



# When the Easy Becomes Difficult: Factors Affecting the Acquisition of the English /i:/-/ɪ/ Contrast

Juli Cebrian<sup>1\*</sup>, Celia Gorba<sup>1</sup> and Núria Gavalda<sup>2</sup>

<sup>1</sup>Departament de Filologia Anglesa i Germanística, Facultat de Filosofia i Lletres, Universitat Autònoma de Barcelona, Barcelona, Spain, <sup>2</sup>Facultad de Educación, Universidad Internacional de la Rioja, Logroño, Spain

The degree of similarity between the sounds of a speaker's first and second language (L1 and L2) is believed to determine the likelihood of accurate perception and production of the L2 sounds. This paper explores the relationship between cross-linguistic similarity and the perception and production of a subset of English vowels, including the highly productive /i:/-/ɪ/ contrast (as in “beat” vs. “bit”), by a group of Spanish/Catalan native speakers learning English as an L2. The learners' ability to identify, discriminate and produce the English vowels accurately was contrasted with their cross-linguistic perceived similarity judgements. The results showed that L2 perception and production accuracy was not always predicted from patterns of cross-language similarity, particularly regarding the difficulty distinguishing /i:/ and /ɪ/. Possible explanations may involve the way the L2 /i:/ and /ɪ/ categories interact, the effect of non-native acoustic cue reliance, and the roles of orthography and language instruction.

**Keywords:** vowel contrast, L2 perception, L2 production, cross-linguistic similarity, individual variation

## OPEN ACCESS

### Edited by:

John Archibald,  
University of Victoria, Canada

### Reviewed by:

Phillip J. Monahan,  
University of Toronto, Canada  
Anabela Rato,  
University of Toronto, Canada

### \*Correspondence:

Juli Cebrian  
juli.cebrian@uab.cat

### Specialty section:

This article was submitted to  
Language Sciences,  
a section of the journal  
Frontiers in Communication

**Received:** 29 January 2021

**Accepted:** 04 May 2021

**Published:** 10 June 2021

### Citation:

Cebrian J, Gorba C and Gavalda N  
(2021) When the Easy Becomes  
Difficult: Factors Affecting the  
Acquisition of the English /i:/-/ɪ/  
Contrast.  
Front. Commun. 6:660917.  
doi: 10.3389/fcomm.2021.660917

## INTRODUCTION

Learning a second or foreign language (L2) after childhood is a difficult task for a variety of reasons. In addition to learner-related factors such as age of learning and amount of first and second language use (Piske et al., 2001; Flege et al., 2003, among others), the existence of a first language (L1) phonological system in part accounts for the fact that adult L2 speakers rarely sound like native speakers (Trubetzkoy, 1939; Rochet, 1995; Strange, 1995, among others). Variability in L2 performance has been linked to the degree of similarity between first and second language sound systems (Lado, 1957; Best, 1995; Flege, 1995; Kuhl and Iverson, 1995, among others). Models of L2 speech relate the likelihood of accurate perception and production of target language phones to the degree of similarity between L1 and L2 phones. Early models of L2 speech, such as the contrastive analysis hypothesis (Lado, 1957), suggested that the closer a target language phone was to a L1 category, the easier it was to perceive and produce it accurately. Escudero's (2005) Second Language Linguistic Perception Model (L2LP) posits that L2 phones that have a similar L1 counterpart are easier to learn than new phones with no clear counterpart in the L1. The L2LP claims that the L2 is, originally, a copy of the L1, and this copy may then evolve toward more target-like values. Thus, learning a similar sound involves an adjustment of the category boundaries and their acoustic properties, whereas acquiring a new sound requires establishing a new L2 to L1 mapping prior to the process of phonetic adjustment. Other recent theories link successful L2 category creation to the ability to distinguish between L1 and L2 sound categories. Best and colleagues' Perceptual Assimilation Model (PAM, Best, 1995), and its adaptation to L2 speech learning PAM-L2 (Best and Tyler, 2007), propose that L2 learners' ability to perceive L2

contrasts depends on the degree to which L2 phones are heard as, or assimilated to, an exemplar of an L1 category. The model describes several ways in which non-native or L2 sound categories can be assimilated to L1 categories and predicts discrimination ability of non-native (PAM) or L2 phones (PAM-L2), accordingly. Thus, for instance, if two non-native phones are perceptually mapped onto two separate L1 categories (two category assimilation type), their discrimination will be easier than if the two phones are perceived as exemplars of the same L1 category (Best, 1995). In turn, if the two phones perceived as belonging to the same L1 category differ in how closely they match the L1 category (category goodness assimilation type), discrimination will be better than if the two phones are perceived as equal matches to the L1 category (single category assimilation type). Other assimilation types involve combinations of categorized and uncategorized sounds, whose discrimination accuracy may vary depending on the degree of cross-language assimilation overlap (Levy, 2009a), or perceived phonological overlap (Faris et al., 2018), that is, the overlap in the categorizations to native phones (see Tyler, 2021 for a description of all types and their predictions).

The Speech Learning Model (SLM, Flege, 1995; Flege, 2003; see Flege and Bohn, 2021 for its recently revised version SLM-r) proposes that the greater the perceived dissimilarity between L1 and L2 phones, the greater the likelihood that learners establish target-like categories. Like the PAM, the SLM proposes that the ability to discern between L1 and L2 categories is expected to improve with increased exposure to target-language input. Therefore, phones that are perceived as different from any L1 phone may eventually be categorized more accurately than those that are readily mapped onto L1 categories and are consequently perceived and produced in terms of that L1 category (Flege, 1995; Flege and Bohn, 2021). Still, some target language phones may in fact be sufficiently close to L1 counterparts that their perception and production in terms of the L1 category may be undetected by native listeners (Flege, 1992). These can be referred to as identical or near identical L1-L2 pairs and would not require a separate L2 category (e.g., Spanish and English [f]).

The prediction that target phones with no clear match in the L1 will be learned more accurately than target phones with a counterpart in the L1 is supported by some studies. Busà (1992) examined the production of English vowels by Italians living in the United States and found that most of them produced English /u:/, classified as similar to Italian /u/, with lower F2 values than native English speakers did (that is, closer to Italian /u/). By contrast, a larger number of Italians produced the new L2 vowel /ʊ/ more accurately than the similar vowel /u/. Similarly, research has found that Japanese learners of English show greater improvement in their perception and production of English /r/ than of English /l/, where the former is more dissimilar from Japanese /r/ than the latter (Bradlow et al., 1997; Aoyama et al., 2004). Regarding the effect of L2 experience on the categorization of dissimilar phones, Jun and Cowie (1994) observed that experienced Korean learners of English with a length of residence of 16–31 years in the United States produced the new vowel /ɪ/ more accurately than inexperienced learners with a length of residence of 1.3–5.3 years. Further, Bohn and

Flege (1992) also found that increased experience (i.e., longer length of residence in the US) resulted in a more accurate production of the new vowel /æ/ by German speakers of L2 English, while the more similar English vowels /i:/, ɪ, ε/ were equally accurately produced by all learners.

The studies reviewed above generally support the SLM's prediction that increased L2 experience, particularly when experience involves exposure to authentic L2 input and predominant L2 use (Flege and Bohn, 2021, SLM-r), results in more authentic categories for target sounds without a clear match in the L1. By contrast, L2 phones that are readily assimilated to L1 phones may be less accurately categorized, and may not show improvement over time. However, there are two types of cases that do not follow these predictions. First, the case of target phones that are similar to L1 phones and are nevertheless accurately produced or perceived. For instance, Fullana-Rivera and MacKay (2003) tested the effect of starting age of learning and years of foreign language instruction on Catalan/Spanish bilinguals' production of English vowels. They observed that English /i:/, with a clear counterpart in the L1, was always produced more accurately than the more dissimilar /ɪ/, although an improvement with /ɪ/ was observed with more years of learning, consistent with the SLM's expectations. Further, in a study involving 240 Italian immigrants in Canada, Munro et al. (1996) found that degree of L2 English accentedness increased with age of arrival to the target language country (AOA), which varied from 3.1 to 21.5 years. However, the English vowel /i:/, judged on the basis of an acoustic comparison to be similar to Italian /i/, was more accurately produced by the Italian L2 speakers than more dissimilar vowels like /ʌ/ and /ɜ:/. The second type of outcome that does not follow from the predictions is the case of pairs of target phones involving a new and a similar phone that obtain comparable results. For example, Munro et al. (1996) also report that English /i:/ and /æ/, considered similar and dissimilar to L1 categories, respectively, were comparably accurately produced. Flege et al. (1997) tested the production and perception of L2 speakers from four L1 backgrounds and found that two groups of L1 Spanish L2 English speakers differing in length of residence in the United States (9 years vs. less than 6 months) produced the similar English vowel /ε/ more accurately than the new vowel /æ/ (91–99% vs. 70–73% correct identification of the intended vowel by native English listeners). In addition, their scores for English /i:/ and /ɪ/, classified on the basis of acoustic measurements as similar and new, respectively, were comparable (57–69% for /i:/ and 51–61% for /ɪ/). The question that remains is why similarly classified target phones (in terms of their similarity to L1 categories) are not always equally learned, while differently classified target phones sometimes show comparable results.

The inconsistencies found in some studies thus underscore the need for a consistent and reliable method of measuring cross-linguistic similarity. One of the most common approaches involves the use of perceptual cross-language mapping tasks (Best, 1995; Flege et al., 1997; Strange et al., 2009; Tyler et al., 2014, among many others). In these tasks, commonly referred to as categorization tasks or perceptual assimilation tasks (PAT), listeners are presented with non-native or L2 speech stimuli, and

asked to 1) indicate to which L1 phonetic category each token is most similar, and 2) rate its "goodness of fit" as an exemplar of that category (see *Materials and Methods* section below). Tyler (2021) argues that while there are some limitations with these tasks, PATs are currently the most suitable method for measuring cross-linguistic similarity given that they evaluate different possible sources of information that listeners attend to when perceiving non-native sounds. This measure of perceived similarity has been found to be a good predictor of L2 learning difficulty (e.g., Best and Strange, 1992; Guion et al., 2000; Strange et al., 2011; Tyler et al., 2014, among others).

The current paper aims to explore the link between similarity relations and L2 performance further by examining the perception and production of the English /i:/-/ɪ/ contrast by a group of L1 Spanish/Catalan speakers. Previous studies have found that L1 Spanish/Catalan speakers have difficulty with the English /i:/-/ɪ/ contrast but have not been able to relate it to the degree to which each vowel is assimilated to a L1 category. Cebrian (2006) found that two groups of L2 English speakers (20 Catalan speakers living in Canada and 30 Catalan undergraduate students of English living in Spain) perceptually assimilated Canadian English /i/, ε, e/ to Catalan /i/, ε, ei/, respectively, whereas English /ɪ/ obtained lower assimilation scores and goodness ratings as Catalan /e/. Perception of L2 sounds was examined with an identification test involving a synthetic continuum from /i:/ to /ɪ/ to /e/ varying in vowel quality and vowel duration. The two groups performed like native speakers in their perception of English /e/, but differed from native speakers in their perception of English /i:/ and /ɪ/, which showed a predominant reliance on temporal cues, unlike native Canadian English speakers who relied mostly on spectral differences. These results do not follow from the results of the PAT since two English vowels that were strongly assimilated to Catalan vowels, English /i:/ and /e/, obtained different results, and the former patterned like the weakly assimilated vowel /ɪ/. The production of the target English vowels by the 30 Catalan undergraduate students of English was examined in a subsequent study (Cebrian, 2007). Data from eight native English speaker judges indicated that both English /e/ and /ɪ/ were accurately produced (correctly identified 99 and 86% of the time, with goodness ratings reaching 5.5 and 5.2 out of 7, respectively). By contrast, English /i:/ and /ɪ/ obtained similarly lower rates of correct identification (73 and 71%, respectively), and /i:/ obtained higher goodness ratings than /ɪ/ (4.5 vs. 3.9) but still lower than /e/ and /ɪ/. Acoustic measurements confirmed the native speaker rating data and showed that the Catalan learners produced English /e/ distinctly from English /i:/ and /ɪ/ and produced the latter two with a large amount of spectral overlap in terms of first and second formant. The learners, however, produced a temporal difference between the tense and the lax vowel (/i:/: 243 ms, /ɪ/: 153 ms), resulting in a ratio (1.62) that falls within the values reported for native English speakers (Hillenbrand et al., 2000).

