., seat-SIT; Mismatch condition). The task had a total of 48 trials, half of them corresponding to the Match condition and the other half to the Mismatch condition. Importantly, all /ɛ/-/æ/ and /i/-/ɪ/ items (six each, see columns one and two in Table 2) were presented once in the Match condition and once in the Mismatch condition. The remaining 24 trials presented words in the other minimal pairs (i.e., columns 3 to 6 in Table 2), also keeping the number of Match and Mismatch configurations balanced. The order of trials was randomized. The WPM task took approximately 10 minutes to complete.

Data analysis and results

In our analyses, we focused on the accuracy with which participants identified the vowels of only the /ɛ/-/æ/ ("easy") and /i/-/ɪ/ ("difficult") items in the VI task, and on their accuracy with (auditory) word-picture associations in the WPM task for the same

[2]This order was chosen to avoid that participants already knew the vowel targets, which are made explicit in the VI task, when completing the WPM task.

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

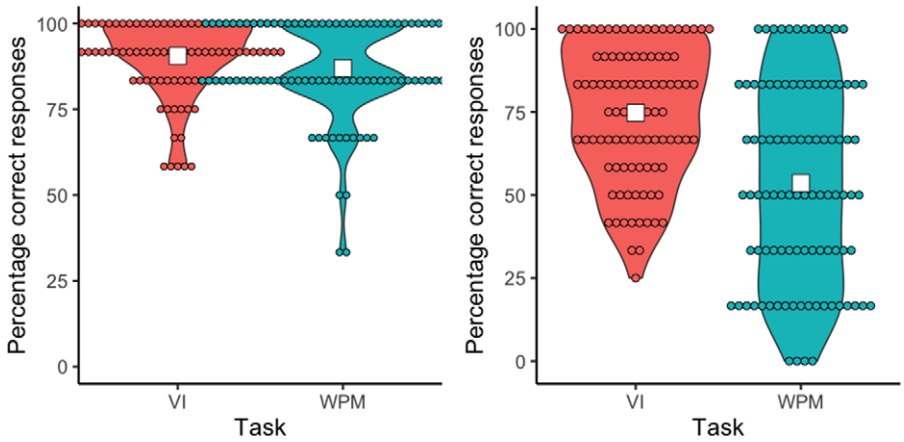


Figure 1. Percentage of correct responses for /ɛ/-/æ/ (left panel) and /i/-/ɪ/ (right panel) in the two tasks. Filled circles represent individual scores, and white squares signal the group means.

contrasts and in mismatch trials only. This is because this kind of task, when combining similar-sounding stimuli that vary slightly in their form, like the minimal pairs here, triggers a bias to answer YES whenever there is a reasonable match between the signal and the listeners' stored representation. Therefore, match trials usually result in ceiling or close–to–ceiling performances that do not necessarily reflect listeners' actual acuity. Mismatch trials, by contrast, even if potentially affected by the same bias, provide information that is easier to interpret, as they are a reflection of listeners' ability to identify a disconnect between the phonetic substance of the stimulus (*seat* (s[i]t)) and the stored phonological representation for the word in question (SIT (s/ɪ/t)). This same issue applies to lexical decision tasks containing phonological substitutions, and in many cases, similar approaches have been implemented whereby the focus has been restricted to the mismatching nonword stimuli (e.g., *dr[ɛ]gon in Llompart [2021], see also Amengual [2016], Darcy & Thomas [2019], and Llompart [2024]).

Figure 1 showcases L2 learners' percentage of correct responses for the /ɛ/-/æ/ ("easy"; left panel) and /i/-/ɪ/ contrasts ("difficult"; right panel) in the two tasks (VI to the left, WPM [mismatch trials only] to the right). Individual scores (filled circles) and group means (white squares) are provided. We observe that accuracy with /ɛ/-/æ/ stimuli was similarly high in the two tasks, with group means of 91% in the VI task and 87% in the WPM task. There is not much individual variation, and there are only three participants who are at or below chance, all three in the WPM task. This contrasts with the results for /i/-/ɪ/. For the difficult L2 contrast, there appears to be a large difference in mean accuracy between the two tasks (VI = 75%; WPM = 54%) and massive individual differences, especially for the WPM task, where the group mean is around chance level but we can find learners with perfect accuracies and others who did not respond correctly to any stimuli.

