

Article

Online Assessment of Cross-Linguistic Similarity as a Measure of L2 Perceptual Categorization Accuracy

Juli Cebrian ^{1,*}  and Joan C. Mora ² 

¹ Department of English Philology and German Studies, Universitat Autònoma de Barcelona (UAB), Campus UAB, 08193 Bellaterra, Spain

² Department of Modern Languages and Literatures and English Studies, Universitat de Barcelona, 08007 Barcelona, Spain; mora@ub.edu

* Correspondence: juli.cebrian@uab.cat

Abstract: The effect of cross-linguistic similarity on the development of target-like categories in a second or additional language is widely attested. Research also shows that second-language speakers may access both their native and the second-language lexicons when processing second-language speech. Forty-three Catalan learners of English performed a perceptual assimilation task evaluating the perceived similarity between English and Catalan vowels and also participated in a visual world eye-tracking experiment investigating between-language lexical competition. The focus of the study was the English vowel contrasts /i:/-/ɪ/ and /æ/-/ʌ/. The perceptual task confirmed that English /i:/ and /æ/ were perceptually closer to native Catalan categories than English /ɪ/ and /ʌ/. The results of the spoken word recognition task indicated that learners experienced greater competition from native words when the target words contained English /i:/ and /æ/, illustrating a close link between the two types of tasks. However, differences in the magnitude of cross-language lexical competition were found to be only weakly related to learners' degree of perceived similarity to native categories at an individual level. We conclude that online tasks provide a potentially effective method of assessing cross-linguistic similarity without the concerns inherent to more traditional offline approaches.

Keywords: cross-linguistic similarity; perceptual assimilation; lexical access; cross-language competition; interlingual homophones; visual world paradigm; L2 vowel learning; English vowel contrasts



Citation: Cebrian, Juli, and Joan C. Mora. 2024. Online Assessment of Cross-Linguistic Similarity as a Measure of L2 Perceptual Categorization Accuracy. *Languages* 9: 152. <https://doi.org/10.3390/languages9050152>

Academic Editors: Paolo Mairano and Sandra Schwab

Received: 30 September 2023

Revised: 4 April 2024

Accepted: 12 April 2024

Published: 23 April 2024



Copyright: © 2024 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The study of the acquisition of second or foreign language (L2) speech has experienced great development in the last few decades. A subfield within the broad area of the speech sciences, research in L2 learning has generally drawn from speech analysis techniques used to analyze native speech. However, these methodologies may not always be well suited to the study of L2 learning given the latter's distinct characteristics, such as the large amount of variability across and within speakers, the dynamic nature of the learning process, and the need to consider pedagogical applications for the teaching and learning of pronunciation.

One crucial aspect of non-native speech is the fact that the L2, by definition, develops alongside an already acquired language (or languages), which inevitably affects the development of the L2 under the assumption that the L1 and subsequently learned languages coexist in the same phonological space (Kuhl et al. 2008; Flege and Bohn 2021). One dimension in which this is evident is the way in which the sounds of the language being learned, the target language (TL), are perceived in terms of native sound categories, at least at the initial stages of L2 acquisition (Best 1995; Flege 1995). When processing L2 speech sounds, learners initially rely on the speech processing skills and the language-specific phonetic cue-weightings they automatized to process L1 speech in a maximally efficient, robust, effortless manner (Best and Tyler 2007; Flege and Bohn 2021).

A crucial component of L1-based processing that has been shown to be an essential determinant of L2 segmental learning is cross-language perceptual similarity. Differences in perceptual similarity between L2 and L1 sounds determine, to a large extent, the difficulty in the acquisition of L2 sound categories and have been shown to predict the perception and production accuracy of L2 sounds. For example, the English low vowels in *cat* /kæt/, *cut* /kʌt/, and *cart* /kɑ:t/, all of which tend to be perceptually mapped onto the only Spanish low vowel category /a/ by Spanish learners of English, do not present the same degree of perceptual similarity with respect to Spanish /a/. English /æ/ is perceived to be much closer to Spanish /a/ than English /ɑ:/, which is perceived to be a worse match to Spanish /a/ than English /ʌ/ (Cebrian 2019)¹. The larger the perceived distance between an L2 sound and the L1 category it is mapped onto, the more likely it will be for the learner to identify the phonetic properties that distinguish it from the perceptually closest L1 sound. Consequently, the creation of a new category for the L2 sound distinct from the closest L1 category (e.g., English /ɑ:/ with respect to Spanish /a/) will also be more likely. Conversely, the smaller the perceptual distance between an L2 sound and the perceptually closest L1 category, the less likely it will be that the learner distinguishes them and, consequently, the harder it will be to create a new category for the L2 sound (e.g., English /æ/ with respect to Spanish /a/), as the L1 sound will tend to be used for the L2 sound. Understanding the way non-native sounds are mapped onto existing L1 categories and how these mappings are modified (or not) with increasing experience with the TL is, thus, a crucial issue in L2 speech research. This paper explores two different methods of analyzing cross-linguistic perceived similarity: a more common perceptual approach involving the use of perceptual tasks where learners consciously classify L2 sounds in terms of L1 sound categories and an online approach based on the analysis of eye-gaze measures during real-time spoken word recognition. The objective of the current study is, thus, to explore the relationship between cross-linguistic perceived similarity and L2 speech processing and to investigate the potential of an online measure of spoken-word recognition as an alternative approach to measuring cross-linguistic similarity. The two methodological approaches, their motivation, and the findings from previous studies are reviewed next.

2. Background

2.1. Measuring Cross-Language Similarity

Assessing the perceived similarity between native and TL sounds is a crucial first step in all studies on L2 speech learning, as discussed above. Despite its importance, however, there is still no clear consensus as to what is the best method of measuring cross-linguistic similarity (Strange 2007). The most common methods include articulatory descriptions, acoustic analyses, and perceptual measures. L2 speech researchers generally advocate for the use of the latter or a combination of measures (Bohn 2002; Strange 2007). Two perceptual ways of evaluating cross-linguistic similarity are interlingual identification and paired comparisons (Bohn 2017). The former is the most common approach and is exemplified by perceptual tasks known as forced choice categorization tasks or perceptual assimilation tasks (PAT). In a PAT, listeners identify TL stimuli in terms of native categories by matching a given aural stimulus with one of several possible L1 alternatives and providing a goodness of fit rating (e.g., Best 1995; Strange et al. 2009; Faris et al. 2018; Cebrian 2021). A paired comparison technique, such as a rated dissimilarity task, involves using a Likert scale to indicate the degree of (dis)similarity between two stimuli presented aurally, typically a native and a non-native stimulus (e.g., Flege et al. 1994; see Cebrian 2022 for a comparison of both perceptual methods). Although these perceptual tasks provide valuable information about cross-linguistic perceived similarity, they have some methodological drawbacks. For example, perceptual tasks require a large number of stimuli, as multiple exemplars of each TL sound are necessary in order to obtain reliable results. In addition, these tasks are rather repetitive in nature precisely because of the large number of stimuli and trials that are needed. Finally, particularly in a PAT, the listeners perform

the task by choosing among a number of typically orthographic representations of the L1 categories, resulting in a potential influence of the spelling on listeners' judgments (Strange 2007; Cebrian 2022). A possible way of bypassing these shortcomings may be to use a task that explores the role of cross-linguistic perceived similarity during real-time spoken word recognition. Such a methodology is common in studies on online speech processing using a visual world paradigm (Tanenhaus et al. 1995). The relationship between perceived similarity and interlingual word processing is discussed in the next section.

2.2. Cross-Language Lexical Competition

Previous studies on L2 processing in bilinguals investigating the time course of lexical activation through word recognition in the visual world paradigm (e.g., Weber and Cutler 2004; Cutler et al. 2006; Escudero et al. 2008) make use of images representing target lexical items (targets) that overlap phonologically with other lexical items (competitors). Target and competitor images compete for visual attention when presented simultaneously, which is reflected in the proportion of eye gaze fixations occurring within the competing areas of interest (AOI) defined by the images presented on the computer screen. Proportions of looks to the image of a competitor lexical item while the image of a target lexical item is being visually fixated is meant to index the amount of competition for lexical activation between target and competitor, which also reflects the amount of phonological overlap between the lexical items that compete for activation (Spivey and Marian 1999; Marian and Spivey 2003; Chambers and Cooke 2009).