The inability to produce a spectral difference between English /i:/ and /ɪ/ thus seems to be linked at least in part to the fact that Catalan and Spanish learners of English tend to base the /i:/-/ɪ/ distinction on a temporal difference (e.g., Kondaurova and Francis, 2008; Cerviño and Mora, 2009). In fact, L2 English learners have been found to exploit temporal cues in their categorization of the

English /i:/-/ɪ/ contrast to a greater extent than native English speakers, who appear to rely mostly on spectral cues (Hillenbrand et al., 2000; Escudero and Boersma, 2004; Cebrian, 2006). This is found with learners whose L1 has temporal contrasts (e.g., /i:/-/ɪ/ contrast), such as Japanese and Finnish (Ylinen et al., 2009; Grenon et al., 2019), and importantly also with speakers whose L1 has no vowel duration contrast, such as Mandarin Chinese, Korean, Russian, Catalan, Portuguese and Spanish (Bohn, 1995; Flege et al., 1997; Wang and Munro, 1999; Escudero and Boersma, 2004; Cebrian, 2006; Mora and Fullana, 2007; Kondaurova and Francis, 2008; Morrison, 2008; Aliaga-García, 2011; Kivistö de Souza et al., 2017). These results lend support to Bohn's (1995) desensitization hypothesis, which claims that L2 learners may not be sensitive to L2 spectral distinctions that are not exploited in their L1, to which they have become desensitized. In these cases, duration may be used to differentiate L2 pairs with no parallel L1 spectral distinction. Further support comes from studies that show that learners exploit duration for specific contrasts only. Flege and Bohn (1989) found that Spanish learners exploited temporal cues to a greater extent than spectral cues in their identification of English /i:/ and /ɪ/, which do not have two clear L1 counterparts, but overreliance on duration was not evident with /e/ and /æ/, which may be identified in terms of Spanish /e/ and /a/. Hence, duration is used to distinguish the pair of target vowels for which the L1 does not have a comparable spectral contrast. Kondaurova and Francis (2008) conclude that several factors play a role in the categorization of the /i:/-/ɪ/ contrast as a mostly temporal contrast, including the role of duration as a phonetic cue in the learners' L1 (e.g., as a cue to stress differences), experience with different target-language varieties, psychoacoustic explanations based on the desensitization to spectral differences not exploited in the L1, and a possible salience of duration and consequent ease of learning (Holt and Lotto, 2006; Hacquard et al., 2007; Kivistö de Souza et al., 2017).

This study examines the perception and production of the English /i:/-/ɪ/ contrast by Spanish/Catalan bilinguals in relation to their perceived similarity to L1 categories. The English high front tense-lax pair /i:/-/ɪ/ has received a lot of attention in the literature both because it poses a problem to learners from different language backgrounds and also because it has a very high functional load in English (with more than 450 minimal pairs according to Higgins, 2018). The fact that a large number of these minimal pairs involve high-frequency words, and that many share the same grammatical category (e.g. feel-fill, leave-live), contribute to the high functional load of this contrast [Munro, 2021 (this volume)]. Thus, failure to distinguish these vowels can result in intelligibility problems. The perception and production of English /e/ and non-rhoticized /ɜ:/ (as in bed and bird in Southern British English, respectively) will also be analyzed for comparison purposes as these vowels represent a clear case of a similar and a dissimilar vowel. **Table 1** shows the assimilation scores (percent identification in terms of Spanish vowel categories) reported for English /i:/, ɪ, ε, ɜ:/ in a few previous studies involving Spanish speakers. Generally, English /i:/ and /e/ seem to be strongly assimilated to an L1 category (/i/ and /e/, respectively), while /ɪ/ is more weakly assimilated to Spanish /e/. Cebrian (2019) also examined English /ɜ:/, which patterned with

**TABLE 1** | Percent assimilation of English /i:/, i, ε, ɜ:/ to Spanish vowels reported in recent studies (Goodness of fit ratings are given in parentheses when available). SSBE, GA (General American) and Can. Eng. (Canadian English) indicate the English variety of the stimuli.

English stimuli	L1 Spanish response	Cebrian (2019), <sup>a</sup> SSBE	Baigorri et al. (2019), <sup>b</sup> GA	Morrison (2012), <sup>c</sup> Can. Eng.	Escudero and Chládková (2010), <sup>d</sup>		Flege (1991), <sup>e</sup> GA
					SSBE	GA	
/i:/	/i/	93 (5.9/7)	95–96 (8/9)	98–99	100	99	84–94
/ɪ/	/e/	66 (5.3/7)	50 (3–6/9)	95–96	35	90	19–39
	/i/	33 (5/7)	49	3	21	10	36–68
/e/	/e/	97 (5.3/7)	87–94 (6/9)	95–96	77	96	44–81
/ɜ:/	/e/	64 (3/7)					
	/a/	14 (3.4/7)					
	/o/	13 (3.9/7)					

<sup>a</sup>Monolingual European Spanish speakers.

<sup>b</sup>Late and early L2 English learners.

<sup>c</sup>Monolingual European Spanish and Mexican Spanish speakers.

<sup>d</sup>Monolingual Peruvian Spanish speakers.

<sup>e</sup>L2 English speakers and monolingual Spanish speakers.

/ɪ/ being weakly assimilated with Spanish /e/, with comparatively low goodness of fit ratings.<sup>1</sup>

Another reason for the lack of a consistent link between cross-linguistic perceived similarity and L2 performance may stem from the amount of individual variation potentially found in L2 performance. For instance, Munro et al. (1996) study involving 240 Italian L2 English speakers reported a lack of consistent results per vowel across L2 speakers as not all L2 speakers of similar experience produced the same vowels accurately. In a study on Cantonese speakers' production of English tense and lax high vowels, Munro (2021 (this volume)) also underscores the high degree of inter-speaker variability, which could not be related to specific linguistic (word, phonetic context) or individual (language use, length of residence) factors. Still, the majority of studies make predictions about L2 performance for a specific group of learners based on the average similarity judgements obtained for that group's performance (e.g., Cebrian, 2006; Rallo Fabra and Romero, 2012; Baigorri et al., 2019). This assumes that group tendencies are representative of individuals' perceptual judgements. Flege and Bohn's (2021) SLM-r claims that individual differences in L1 category precision may influence the discernment of cross-linguistic phonetic differences. In addition, Flege and Bohn caution that individuals may differ in the way they map L2 sounds onto L1 categories particularly in a PAT. This is because listeners' ratings may be affected by the interval between accessing long term memories of their native sounds in order to label the auditory stimulus and judging the degree of similarity between that stimulus and the selected native category. In fact, studies have shown evidence of variability among individuals sharing the same L1 in perceived similarity judgments

(Strange et al., 2009; Gilichinskaya and Strange, 2010; Tyler et al., 2014; Faris et al., 2018). For example, an individual analysis of the perceptual assimilation data collected by Cebrian (2019) showed that 15 out of the 29 Spanish speakers mapped English /ɪ/ to Spanish /e/ 75% of the time or more, eight participants chose Spanish /i/ as the closest L1 vowel at a similar rate, and six participants divided their responses between Spanish /i/ and /e/ (67% or less for each vowel). Thus, a fair amount of variability was found among speakers of the same L1 background.

Therefore, another goal of the current study is to investigate the relationship between perceived similarity and likelihood of accurate L2 category creation at the individual level. The cross-linguistic perceived similarity judgements of each individual are compared to their ability to produce and perceive the two vowels distinctly. The study thus examines if previous findings that English /i:/ and /ɪ/ are perceived and produced with similar degrees of accuracy, despite differing in the extent to which they are mapped onto L1 vowels, are related to individual differences in cross-linguistic perception, or to other factors (e.g., the relevance of different acoustic cues, the role of language instruction or the influence of orthography). The focus of this paper is the English high front tense-lax vowel contrast (/i:/-/ɪ/), and two additional vowels are also examined, namely a vowel often reported to be strongly assimilated to an L1 vowel (English /ε/) and one that lacks a consistent counterpart in the L1 (English /ɜ:/) in order to understand what is specific about the /i:/-/ɪ/ contrast and what is determined by perceived similarity relationships.

## MATERIALS AND METHODS

A group of L2 English speakers performed a series of tests including a perceptual assimilation task, a vowel identification task, a vowel discrimination task and a picture naming production task. The data here presented are part of a larger study evaluating the perception and production of a number of English vowels. This paper focuses on the English /i:/-/ɪ/ and /ε/-/ɜ:/ contrasts.

<sup>1</sup>Similar results have been obtained for similarity measurements involving Catalan and English despite differences in vowel inventory between the two languages (Catalan has seven vowel phonemes (/i e ε a o u/, plus [ə] in unstressed position) while Spanish has five (/i e a o u/). For instance, Cebrian (2006) found that English /i:/ was assimilated to Catalan /i/ (99%, GR: 6.2/7), while /ɪ/ was split between Catalan /e/ (66%, GR: 3.5), /ε/ (20%, GR: 3/7) and /i/ (14%, GR: 2.5). See also Cebrian (2021).



## Participants

The participants were 43 Catalan/Spanish bilinguals (37 females, mean age 19.2, standard deviation 1 year) who were first-year English Studies undergraduate students at Universitat Autònoma de Barcelona (UAB), Barcelona, Spain. They reported having started learning English in school before the age of 12 (the average starting age of learning was six; average length of learning 13 years), and were at the time attending classes in English at university between 6 and 15 h per week. Most participants reported speaking English at university (with classmates, foreign students, and teachers) at least half the time, although they used English less often with family, friends and at work. The majority had not lived in an English-speaking country and those who had, had spent less than 6 months (an average of 2 weeks), generally as part of a summer stay. Participants reported having no hearing impairments. All 43 participants performed the three perception tasks, but five participants did not complete the production task and, thus, only the production of the remaining 38 participants was analyzed (33 females, mean age 19.4, standard deviation 0.9 years). The participants were all native speakers of Spanish although some spoke Catalan as well. Participants completed an online personal and language background questionnaire based on the Bilingual Language Profile (BLP; Birdsong et al., 2012). Among other information, the BLP renders a score that ranges from 218 to -218 representing the two monolingual endpoints, with values close to zero indicating balanced bilingualism. According to the questionnaire, 22 participants were Spanish-dominant, 11 appeared to be balanced in Spanish and Catalan, and 10 were more dominant in Catalan. This difference among participants need not be problematic as one of the goals of this paper is precisely to examine how individual variation in cross-linguistic perception may affect L2 performance. Nevertheless, an analysis of the Catalan-dominant bilinguals' results in the PAT revealed that their responses for the vowels under study did not differ significantly from the remaining Spanish speakers' performance (see *Perceptual Assimilation Task*). All tests took place at the Speech Laboratory at UAB on two different days. Participants performed the production task and the perceptual assimilation task in one session and the identification and discrimination tasks in a second session.

## Cross-Language Perception Stimuli

The stimuli used in the perceptual assimilation task (PAT) were a subset of the stimuli used in a previous study (Cebrian, 2019) and consisted of the Standard Southern British English (SSBE) vowels /i:/, ɪ, ɛ, eɪ, aɪ, æ, ɑ:/, ɜ:/, ʌ/ produced in /bVt/ words.<sup>2</sup> These words were elicited from three monolingual male speakers of SSBE (mean age 35), who had spent most of their lives in the South of England and were living in London. The recordings were

carried out in a soundproof booth at the speech laboratory at University College London, in London, United Kingdom, and were digitized at a 44.1 kHz-sampling rate and normalized for intensity [70 dB in sound pressure level (SPL)]. The best tokens per talker were selected, based on auditory judgements and spectrographic analysis. Neighboring sounds have been found to affect perceptual assimilation judgements (e.g., Bohn and Steinlen, 2003; Levy, 2009b). Hence stimuli were edited to include from the release of the /b/ until the closure of /t/, thus eliminating potential cross-linguistic and individual differences in stop production (prevoicing of /b/, /t/ release), while maintaining intact the cues to the vowel.

## Perceptual Assimilation Task

In the perceptual assimilation tasks, listeners were required to identify each English vowel stimulus in terms of one of several possible L1 categories by clicking on one of the response options presented on a computer screen. Upon selecting an L1 category, listeners provided a goodness of fit rating on a 7-point scale, where 1 meant a poor example of the selected vowel and 7 meant a good, native-like example. The response options consisted of the most common and unequivocal spelling for each Spanish vowel and diphthong (<i, e, a, o, u, ai, ei, oi>) together with a monosyllabic word illustrating that vowel, namely di (for /i/), se (/e/), da (/a/), do (/o/), tu (/u/), hay (/ai/), rey (/ei/), hoy (/oi/), meaning say, self (reflexive pronoun), give, do (musical note), you, there is, and today, respectively. Every vowel appeared in 12 trials (3 talkers × 2 tokens × 2 repetitions). This paper reports the results for the vowels under study, namely /i:/, ɪ, ɛ, ɜ:/ . See Cebrian (2019) for a study covering a near complete set of SSBE vowels.

## Vowel Identification and Vowel Discrimination Tasks Stimuli

The stimuli used in the identification and discrimination tasks consisted of high-frequency monosyllabic English words. Each word shared the initial and final consonants with at least one other word, thus resulting in a series of minimal pairs (e.g., bead, bid, bed, bird). Half the words ended in a voiced stop and the other half ended in a voiceless stop. The words were produced by two native speakers of SSBE, namely a 23 year-old female and a 33 year-old male, who were different from the speakers who produced the stimuli for the PAT. The speakers were recorded in a soundproof booth at the speech laboratory at University College London. The recordings were digitized at 44.1 kHz sampling rate and normalized for intensity (70 dB in SPL). The same stimuli were used in the identification task, presented individually, and the discrimination task, arranged in pairs (e.g., bead-bid, bead-bead), as explained below. These tasks were part of a larger study examining the perception of seven English vowels (/i:/, ɪ, ɛ, æ, ɑ:/, ɜ:/, ʌ/).

## Identification Task

Participants were asked to identify the vowel in the stimulus using one of seven response options appearing on a computer screen. The options consisted of a phonetic symbol together with two common words representing each sound, namely /æ/ ash, mass; /ʌ/

<sup>2</sup>Participants were exposed to different English varieties through music, TV and cinema, and the internet (according to the language background questionnaire, their average exposure to British and American English was 54 and 46%, respectively). Still, British English is the main variety taught in schools and language centers in Spain and it is the variety described in most textbooks used at UAB.

sun, thus; /ɪ/ fish, his; /i:/ cheese, leaf; /ɜ:/ earth, first; /ɛ/ less, west; /ɑ:/ arm, palm. There were eight stimuli per vowel (four words produced by two different speakers) and each stimulus appeared twice throughout the task, resulting in a total of 16 trials per vowel.