Building both on previous findings and the theoretical motivation behind this paper, in our analyses we aimed to quantify i) the extent to which there is a difference in accuracy between VI and WPM for /i/-/ɪ/ and /ɛ/-/æ/, ii) whether VI scores predict WPM accuracy for /i/-/ɪ/, and iii) the strength of the relationship between VI and WPM, also for /i/-/ɪ/, when assessed through individual scores.

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

First, to assess the effect of task and contrast on L2 learners' accuracy, data from the two tasks were submitted to a generalized linear mixed-effects regression model (GLMM; lme4 package 1.1–31, Bates et al., 2015) in R (Version 4. 2. 2) with a logit linking function and Response (0 = incorrect, 1 = correct) as categorical dependent variable, and Task (VI, WPM), Contrast (/ɛ/-/æ/, /i/-/ɪ/) and their interaction as predictors. The random-effects structure of the model included random intercepts by Participants and Items and random slopes for Task and Contrast over Participants and for Task over Items, as the items were the same in the two tasks. The model revealed significant effects of Task (b = -1.18; SE = .39; z = -3.02; $p < .01$) and Contrast (b = 1.46; SE = .41; z = 3.57; $p < .001$), as well as a significant interaction between the two (b = 1.61; SE = .57; z = 2.82; p = .01). Participants were overall more accurate in the VI than in the WPM task and with /ɛ/-/æ/ than with /i/-/ɪ/ stimuli. The interaction, together with the patterns outlined in Figure 1, suggests that the difference between the two tasks is larger for /i/-/ɪ/ than for the easier /ɛ/-/æ/ distinction. To follow up on this interaction, we split the dataset by contrast and examined the effect of Task for each of them. GLMMs with Response as dependent variable, Task as predictor, random intercepts by Participants and Items and random slopes for Task over both showed that the effect of Task was indeed significant for /i/-/ɪ/ (b = –1.10; SE = .42; z = –2.64; $p < .01$) but not for /ɛ/-/æ/ (b = .31; SE = .49; z = .63; p = .53). Hence, accuracy was significantly higher in the VI task than the WPM task for /i/-/ɪ/, but that was not the case for /ɛ/-/æ/.

Secondly, we assessed whether, for the difficult L2 contrast (/i/-/ɪ/), individual scores in the VI task predict accuracy in mismatch detection (i.e., *pick* is not the upper part of a mountain) in the mismatch trials of the WPM task. For this purpose, another GLMM was constructed, this time with Response as the dependent variable and individual VI scores (entered as proportions of correct responses) as the predictor. Random-effects included intercepts by Participants and Items. The model rendered a significant effect of VI scores (b = 3.01; SE = .88; z = 3.43; $p < .001$). L2 learners who scored higher in their identification of /i/ and /ɪ/ in the words of the VI task did indeed perform better when tasked with linking the same words to meanings in the WPM. Finally, a correlational analysis (Pearson's) of individual scores revealed a weak but significant correlation ($r(95)$ = .32, $p < .01$) between the two abilities, suggesting that the strength of this association in the present data is relatively low.

Discussion and conclusions

The present study probed the connection between L2 sound identification and word recognition when the latter depends on these particular sounds. Previous studies point towards a somewhat loose relationship between the two, with substantial differences in group-level performance and much individual variation that does not necessarily pattern in a similar way across the two abilities. As we elaborated in the introduction, while fitting theoretically-based accounts are in place, touching upon critical differences in cognitive demands in processing as well as in the representational levels involved in the two tasks, a more prosaic explanation rooted in methodological dissimilarity in the testing instruments could, to date, be entertained. In this study, we addressed this last possibility by drawing comparisons between tasks that were devised to be more similar to each other than has been the case in previous research[3].

[3]We acknowledge that, considering the methods used in prior work, different directions could have been taken when constructing the experimental tasks with the goal of increasing their comparability. For example, an anonymous reviewer suggested using a 4-AFC picture-identification task as the word-level task to better

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

We designed a vowel identification task (VI) and a word-picture matching task (WPM) that targeted a difficult English vowel contrast for L1-Spanish-Catalan learners (i.e., /i/-/ɪ/) and a second, less challenging contrast (i.e., /ɛ/-/æ/) that in effect worked as a baseline. Crucially, the tasks shared auditory stimuli and elicited similar closed-ended responses (see Methods).