When L2 learners perform online word recognition tasks in L2, such as when identifying the word *sheep* /ʃi:p/ (target) through a mouse click on its image (a picture of one *sheep*), upon hearing the instruction "Click on the *sheep*", the presence of an image with two *ships* /ʃɪps/ (competitor) is likely to capture the learners' gaze due to its phonetic overlap with *sheep*. As a result, the image of *ships* receives a certain proportion of the eye gaze fixation time as the auditory presentation of the string of phonemes /ʃi:p/ unfolds in time until the absence of /s/ at the end of *sheep* /ʃi:p/ makes it clear to the learner which of the two pictures (*sheep* or *ships*) is to be clicked on. For L2 words that contain phonologically distinct sounds, such as *panda* and *pencil* in English, competition for lexical activation during word recognition may occur in L2 learners if the L2 words contain confusable sounds, such as /æ/-/e/ in the first syllable of *panda* and *pencil* (/pæn-/ and /pen-/ , respectively) for Dutch learners of English (Weber and Cutler 2004). Crucially, the amount of competition for activation between two (near-)minimal-pair lexical items like *sheep* /ʃi:p/ and *ships* /ʃɪps/ in L2 learners was found to depend on the degree of overlap between their L2 vowel category /i:/ in *sheep* and their L2 vowel category /ɪ/ in *ship*. That is, less lexical competition between *sheep* and *ships* is expected in learners who developed separate L2 phonetic categories for /i:/ and /ɪ/ and for whom the /i:/-/ɪ/ contrast is lexically encoded in their L2 mental lexicon. However, for L2 learners for whom lexically contrastive word pairs like *sheep* and *ship* are phonetically confusable (e.g., Spanish or Catalan learners of English), lexical competition is expected to be maximal until the final disambiguating /s/ in *ships* is heard. Thus, the amount of lexical competition indexes the degree of phonetic confusability between contrastive L2 sounds and, at the same time, indicates the extent to which learners acquired an L2 phonological contrast.

Catalan learners of English generally find the English low vowels /æ/ and /ʌ/ confusable because these vowels tend to be assimilated perceptually to the Catalan low vowel category /a/ (Cebrian 2021). Similarly, they tend to find the English high front vowels /i:/ and /ɪ/ confusable because these two vowels are perceptually close to Catalan /i/ (although /ɪ/ is also perceived as Catalan /e/ (Cebrian 2021)). Confusable contrasts like English /æ/-/ʌ/ and /i:/-/ɪ/ that show a perceptual overlap in their assimilation to an L1 category pose a learning challenge for L2 learners. Just as two lexical items like *panda* and *pencil* (/æ/-/e/) for Dutch learners of English compete for lexical activation during online word recognition based on their phonetic overlap, L2 and L1 words that overlap phonetically can also compete for lexical activation. For example, English learners of French

activate the English word *pool* /pu:l/ upon hearing the French word “poule” /pul/ *hen* because they perceive French /u/ as being phonetically close to English /u:/ (Chambers and Cooke 2009). Similarly, Marian and Spivey (2003) found that in the course of hearing the English word *marker*, with the target word presented visually in a four-object visual display, Russian learners of English looked at a *stamp* (Russian “marku”) more often than to the other two objects, whose names were phonologically unrelated to either the target or the competitor objects (distractors).

Thus, for example, for Catalan learners of English, cross-language near minimal-pair words like English *pillow* /'piləʊ/ and Catalan “pila” /'pilə/, *battery*, (interlingual near homophones) are likely to compete for lexical activation as a function of the extent to which they perceptually overlap. The degree of cross-language lexical competition between words in the L2 and the L1 sharing similar sounds like English /ɪ/ and Catalan /i/ would, therefore, be larger for learners who perceive English /ɪ/ to be a good perceptual match to Catalan /i/ than for those who learned to discern the phonetic differences between English /ɪ/ and Catalan /i/ and do not perceive English /ɪ/ to be a good perceptual match to Catalan /i/. Measures of the amount of lexical competition during online word recognition can, therefore, be used to obtain within-language online estimates of L2 phonological competence based on the categorical perception of L2 segmental contrasts in a lexical context, as well as to obtain a cross-language online estimate of the degree of perceptual assimilation of L2 sounds to L1 sound categories.

2.3. Asymmetrical Lexical Access

A well-established finding from studies investigating visual word recognition involving confusable L2 sound categories through within- and cross-language lexical competition is the asymmetric nature of the competition. For example, in their study on Dutch speakers of L2 English, Weber and Cutler (2004) reported that the amount of lexical competition (% of looks to a non-target object) of /e/-word competitors (e.g., *pencil*) in the process of visually recognizing target /æ/-word objects (e.g., *panda*) was larger than the amount of lexical competition of /æ/-word competitors (e.g., *panda*) in the process of visually recognizing target /e/-word objects (e.g., *pencil*). This asymmetry can be explained by the fact that English /e/ is a perceptually better match to the closest L1-Dutch vowel category /ɛ/ (the dominant L2 category) than English /æ/. That is, although both the L2-English /æ/ and /e/ in the input activate the L1 category /ɛ/ in Dutch listeners, the dominant category is lexically coded as a better match than the non-dominant one. Cutler et al. (2006) replicated this asymmetric mapping of a phonological contrast to lexical representations with L1-Japanese listeners and the English /ɪ/-/i/ contrast. Because Japanese listeners map both English prevocalic /ɪ/ and /i/ to Japanese /ɪ/, but English /i/ is perceptually closer than English /ɪ/ to their native /ɪ/ (Iverson et al. 2003), /i/ acted as the dominant category. Consequently, /i/-initial words competed for activation with /ɪ/-initial targets but not the other way around. Although orthography (distinct orthographic representation of confusable L2 sounds) likely plays a role in determining dominance, such findings suggest that what mainly determines which of two confusable L2 sounds is dominant in a lexical context is how close it is phonetically (and perceptually) to the closest L1 category. However, how gradient such dominance effects are as a function of listeners' perception of phonetic closeness of L2 sounds to L1 categories remains an empirical question both in within- and between-language competition.

In the case of the cross-language activation of lexical competitors (looking at a *battery*, “pila” /'pilə/ in Catalan, when listening to English target *pillow* /'piləʊ/ or looking at “thread”, *fil* /fil/ in Catalan, when listening to English target *field* /fi:ld/), potential dominant category effects (i.e., asymmetries in cross-language lexical competition) may be determined by differences in the cross-language phonetic similarity between confusable L2 sounds (English /i:/ and /ɪ/) and a perceptually close native Catalan vowel category (/i/). Therefore, given the English /i:/-/ɪ/ pair, one would expect English /i:/ to be the dominant category, as English /i:/ is phonetically closer to Catalan /i/ than English /ɪ/

is. Consequently, L1 lexical competitors with Catalan /i/ (/fil/, /'pilə/) are predicted to attract a larger proportion of looks when the target English word contains the dominant /i:/ (field /fi:ld/) than when the target English word contains the non-dominant /ɪ/ (pillow /'pɪləʊ/). However, the extent to which confusable L2 sounds are perceived to be good or poor exemplars of the L1 sound categories they are mapped onto is likely to change over time within learners as a function of increased experience with L2 and across learners as a function of their level of L2 phonological proficiency.

Online tasks within the visual world paradigm (Tanenhaus et al. 1995) offer the possibility of overcoming the limitations of perceptual assimilation tasks in assessing cross-linguistic similarity. No research, to date, specifically examined whether varying degrees of cross-language perceptual assimilation of confusable L2 sounds to L1 categories is related to varying degrees of cross-language lexical competition between L2 words containing confusable L2 sounds and L1 words containing the L1 categories onto which these L2 sounds are perceptually mapped.

2.4. The Current Study

The current study explores the relationship between asymmetrical lexical access and cross-linguistic similarity and evaluates its effectiveness in measuring cross-linguistic similarity and assessing accurate target-like categorization. We aimed to determine to what extent two pairs of English vowels (/i:/-/ɪ/ and /æ/-/ʌ/) that are confusable for Catalan learners of English display asymmetric mapping to their closest L1 vowel categories /i/ and /a/, respectively, and if so, whether asymmetries shown in differing strengths of online cross-linguistic lexical competition for the dominant and the non-dominant L2 sounds are related to the degree of cross-linguistic similarity as assessed offline through a cross-language perceptual assimilation task. More specifically, we asked the following research questions:

RQ1: Do Catalan learners of English exhibit asymmetric levels of cross-language lexical competition between the dominant and the non-dominant L2 vowel categories in the confusable English vowel pairs /i:/-/ɪ/ and /æ/-/ʌ/ and their corresponding L1 vowel categories /i/ and /a/?

RQ2: To what extent are asymmetries in the strength of cross-language lexical competition related to differences in cross-linguistic perceptual similarity?