### Discrimination Task

The discrimination task was a categorical AX same/different discrimination task, including several pairs of English vowels. The data presented in this paper focuses on the /i:/-/ɪ/ and /ɛ/-/ɜ:/ pairs. For each pair, several stimuli were created by including the two possible speaker orders (speaker 1-speaker 2, speaker 2-speaker 1) and vowel orders (e.g., /i:/-/ɪ/ and /ɪ/-/i:/). Half the stimuli included the same vowel category (same-category trials, /i:/-/i:/), and half were different-category trials (/i:/-/ɪ/). This resulted in a total of 16 different-category trials per vowel pair, and 8 same-category trials per vowel. The interstimulus interval was 1.15 s so that participants used phonetic information stored in long term memory instead of relying on sensory memory (e.g. Højen and Flege, 2006). Participants responded by clicking on the words “same” or “different” displayed on a computer screen.

### Production Task

Thirty-eight of the 43 participants who performed the PAT and the identification and discrimination tasks also completed a picture naming task in which they were asked to name 44 different pictures twice. The list included the four words containing each target vowel. Specifically, the words were bed, bet, head, pet for /ɛ/, bird, heard, hurt, dirt for /ɜ:/, bid, bit, dip, lid for /ɪ/, feed, feet, league, leak, for /i:/. The words were presented in a random order, which was the same for all participants. The first four words were fillers used to familiarize participants with the task. For each word, a picture was displayed on a computer screen (MS PowerPoint was used). In order to ensure that the right word was produced, the target word was shown at the bottom of the screen in written form. The written word disappeared after two seconds, then participants were instructed to name the picture twice, prompted by the appearance of a dialogue balloon next to the picture. Participants were recorded in a soundproof chamber at the speech laboratory at UAB, Spain. In addition, the production of 13 native English speakers (seven female) was also collected using the same task to provide native English values to compare the L2 production to. These were native speakers of Southern British English in their twenties and early thirties. The group consisted of a mixture of international students and English teachers residing in Barcelona (for 2–10 years; reported weekly use of English: 75% of the time), who were recorded at the speech laboratory at UAB, and five undergraduate students at Queen Mary University, London (monolingual English speakers, three of whom had taken Spanish lessons but barely used Spanish outside the classroom), recorded in a sound attenuated room at that institution. Recordings were digitized at 44.1 kHz sampling rate and normalized for intensity (70 dB in SPL).

## RESULTS AND INTERIM DISCUSSIONS

This section presents the results of the perceptual similarity task, followed by the L2 perception and L2 production results. The section

ends with a correlational analysis contrasting the results of the different tasks. Each set of results is followed by a brief interim discussion.

## Perceptual Assimilation Task

### Perceptual Assimilation Tasks Results

The percentage of times participants identified each target vowel as one of the Spanish responses (i.e., assimilation scores) and the corresponding average goodness of fit ratings (GR) were calculated. In addition, following Guion et al. (2000), a fit index score (FI) was calculated by multiplying the identification proportion by the goodness of fit rating. This composite score is meant to distinguish between cases of comparable assimilation percentages that differ in GR, and was used by Guion et al. (2000) to determine the likelihood of accurate discrimination for pairs of L2 vowels. Further, given that individual differences in perceived similarity are often observed, the number of participants who selected a given response at a given level of consistency was tallied in order to assess the degree of agreement among listeners (Strange et al., 2009; Gilichinskaya and Strange, 2010; Cebrian, 2021). To that effect, the categorization thresholds used in some previous studies were used, i.e., 70% (Tyler et al., 2014) and 50% (Bundgaard-Nielsen et al., 2011; Faris et al., 2018). The results are presented in **Table 2**.

Recall from Section *Participants* that many of the participants spoke Catalan in addition to Spanish and participants differed in how strongly dominant in Spanish they were. An analysis was conducted to see if responses in the PAT were influenced by the bilinguals' language dominance. The results for the subgroup with a clear dominance in Spanish ( $n = 22$ ) were compared to those who showed a balanced dominance ( $n = 11$ ) or a more Catalan dominance ( $n = 10$ ). The results for English /ɛ, i:, ɪ/ were practically identical across groups, while in the case of English /ɜ:/, a greater number of Spanish /a/ responses were obtained as dominance in Catalan increased, reaching 20%. A series of Mann-Whitney U-tests yielded no significant effect of language dominance on assimilation scores, confirming that the general pattern of results was very comparable across dominance groups.

In terms of individual variation, out of the 43 participants who performed the PAT, 23 assimilated English /ɪ/ to Spanish /e/ at least 70% of the time, six learners chose Spanish /i/ as the best L1 match 70% of the time or more, and the remaining 14 learners chose either option between 25 and 67% of the time. This shows that, though the predominant pattern was to perceive English /ɪ/ as most similar to Spanish /e/, there was a certain degree of variability among participants. Responses for English /ɛ/ and /i:/ were very consistent, with all participants identifying English /ɛ/ with Spanish /e/ 70% of the time or more, and almost all participants (39) assimilating English /i:/ consistently to Spanish /i/, while four yielded responses split between Spanish /i/ and /ei/. Regarding /ɜ:/, 28 participants chose Spanish /e/ as the closest vowel at least 70% of the time, three participants selected Spanish /a/, two other participants selected Spanish /o/ and the remaining 10 participants did not choose a specific L1 vowel more than 67% of the time.

**TABLE 2** | Perceived similarity between nonnative (English) vowels and L1 (Spanish) vowels. (Num. of Ss assim.  $\geq 70\%/50\%$  = number of subjects who selected a given response 70%/50% of the time or more). The total number of participants was 43.

English stimuli	L1 Spanish response	% Assimilation	Goodness Rating	Fit Index	Num. of Ss assimn. $\geq 70\%$	Num. of Ss assimn. $\geq 50\%$
/i:/	/i/	94	6.5	6.1	39	43
	/ej/	5.8	4.4	0.3		1
/ɪ/	/e/	67	5.9	4.0	23	32
	/i/	32	5.4	1.6	6	12
/ɛ/	/e/	99	5.3	5.2	43	43
/ɜ:/	/e/	72	4.3	3.1	28	34
	/a/	15	2.8	0.4	3	4
	/o/	11	4.3	0.5	2	5

## Discussion of Perceptual Assimilation Tasks Results

Assuming a categorization threshold of 70% (Tyler et al., 2014), English /i:/, /ɛ/ and /ɜ:/ were categorized as Spanish /i/, /e/ and /e/, respectively. English /ɪ/ did not reach the categorization threshold of 70%, but was assimilated above chance to more than one L1 phone (Spanish /e/ and /i/) and thus illustrates an uncategorized-clustered type of assimilation (Faris et al., 2018). The current results are on the whole in agreement with the previous studies showing that Spanish speakers consistently assimilate English /i:/ and /ɛ/ to Spanish /i/ and /e/, and that English /ɪ/ displays a less consistent pattern (Escudero and Chládková, 2010; Morrison, 2012; Baigorri et al., 2019; Cebrian, 2019, see **Table 1** above). In fact, the results closely replicate those obtained by Cebrian (2019), which involved a near-complete set of SSBE and Spanish vowels and diphthongs and tested a group of Spanish speakers with little or no knowledge of English. The only difference is that English /ɜ:/ fell short of a 70% categorization threshold in the previous study (64% as Spanish /e/ vs. 72% in the current study) and was thus classified as uncategorized-clustered. Setting the categorization threshold at 70% (or 50%) is an arbitrary decision, as there are no clear criteria supporting one or another cut-off (see Tyler, 2021, for discussion). In fact, in both studies English /ɪ/ obtained a higher GR as Spanish /e/ than English /ɜ:/ did, indicating that the latter was perceived as a more dissimilar vowel. The lower GRs for /ɜ:/ may also be related to the greater differences across English varieties regarding this vowel, which may have affected listeners' familiarity with SSBE /ɜ:/.

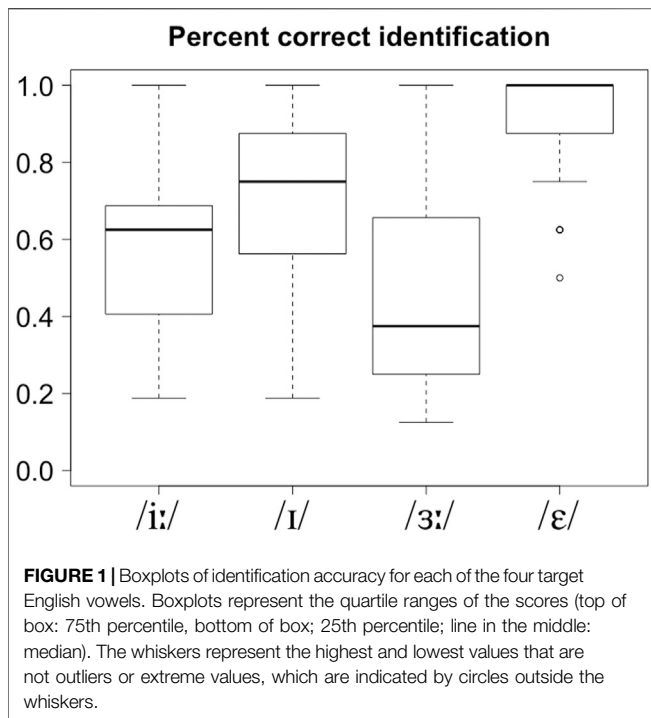
Considering both the degree of categorization and the FI, what the data show is that English /ɛ/ and /i:/ are strongly assimilated to Spanish /e/ and /i/, while English /ɪ/ and /ɜ:/ show weaker degrees of assimilation to Spanish /e/. The main focus of the current study is the English /i:/-/ɪ/ contrast. The PAT results show that the two English vowels differ notably in degree of assimilation to an L1 category (FIs of 6.1 vs. 4, respectively). The pair patterns as an uncategorized-categorized type of assimilation in PAM-L2's terms (as Spanish /i/ and /e/, respectively), with a cross-language assimilation or perceived phonological overlap (Levy, 2009a; Faris et al., 2018) of 32% as Spanish /i/. The English /ɛ/-/ɜ:/ contrast patterns as a category-goodness assimilation type, with a considerable perceived phonological overlap (72%) and a notable difference in degree of assimilation to Spanish /e/, as English /ɛ/ is perceived as a better match for Spanish /e/ than English /ɜ:/ (FIs:

5.3 vs. 3.1). PAM-L2 predicts that discrimination of these two types of contrasts by L2 speakers will be more difficult than between pairs that assimilate to two different non-native vowels (e.g., English /i:/ and /ɛ/) and their discrimination accuracy may depend on the degree of perceived phonological overlap. Thus, we may expect that English /i:/-/ɪ/ will be more accurately discriminated than English /ɛ/-/ɜ:/.

In terms of the SLM (Flege, 1995; and its recently revised version, the SLM-r; Flege and Bohn, 2021), all four English vowels will be initially categorized in terms of the closest L1 phones. Exposure to authentic target language input may enhance the ability to differentiate the L2 from L1 sounds and thus establish separate categories for the L2 sounds. This is more likely for English /ɪ/ and /ɜ:/, judged to be more different from L1 categories, than for English /ɛ/ and /i:/, which have a clear match in the L1. In fact, the latter two may pattern as near-identical to L1 categories. Flege (1992) suggests that in order to know if a non-native sound is perceived to be indistinguishable from a native category, this non-native phone should be undetectable when produced in a native context. In brief, English /i:/ and /ɛ/ pattern as highly assimilated to Spanish /i/ and /e/, while English /ɪ/ and /ɜ:/ are weakly assimilated. The prediction is that the level of accuracy in the perception and production of English /i:/ and /ɛ/ by Spanish L2 speakers will be similarly high, while for /ɪ/ and /ɜ:/ performance may be originally worse and accuracy will depend on the extent to which learners detect differences between L1 and L2 sounds thanks to continuous authentic input.

## L2 Perception Identification Results

The results of the identification task are presented in **Figure 1** and **Table 3**. **Figure 1** shows the median and distribution of the percent correct identification for each of the four target vowels. **Table 3** shows the average correct identification and misidentification responses (scores below 3% are omitted). As can be observed, English /ɛ/ was the most accurately identified, followed by /ɪ/, while identification was worse with /i:/ and /ɜ:/. The identification results were submitted to a series of generalized linear mixed models (GLMM) with score (correctly or incorrectly identified) as the dependent variable. The best fitting model was a GLMM with Target vowel as fixed factor, and participant as

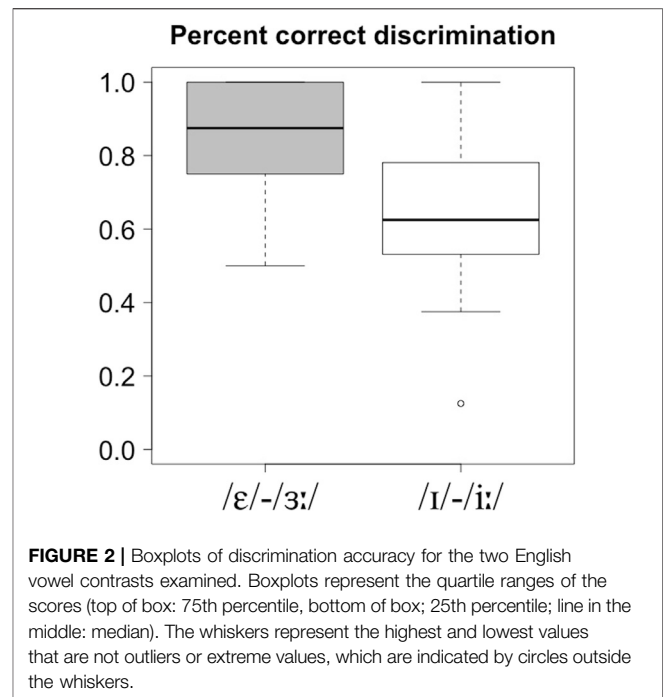


**TABLE 3 |** Identification of the four English target vowels by 43 Spanish/Catalan learners of English. Mean correct identification percentages are highlighted in bold, misidentifications equal to or lower than 3% have been omitted.