The study provided three main findings of interest. First, as expected, we found better accuracies across the board for the easy than for the difficult L2 contrast. Learners were found to struggle with /i/-/ɪ/, particularly in the WPM task, while accuracy levels for /ɛ/-/æ/ words were very high and, importantly, not different between the VI and the WPM tasks (see Figure 1). This is relevant because learners' response to the easy contrast provides us with information about how similar or different the two tasks are in terms of task demands and overall difficulty in the absence of perceptual problems with the vowels. The pattern of results obtained can thus be taken as evidence that our attempt at levelling the playing field for sound and word identification by limiting the extent to which the lexical task was inherently more demanding was successful. While further testing is necessary, we consider that this speaks to the promise of this paradigm to tackle similar questions in the future.

In second place, for the difficult L2 contrast, /i/-/ɪ/, we found large task effects that are very much in line with those previously reported in the literature (e.g., Darcy et al., 2013; Díaz et al., 2012; Sebastián-Gallés & Baus, 2005). VI performance was well above chance and reasonably good overall, while learners as a group showed chance levels of accuracy in the WPM task. There was much individual variation in this task, and many more learners scored at or below chance than in VI (see Figure 1). Crucially, this asymmetry cannot be attributed to task demands independent from the critical sounds if we take the results for /ɛ/-/æ/ into consideration, nor to the use of different stimuli in the two tasks—as these were the same—or the fact that the responses elicited were vastly different, for the two tasks involved close-ended associations of two percepts (i.e., word-phonetic symbol and word-image, respectively). Hence, the results for /i/-/ɪ/ show that task effects remain even when the methodological alignment between tasks is prioritized in the experimental design.

Finally, in the third place, an examination of the relationship of interest across individuals rendered results that are again in line with prior findings outlining a loose link between L2 sound and word perceptual processing (Darcy & Holliday, 2019; Díaz et al., 2012; Llompart, 2021; Simonchyk & Darcy, 2017). Like in previous studies, we see that sound identification scores can predict identification-dependent word recognition to an extent (e.g., Llompart, 2021; Rocca et al., 2025), yet the relationship is relatively weak as far as patterns of individual variation are concerned. VI and WPM scores exhibited a small–to–medium–sized correlation of $r = .32$, which fits well within the range of values available in the literature (see Table 1) and is actually even numerically

match it to the present VI task in terms of response type and number of options. Although we understand the rationale for this, our goal here was to improve methodological alignment while ensuring that the tasks still tapped into the two constructs that we were interested in. Considering learners' metalinguistic knowledge at these levels of proficiency, a 4-AFC task in which pairs of targets and competitors (e.g., *peak-pick*) are regularly presented together has the risk of becoming an exercise in vowel identification in disguise instead of a task examining how reliably particular auditory forms are recognized as the intended lexical representation. In spite of this, we believe that the extent to which additional modifications to the sound- and word-recognition paradigms may further modulate the relationship to be observed remains an open and very interesting question to address in future work.

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

smaller than several of these in spite of our explicit goal of controlling for methodological differences.

In sum, the results of this study do very little to change the *status quo* concerning our understanding of the relationship between sound-level perceptual identification and the retrieval of form-meaning mappings when the sounds in question are involved. On the contrary, they can be taken to solidify the vision of it as variable, nuanced and far removed from a simple one–to–one correspondence. Our results (at least partly) dispel a purely methodological account of the asymmetries and variability that concern the relationship of interest and, as discussed in more depth in the introduction, point to the additional cognitive and attentional demands of lexical processing (Llompart, 2024; Pajak et al., 2016) and the involvement of the lexicon therein (see Darcy et al., 2025) as key contributing factors to the disconnects observed.