Two experiments were carried out in the present study with the same participants. In Experiment 1, participants were asked to perform a cross-linguistic perceptual assimilation task targeting the confusable English vowel pairs /i:/-/ɪ/ and /æ/-/ʌ/. Experiment 2 was a visual world paradigm experiment in which the same participants performed a simple picture selection task within a four-picture display following auditory instructions while their eye movements were being recorded. From this experiment, we obtained measures of cross-language competition (proportions of eye-gaze fixation to competitors and eye-gaze fixation durations) between L2 dominant and non-dominant sounds and perceptually close L1 sounds.

3. Experiment 1: Perceived Similarity between English and Catalan Vowels

The goal of the first experiment was to assess the perceived similarity between English and Catalan vowels, particularly SSBE /i ɪ æ ʌ/, and their closest L1 vowels in order to contrast the similarity measure obtained in this experiment with the results of Experiment 2.

3.1. Methods and Materials

3.1.1. Participants

The participants were 43 Spanish/Catalan bilingual learners of English (36 females) who were second-year undergraduate students in an English Studies degree at a public university in Barcelona with an upper-intermediate-to-advanced level of proficiency in English according to self-reported proficiency and a 120-item yes/no vocabulary size test

(X/Y Lex, [Meara and Milton 2003](#)). Their level of bilingualism ranged from equally dominant in both languages to Catalan-dominant (the majority), according to their responses to the Bilingual Language Profile ([Birdsong et al. 2012](#)). Participants were given course credit for their participation. These participants also completed experiment 2.

3.1.2. Target Sounds

The target contrasts were SSBE /i/-/ɪ/ and /æ/-/ʌ/, which were found to pose a problem to both Catalan and Spanish learners of English ([Cebrian 2006](#); [Mora and Mora-Plaza 2019](#)). Previous studies showed that SSBE /æ/ and /ʌ/ are both consistently assimilated to Catalan and Spanish /a/, with English /æ/ obtaining higher goodness ratings as the L1 /a/ than English /ʌ/ ([Cebrian 2019, 2021](#)). Regarding the SSBE high front vowels, /i:/ is consistently assimilated to Spanish/Catalan /i/, whereas SSBE /ɪ/ is split between the L1 high and mid vowels (Catalan /i/, /e/ and /ɛ/, Spanish /i/ and /e/; [Cebrian et al. 2011](#); [Cebrian 2019, 2021](#)). Thus, these target vowels were selected regardless of the level of Catalan/Spanish dominance of the learners, as the target SSBE /i/-/ɪ/ and /æ/-/ʌ/ pairs are similarly assimilated to the Spanish and Catalan L1 categories /i/ and /a/, respectively.

3.1.3. Stimuli

The stimuli included a subset of SSBE vowels, namely /i: ɪ e æ ʌ eɪ aɪ/, for the sake of brevity and simplicity (see [Cebrian et al. 2011, Cebrian 2021](#) for a PAT involving a near-complete set of Catalan and English vowels and diphthongs). The vowels were elicited in /b/+vowel+/t/ sequences, embedded in a carrier phrase that included a rhyming word so as to ensure the correct pronunciation of each vowel: “It rhymes with *hot*. I say *bot*. I say *bot* again.” Three male native English speakers (mean age: 35, range 29–44), who spent most of their lives in the South of England, produced the stimuli. They were recorded in a soundproof booth at the research laboratory at a university in London, United Kingdom. The best two productions of each stimulus per speaker were selected on the basis of spectrographic analyses and auditory judgments. The stimuli were then edited to include anything from the release of the /b/ until the end of the vowel portion to the beginning of the /t/ closure in order to avoid any potential influence of the consonantal context on the listeners’ judgments (e.g., [Levy 2009](#)).

3.1.4. Procedure

In the PAT, the SSBE stimuli were presented to the listeners in a random order, and listeners had to identify each stimulus as one of six possible L1 categories (following the findings of previous studies) and subsequently provided a goodness of fit rating on a seven-point scale, where 1 meant a bad exemplar of the chosen vowel and 7 meant a good, i.e., native-like, exemplar. The L1 categories were represented by an example word containing the L1 vowel alongside the most common spelling for that vowel, namely “xai” (ai), “rei” (ei), “set” (è), “fet” (é), “dit” (i), and “gat” (a) (meaning *sheep, king, seven, fact, finger, and cat*, respectively) and representing Catalan /aᵢ/, /eᵢ/, /ɛ/, /e/, /i/, /a/, respectively. The total number of trials was 84 (7 vowels × 3 talkers × 2 tokens × 2 repetitions). The main task was preceded by a small practice session containing six trials to familiarize listeners with the task and adjust the volume. Participants completed the task in approximately 10 min (see Experiment 2 for more details about the general procedure).

3.2. Results

The percentage of times each non-native stimulus was identified in terms of one of the L1 response categories (i.e., percent assimilation as an L1 vowel) and the corresponding mean goodness of fit ratings (GR) were calculated for each listener. In addition, following [Guion et al. \(2000\)](#), a fit index score was calculated, that is, a composite measure calculated by multiplying the percent assimilation by the GR obtained for each L2 to L1 assimilation pattern. The results for the two vowel contrasts that are the focus of the study of this paper

are presented in Table 1². The majority of SSBE vowels tested were assimilated to a single L1 vowel above 70% of the time (a suggested categorization threshold, Tyler et al. 2014). The only exception was English /ɪ/, whose assimilation was split between Catalan /e/ (47%) and /i/ (46%). By contrast, English /i:/ was consistently assimilated to Catalan /i/ (96%). With respect to the low vowel pair, /æ/ and /ʌ/ were very consistently assimilated to the same L1 vowel, /a/, above 95% of the time, but /æ/ obtained a higher mean GR than /ʌ/ (5.4 vs. 4.5 out of 7). In fact, this difference reached significance on a paired samples *t*-test ($t(41) = 7.353, p < 0.001$). Such a statistical outcome was suggested as evidence that one of the two L2 vowels (/æ/ in this case) is perceived to be significantly closer to the L1 vowel (Catalan /a/) than the other L2 vowel (/ʌ/) (Tyler et al. 2014; Tyler 2021).

Table 1. Perceived similarity between SSBE /æ ʌ i: ɪ/ and L1 Catalan vowels. (GR = goodness rating). Assimilations below chance level (16.7%) were excluded.

SSBE Vowel	L1 Vowel	% Assimilation	GR (Out of 7)	Fit Index (Max 7)
/æ/	/a/	100	5.4	5.4
/ʌ/	/a/	97.5	4.5	4.4
/i:/	/i/	96	5.5	5.3
/ɪ/	/i/	46.2	4.5	2.1
	/e/	46.6	5.2	2.4

3.3. Interim Discussion

Experiment 1 examined the perceived similarity between a subset of SSBE vowels and Catalan vowels by a group of Catalan learners of English. The results are in agreement with the outcomes of previous studies involving an analogous population (Rallo Fabra 2005; Cebrian et al. 2011; Cebrian 2021). This study focuses on the SSBE pairs /i:/ and /ɪ/ and /æ/-/ʌ/. In both cases, the results show that one L2 vowel has a clearer match in the L1 than the other, but the pairs differ with regard to the degree of that difference. For the low vowels, both SSBE /æ/-/ʌ/ are strongly assimilated to their L1 counterpart, /a/, but the significant difference in GRs reveals that SSBE /æ/ is perceptually closer to Catalan /a/ than SSBE /ʌ/. In other words, /æ/ and /ʌ/ constitute a category-goodness assimilation type in PAM’s terms (Best and Tyler 2007). By contrast, in the case of /i:/ and /ɪ/, the former is clearly categorized in terms of an L1 vowel (/i/) and the latter patterns as uncategorized, as it is perceived as more than one L1 vowel (/i/ and /e/). Experiment 2 will explore the effect of L1-L2 perceptual similarity on the degree of L1 activation during L2 speech processing. The results of the PAT indicate that one of the L2 vowels in each pair is more likely than the other to trigger the activation of interlingual homophones during online word recognition in the L2, and this difference between the two vowels is greatest in the case of the high front vowels, given the greater difference in their assimilation to the L1 category (/i/).

4. Experiment 2: Cross-Language Lexical Competition

In this experiment, we adopted a visual world paradigm (Tanenhaus et al. 1995) using an eye-tracking technique to measure interlingual competition during sentence comprehension. The goal of this experiment was to assess the amount of cross-language lexical competition Catalan learners of English would experience during online word recognition when following auditory instructions to click on target images representing English words containing the SSBE vowels /i ɪ æ ʌ/ in the presence (or absence) of a competitor image representing an L1 near-homophonous word containing the Catalan vowels /i a/.