Responses	English target vowels			
	/i:/	/ɪ/	/ɛ/	/ɜ:/
/i:/	<b>58</b>	21		
/ɪ/	39	<b>72</b>		
/ɛ/		4	<b>90</b>	4
/ɜ:/			7	<b>47</b>
/ʌ/				23
/ɑ:/				19
/æ/				4

random intercept.<sup>3</sup> The results showed a significant effect of Target vowel ( $F(3, 2,444) = 65.3; p < 0.001$ ). Bonferroni-adjusted pairwise comparisons showed that all between-vowel differences were significant ( $p < 0.001$ ). Regarding the pattern of misidentifications, English /i:/ was most frequently misheard as /ɪ/ and vice versa, while /ɜ:/ was most often misidentified as /ɑ:/ or /ʌ/.

<sup>3</sup>For the statistical analyses on the L2 perception and production data, several generalized linear mixed models (GLMM) were considered, which included the independent variables in each case as the fixed effects, and different combinations of random intercepts and slopes for participant and stimulus. The selection of the best model was based on the lowest Akaike Information Criterion for a model that included participant as random intercept or random slope, given the theoretical relevance of participant variability in the current study. In fact, in every case, the results for the different models were very consistent and the levels of significance for each factor and pairwise comparison were very similar across models. IBM Corp (2017) software was used.



### Discrimination Results

The results of the discrimination task are presented in **Figure 2**. The /ɛ/-/ɜ:/ pair was discriminated more successfully than the /i:/-/ɪ/ pair, with mean percentages of correct discrimination reaching 87% for /ɛ/-/ɜ:/ and 66% for /i:/-/ɪ/. The results of a GLMM with Vowel pair as fixed effect and a random slope for participant<sup>3</sup> confirmed that the difference between the two vowel pairs was significant ( $F(1, 1,030) = 38.15, p < 0.001$ ).

### Discussion of L2 Perception Results

The finding that /ɛ/ was very successfully identified and /ɜ:/ was poorly identified can be explained in terms of the PAT results, as the former was strongly assimilated to an L1 category, while the latter barely reached a categorization threshold of 70% (99 and 72%, respectively, as Spanish /e/). Two studies testing the effect of high variability perceptual training on SSBE vowel identification by Catalan/Spanish learners of English (Carlet and Cebrian, 2019) and Spanish monolingual learners of English (Fouz-González and Mompean, 2020) also found that /ɜ:/ was the most poorly identified vowel, but its identification tended to improve the most as a result of increased L2 experience (namely perceptual training), as expected for a dissimilar vowel (e.g., Flege, 1995). On the other hand, the fact that /ɪ/ was better identified than /i:/ (72 vs. 58%, respectively) cannot be equally explained by the PAT results. The expectation would be for /i:/ to pattern with /ɛ/ and be successfully identified given that both English vowels were strongly assimilated to L1 vowels. Still, better identification of /ɪ/ than of /i:/ has been reported for similar populations (Cebrian and Carlet, 2014; Carlet and Cebrian, 2019; Fouz-González and Mompean, 2020).

When misidentified, SSBE /i:/ was usually perceived as /ɪ/. The results of the PAT showed that /ɪ/ was assimilated to Spanish /i/



about a third of the time (32%), that is, there was a 32% overlap in the assimilation of English /i:/ and /ɪ/ to Spanish /i/. This overlap, and the relatively high goodness of fit ratings (GR) obtained for English /ɪ/ as Spanish /i/ (5.4/7), indicate a degree of perceptual confusion that may explain the misidentifications of English /i:/ as /ɪ/. On the other hand, the pattern of misidentifications of /ɜ:/, mostly as /ɑ:/ or /ʌ/ but hardly as /ɛ/, may be linked to the comparatively low GRs for /ɜ:/ as Spanish /e/ (4.3). Thus, learners tended to identify as English /ɛ/ only stimuli that were clear exemplars of this vowel. In brief, the identification results are not expected from the PAT results, as the two strongly assimilated vowels (/i:/ and /ɛ/) obtained very different results, and the two weakly assimilated vowels (/ɪ/ and /ɜ:/) also patterned very differently.

Regarding the discrimination task, the results seem to show more difficulty with the /i:/-/ɪ/ contrast than some previous studies, which reported discrimination accuracy rates close to or over 75% or A' scores denoting sensitivity to the contrast, i.e., over 0.6 (Mora and Fullana, 2007; Rallo Fabra and Romero, 2012; Baigorri et al., 2019). Comparisons across studies are complicated, however, by methodological differences such as the selection of English contrasts examined and the variety of English tested (SSBE vs. American English). In terms of the PAM, /i:/-/ɪ/ was classified as a categorized-uncategorized clustered type of assimilation, with partial perceptual overlap. The English pair /ɛ/-/ɜ:/ was classified as a category-goodness difference type of assimilation (and as a categorized-uncategorized clustered type in a previous study, Cebrian, 2019). According to Faris et al. (2018), the two types of assimilation pose a similar degree of discrimination difficulty, and thus discrimination accuracy depends on the amount of phonological overlap. The /ɛ/-/ɜ:/ pair, with a greater phonological overlap (72% as Spanish /e/), was expected to pose a greater difficulty than the /i:/-/ɪ/ contrast (32% overlap as /i/), but the results showed the opposite pattern.

## L2 Production

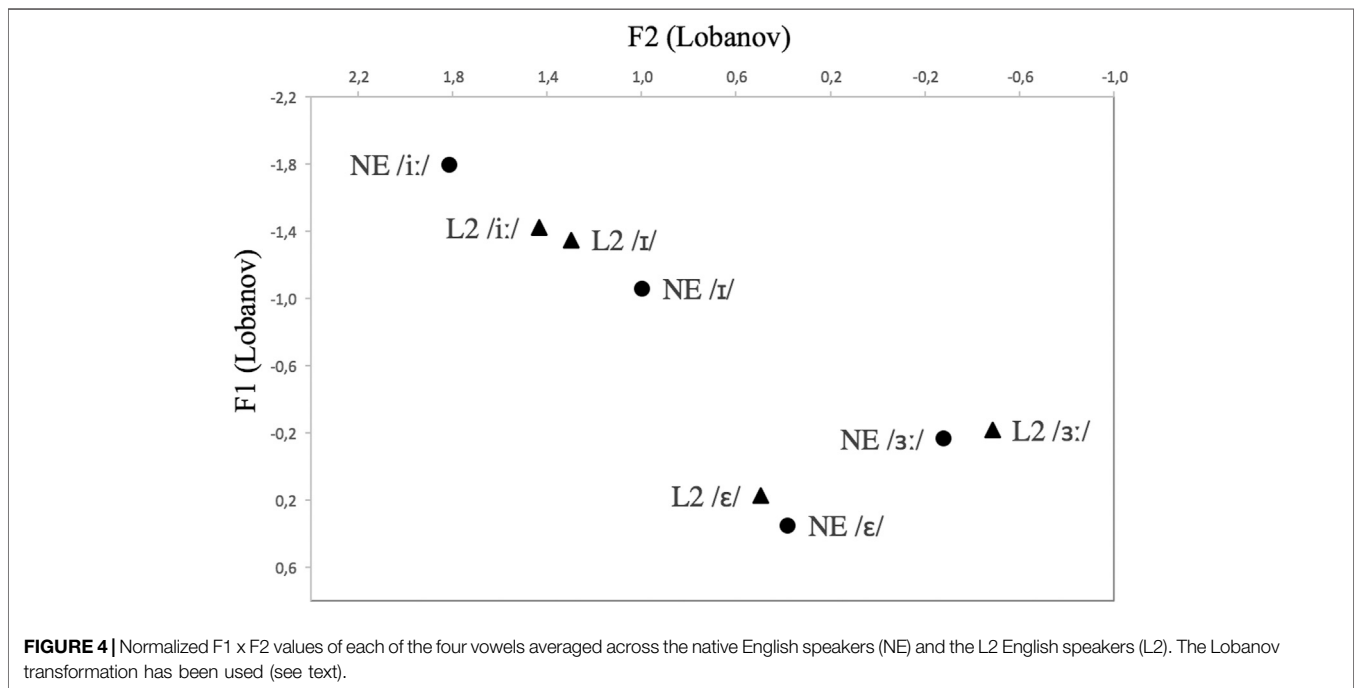
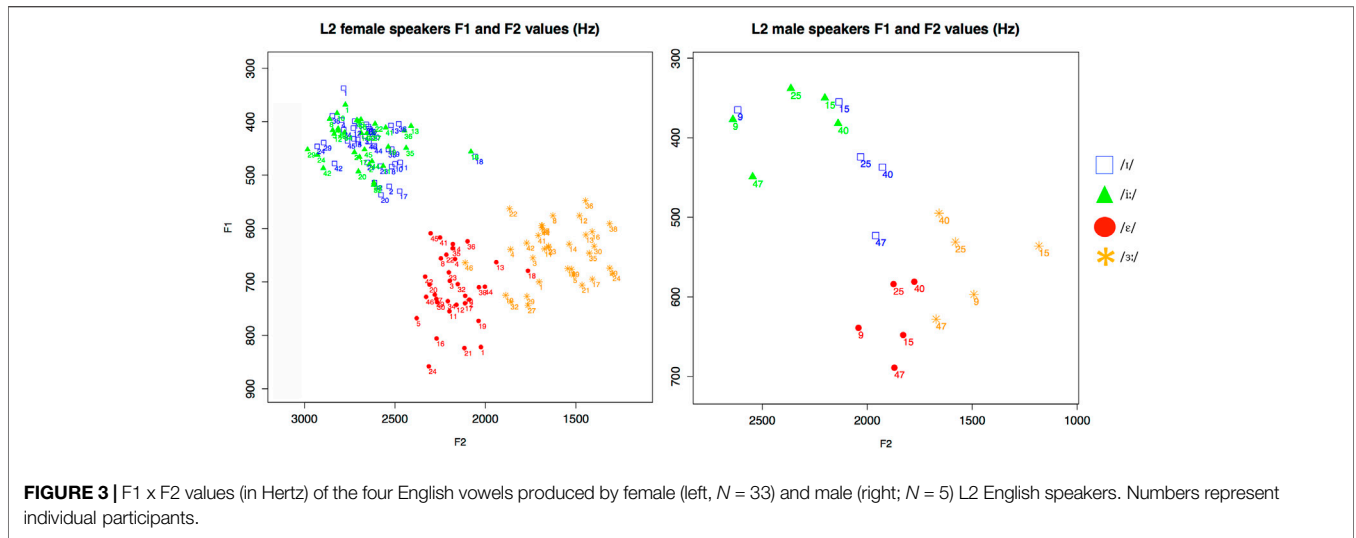
### Production Results

The vowel production of the 38 L2 speakers (recall that five participants only completed the perception tasks) and 13 native English speakers was analyzed acoustically in terms of first and second formant (F1 and F2) and vowel duration. Vowel formants were measured from a 25 ms window located manually at a steady-state portion between one third and two fourths into the vowel, using Praat (Boersma and Weenink, 2018). **Figure 3** displays the mean F1 and F2 values in Hertz per vowel for each of the L2 speakers, showing the data for female and male L2 speakers separately. Each individual is represented by a number. As can be observed from the distribution of the individuals' production, there is considerable spectral overlap between /i:/ and /ɪ/, while /ɛ/ and /ɜ:/ appear to be distinguished more clearly. As expected, a fair amount of individual variation can be observed. For example, male speaker 15 produced practically indistinguishable /i:/ and /ɪ/ (see **Figure 3**). By contrast, male speaker 47 produced these two vowels with very different spectral values. The issue of whether this difference in performance can be predicted from individual differences in

perceptual assimilation patterns is analyzed in *Cross-Task comparison. Correlational analysis*.

In order to compare the L2 data to native English data, and to group males and female speakers together, the formant values were normalized following the Lobanov method (Lobanov, 1971). A Lobanov transformation is a vowel extrinsic normalization method that has been found to render superior normalization outcomes to other methods, particularly if the data set examined includes a large set of vowels (Adank et al., 2004; Recasens and Espinosa, 2009). Recall that, although this paper focuses on four vowels, the current data is a subset of a larger set including a larger number of vowels, which were taken into account in the normalization procedure. Normalization was carried out using NORM (Thomas and Kendall, 2007). **Figure 4** displays the spectral characteristics of the native and L2 productions, averaged across the 38 L2 speakers (L2S; 33 females) and the 13 native English speakers (NES; seven females). As can be observed, as a group the L2 speakers seemed to produce English /ɜ:/ and /ɛ/ as two separate vowels and with values that were close to those of NES's. By contrast, the learners' production of /i:/ and /ɪ/ were much closer than the respective productions by native speakers.