Even though this article may be considered to be mostly methodological, we believe that its findings can contribute relevant theoretical implications and follow-up questions for the study of the acquisition of L2 speech as a whole, as well as, more specifically, for work on L2 pronunciation training. Regarding the former, we provide additional evidence that phonological mastery of difficult L2 contrasts is not a monolithic, unidimensional trait that can be solely captured by sound-level identification/discrimination tasks by themselves. For this reason, learners' actual usage of the contrasts must be tested under varying processing conditions and embedded in units of different grain-sizes (i.e., on their own, in words, in words in sentences) to obtain a fuller picture of their phonological development. Relatedly, given that many pronunciation training paradigms target improvements in perception, the inconsistencies in learners' behavior in sound-focused vs. word-focused tasks also spotlight a central question for this line of work. Many perceptual training regimes (e.g., High Variability Phonetic Training) usually rely on identification and discrimination tasks with nonwords or reduced sets of words, and exhibit very little focus on meaning. In light of the recurring patterns that are instantiated once again in this study, and the limited evidence available that sound-level training transfers to L2 spoken word recognition (but see Melnik & Peperkamp, 2021), it becomes crucial to assess whether training paradigms with a stronger emphasis on lexical retrieval (e.g., a "feedbacked" picture-word matching task like the one used in this study) may actually be more effective at triggering generalized long-lasting improvements in L2 perception and pronunciation.

Data availability statement. The dataset analyzed in this article and the code to reproduce the analyses reported are available at https://osf.io/awj4f (Open Science Framework).

Acknowledgements. The authors would like to acknowledge the financial support awarded by the Generalitat de Catalunya (2021 SGR 00922) and the Spanish Ministry of Science and Innovation (PID2021-123823NB-I00). The first author's contribution was additionally supported by the Spanish Ministry of Universities (Beatriz Galindo Junior Research Fellowship, BG22/00161), as was the second author's (Margarita Salas Fellowship, ID-706211). We would also like to thank the editor and three anonymous reviewers for their helpful suggestions.

Competing interests. The authors declare none.

References

Aliaga-Garcia, C. (2017). *The effect of auditory and articulatory phonetic training on the perception and production of L2 vowels by Catalan-Spanish learners of English* [Doctoral dissertation, University of Barcelona].

Amengual, M. (2016). The perception of language-specific phonetic categories does not guarantee accurate phonological representations in the lexicon of early bilinguals. *Applied Psycholinguistics*, *37*(5), 1221–1251. https://doi.org/10.1017/S0142716415000557

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1). https://doi.org/10.18637/jss.v067.i01

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech perception: Commonalities and complementarities. In O.S. Bohn & M. J. Munro (Eds.), *Language experience in second language speech learning: In honor of James Emil Flege*, pp. 13–34. Amsterdam: John Benjamins. https://doi.org/10.1075/lllt.17.07bes

Broersma, M. (2005). *Phonetic and lexical processing in a second language* [Doctoral dissertation, Radboud University Nijmegen].

Cebrian, J. (2019). Perceptual asymmetries and lexical effects in L2 vowel discrimination. In *Proceedings of the 19th International Congress of Phonetic Sciences* (pp. 2590–2594). International Congress of Phonetic Sciences (19è: 2019: Melbourne).

Cebrian, J. (2021). Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *The Journal of the Acoustical Society of America*, *149*(4), 2671–2685. https://doi.org/10.1121/10.0004257

Cebrian, J., Gorba, C., & Gavaldà, N. (2021). When the easy becomes difficult: Factors affecting the acquisition of the English /iː/-/ɪ/ contrast. *Frontiers in Communication*, *6*. https://doi.org/10.3389/fcomm.2021.660917

Cook, S. V., Pandža, N. B., Lancaster, A. K., & Gor, K. (2016). Fuzzy nonnative phonolexical representations lead to fuzzy form-to-meaning mappings. *Frontiers in Psychology*, *7*. https://doi.org/10.3389/fpsyg.2016.01345

Darcy, I., Daidone, D., & Kojima, C. (2013). Asymmetric lexical access and fuzzy lexical representations in second language learners. *The Mental Lexicon*, *8*(3), 372–420. https://doi.org/10.1075/ml.8.3.06dar

Darcy, I. & Holliday, J. J. (2019). Teaching an old word new tricks: Phonological updates in the L2 mental lexicon. In J. Levis, C. Nagle, & E. Todey (Eds.), *Proceedings of the 10th Pronunciation in Second Language Learning and Teaching Conference* (pp. 10–26). Ames, IA: Iowa State University.