4.1. Methods and Materials

4.1.1. Participants

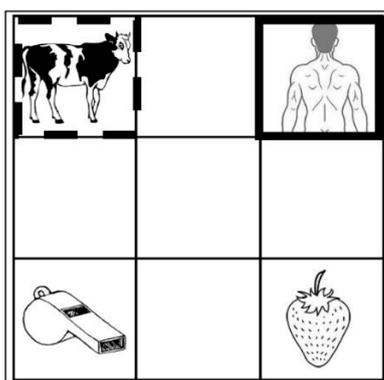
The participants were the same 43 Spanish/Catalan bilingual learners of English who completed Experiment 1 and a group of eighteen native English speakers (10 males). The latter were university students and English teachers residing in Barcelona at the time of the

experiment (mean age: 26, range: 20–44). Most were from England, one from Ireland, and one from Scotland. They had spent between one month and six years in Barcelona. Their knowledge of other languages varied, but only five participants indicated that they could speak Spanish fluently. Most of them reported a weekly use of English of about 75–100% of the time, with four of them reporting a 50–74% use and one under 50% (the other language used was typically Spanish, sometimes Catalan or other).

4.1.2. Cross-Language Lexical Competition Task

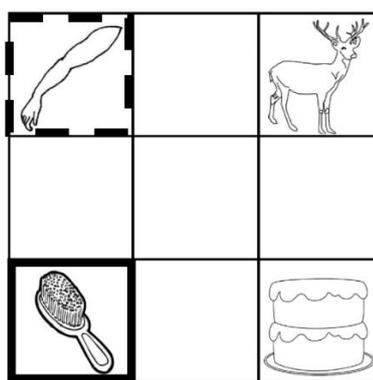
Participants’ eye movements were recorded using a Tobii T120 eye-tracker integrated into a 17” monitor with a resolution of 1024 × 768 pixels. The eye-tracker worked at a sampling rate of 120 Hz. Although this resolution was relatively low, it was deemed sufficiently accurate to register fixation counts and durations within the AOIs drawn around the picture stimuli (see Figures 1 and 2 below). A 9-point calibration and validation procedure was performed individually before the task. Participants’ eye movements were recorded as they followed spoken instructions to click on one of four objects presented on the computer screen. Figures 1 and 2 provide an illustration of the display of the objects on the computer screen for two actual trials. The target sounds were the English vowel pairs in Experiment 1, namely /i:/-/ɪ/ and /æ/-/ʌ/. On critical trials, the target image (for example, a word containing English /æ/ such as *back*, see Figure 1) was accompanied by an interlingual competitor, that is, an English word whose translation into Catalan contains a Catalan vowel that is perceptually close to the English target (e.g., the word *cow*, “*vaca*” in Catalan, containing Catalan /a/, /bakə/). Comparing the results for this type of trial with trials involving /ʌ/ as the target vowel (e.g., *brush*, in competition with *arm*, Catalan “*braç*” /bras/) provided an online measure of which of the two English vowels, /æ/ or /ʌ/, is perceptually closer to Catalan /a/. In the case of English /i:/-/ɪ/ (see Figure 2), the interlingual competitor was an English word whose Catalan translation contains the Catalan vowel /i/. For example, *thread* (“*fil*”, /fil/, in Catalan) was the competitor for the target word *field*, and *battery* (“*pila*”, /'pilə/, in Catalan) was the competitor for the target word *pillow*. Target words were embedded in the sentence “Click on the....”, which was followed by a filler sentence (“and now click on the....”) that prompted participants to click on a second picture, which was always a distractor. The filler sentence was included in order to divert participants’ attention from the focus of the experiment.

A. L2 /æ/ - L1 /a/



Click on the **back**. Now click on the *strawberry*.
 Target: *back* /bæk/
 Competitor: *cow* (*vaca* /'bakə/)

B. L2 /ʌ/ - L1 /a/



Click on the **brush**. Now click on the *cake*.
 Target: *brush* /brʌʃ/
 Competitor: *arm* (*braç* /'bras/)

Figure 1. Example of visual display for a trial with English /æ/ (A) and English /ʌ/ (B) as the target vowels and an interlingual competitor with the Catalan vowel /a/. The target and competitor words are indicated with a solid and a dashed line, respectively.

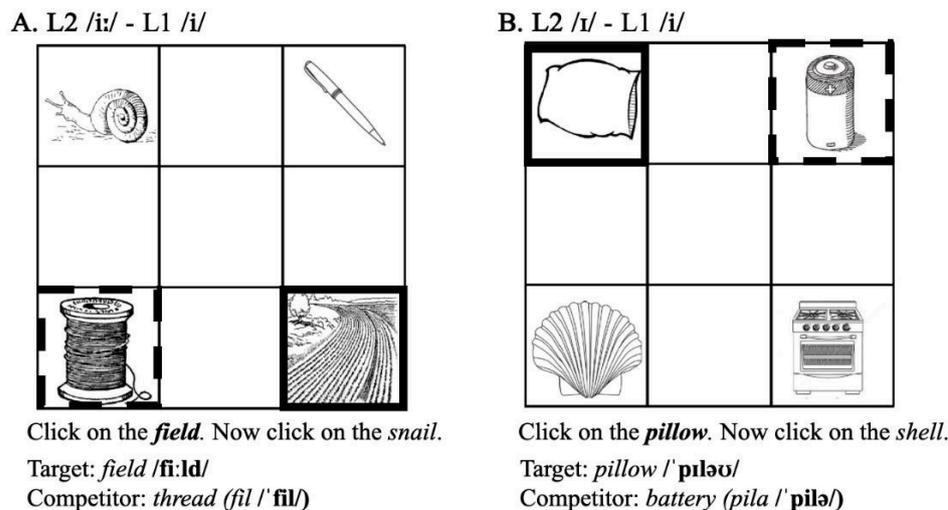


Figure 2. Example of visual display for a trial with English /i:/ (A) and English /ɪ/ (B) as the target vowels and an interlingual competitor with the Catalan vowel /i/. The target and competitor words are indicated with a solid line and a dashed line, respectively.

4.1.3. Stimuli and Task Design

The whole experiment consisted of 48 trials. Twelve trials involved target words that included the target English vowels (/i: ɪ æ ʌ/, three trials per vowel) alongside an interlingual competitor. Of the remaining 36 trials, 12 contained the target vowels without an interlingual competitor, and the remaining 24 involved English vowels other than /i: ɪ æ ʌ/ with an interlingual competitor (12) or without an interlingual competitor (12). The non-crucial trials were used as fillers and for control purposes.

Table 2 presents the L2 target–L1 competitor pairs used in the twelve critical trials. In all cases, the competitor and the target overlapped in the first two or three sounds, crucially including the stressed vowel, that is, the target vowel under study. This amount of overlap was demonstrated to trigger interlingual competition (e.g., Weber and Cutler 2004). All the words were nouns representing relatively common and similarly frequent words. The comparable word frequency of targets and competitors was confirmed by the results of two independent sample *t*-tests conducted on log-transformed frequency scores.³ The results revealed no significant difference in word frequency between the English target and the competitor words ($t(22) = 0.488, p = 0.63$), nor between the corresponding Catalan translations of the target and competitor words ($t(22) = 0.451, p = 0.656$). In addition, two paired-samples *t*-tests yielded no significant difference between the frequency index of the English competitor words and their corresponding Catalan translations ($t(11) = 0.316, p = 0.758$), nor between the English target words and the Catalan translations of the corresponding competitor words ($t(11) = 0.494, p = 0.631$). The lexical frequency of the distractors was not controlled as previous studies indicate that the probability of fixating distractors that do not match the acoustic information of the target word does not vary with lexical frequency (Dahan et al. 2001).

The sentences were recorded by a 29-year-old male native speaker of Standard Southern British English in a soundproof booth on a Marantz PMD-661 solid-state digital recorder with an external Shure SM58 voice microphone at a sampling frequency of 44.1 KHz. The visual stimuli were clip-art black and white pictures obtained from free Internet resources and standardized picture databases (Alario and Ferrand 1999; Bonin et al. 2003; Snodgrass and Vanderwart 1980). To ensure that pictures elicited the expected words, participants performed a picture familiarization task prior to the eye-tracking experiment (see the procedures section). The position of the four pictures on the display (target, competitor, and two distractors) was pseudo-randomized and counterbalanced across all trials so that the target and competitors appeared in all four positions (top right, top left, bottom right, bottom left) and all possible combinations, and no two trials in a row contained the same target vowel.

Table 2. Target–competitor pairs used on critical trials.