A series of GLMMs were conducted<sup>3</sup> with F1 and F2 (Lobanov-normalized values) as the dependent variables to assess if the L2 speakers produced the English vowels differently and if the L2 production differed from NES production. In both cases the best fitting model was a GLMM with Group (learners vs. NES) and Vowel (the four target vowels) as fixed effects, and random intercepts for participant and for word. Regarding F1 values, the model revealed a significant effect of Vowel ( $F(3, 1,801) = 238.02, p < 0.001$ ) and a significant Group by Vowel interaction ( $F(3, 1,801) = 82.4, p < 0.001$ ), but no significant effect of Group ( $F(1, 1,801) = 2.488, p = 0.115$ ). Bonferroni-adjusted pairwise comparisons indicated that the L2 speakers differed from the native speakers with respect to /ɛ/, /i:/ and /ɪ/ ( $p < 0.001$ ), but not to /ɜ:/ ( $p = 0.15$ ), which explains the significant interaction. NES produced significantly different F1 values for all four vowels ( $p < 0.001$ ), while L2S produced significant differences between all vowels ( $p < 0.001$ ) except for the /i:/-/ɪ/ contrast ( $p = 0.335$ ). With respect to F2, all main effects and the interaction reached significance (Vowel:  $F(3, 1,801) = 235.92, p < 0.001$ ; Group:  $F(1, 1,801) = 4.6, p = 0.031$ ; Group x Vowel:  $F(3, 1,801) = 105.87, p < 0.001$ ). Pairwise comparisons in this case indicated that the L2 speakers differed from NES with respect to /ɜ:/, /i:/ and /ɪ/ ( $p < 0.001$ ), and /ɛ/ ( $p < 0.01$ ). The smaller level of significance for the group comparison involving /ɛ/ may explain the significant interaction. NES produced significantly different F2 values for all four vowels ( $p < 0.001$ ). As was found for F1, L2S produced a significant F2 difference between all vowels ( $p < 0.001$ ) except for the /i:/-/ɪ/ contrast ( $p = 0.08$ ). In brief, although L2 productions generally differed from NES's, the production of /ɛ/ and /ɜ:/ by L2 speakers deviated to a lesser extent from NES's production. In addition, learners produced significant spectral differences between all vowels except for the /i:/ and /ɪ/ pair.



Given previous findings about Spanish/Catalan speakers' reliance on temporal cues for the English /i:/-/ɪ/ contrast (Cebrian, 2006; Rallo Fabra and Romero, 2012), vowel duration was also examined. **Table 4** provides the means and standard deviations for each group and vowel, as well as the duration ratio for /i:/-/ɪ/ and /ɜ:/-/ε/. As can be observed, both groups produced the tense vowels with greater duration than lax vowels, but the duration difference was greater for NES. The results of a GLMM exploring vowel duration with Group and Vowel as fixed effects and a random slope for participants revealed a significant effect of Vowel ( $F(3, 1,802) = 469.25, p < 0.001$ ) and a significant Group by Vowel interaction ( $F(3, 1,802) = 58.41, p < 0.001$ ), but no significant effect of Group ( $F(1, 1,802) = 3.176, p = 0.075$ ).

Bonferroni-adjusted pairwise comparisons indicated that the L2 speakers differed from the native speakers in the duration of the tense vowels /ɜ:/ and /i:/ ( $p < 0.001$ ), but not regarding the lax vowels /ɪ/ ( $p = 0.304$ ) and /ε/ ( $p = 0.249$ ), which explains the significant interaction. Both groups produced highly significant duration differences between all vowels ( $p < 0.001$ ), except for the difference between /ε/-/ɪ/ for NES, which was significant at the  $p < 0.05$ , and the difference between /ε/-/i:/, which was not significant for the L2 speakers ( $p = 0.747$ ).

It has been argued that perceptual measures such as native speaker judgements are a more appropriate method for assessing L2 production and accentedness than acoustic analyses (e.g., Munro, 2008; Derwing and Munro, 2015). A preliminary

**TABLE 4 |** Mean vowel duration (in ms) and duration ratio for L2 speakers (L2S) and English native speakers (NES).

	L2S (n = 38)	NES (n = 13)
/i:/	154 (38)	190 (29)
/ɪ/	140 (29)	129 (24)
/ɜ:/	209 (38)	257 (35)
/ɛ/	153 (28)	141 (24)
/i:/-/ɪ/ ratio	1.11 (0.22)	1.50 (0.24)
/ɜ:/-/ɛ/ ratio	1.38 (0.18)	1.88 (0.33)

perceptual analysis of the production data was carried out. Four judges (native speakers of SSBE aged between 25 and 36) rated a selection of all the words produced by the L2 speakers and the native English controls (recall that the current data form part of a larger study involving a greater number of vowels). On a given trial raters first identified the vowel in the stimulus in terms of one of ten English words presented on a computer screen (including meet, sit, set and shirt) and then provided a goodness rating on a 7-point scale where 7 meant native-like. The results show that the L2 speakers' production of English /ɛ/ was very accurate, with 100% identification and an average goodness rating (GR) of 5.8 out of 7. Vowel /ɜ:/ received very high identification scores too, 89%, and a relatively high GR of 5. Vowel /i:/ was correctly identified 84% of the time with a GR of 4.9, and finally vowel /ɪ/ was the least accurately produced, reaching only 50% correct identification, with a GR of 4.3, and being heard as /i:/ 43% of the time with a GR of 4.3. Although these results are preliminary as so far only two judgements per stimulus have been obtained, they appear to confirm the acoustic analysis and point to the strongly assimilated vowel /ɛ/ as the most accurately produced vowel, followed by /ɜ:/ and /i:/, while /ɪ/ was the least accurately produced.

### Discussion of L2 Production Results

The production results show better production of /ɛ/ and /ɜ:/ than of /i:/ and /ɪ/. These results are consistent with previous studies involving Spanish learners of English. Flege et al. (1997) reported that two groups of Spanish speakers differing in the amount of time spent in the United States (0.5 vs. 9 years) produced English /ɛ/ very accurately (correctly identified between 91 and 99% of the time by native English judges), while /i:/ and /ɪ/ obtained similarly lower identification scores (51–61 and 57–69%, respectively). Cebrian (2007) also found that English /ɛ/ was the most accurately produced vowel by a group of 30 Spanish/Catalan speakers of English (86% correct identification and a 5.3/7 goodness rating by native English speakers), while /i:/ and /ɪ/ were less accurately produced (/i:/ 73% and GR 4.5/7; /ɪ/: 71%, GR: 3.9). The native speaker judgements were consistent with the acoustic analysis of L2 data, which revealed a large overlap in F1 and F2 space between /i:/ and /ɪ/, while /ɛ/ was more distinctly produced. Further, in a study involving Catalan learners of English, Rallo Fabra and Romero (2012) reported that the learners produced /ɪ/ more accurately than /i:/ (71% vs. 49% correct identification by native speakers). Regarding /ɜ:/, the current results are in agreement with Carlet and Cebrian's (2019) study involving

Spanish/Catalan learners of English, whose /ɜ:/ was judged to be comparatively accurate by native English judges. Raters in that study also judged /i:/ to be better produced than /ɪ/, in line with our preliminary native speaker judgements. Similarly, Carlet and Kivistö-de Souza (2018) found that the productions of /i:/ by Spanish/Catalan learners of English were judged to be more accurate (5.1 on 9-point Likert scale) than those of /ɪ/ (4.8). Overall, then, previous studies tend to find an accurate production of English /ɛ/, and a spectral overlap in the production of /i:/ and /ɪ/, which are found to be less accurately produced than /ɛ/.

Regarding vowel duration, the results of the current study show a smaller reliance on temporal cues for the /i:/-/ɪ/ contrast than previous studies involving Catalan/Spanish L2 English speakers (Cebrian, 2007; Aliaga-Garcia and Mora, 2009; Rallo Fabra and Romero, 2012), as the tense vowel was only 11% longer than the lax vowel (compared to 50% for NES). Still, there was variability among the participants. One of the SSBE native speakers did not produce a temporal difference between /i:/-/ɪ/. Regarding the L2 speakers, eight had duration ratio values for /i:/-/ɪ/ within the range of the SSBE speakers' values, while 15 had ratio of 1 or lower. In order to see if there was an inverse relationship between producing a temporal difference and a spectral difference, the relationship between the /i:/-/ɪ/ duration ratio and the spectral Euclidean distance between their production of /i:/ and /ɪ/ was examined. No significant correlation was found, indicating that those participants who relied more on spectral differences were not necessarily those who relied less on temporal differences.

In summary, the production of /ɛ/ follows predictions for a highly assimilated vowel that may pattern as a near identical vowel. Results for /ɜ:/ are unexpected for a comparatively new vowel that was poorly identified. Results for /i:/ and /ɪ/ show that, while /i:/ tends to be better produced, particularly as judged from a native perceptual perspective, it is not as accurately produced as /ɛ/ despite comparable assimilation scores for English /ɛ/ and /i:/ as Spanish /e/ and /i/, respectively. In fact, /i:/ patterns more with /ɪ/ than with /ɛ/ in production accuracy. The analysis also revealed that learners made use of temporal cues, but to a lesser extent than what has been reported in previous studies.

### Cross-Task Comparison. Correlational Analysis

One of the goals of this paper is to examine if individual variation in the perceived similarity between L1 and L2 sounds is related to perception and production accuracy, as would be predicted by the SLM-r (Flege and Bohn, 2021). In order to answer this question, a number of Spearman correlations were conducted relating cross-linguistic perceptual similarity measures and L2 perception and production measures. Specifically, the perceptual assimilation measures included in the analysis were percent assimilation, goodness ratings, and the composite fit index score for each vowel. Identification accuracy for each L2 vowel and discrimination accuracy for each vowel contrast were the L2 perception measures. Finally, two types of production measures were explored. First, the extent to which speakers distinguish the

two vowels in each contrast, that is, the Euclidean distances between /i:/ and /ɪ/, and between /ɜ:/ and /ɛ/. In addition, to explore the degree to which L2 speakers approximated native production, the Euclidean distance between each L2 individual's production of each vowel and the native English speakers' average production was also examined.

Only a subset of the possible correlations reached significance (see **Supplementary Material** in the Appendix for plots of the significant correlations). Accuracy in the identification of vowel /i:/ was moderately correlated with the percent assimilation of English /ɪ/ to Spanish /e/ ( $r_s = 0.331$ ,  $p = 0.036$ ,  $n = 43$ ) and negatively correlated with the fit index of English /ɪ/ to Spanish /i/ ( $r_s = -0.333$ ,  $p = 0.029$ ,  $n = 43$ ). Accurate /i:/-/ɪ/ discrimination was moderately correlated with the fit index of English /i:/ to Spanish /i/ ( $r_s = 0.364$ ,  $p = 0.016$ ,  $n = 43$ ). Similarly, /ɛ/-/ɜ:/ discrimination was related to a greater assimilation percentage of English /ɛ/ to Spanish /e/ ( $r_s = 0.335$ ,  $p = 0.028$ ,  $n = 43$ ). These results show that subjects who identified English /i:/ more successfully tended to assimilate English /ɪ/ to Spanish /e/ rather than to Spanish /i/, and that subjects who were more successful at discriminating English /i:/-/ɪ/ and English /ɛ/-/ɜ:/ tended to assimilate English /i:/ to Spanish /i/ and English /ɛ/ to Spanish /e/, respectively, more consistently. However, there was no correlation involving the perceived similarity results obtained for /ɪ/ and /ɜ:/ and the identification scores obtained for these vowels. No correlations were found between the cross-linguistic similarity measures and the production measures. On the other hand, moderate correlations were found between the results for different vowels in the same task. Identification of English /ɛ/ was correlated with the identification of /ɜ:/ ( $r_s = 0.323$ ,  $p = 0.035$ ,  $n = 43$ ) and of /ɪ/ ( $r_s = 0.329$ ,  $p = 0.031$ ,  $n = 43$ ). Within production, a greater Euclidean distance between /i:/ and /ɪ/ was negatively correlated with the degree of difference between an L2 speaker's production and the native English mean in the case of /i:/ production ( $r_s = -0.370$ ,  $p = 0.022$ ,  $n = 38$ ) and /ɪ/ production ( $r_s = -0.639$ ,  $p = 0.000$ ,  $n = 38$ ). This means that the more differently participants produced /ɪ/ and /i:/, the more their productions resembled the native speakers' production. Those participants who produced /ɪ/ more accurately (that is, with values closer to those of native speakers') also tended to produce /i:/ more accurately, though the correlation in this case was marginal ( $r_s = 0.320$ ,  $p = 0.05$ ,  $n = 38$ ).

In conclusion, perceived similarity does not seem to determine accuracy of L2 vowel perception or production at an individual level. The difficulty with the /i:/-/ɪ/ contrast, in particular, does not seem to follow from the pattern of assimilation of each of these two vowels to L1 categories. Some possible explanations for this fact are discussed next.