Darcy, I., Llompart, M., Hayes-Harb, R., Mora, J. C., Adrian, M., Cook, S., & Ernestus, M. (2025). Phonological processing and the L2 mental lexicon: Looking back and moving forward. *Studies in Second Language Acquisition. 47*(1), 361–387. https://doi.org/10.1017/S0272263124000482

Darcy, I., & Thomas, T. (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. https://doi.org/10.1017/S1366728918001050

Díaz, B., Mitterer, H., Broersma, M., & Sebastián-Gallés, N. (2012). Individual differences in late bilinguals' L2 phonological processes: From acoustic-phonetic analysis to lexical access. *Learning and Individual Differences*, *22*(6), 680–689. https://doi.org/10.1016/j.lindif.2012.05.005

Elvin, J. D. (2016). *The role of the native language in non-native perception and spoken word recognition: English vs. Spanish learners of Portuguese* [Doctoral dissertation, Western Sydney University].

Gorba, C., Llompart, M., & Prieto, P. (in preparation). Are non-words really superior to real words to train L2 sounds?

Llompart, M. (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. https://doi.org/10.1017/S1366728920000656

Llompart, M. (2024). On the effects of task focus and processing level on the perception–production link in second-language speech learning. *Studies in Second Language Acquisition*, *46*(1), 214–226. https://doi.org/10.1017/S0272263123000414

Llompart, M., & Reinisch, E. (2019). Robustness of phonolexical representations relates to phonetic flexibility for difficult second language sound contrasts. *Bilingualism: Language and Cognition*, *22*(5), 1085–1100. https://doi.org/10.1017/S1366728918000925

Llompart, M., & Reinisch, E. (2020). The phonological form of lexical items modulates the encoding of challenging second-language sound contrasts. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *46*(8), 1590–1610. https://doi.org/10.1037/xlm0000832

Melnik, G. A., & Peperkamp, S. (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. https://doi.org/10.1017/S1366728920000644

Mora, J. C., & Darcy, I. (2023). Individual differences in attention control and the processing of phonological contrasts in a second language. *Phonetica*, *80*(3–4), 153–184. https://doi.org/10.1515/phon-2022-0020

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press

Nagle, C. L., & 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*(2), 580–605. https://doi.org/10.1017/S0272263121000371

Pajak, B., Creel, S. C., & Levy, R. (2016). Difficulty in learning similar-sounding words: A developmental stage or a general property of learning? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *42*(9), 1377–1399. https://doi.org/10.1037/xlm0000247

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

Rallo Fabra, L., Tyler, M.D. (2021) Discrimination of Californian central vowel contrasts by Spanish-Catalan EFL learners. *Proc. 3rd International Symposium on Applied Phonetics (ISAPh 2021)*, 59–62, doi: 10.21437/ISAPh.2021-10

Rocca, B., Llompart, M., & Darcy, I. (2025). Phonological neighborhood density, phonetic categorization, and vocabulary size differentially affect the phonolexical encoding of easy and difficult L2 segmental contrasts. *Bilingualism: Language & Cognition*, *28*(3), 662–675. https://doi.org/10.1017/S1366728924000865

Saito, K., & Plonsky, L. (2019). Effects of second language pronunciation teaching revisited: A proposed measurement framework and meta-analysis. *Language Learning*, *69*(3), 652–708. https://doi.org/10.1111/lang.12345

Sebastián-Gallés, N., & Baus, C. (2005). On the relationship between perception and production in L2 categories. In A. Cutler (Ed.) *Twenty-first century psycholinguistics: Four cornerstones* (pp. 279–292). Hillsdale: LEA.

Simonchyk, A., & Darcy, I. (2017). Lexical encoding and perception of palatalized consonants in L2 Russian. In M. O'Brien & J. Levis (Eds.), *Proceedings of the 8th Pronunciation in Second Language Learning and Teaching Conference* (pp. 121–132). Ames, IA: Iowa State University.

Cite this article: Llompart, M., Gorba, C., & Prieto, P. (2025). Revisiting the link between second-language sound identification and word recognition with an eye on methodological similarity. *Studies in Second Language Acquisition,* 47: 1456–1468. https://doi.org/10.1017/S0272263125101113

https://doi.org/10.1017/S0272263125101113 Published online by Cambridge University Press