English Vowel	Target Word	Competitor Object	Catalan Word
/æ/	back	cow	vaca /'bakə/
/æ/	carrot	face	cara /'karə/
/æ/	man	sleeve	màniga /'manigə/
/ʌ/	brush	arm	braç /bras/
/ʌ/	money	blanket	manta /'mantə/
/ʌ/	nuts	whipped cream	nata /'natə/
/i:/	field	thread	fil /fil/
/i:/	bee	wine	vi /bi/
/i:/	beetle	glass	vidre /'biðrə/
/ɪ/	pillow	battery	pila /'pilə/
/ɪ/	sink	five	cinc /sɪŋ/
/ɪ/	fish	poker chips	fitxa /'fitʃə/

4.1.4. Procedures

Participants first completed a language background questionnaire, read an information sheet, and signed a consent form. The perceptual assimilation task was administered next (Experiment 1). Then, participants completed an inhibitory task and a vocabulary task (not reported in this paper), followed by a picture familiarization task and the cross-language lexical competition task. The familiarization task presented participants with the pictures of all target and competitor objects used in the cross-language lexical competition task to ensure that participants interpreted the pictures correctly (Ju and Luce 2004). On each trial, participants were asked to mentally name the object in the picture when appearing on the screen. The name of the object in English and in Catalan was then displayed 1500 ms after the picture appeared on the screen. Participants read the words displayed on the screen and activated their phonolexical representations via orthography. Participants took about ten minutes to complete the familiarization phase, after which the eye-tracking experiment was administered and completed in about eight minutes (see previous subsections).

4.1.5. Data Analysis

The eye-gaze data from this experiment was first analyzed descriptively in terms of the time course of lexical activation across the various conditions: proportion of fixations to target, competitor, and distractors for critical trials with target words containing dominant (/i:/, /æ/) and non-dominant (/ɪ/, /ʌ/) vowel categories (see Figures 3 and 4 below). We then tested the extent to which the amount of cross-language competition from interlingual homophones (indexed through fixation duration measures) was differentially affected by the presence of a dominant versus a non-dominant vowel in the target word. In order to do so, we built two sets of linear mixed-effects models in R using the glmmTMB package (Brooks et al. 2017). The dependent measure was the duration of fixations occurring within each AOI within a 300–800 ms time window after the auditory onset of the target word, which was previously normalized using the function orderNorm in the bestNormalize Package in R (Peterson 2021; Peterson and Cavanaugh 2020). In the first set of models, we compared fixation durations in target and competitor AOIs as a function of the dominance of vowel categories in the target (Figures 5 and 6). Previous studies on spoken word recognition showed that a measure that considers fixations to distractors relative to competitor captures information not provided in measures contrasting target and competitor fixations (e.g., Marian and Spivey 2003; Dahan and Tanenhaus 2005; Tsang and Chambers 2011). This is motivated by the fact that native speakers generally have longer fixation durations to targets, so the relevant comparison may, in fact, involve the items that are not the target, that is, competitor vs. distractor words. If there is no interlingual competition, the interlingual competitors should receive the same average fixation duration as the distractors. Thus, in the second set of models, we compared fixation durations in competitor and distractor AOIs also as a function of vowel dominance (Figure 7). Both sets of models included fixed and random-effect structures based on the best-fitting model

by comparing estimators across converging models. The significance of tests of pairwise contrasts for planned comparisons were all Bonferroni-adjusted. The best-fitting models were those that included *group*, *dominance*, and *AOI* and their interactions as fixed effects factors with random intercepts for *subject* and *item*. Previous models, including *contrast* (high vowels /i:/ and /ɪ/; low vowels /æ/ and /ʌ/) and interactions involving *contrast* as fixed effects were discarded, as none of the effects involving *contrast* reached significance and its inclusion did not improve the model fit.

4.2. Results

Our first research question asked whether Catalan learners of English would exhibit asymmetric levels of cross-language lexical competition between the dominant (/i:/ and /æ/) and the non-dominant (/ɪ/ and /ʌ/) L2 vowel categories in the confusable English vowel pairs /i:/-/ɪ/ and /æ:/-/ʌ/ and their corresponding L1 vowel categories /i/ and /a/. If cross-linguistic similarity affects L1-L2 lexical competition, we predict that learners will experience greater L1 competition from words with Catalan /i/ for target English words with /i:/ than for target English words with /ɪ/. Similarly, learners will experience greater L1 competition from words with Catalan /a/ for target English words with /æ/ than for target English words with /ʌ/. These predictions are derived from the closer perceptual similarity between Catalan /i/ and /a/ and English /i:/ and /æ/ (the dominant categories), respectively, than between Catalan /i/ and /a/ and English /ɪ/ and /ʌ/ (non-dominant categories), as well as from the assumption that asymmetries in cross-language competition are determined by degree of cross-language perceptual similarity between phonetically similar L2-L1 sound pairs. We are expecting such asymmetries to be reflected mainly in learners (but not in native controls) showing higher fixation durations to the L1 competitor with targets /i:/ and /æ/ than with targets /ɪ/ and /ʌ/ and consequently higher fixation durations to targets /ɪ/ and /ʌ/ than to targets /i:/ and /æ/ resulting from non-dominant categories generating fewer looks at L1 competitors.

In order to assess the prevalence of asymmetric lexical competition, we first explored the time course of lexical activation across the various conditions. More specifically, we examined whether the resolution of the competition between target and competitor occurred earlier in a 1000 ms time window after the auditory onset of the target word for target L2 words with an L1 interlingual competitor with a non-dominant vowel (/ɪ/, /ʌ/) than for L2 words with an L1 interlingual competitor with a dominant vowel (/i:/, /æ/). The latter is predicted to cause greater competition and result in a slower resolution phase. This is shown in Figures 3 and 4, which present the proportion of fixations to target and competitor averaged over participants for trials with target words containing a dominant vowel (Figure 3) and with target words containing a non-dominant vowel (Figure 4). Approximately 300 ms after the auditory onset of the target, the processing of the acoustic signal results in an increase in fixations to the target object and a decrease in fixation to non-target objects (Allopenna et al. 1998). A delay in the increase in target fixations indicates that another object in the display is exerting lexical competition. As shown in Figures 3 and 4, this is what is found in trials with a dominant target vowel, in contrast with trials with a non-dominant target vowel. There was a larger proportion of earlier looks to targets when the interlingual homophone contained a non-dominant category (/ɪ/, /ʌ/) than when the interlingual homophone had a dominant category (/i:/, /æ/). This suggests that in the time course of lexical activation, the cross-language competition was resolved earlier (around 250 ms after auditory onset) when the target word contained a non-dominant L2 vowel (less similar to the vowel of the L1 competitor) than when the target word contained a dominant L2 vowel (around 450 ms after auditory onset). Unexpectedly, L2 learners and native English controls did not seem to differ with respect to when, after the auditory onset of the stimulus, the lexical competition between target and competitor was resolved in the case of both dominant and non-dominant vowels. This may indicate that dominance also affected native English speakers, possibly due to the fact that they were not unfamiliar with Spanish or Catalan, as discussed below. Still, both when the target contained a dominant

and a non-dominant vowel, the proportion of looks to targets was always higher for native English controls than it was for L2 learners, whereas the proportion of looks to competitors was always higher for L2 learners than it was for native English controls. The extent to which this observed difference between the two groups was significant was explored by comparing the two groups on the duration of fixations on different areas of interest, as explained next.

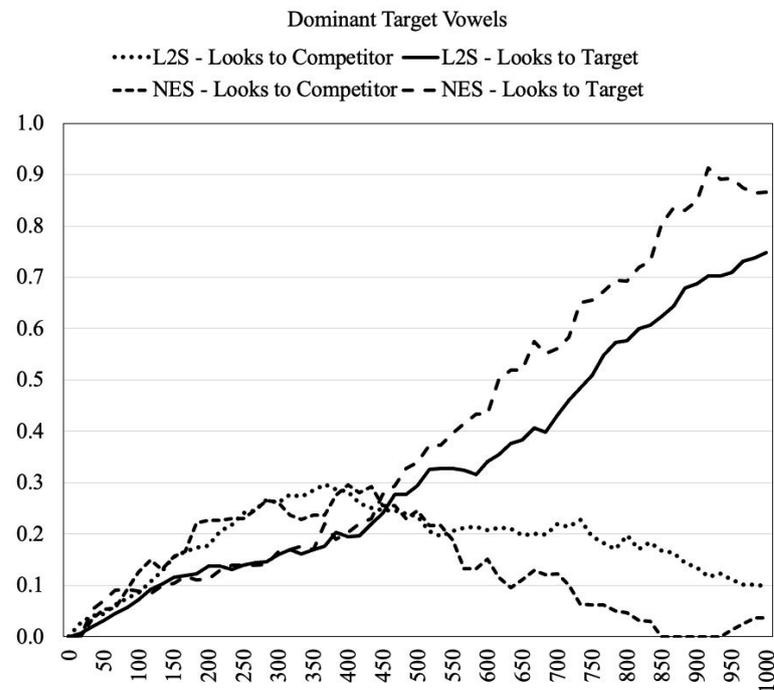


Figure 3. Proportion of looks to competitor and target by group with dominant target vowels (/i:/ and /æ/); L2S = L2 speakers, NES = native English speakers.