## GENERAL DISCUSSION

Previous studies have found that Catalan/Spanish learners of English, in addition to learners of other language backgrounds, have difficulty distinguishing the English /i:/-/ɪ/ contrast both in perception and in production (e.g., Flege, 1991; Flege et al., 1997; Escudero and Boersma, 2004; Cebrian, 2006; Cebrian,

2007; Kondaurova and Francis, 2008; Morrison, 2008; Morrison, 2009). L2 speech theories relate accurate perception and production of target language phones to the ability to discern differences between L1 and L2 sounds (Best, 1995; Flege, 1995; Escudero, 2005; Best and Tyler, 2007; Flege and Bohn, 2021). This study set out to examine if this difficulty is related to the degree of perceived similarity between L1 and L2 vowels, both in terms of group results and individual judgements of cross-linguistic similarity. To that effect, a group of 43 Catalan/Spanish bilinguals, who were undergraduate students in English Studies at a Spanish university, performed a perceptual assimilation task evaluating the perceived similarity between English /i:/, /ɪ/, /ɛ/, /ɜ:/ and the perceptually closest Spanish vowels. English /ɛ/, /ɜ:/ were included for comparison purposes. Previous studies had found that English /i:/ and /ɛ/ have a clear counterpart in the L1, Spanish /i/ and /e/, respectively, while English /ɪ/ and /ɜ:/ are less consistently mapped onto an L1 vowel (see **Table 1** above). The goal of the current study was thus to contrast the perceived similarity judgements obtained by a group of Spanish speakers with these same speakers' ability to identify the English vowels, discriminate each vowel pair (/i:/, /ɪ/, /ɛ/, /ɜ:/) and produce each vowel accurately. In addition, this study investigated if the individual variation often observed in perceived similarity was related to the variability found in L2 performance.

The results of the perceptual assimilation task confirmed previous findings (e.g. Cebrian, 2019) and showed that the English vowels differed in the degree to which they assimilated perceptually to Spanish categories. English /i:/ and /ɛ/ were strongly assimilated to Spanish /i/ and /e/ (93 and 99%, respectively). By contrast, English /ɪ/ and /ɜ:/ patterned as more dissimilar to an L1 vowel, with /ɜ:/ barely reaching a categorization threshold of 70% (72%) and /ɪ/ being slightly below that threshold (67%). Both /ɪ/ and /ɜ:/ obtained low goodness of fit scores, and consequently comparatively low fit indices (4/7 and 3.1/7, respectively; see **Table 2** above for all results). Based on their strong assimilation to Spanish /i/ and /e/, it was hypothesized that English /i:/ and /ɛ/ would be perceived and produced by the L2 learners with similarly high levels of accuracy. Regarding the poorly assimilated vowels, perception and production accuracy may depend on the extent to which the L2 speakers have detected differences between the L1 and the L2 phones to establish separate categories for these vowels (Flege, 1995). This in turn may depend on the amount of authentic L2 input received as well as the amount of L2 use (Flege and Bohn, 2021). Given that L2 learning took place in an instructional setting in the L1 country, with overall limited amount of L2 exposure and use, it is likely that learners were at a stage when the most dissimilar vowels are still challenging. In terms of discrimination ability, the PAM-L2 predicted that for L2 contrasts classified as uncategorized-categorized with partial overlap (/i:/-/ɪ/) and category-goodness assimilation (/ɛ/-/ɜ:/), discrimination accuracy depends on the degree of cross-language assimilation overlap (Faris et al., 2018; Tyler, 2021). Thus, the expectation was that /i:/-/ɪ/ would be better discriminated than /ɛ/-/ɜ:/, given the greater overlap



displayed by the latter (72% as Spanish /e/) than the former (32% as Spanish /i/).

The results of the vowel identification test showed that English /e/ was the most accurately identified vowel, while /ɜ:/ was the worst (90 and 47%, respectively). This result may follow from the assimilation PAT results. However, the results for /i:/ and /ɪ/ did not follow the expectations, and, in fact, the latter was better identified than the former (/ɪ/: 72%, /i:/: 58%). This outcome is in line with previous research on the effect of perceptual training that found that Catalan/Spanish learners of English identified /ɪ/ more successfully than /i:/ and /ɜ:/ (Cebrian and Carlet, 2014; Carlet and Cebrian, 2019; Fouz-González and Mompean, 2020) and that identification of /i:/ and /ɜ:/ improved the most after participants underwent perceptual training. The discrimination results also failed to support the predictions, as /e/-/ɜ:/ was more accurately discriminated than /i:/-/ɪ/ (87% vs. 66%, respectively). The discrimination results were replicated in production, as /e/ and /ɜ:/ were also more accurately produced than /i:/ and /ɪ/, which tended to be produced as /i:/. In the case of /e/, results were consistent across tasks and in agreement with previous studies involving Spanish L2 English speakers (Flege et al., 1997; Cebrian, 2007). Further studies could investigate if this Spanish vowel would be undetectable by native English listeners when produced in a native English context (Flege, 1992). In fact, our preliminary data from native English listeners yielded very high correct identification scores for the learners' /e/ production (100%, with an average goodness rating of 5.8/7). The results for /ɜ:/ showed a surprisingly high level of accuracy, in contrast with its poor identification results. Few studies have examined the production of this vowel by Spanish/Catalan learners, but Carlet and Cebrian (2019) obtained a similar result with a similar population (second-year undergraduate students in an English Studies degree, as opposed to first-year undergraduates in the current study). The relative success in the production of /ɜ:/ may indicate that learners are capable of producing a vowel that is different from other L2 (and L1) vowels, even if they cannot successfully identify it. This difference may be methodologically based, as identification involves the ability to distinguish the target sound from other potentially conflicting sounds (e.g., other response options present in the identification task), while production involved the articulation of the sound in a high frequency word, which may make production more successful. The idea that accurate perception precedes accurate production in L2 (e.g., models like Flege (1995) SLM or Best (1995) PAM) is not always supported by the findings (e.g., Llisterra, 1995), and more recent proposals suggest that L2 perception and production may develop without the requirement that one modality precedes the other (Flege and Bohn, 2021, SLM-r).

The more striking results involve the vowel contrast that was the focus of this study, the English /i:/-/ɪ/ contrast. In line with some previous studies, the results showed a high degree of perceptual confusion between /i:/ and /ɪ/, which achieved only 66% correct discrimination, and were often misidentified as one another in the identification task. In addition, the two vowels were produced with a large amount of spectral overlap. In terms of duration, L2 learners' tense vowels were longer than the lax vowels, but this difference in duration was smaller than the

duration difference produced by NES and also by other learners of English of a similar background reported in the literature (Cebrian, 2007; Rallo Fabra and Romero, 2012). The possible role of duration in the categorization of the English /i:/-/ɪ/ contrast is further discussed below.

One of the goals of this study was to assess if variability in L2 production and perception was related to individual differences in cross-linguistic perceived similarity. The acquisition of the English /i:/-/ɪ/ contrast by Spanish/Catalan learners of English provides a good ground to explore this issue as previous studies show inconsistent patterns of assimilation of English /ɪ/ to L1 vowels. Recall that English /ɪ/ is generally assimilated to Spanish /i/ and /e/ with varying degrees across studies. Assimilation of English /ɪ/ to Spanish /e/ ranges from 50% or less to 90% or more (see **Table 1** above). The current study found that English /ɪ/ was identified with Spanish /e/ two thirds of the time, and with /i/ one third of the time. An inspection of individual data showed that more than half the participants selected Spanish /e/ as the closest L1 vowel to English /ɪ/, while about a third selected Spanish /i/ and the remaining were not consistent and selected both Spanish /i/ and /e/. Therefore, it was possible that these differences in perceptual assimilation influenced the way learners perceived and produced the English vowels. For instance, participants who predominantly perceived English /ɪ/ as Spanish /e/ may be more successful at discriminating between /ɪ/ and /i:/ than those who assimilate both English /i:/ and /ɪ/ as Spanish /i/. However, the correlational analysis involving the perceptual similarity measures, the identification and discrimination results and the production results yielded very little evidence of a relationship between individuals' cross-linguistic perceived similarity and their L2 performance. Those learners who assimilated English /ɪ/ to Spanish /e/ and who judged English /i:/ to be closest to Spanish /i/, tended to be better at identifying English /i:/ and discriminating between /i:/ and /ɪ/. Still, no correlation involving the assimilation patterns obtained for /ɪ/ and /ɜ:/ and their identification or production accuracy were found. Therefore, the current study presents only a weak relationship between cross-linguistic perceived similarity and L2 performance at an individual level and hence it does not provide evidence to the claim that learners' accuracy in perception and production is related to individual variation in the perceived similarity between L1 and L2 sounds (Flege and Bohn, 2021). Future research is necessary to examine if a relationship can be found with other measures of similarity and of production and perception and with learners of different levels of experience.

In summary, the perception and production of the English /i:/-/ɪ/ contrast by the L2 speakers in the current study does not follow from the perceived similarity relationships between the target and the native vowels, neither at a group level nor at an individual level. This outcome is consistent with the results reported in previous studies involving Spanish and Catalan learners of English showing comparatively poor perception and production of both /i:/-/ɪ/, as discussed in *Results and Interim Discussions*. In fact, similar results are reported for learners of other L1 backgrounds like Italian and Japanese (e.g., Flege et al., 1998; Grenon et al., 2019).

What is it then about the English /i:/-/ɪ/ contrast that makes it difficult for L2 speakers to acquire? And why is it that a target vowel like English /i:/ that is perceptually very close to an L1 vowel is often found to be poorly perceived and produced? Flege (2018) explains that cases of inaccurate L2 performance that have been attributed to variables such as a late starting age of learning or number of years of L2 use can be accounted for in terms of the quality or quantity of the input, that is, whether learners are exposed to authentic native-like input or to variable and often accented input (see also Flege and Bohn, 2021). For example, Casillas (2015) reports that early Spanish-English bilinguals, who had been exposed to English since age 4–6 and were dominant in English as adults still differed from native English speakers in their use of spectral and temporal cues in their perception (but not the production) of English /i:/-/ɪ/. A possible explanation offered is exposure to Spanish-accented English. Morrison (2012), who also found that Spanish speakers identified English /i:/ and /ɪ/ with Spanish /i/ and /e/, respectively, argued that a perceptual explanation for Spanish speakers' difficulty with the /i:/-/ɪ/ contrast is not satisfactory and points to the effects of pronunciation instruction and orthography as possible causes. Indeed, orthography has been found to influence the pronunciation of L2 words (e.g., Morrison, 2009; Escudero and Wanrooij, 2010). This influence may be particularly notable in foreign language learning contexts where L2 words are often first encountered in writing. For example, the fact that the English lax vowel /ɪ/ is typically spelt with the letter <i>, which represents the sound /i/ in many languages like Spanish, Catalan and Italian, may result in the mispronunciation of this vowel (Morrison, 2008). English /e/, on the other hand, is often spelt with the same letter as in Spanish, <e>. The effect of orthography is likely enhanced by the presence of many cognate words, particularly for speakers of Romance languages (e.g., words containing <i> such as *cinema*, *city*, or *minimum*). Further, the tense vowel /i:/ is often represented by a two-letter grapheme, e.g., <ee> in *feet* or <ea> in *beat*, which may suggest a longer duration. In addition, the effect of explicit instruction that describes the contrast as merely a duration contrast (e.g., long /i/ vs. short /ɪ/) may result in a misrepresentation of the tense-lax contrast as a purely duration contrast (Flege et al., 1997; Wang and Munro, 1999).

As discussed in *Introduction*, L2 English speakers' have been found to rely on duration to implement the English /i:/-/ɪ/ (e.g., Escudero and Boersma, 2004; Cebrian, 2006; Kondaurova and Francis, 2008). A combination of factors may account for this fact. In addition to the issues described above, there are linguistic factors such as sensitivity to phonetic temporal distinctions in the L1, e.g., as a result of stress or final obstruent voicing (Kondaurova and Francis, 2008), desensitization to spectral contrasts not used in the L1 (Bohn, 1995), and statistical learning of the characteristics of each vowel with the consequent detection that /i:/ tends to be longer than /ɪ/ (Escudero and Boersma, 2004; Morrison, 2008). For example, Morrison suggests that learners go through different stages in the process of learning the /i:/-/ɪ/ contrast which go from the inability to discern the two vowels, to establishing a duration contrast, to gradually detecting and incorporating spectral differences. Following this interpretation, we can speculate that when learners start to detect spectral differences between English

/ɪ/ and Spanish or Catalan /i/, their L2 category starts to drift toward a more /ɪ/-like vowel. However, since at this stage both L2 vowels, /i:/-/ɪ/, share a single spectral category, possibly distinguished temporally, the drift toward /ɪ/ affects the perception and production of not only /ɪ/, but also /i:/. This would result in a "deterioration" of /i:/, a vowel that by itself should not be problematic given the high assimilation rates to an L1 vowel. In a similar vein, Major (1987) found that Brazilian Portuguese speakers produced the similar English vowel /e/ with increasing less accuracy as their production of the new vowel /æ/ improved. In any event, the comparatively poor performance with English /i:/ shows that L2 vowels are not learned individually but as part of a system of contrasting phones. This factor, together with the effect of non-linguistic factors like the roles of orthography and explicit instruction, may account for why cross-linguistic similarity may not always make the right predictions, and why a target vowel that should in principle be "easy" to learn on the basis of its strong perceptual assimilation to an L1 vowel, like English /i:/ for Spanish speakers, is not accurately perceived and produced in the L2.