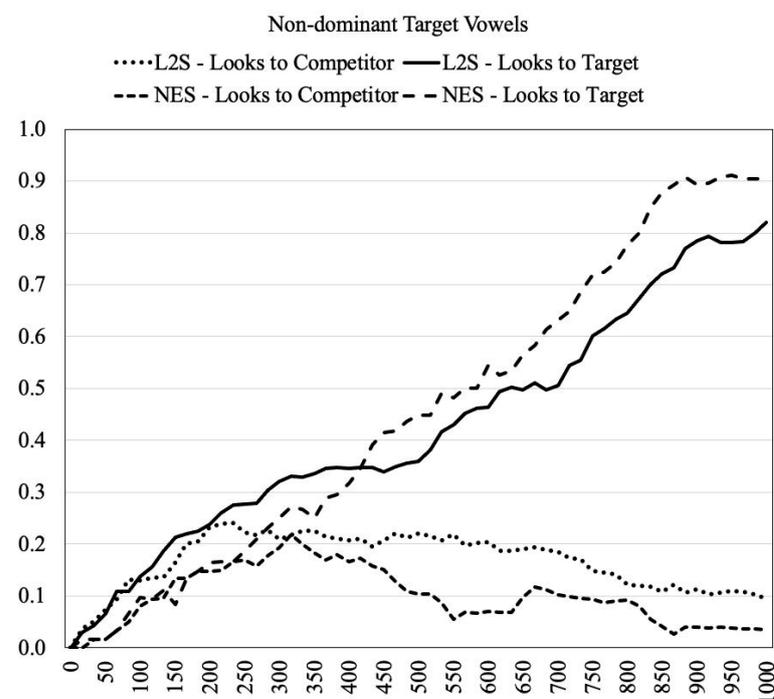


Figure 4. Proportion of looks to competitor and target by group with non-dominant target vowels (/ɪ/ and /ʌ/); L2S = L2 speakers, NES = native English speakers.

In order to assess the extent to which the English dominant vowels (/i:/, /æ/) and the non-dominant vowels (/ɪ/, /ʌ/) differentially affected the degree of cross-language lexical competition exerted by competitor objects whose names contained the Catalan vowels /i/ and /a/, we carried out two kinds of analysis. In the first one we compared the fixation durations (in milliseconds) to targets and competitors as a function of whether the vowel in the target was dominant or non-dominant. As the outcome of the PAT suggested there might be differences between the high (/i:/, /ɪ/) and the low (/æ/, /ʌ/) vowel contrasts in the amount of cross-language competition learners would experience from L1 homophones (see Section 3.3 above), we initially included *contrast* (/i:-/ɪ/, /æ:-/ʌ/) as a fixed effect in a linear mixed-effects model that also included *group* (L2 learners, native English speakers), *dominance* (dominant vowels /i:/ and /æ/; non-dominant vowels /ɪ/ and /ʌ/) and *AOI* (target, competitor), their interactions as fixed effects, and random intercepts for *subject* and *item*. As shown in Figure 5 below, such *contrast* effects were not clearly observable in the data, as there were no notable differences in the patterns of fixation durations for the two vowel contrasts for either dominant (/i:/: $M = 104.7$ vs. /æ/: $M = 103.8$) or non-dominant targets (/ɪ/: $M = 140.7$ vs. /ʌ/: $M = 144.9$), and consequently, differences between contrasts in fixation durations to competitors were also relatively small with dominant (/i:/: $M = 89.9$ vs. /æ/: $M = 72.8$) and non-dominant targets (/ɪ/: $M = 56.4$ vs. /ʌ/: $M = 66.4$). This was confirmed by the outcome of this model (see Appendix B Table A1 for the model's parameter estimates), which yielded a significant main effect of *AOI* ($\chi^2 = 104.098$, $p < 0.109$) but no significant main effects of *group* ($\chi^2 = 0.017$, $p = 0.898$), *dominance* ($\chi^2 = 0.279$, $p = 0.597$), or *contrast* ($\chi^2 = 0.110$, $p = 0.741$). *AOI* significantly interacted with *group* ($\chi^2 = 10.337$, $p = 0.001$) because whereas fixation durations on competitors were significantly longer for L2 learners than for native controls ($p = 0.019$, $d = 0.203$), learners' fixation durations on targets were significantly shorter ($p = 0.027$, $d = 0.190$). The *AOI* \times *dominance* interaction was also significant ($\chi^2 = 20.450$, $p < 0.001$) because whereas fixation durations on targets were significantly shorter for dominant than for non-dominant vowels ($p = 0.015$, $d = 0.276$), fixation durations on competitors tended to be longer for dominant than non-dominant vowels ($p = 0.071$, $d = 0.204$). The interactions of *contrast* with *group* ($\chi^2 = 2.576$, $p = 0.109$), *dominance* ($\chi^2 = 0.068$, $p = 0.794$), and *AOI* ($\chi^2 = 0.183$, $p = 0.669$) did not reach significance. Therefore, at least in terms of the average duration fixations we obtained in the eye-tracking experiment, no *contrast* effects could be attested.

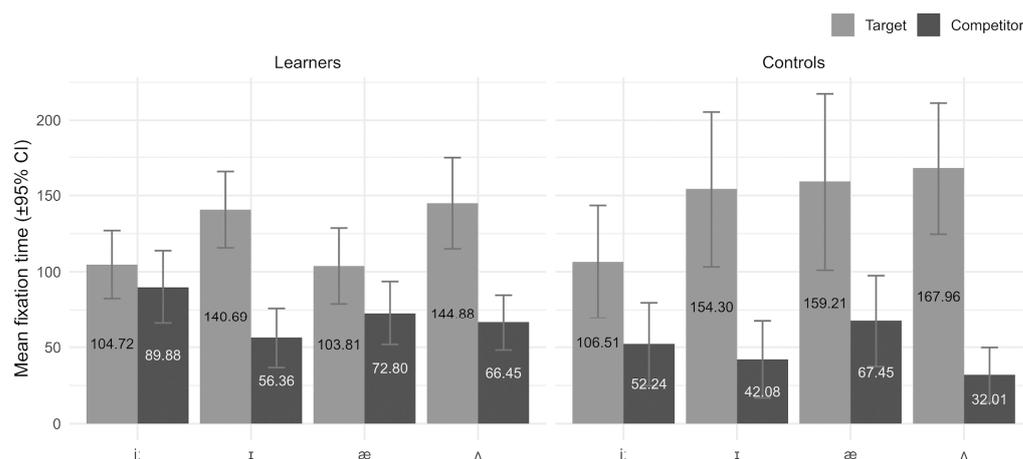


Figure 5. Mean fixation durations to target and competitor words by vowel and group.

We next assessed the effects of dominance in a mixed-effects model that included *group*, *dominance*, and *AOI* and their interactions as fixed effects, and random intercepts for *subject* and *item*. Figure 6 shows the average fixation duration in milliseconds to target and competitor words for the dominant and non-dominant vowel conditions across vowel contrasts. For learners, the amount of cross-language competition was higher with dominant ($M = 81.3$) than with non-dominant vowels ($M = 61.4$), and overall, competition

for activation from competitor objects was higher for learners ($M = 61\text{--}81$) than for native controls ($M = 37\text{--}59$) who, as expected, consequently fixated longer on targets ($M = 132\text{--}161$) than learners did ($M = 104\text{--}143$). The outcome of this model (see Appendix B Table A2 for the model's parameter estimates) revealed no significant main effects of *group* ($\chi^2 = 0.016$, $p = 0.898$) and *dominance* ($\chi^2 = 0.288$, $p = 0.591$) and a significant main effect of *AOI* ($\chi^2 = 103.624$, $p < 0.001$). This model also revealed significant *AOI* interactions with *group* ($\chi^2 = 10.246$, $p = 0.001$) and *dominance* ($\chi^2 = 20.434$, $p < 0.001$). The *AOI* \times *group* interaction arose because whereas fixation duration to targets was significantly longer for native controls ($M = 146.7$) than for L2 learners ($M = 123.5$; $p = 0.029$, $d = 0.188$), fixation durations to competitors were significantly longer for L2 learners ($M = 71.3$) than for native controls ($M = 48.4$; $p = 0.019$, $d = 0.203$). The *AOI* \times *dominance* interaction indicated that fixation durations to targets were longer for non-dominant than for dominant vowels ($p = 0.014$, $d = 0.278$), whereas fixation durations to competitors were marginally longer for dominant than for non-dominant vowels ($p = 0.072$, $d = 0.202$). The *group* \times *dominance* \times *AOI* interaction did not reach significance ($\chi^2 = 0.050$, $p = 0.823$). Still, given the prediction that L2 learners' fixations to competitors would be longer in the dominant vowel condition, it was important to examine the results of the relevant pairwise contrasts, which constituted planned comparisons in the current study. Learners obtained longer fixation durations to competitors with a dominant ($M = 81.3$) than a non-dominant vowel ($M = 61.4$), approaching significance ($p = 0.098$, $d = 0.192$) and significantly longer fixation durations to targets with a non-dominant vowel ($M = 142.8$) than with a dominant vowel ($M = 104.3$; $p = 0.006$, $d = 0.316$), but this was not the case for native speakers ($p = 0.193$, $d = 0.213$, and $p = 0.144$, $d = 0.239$, respectively). Nevertheless, the magnitude of these effects was relatively small, hence the lack of significant interaction (see Figure 6).