## IMPLICATIONS, LIMITATIONS AND CONCLUSION

The observations made in the previous section have some pedagogical implications for the teaching and learning of this vowel contrast. First of all, the contrast between English /i:/ and /ɪ/ should be set as a priority in L2 speech teaching given the contrast's high functional load, and, consequently, the fact that mispronunciation may contribute to loss of intelligibility and reduction of comprehensibility [Brown, 1988; Levis, 2018; Munro, 2021 (this volume); O'Brien, 2021 (this volume)]. Secondly, instructors should avoid characterizing the English vowels /i:/ and /ɪ/ as "long" and "short" respectively, as this may lead to an inaccurate simplification of the difference between them and may make learners ignore vowel quality differences. Although the duration difference is an important feature to illustrate, the focus should be on the qualitative difference. In that sense, the results of cross-linguistic perceived similarity tasks could be useful. For instance, given that English /ɪ/ is perceptually closer to Spanish /e/ than to Spanish /i/, learners' attention could be drawn to this L1 vowel quality difference as a way of helping learners detect differences between English /i:/ and /ɪ/. On a similar note, cognate words could be used to illustrate differences in L1 and L2 pronunciation (e.g., Spanish "lista, video" and English "list, video"). Further, instruction should avoid exercises involving the written form of words, at least at initial stages, and concentrate on auditory input as much as possible, in order to avoid the confusion that spelling might cause in the acquisition of this contrast. The use of phonetic symbols is a possible option; Fouz-González and Mompean (2020) report that both the use of keywords and phonetic symbols in high variability phonetic training enhances identification accuracy of L2 vowels. See O'Brien (2021) (this volume) for an overview of approaches to and suggestions for pronunciation teaching. Finally, note that, as Munro (2021) (this volume) demonstrates, while there is some systematicity in the L2 pronunciation difficulties observed for a given L1 group, there is

considerable variation in the pronunciation difficulties experienced by different individuals. This fact underscores the importance of individual pronunciation diagnosis and pronunciation practice targeting individual difficulties.

The current study has some limitations that should be acknowledged. The first limitation is the cross-sectional nature of the current study. A single group of learners with a similar level of proficiency was tested at a single point in time. Additional groups differing in amount of experience or a longitudinal approach would be necessary to fully evaluate the relationship between perceived similarity and L2 performance. In fact, L2 experience has been found to influence the way learners categorize the /i:/-/ɪ/ contrast (Aliaga-García and Mora, 2009; Ylinen et al., 2009; Grenon et al., 2019). For instance, Grenon et al. found that a subset of the inexperienced Japanese learners of English trained to attend to spectral differences between English /i:/ and /ɪ/ managed to modify their original reliance on duration and to attend to spectral cues, although their perception of the English contrast was not completely native-like. Another shortcoming involves the lack of L1 production data to compare to the L2 data so as to have a measure of how differently the L1 and L2 vowels are produced, particularly when evaluating cases of near-identical L1-L2 vowels. An additional limitation is related to the bilingual nature of the participants in this study, as, in addition to speaking Spanish, many participants spoke Catalan, with varying degrees of dominance (see *Participants*). Given the difference in vowel inventory between the two languages (five Spanish vowels vs. seven Catalan vowels), it is difficult to determine, for instance, if the successful perception and production of English /ɛ/ is related to the perceived similarity between this English vowel and Spanish /e/ or Catalan /ɛ/<sup>4</sup>. Even if differences in Catalan/Spanish dominance were not found to affect similarity judgements (see *Perceptual Assimilation Tasks Results*), and the results of the PAT in the current study closely replicated the results obtained by Cebrian (2019) with monolingual Spanish speakers, there is a potential effect of speaking more than one language on L2 perception and production. Further, production data was mainly measured acoustically instead of by means of native speaker judgements, a measure that has been argued to be more appropriate for measuring L2 production and accentedness (Munro, 2008). Finally, the limited number of vowels examined, two vowel contrasts, prevents the comparison of the current results with those for other vowel combinations, for example in the discrimination task. These issues are left for further research.

To conclude, the perception and production of L2 phones do not always follow from cross-linguistic similarity relations, contrary to the predictions of most current L2 speech models. In addition, no clear explanation has been found based on individual variation, as variability in L2 to L1 mapping does not seem to be related to the degree of L2 perception or production accuracy, at least with the measures and the level of proficiency evaluated in this study. Still, some sounds may be

readily perceived and produced successfully in terms of the L1 category (e.g., English /ɛ/ by Spanish speakers). Further, similarity predictions may differ for production and perception accuracy, as was the case of /ɜ:/, a vowel that was very poorly perceived but relatively accurately produced, which may be related to different demands for perception and production tasks. The results for the /i:/-/ɪ/ contrast suggest that different factors are at play in the acquisition of this L2 contrast, including non-native overreliance on acoustic cues (triggered by statistical learning or by metalinguistic information present in pronunciation instruction), the effect of orthography, and the need to maintain an L2 contrast with a high functional load. These factors may override the effect of cross-linguistic similarity, as the need to categorize the most dissimilar sound may bring with it the deterioration of the more similar sound, at least at some stages. These issues may be behind the reason why a presumably “easy” sound like English /i:/ may be more challenging than expected.

## DATA AVAILABILITY STATEMENT

Data from this study are available by request if required to form part of a metaanalysis or similar form of research. Requests to access the datasets should be directed to juli.cebrian@uab.cat.

## ETHICS STATEMENT

Ethical review and approval was not required for the study on human participants in accordance with the local legislation and institutional requirements. The patients/participants provided their written informed consent to participate in this study.

## AUTHOR CONTRIBUTIONS

JC contributed to conception and design of the study. JC, CG and NG collected the data, analyzed the results and performed the statistical analysis. JC wrote the first draft of the manuscript. CG and NG wrote sections of the manuscript.

## FUNDING

This work was supported by research grants FFI2017-88016-P the Spanish Ministry of Economy and Competitiveness, and by the research group Grant 2017SGR34 from the Catalan Government.

## SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcomm.2021.660917/full#supplementary-material>

<sup>4</sup>See Cebrian (2021) for a study on the perceived similarity between a near-complete set of Catalan and SSBE vowels.

## REFERENCES

- Adank, P., Smits, R., and van Hout, R. (2004). A Comparison of Vowel Normalization Procedures for Language Variation Research. *J. Acoust. Soc. Am.* 116, 3099–3107. doi:10.1121/1.1795335
- Aliaga-García, C. (2011). “Measuring Perceptual Cue Weighting after Training: A Comparison of Auditory vs Articulatory Training Methods,” in *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Editors M. Wrembel, M. Kul, and K. Dziubalska-Kolaczyk (Frankfurt am Main, Germany: Peter Lang), 15–28.
- Aliaga-García, C., and Mora, J. C. (2009). “Assessing the Effects of Phonetic Training on L2 Sound Perception and Production,” in *Recent Research in Second Language Phonetics/Phonology: Perception and Production*. Editors M. A. Watkins, A. S. Rauber, and B. O. Baptista (Newcastle upon Tyne, United Kingdom: Cambridge Scholars Publishing), 2–31.
- Aoyama, K., Flege, J. E., Guion, S. G., Akahane-Yamada, R., and Yamada, T. (2004). Perceived Phonetic Dissimilarity and L2 Speech Learning: The Case of Japanese /r/ and English /l/ and /r/. *J. Phon.* 32 (2), 233–250. doi:10.1016/S0095-4470(03)00036-6
- Baigorri, M., Campanelli, L., and Levy, E. S. (2019). Perception of American-English Vowels by Early and Late Spanish-English Bilinguals. *Lang. Speech.* 62 (4), 681–700. doi:10.1177/0023830918806933
- Best, C. T. (1995). “A Direct Realist View of Cross-Language Speech Perception,” in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Editor W. Strange (Timonium, MD: York Press), 171–204.
- Best, C. T., and Strange, W. (1992). Effects of Phonological and Phonetic Factors on Cross-Language Perception of Approximants. *J. Phon.* 20 (3), 305–330. doi:10.1016/S0095-4470(19)30637-0
- Best, C. T., and Tyler, M. D. (2007). “Nonnative and Second-Language Speech Perception,” in *Honor of James Emil Flege*. Editors O. S. Bohn and M. J. Munro (Amsterdam, Netherlands: John Benjamins), 13–34. doi:10.1075/llt.17.07bes
- Birdsong, D., Gertken, L. B., and Amengual, M. (2012). *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism*. Austin, TX: COERLL, University of Texas at Austin.
- Boersma, P., and Weenink, D. (2018). Praat: Doing Phonetics by Computer [Computer Program]. version 6.0.43. Available at: <http://www.praat.org/> (Accessed September 8, 2018).
- Bohn, O.-S. (1995). “Cross-Language Speech Perception in Adults: First Language Transfer Doesn’t Tell it All,” in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Editor W. Strange (Timonium, MD: York Press), 279–304.
- Bohn, O.-S., and Flege, J. E. (1992). The Production of New and Similar Vowels by Adult German Learners of English. *Stud. Second Lang. Acquis.* 14 (02), 131–158. doi:10.1017/S0272263100010792
- Bohn, O.-S., and Steinlen, A. K. (2003). “Consonantal Context Affects Cross-Language Perception of Vowels,” in Proceedings of the XV International Congress of Phonetic Sciences. Editors D. Recasens, M. J. Solé, and J. Romero (Barcelona, Spain: Causal Productions), 2289–2292.
- Bradlow, A. R., Pisoni, D. B., Akahane-Yamada, R., and Tohkura, Y. i. (1997). Training Japanese Listeners to Identify English /r/ and /l/: IV. Some Effects of Perceptual Learning on Speech Production. *J. Acoust. Soc. Am.* 101, 2299–2310. doi:10.1121/1.418276
- Brown, A. (1988). Functional Load and the Teaching of Pronunciation. *TESOL Q.* 22 (4), 593–606. doi:10.2307/3587258
- Bundgaard-Nielsen, R. L., Best, C. T., and Tyler, M. D. (2011). Vocabulary Size Matters: The Assimilation of Second-Language Australian English Vowels to First-Language Japanese Vowel Categories. *Appl. Psycholinguist.* 32, 51–67. doi:10.1017/S0142716410000287
- Busà, M. G. (1992). “On the Production of English Vowels by Italian Speakers with Different Degrees of Accent,” in *New Sounds 92*, Proceedings of the 1992 Amsterdam Symposium on the Acquisition of Second-Language Speech. Editors J. Leather and J. Allan (Amsterdam, Netherlands: University of Amsterdam), 47–63.
- Carlet, A., and Cebrian, J. (2019). “Assessing the Effect of Perceptual Training on L2 Vowel Identification, Generalization and Long-Term Effects,” in *A Sound Approach to Language Matters. In Honor of Ocke-Schwen Bohn*. Editors A. M. Nyvad, M. Hejná, A. Højen, A. B. Jespersen, and M. H. Sørensen (Aarhus, Denmark: Department of English, School of Communication and Culture, Aarhus University), 91–119.
- Carlet, A., and Kivistö-de Souza, H. (2018). Improving L2 Pronunciation Inside and Outside the Classroom. *Ilha do Desterro.* 71 (3), 99–124. doi:10.5007/2175-8026.2018v71n3p99
- Casillas, J. (2015). Production and Perception of the /i/-/I/ Vowel Contrast: The Case of L2-Dominant Early Learners of English. *Phonetica* 72, 182–205. doi:10.1159/000431101
- Cebrian, J. (2006). Experience and the Use of Non-Native Duration in L2 Vowel Categorization. *J. Phon.* 34, 372–387. doi:10.1016/j.wocn.2005.08.003
- Cebrian, J. (2007). “Old Sounds in New Contrasts: L2 Production of the English Tense-Tax Vowel Distinction,” in Proceedings of the 16th International Congress of Phonetic Sciences. Editors J. Trouvain and W. J. Barry (Saarbrücken, Germany: University of Saarland), 1637–1640.
- Cebrian, J. (2019). Perceptual Assimilation of British English Vowels to Spanish Monophthongs and Diphthongs. *J. Acoust. Soc. Am.* 145, EL52–EL58. doi:10.1121/1.5087645
- Cebrian, J. (2021). Perception of English and Catalan Vowels by English and Catalan Listeners: A Study of Reciprocal Cross-Linguistic Similarity. *J. Acoust. Soc. Am.* 149 (4), 2671–2685. doi:10.1121/10.0004257
- Cebrian, J., and Carlet, A. (2014). Second-Language Learners’ Identification of Target-Language Phonemes: A Short-Term Phonetic Training Study. *Can. Mod. Lang. Rev.* 70 (4), 474–499. doi:10.3138/cmlr.2318
- Cerviño, E., and Mora, J. C. (2009). Duration as a Phonetic Cue in the Categorization of /i:/-/I/ and /s/-/z/ by Spanish/Catalan Learners of English. *J. Acoust. Soc. Am.* 125, 2764. doi:10.1121/1.4784683
- Derwing, T. M., and Munro, M. J. (2015). A Prospectus for Pronunciation Research in the 21st Century. A Point of View. *J. Second Lang. Pronunciation* 1 (1), 11–42. doi:10.1075/jslp.1.1.01mun
- Escudero, P. (2005). Linguistic Perception and Second Language Acquisition: Explaining the Attainment of Optimal Phonological Categorization. PhD Thesis. Amsterdam (Netherlands): Netherlands Graduate School of Linguistics.
- Escudero, P., and Boersma, P. (2004). Bridging the Gap Between L2 Speech Perception Research and Phonological Theory. *Stud. Second Lang. Acquis.* 26 (4), 551–585. doi:10.1017/S0272263104040021
- Escudero, P., and Chládková, K. (2010). Spanish Listeners’ Perception of American and Southern British English Vowels. *J. Acoust. Soc. Am.* 128, EL254–EL260. doi:10.1121/1.3488794
- Escudero, P., and Wanrooij, K. (2010). The Effect of L1 Orthography on Non-Native Vowel Perception. *Lang. Speech.* 53 (3), 343–365. doi:10.1177/0023830910371447
- Faris, M. M., Best, C. T., and Tyler, M. D. (2018). Discrimination of Uncategorised Non-native Vowel Contrasts is Modulated by Perceived Overlap with Native Phonological Categories. *J. Phon.* 70, 1–19. doi:10.1016/j.wocn.2018.05.003
- Flege, J. E. (1991). The Interlingual Identification of Spanish and English Vowels: Orthographic Evidence. *Q. J. Exp. Psychol. A.* 43 (3), 701–731. doi:10.1080/14640749108400993
- Flege, J. E. (1992). “Speech Learning in a Second Language,” in *Phonological Development: Models, Research, Implications*. Editors C. A. Ferguson, L. Menn, and C. Stoel-Gammon (Timonium, MD: York Press), 565–604.
- Flege, J. E. (1995). “Second Language Speech Learning: Theory, Findings and Problems,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*. Editor W. Strange (Baltimore, MD: York Press), 233–277.
- Flege, J. E. (2003). “Assessing Constraints on Second-Language Segmental Production and Perception,” in *Phonetics and Phonology in Language Comprehension and Production*. Editors A. Meyer and N. Schiller (Berlin, Germany: Mouton de Gruyter), 319–355.
- Flege, J. E. (2018). It’s Input That Matters Most, Not Age. *Biling. Lang. Cogn.* 21 (5), 919–920. doi:10.1017/S136672891800010X
- Flege, J. E., Frieda, E. M., Walley, A. C., and Fandazza, L. A. (1998). Lexical Factors and Segmental Accuracy in Second Language Speech Production. *Stud. Second Lang. Acquis.* 20 (2), 155–187. doi:10.1017/S0272263198002034
- Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of Experience on Non-Native Speakers’ Production and Perception of English Vowels. *J. Phon.* 25 (4), 437–470. doi:10.1006/jpho.1997.0052
- Flege, J. E., and Bohn, O.-S. (1989). The Perception of English Vowels by Native Spanish Speakers. *J. Acoust. Soc. Am.* 85 (1), S85. doi:10.1121/1.2027177
- Flege, J. E., and Bohn, O.-S. (2021). “The Revised Speech Learning Model (SLM-R),” in *Second Language Speech Learning: Theoretical and Empirical Progress*. Editor R. Wayland (Cambridge, United Kingdom: Cambridge University Press), 3–83.
- Flege, J. E., Schirru, C., and MacKay, I. R. A. (2003). Interaction Between the Native and Second Language Phonetic Subsystems. *Speech Commun.* 40 (4), 467–491. doi:10.1016/S0167-6393(02)00128-0



- Fouz-González, J., and Mompean, J. A. (2020). Exploring the Potential of Phonetic Symbols and Keywords as Labels for Perceptual Training. *Stud. Second Lang. Acquis.* 1–32. doi:10.1017/S0272263120000455
- Fullana-Rivera, N., and MacKay, I. R. (2003). “Production of English Sounds by EFL Learners: The Case of /i/ and /i:/,” in Proceedings of the 15th International Congress of Phonetic Sciences, ICPhS. Editors M. J. Solé, D. Recasens, and J. Romero (Barcelona, Spain: Universitat Autònoma de Barcelona), 1525–1528.
- Gilichinskaya, Y. D., and Strange, W. (2010). Perceptual Assimilation of American English Vowels by Inexperienced Russian Listeners. *J. Acoust. Soc. Am.* 128, EL80–EL85. doi:10.1121/1.3462988
- Grenon, I., Kubota, M., and Sheppard, C. (2019). The Creation of a New Vowel Category by Adult Learners After Adaptive Phonetic Training. *J. Phon.* 72, 17–34. doi:10.1016/j.wocn.2018.10.005
- Guion, S. G., Flege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). An Investigation of Current Models of Second Language Speech Perception: The Case of Japanese Adults’ Perception of English Consonants. *J. Acoust. Soc. Am.* 107 (5), 2711–2724. doi:10.1121/1.428657
- Hacquad, V., Walter, M. A., and Marantz, A. (2007). The Effects of Inventory on Vowel Perception in French and Spanish: An MEG Study. *Brain Lang.* 100 (3), 295–300. doi:10.1016/j.bandl.2006.04.009
- Higgins, J. (2018). Minimal Pairs for English RP: Lists by John Higgins. Available at: <http://minimal.marlogde.net/minimal.html> (Accessed January 29, 2021).
- Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). Some Effects of Duration on Vowel Recognition. *J. Acoust. Soc. Am.* 108 (6), 3013–3022. doi:10.1121/1.1323463
- Højen, A., and Flege, J. E. (2006). Early Learners’ Discrimination of Second-Language Vowels. *J. Acoust. Soc. Am.* 119 (5), 3072–3084. doi:10.1121/1.2184289
- Holt, L. L., and Lotto, A. J. (2006). Cue Weighting in Auditory Categorization: Implications for First and Second Language Acquisition. *J. Acoust. Soc. Am.* 119 (5), 3059–3071. doi:10.1121/1.2188377
- IBM Corp (2017). *IBM SPSS Statistics for Windows*. Version 25.0. Armonk, NY: IBM Corp.
- Jun, S.-A., and Cowie, I. (1994). Interference for ‘New’ and ‘Similar’ Vowels in Korean Speakers of English. *Ohio State Univ. Working Papers*. 43, 117–130.
- Kivistö-de Souza, H., Carlet, A., Julkowska, I., and Rato, A. (2017). Vowel Inventory Size Matters: Assessing Cue-Weighting in L2 Vowel Perception. *Ilha do Desterro*. 70 (3), 33–46. doi:10.5007/2175-8026.2017v70n3p33
- Kondaurova, M. V., and Francis, A. L. (2008). The Relationship Between Native Allophonic Experience With Vowel Duration and Perception of the English Tense/Lax Vowel Contrast by Spanish and Russian Listeners. *J. Acoust. Soc. Am.* 124 (6), 3959–3971. doi:10.1121/1.2999341
- Kuhl, P. K., and Iverson, P. (1995). “Linguistic Experience and the Perceptual Magnet Effect,” in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Editor W. Strange (Timonium, MD: York Press), 121–154.
- Lado, R. (1957). *Linguistics Across Cultures: Applied Linguistics for Language Teachers*. Ann Arbor, MI: University of Michigan Press.
- Levis, J. M. (2018). *Intelligibility, Oral Communication, and the Teaching of Pronunciation*. Cambridge, United Kingdom: Cambridge University Press.
- Levy, E. S. (2009b). Language Experience and Consonantal Context Effects on Perceptual Assimilation of French Vowels by American-English Learners of French. *J. Acoust. Soc. America*. 125, 1138–1152. doi:10.1121/1.3050256
- Levy, E. S. (2009a). On the Assimilation-Discrimination Relationship in American English Adults’ French Vowel Learning. *J. Acoust. Soc. Am.* 126 (5), 2670–2682. doi:10.1121/1.3224715
- Llisterri, J. (1995). “Relationships between Speech Production and Speech Perception in a Second Language,” in Proceedings of the 13th International Congress of Phonetic Sciences. Editors K. Elenius and P. Branderud (Stockholm, Sweden: KTH and Stockholm University), 92–99.
- Lobanov, B. M. (1971). Classification of Russian Vowels Spoken by Different Speakers. *J. Acoust. Soc. Am.* 49, 606–608. doi:10.1121/1.1912396
- Major, R. C. (1987). English Voiceless Stop Production by Speakers of Brazilian Portuguese. *J. Phon.* 15 (2), 197–202. doi:10.1016/S0095-4470(19)30560-1
- Mora, J. C., and Fullana, N. (2007). “Production and Perception of English /i:/-/i/ and /æ:/- / / in a Formal Setting: Investigating the Effects of Experience and Starting Age,” in Proceedings of the 16th International Congress of Phonetic Sciences 3. Editors J. Trouvain and W. Barry (Saarbrücken, Germany: University of Saarland), 1613–1616.
- Morrison, G. S. (2008). L1-Spanish Speakers’ Acquisition of the English /i /-/i/ Contrast: Duration-Based Perception is Not the Initial Developmental Stage. *Lang. Speech*. 51 (4), 285–315. doi:10.1177/0023830908099067
- Morrison, G. S. (2009). L1-Spanish Speakers’ Acquisition of the English /i /-/i/ Contrast II: Perception of Vowel Inherent Spectral Change. *Lang. Speech*. 52 (4), 437–462. doi:10.1177/0023830909336583
- Morrison, G. S. (2012). Perception of Natural Vowels by Monolingual Canadian-English, Mexican-Spanish, and Peninsular-Spanish Listeners. *Can. Acoust.* 40 (4), 29–40. doi:10.1121/1.3587895
- Munro, M. J. (2008). “7. Foreign Accent and Speech Intelligibility,” in *Phonology and Second Language Acquisition*. Editors Hansen Edwards, J. G., and Zampini, M. L. (Burnaby, Canada: Simon Fraser University), 193–218.
- Munro, M. J. (Forthcoming 2021). On the Difficulty of Defining “Difficult” in Second-Language Vowel Acquisition. *Front. Commun.* doi:10.3389/fcomm.2021.639398
- Munro, M. J., Flege, J. E., and MacKay, I. R. A. (1996). The Effects of Age of Second Language Learning on the Production of English Vowels. *Appl. Psycholinguist.* 17, 313–334. doi:10.1017/s0142716400007967
- O’Brien, M. G. (2021). Ease and Difficulty in L2 Pronunciation Teaching: A Mini-Review. *Front. Commun.* 5, 1–7. doi:10.3389/fcomm.2020.626985
- Piske, T., MacKay, I. R. A., and Flege, J. E. (2001). Factors Affecting Degree of Foreign Accent in an L2: A Review. *J. Phon.* 29 (2), 191–215. doi:10.1006/jpho.2001.0134
- Rallo Fabra, L., and Romero, J. (2012). Native Catalan Learners’ Perception and Production of English Vowels. *J. Phon.* 40, 491–508. doi:10.1016/j.wocn.2012.01.001
- Recasens, D., and Espinosa, A. (2009). Dispersion and Variability in Catalan Five and Six Peripheral Vowel Systems. *Speech Commun.* 51, 240–258. doi:10.1016/j.specom.2008.09.002
- Rochet, B. (1995). “Perception and Production of Second-Language Speech Sounds by Adults,” in *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Editor W. Strange (Timonium, MD: York Press), 379–410.
- Strange, W. (1995). *Speech Perception and Linguistic Experience: Issues in Cross Language Research*. Timonium, MD: York Press.
- Strange, W., Hisagi, M., Akahane-Yamada, R., and Kubo, R. (2011). Cross-Language Perceptual Similarity Predicts Categorial Discrimination of American Vowels by Naïve Japanese Listeners. *J. Acoust. Soc. Am.* 130, EL226–EL231. doi:10.1121/1.3630221
- Strange, W., Levy, E. S., and Law, F. F. (2009). Cross-Language Categorization of French and German Vowels by Naïve American Listeners. *J. Acoust. Soc. Am.* 126 (3), 1461–1476. doi:10.1121/1.3179666
- Thomas, E., and Kendall, T. (2007). NORM: The Vowel Normalization and Plotting Suite. Available at: <http://ncslaap.lib.ncsu.edu/tools/norm/> (Accessed January 29, 2021).
- Trubetzkoy, N. (1939). *Principles of Phonology*. Berkeley, CA: University of California Press.
- Tyler, M. D. (2021). “Phonetic and Phonological Influences on the Discrimination of Non-Native Phones,” in *Second Language Speech Learning: Theoretical and Empirical Progress*. Editor R. Wayland (Cambridge, United Kingdom: Cambridge University Press), 157–174. doi:10.1017/9781108886901.005
- Tyler, M. D., Best, C. T., Faber, A., and Levitt, A. G. (2014). Perceptual Assimilation and Discrimination of Non-Native Vowel Contrasts. *Phonetica* 71, 4–21. doi:10.1159/000356237
- Wang, X., and Munro, M. J. (1999). “The Perception of English Tense-Lax Vowel Pairs by Native Mandarin Speakers: The Effect of Training on Attention to Temporal and Spectral Cues,” in Proceedings of the 14th International Congress of Phonetic Sciences. Editors J. Ohala, Y. Hasegawa, M. Ohala, D. Granville, and A. Baily (San Francisco, CA: University of California, Berkeley), 125–129.
- Ylinen, S., Uther, M., Latvala, A., Vepsäläinen, S., Iverson, P., Akahane-Yamada, R., et al. (2009). Training the Brain to Weight Speech Cues Differently: A Study of Finnish Second-Language Users of English. *J. Cogn. Neurosci.* 22 (6), 1319–1332. doi:10.1162/jocn.2009.21272

**Conflict of Interest:** The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Cebrian, Gorba and Gavalda. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.