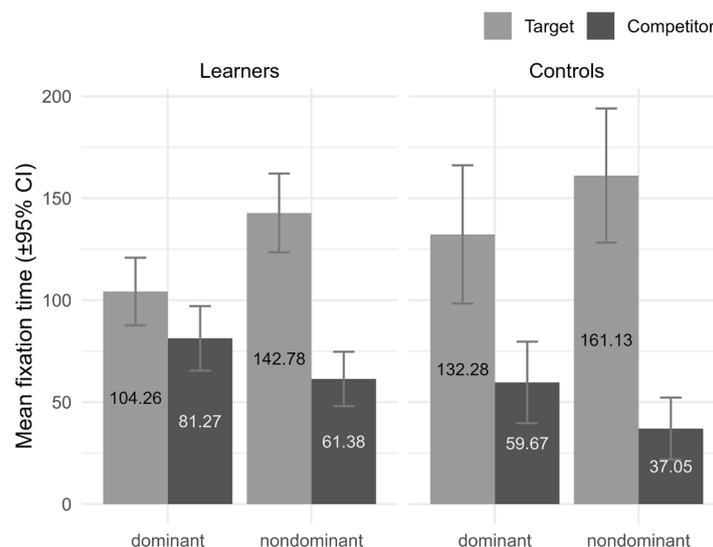


Figure 6. Mean fixation durations to target and competitor words by vowel dominance and group.

In a second analysis, we examined whether the expected longer fixation durations to competitors than to distractors differed as a function of dominance following previous studies (e.g., [Marian and Spivey 2003](#); [Dahan and Tanenhaus 2005](#); [Tsang and Chambers 2011](#)). In the context of our experiment, this analysis is also motivated by the fact that native speakers obtained longer durations to targets than learners. We expected longer fixation durations to competitors than to distractors, especially in the context of a dominant (/i:/, /æ/) than a non-dominant vowel (/ɪ/, /ʌ/) in the target word. Native controls were not expected to show such a difference because, for them, all competitor words should behave as distractors. As shown in Figure 7, this was indeed the case. For L2 learners, fixation durations were longer for competitors than for distractors in the context of both dominant ($M = 81.3$ vs. $M = 61.4$, respectively) and non-dominant vowels ($M = 61.4$ vs.

$M = 51.3$, respectively), but the difference was larger when the target is a dominant vowel (approximately 20 vs. 10 milliseconds on average), as predicted. The native controls did not appear to show such a difference. Finally, average fixation durations to distractors were longer in baseline trials (which contain no interlingual competitors) than in critical trials (which contain interlingual competitors), especially for learners.

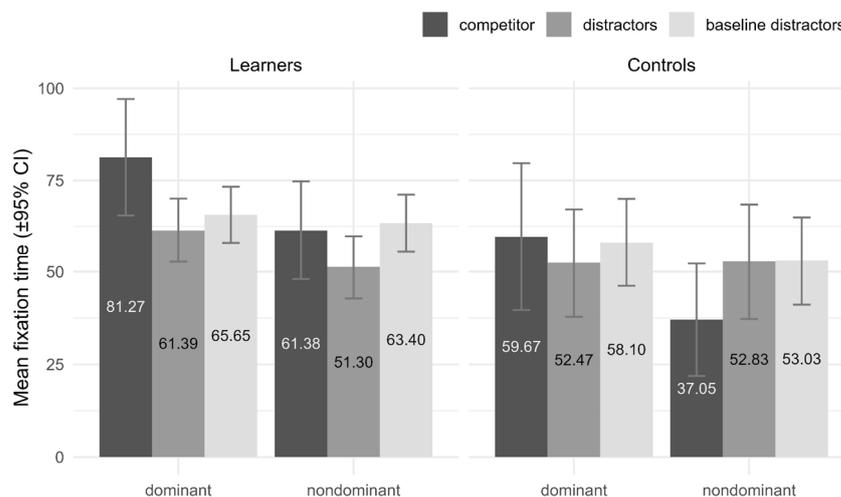


Figure 7. Mean fixation durations to competitors and distractors in critical trials and to distractors in baseline trials by group and vowel dominance.

We submitted the fixation durations to competitors and distractors to a mixed-effects model that included *group* (L2 learners, native English speakers), *dominance* (dominant vowels /i:/ and /æ/; non-dominant vowels /ɪ/ and /ʌ/), *AOI* (competitor, distractor) and their interactions as fixed effects, and random intercepts for *subject* and *item*. This model (see the parameter estimates table in Appendix B Table A3) revealed significant main effects of *group* ($\chi^2 = 7.592$, $p = 0.006$), *dominance* ($\chi^2 = 7.164$, $p = 0.007$), and *AOI* ($\chi^2 = 4.122$, $p = 0.042$). None of the interactions reached significance. Fixation durations in AOIs (competitors and distractors) were overall longer in learners ($M = 63.8$) than in native controls ($M = 50.5$, $p = 0.006$, $d = 0.168$) and longer in the context of a dominant ($M = 71.3$) than a non-dominant vowel ($M = 56.3$; $p = 0.016$, $d = 0.184$). Although the *AOI* \times *group* interaction only approached significance ($\chi^2 = 2.721$, $p = 0.099$), we examined the effect of *AOI* further within each group to assess the outcome of planned comparisons following the study's research questions and predictions. These analyses revealed that fixation durations to competitors were significantly longer for learners ($M = 81.3$) than for native controls ($M = 59.7$; $p = 0.002$, $d = 0.268$); but not fixation durations to distractors: $M = 61.4$ for learners vs. $M = 52.5$ for native controls; $p = 0.428$, $d = 0.068$), as predicted. In addition, it was only for learners ($p = 0.009$, $d = 0.166$) and not for native controls ($p = 0.747$, $d = 0.033$) that fixation durations were significantly longer for competitors than for distractor AOIs ($M = 81.3$ vs. $M = 61.4$, respectively). Hence, the main effect of *AOI* appears to be mainly driven by the cross-language competition that learners (but not native controls) experience through the activation of the L1 homophone competitor. Similarly, tests of pairwise contrasts revealed that the main effect of *dominance* (significantly longer fixation durations with dominant than non-dominant vowels ($p = 0.016$, $d = 0.184$)) was mainly driven by the fact that learners ($p = 0.012$, $d = 0.198$) but not native controls ($p = 0.135$, $d = 0.170$) obtained longer fixation durations with dominant than non-dominant vowels. In line with this result, the planned comparisons investigating the effect of dominance for each *AOI* condition and group revealed that learners' fixations on competitors were longer with dominant targets ($M = 81.3$) than with non-dominant targets ($M = 61.4$, $p = 0.014$, $d = 0.250$), a difference that was marginally significant for native speakers ($p = 0.071$, $d = 0.278$), showing a greater effect of vowel dominance on learners than on native controls.

4.3. Interim Discussion

Both the descriptive data and the outcome of the linear mixed-effects models presented above suggest that L2 learners (but not native controls) experienced asymmetric lexical competition and that this asymmetry could be attributed to the dominance of the target vowels /i:/ and /æ/ in the /i:/-/ɪ/ and /æ:/-/ʌ/ contrasts, respectively, as determined by their closer perceptual similarity to the corresponding Catalan vowels /i/ and /a/ in the competitor interlingual homophones. The native speaker control data did not always show the expected control pattern, as their fixation durations and time course of lexical activation were not completely unaffected by the presence of an interlingual homophone competitor. This could be due to native speakers having been recruited in Barcelona, as their knowledge of Spanish (albeit limited) could have been activated during the experiment. Still, the outcome of Experiment 2 indicates that the presence of a competitor interlingual homophone triggered longer fixation durations to competitors (i.e., higher levels of lexical competition) in the presence of a dominant (/i:/, /æ/) than in the presence of a non-dominant (/ɪ/, /ʌ/) target (Figure 6), as well as longer fixation durations to competitors than to distractors with dominant than with non-dominant targets (Figure 7), thus suggesting an effect of dominance that we expected would be related to the patterns of perceptual assimilation obtained in Experiment 1. Therefore, we next examined to what extent asymmetries in cross-language lexical activation were associated with degrees of perceptual similarity between dominant and non-dominant L2 categories and the corresponding L1 categories /i/ and /a/.

5. Relationship between Perceptual Assimilation and Lexical Competition Measures

The results of Experiment 2 showed that the target vowels that triggered the strongest activation of L1 words were the vowels in each pair that were perceptually closest to the L1 categories, i.e., the dominant vowels. Therefore, the group results for Experiments 1 and 2 indicate that there is indeed a relationship between the degree of perceived similarity between the L1 and L2 sounds and the level of L1 activation during L2 speech processing. Given that individuals sometimes differ in their perception of cross-linguistic similarity, possibly due to differences in L1 development and, consequently, L1 category precision (Flege and Bohn 2021), the relationship between the two types of measures (perceptual and online processing) at an individual level was investigated next. In order to do so, a series of Spearman correlations were conducted relating the perceptual measures obtained in Experiment 1 (namely, percent assimilation, GR, and FI) and the results of the L2 speech-processing experiment (fixation durations to competitor and target). A set of analyses conducted separately on the results for each of the four target vowels yielded no significant correlations. A second examination including all four vowels in a single analysis revealed some significant, albeit weak, correlations between the two types of measures. Specifically, fixation durations to the interlingual competitor were found to be correlated with percent assimilation to an L1 vowel ($r_s = 0.152, p = 0.049, n = 168$) and with fit index as the L1 vowel ($r_s = 0.163, p = 0.035, n = 168$).

The results of the correlational analysis, thus, do not provide consistent support for the relationship between offline and online measures of perceived similarity at an individual level, as the correlations were weak. We expected that if the two measures were strongly related at an individual level, the more similar learners perceived the (dominant) L2 vowel and the L1 counterpart to be, the greater the competition from the interlingual competitor (i.e., the greater the proportion of fixations and fixation durations to the competitor in the word recognition task). No significant correlations emerged in the analysis for individual vowels, separating dominant and non-dominant vowels. Still, some correlations reached significance in the single analysis for all vowels. This shows that the degree of perceived similarity is related to the degree of interlingual competition. Participants who perceive the L2 vowels to be perceptually closest to the L1 category tend to show greater interference from the interlingual homophone in the course of processing the target L2 words. Although the correlations were not very strong, this result is in line with the overall group

performance, which illustrates a clear effect of the dominant category on cross-language activation (longer fixation durations to the interlingual homophone). The possible reasons for the lack of stronger correlations at an individual level are elaborated further in the general discussion section in light of the main research questions of the current study.

6. General Discussion and Conclusions

It is well-known that the degree of perceived similarity between native and target language sounds plays a role in the process of learning L2 sounds (Best and Tyler 2007; Flege and Bohn 2021). However, there is still debate among L2 speech researchers regarding how to measure cross-linguistic perceived similarity appropriately. In addition, research has also shown that in the course of processing L2 speech, the L1 lexicon may be activated (Spivey and Marian 1999; Chambers and Cooke 2009). The degree of L1 activation may be related to the degree of similarity between the sounds that make up the L1 and L2 words (Weber and Cutler 2004). The objective of the current study was to contribute to our understanding of these issues by further investigating the relationship between cross-linguistic perceived similarity and cross-language lexical activation and exploring the efficacy of two different methods of measuring cross-linguistic similarity, a commonly used perceptual approach and an innovative eye-tracking approach based on the visual world paradigm. These two methods were contrasted in two experiments (1 and 2) involving the same group of Catalan learners of English.

The results of the first experiment confirmed that the English vowels /i:/ and /æ/ are perceived to be very close to the L1 categories /a/ and /i/, while English /ɪ/ and /ʌ/ obtained a lower degree of perceived similarity to these L1 categories, particularly in the case of /ɪ/, in line with previous studies (e.g., Cebrian 2021). We can conclude from Experiment 1 that the English vowels /i:/ and /æ/ constitute the dominant vowels in the English /i:/-/ɪ/ and /æ/-/ʌ/ contrasts, respectively, from the perspective of Catalan learners of English. Thus, we expected that in the cross-language lexical competition task (Experiment 2), Catalan learners of English would experience greater competition from an interlingual competitor when the English target word contained a dominant vowel than when the target word contained a non-dominant vowel. We also expected that this asymmetry would be found with the Catalan learners of English but not with the English native speakers, who acted as a control group.

The results of Experiment 2 indicated that the presence of an interlingual competitor affected the time course of the lexical activation of the target word in a spoken word recognition task in agreement with previous studies (Spivey and Marian 1999; Marian and Spivey 2003; Chambers and Cooke 2009). The first research question of the current study addressed the issue of whether Catalan learners of English exhibited asymmetric levels of cross-linguistic lexical competition as a result of differences in dominance between target vowels. An effect of vowel dominance was apparent in the fact that lexical competition was resolved earlier when the target word contained a dominant vowel than when it contained a non-dominant vowel (see Figures 3 and 4 above). The fact that this outcome was observed for both the learners and the native controls indicates that native speakers also showed some effect of cross-language lexical competition, possibly linked to their knowledge of Spanish or Catalan, as discussed below. Still, the analysis of fixation durations to target objects and interlingual competitors revealed that L2 learners fixated on competitors for a longer time than native English speakers, who fixated on the targets longer than the L2 learners. This shows that learners' processing of the target words was affected by the presence of the interlingual competitor in the display in a way that was not observed for native speakers. In addition, the results of planned comparisons indicated that learners, but not native speakers, tended to present longer fixation durations to competitors in the presence of a dominant target than in the presence of a non-dominant target (approaching significance) and had significantly longer fixations to targets under the non-dominant condition. Furthermore, the analysis of fixation durations to competitors relative to distractors revealed additional group differences. The planned comparisons exploring each group's fixation durations on

non-target objects illustrated that L2 learners fixated on competitors significantly longer than on distractors, particularly when the concomitant target English word contained a dominant vowel (/i:/, /æ/). No difference between fixation durations to competitor and distractor was revealed when the target word contained a non-dominant vowel (/ɪ/, /ʌ/). In other words, the competition from the (interlingual) competitor was greater when the target contained the L2 vowel that was found to be perceptually closest to an L1 category in Experiment 1. By contrast, native speakers fixated on competitors and distractors to a similar extent, showing that for native speakers, interlingual competitors did not exhibit the level of lexical competition observed in L2 speakers.

Nevertheless, as mentioned above, the controls appeared to be affected by the nature of the vowel in the target word, as the resolution of the lexical competition between target and competitor also tended to occur earlier when the target contained a non-dominant vowel than when it contained a dominant vowel. This outcome may be explained by the fact that the native English speakers resided in a Spanish/Catalan-speaking setting. Although only five of the eighteen English speakers reported being fluent in Spanish or Catalan, and most of them reported using English at least 75% of the time, it is likely that their knowledge of Spanish and their exposure to Spanish and Catalan, even if small, may have been enough to trigger the activation of their L2 Spanish or Catalan lexicon when processing the English sentences. In fact, some previous studies have shown evidence of L2 lexical competition during L1 processing (Spivey and Marian 1999; Marian and Spivey 2003). Note that the Catalan interlingual competitors were always words that had a phonologically similar cognate in Spanish (e.g., Catalan “cinc”, Spanish “cinco”, meaning *five*, or Catalan “braç”, Spanish “brazo”, meaning *arm*), and most lexical items were relatively common words. Thus, despite using mostly English, English speakers’ passive knowledge of Spanish (and Catalan) possibly resulted in a greater activation of the interlingual competitor than expected. This points to the need to test native English controls who are preferably monolingual and reside in an English-speaking setting, an issue that is left for future research.

The second research question of this study addressed the issue of whether asymmetries in the strength of cross-language competition were explained by differences in cross-linguistic perceptual similarity. Following the results of Experiment 1, the prediction was that the asymmetry would be more evident with the high vowels than with the low vowels, as the non-dominant high vowel (/ɪ/) is perceived to be less similar to its Catalan counterpart (/i/) than the non-dominant low vowel (/ʌ/) to Catalan /a/. In other words, there is a greater difference in perceived similarity to an L1 vowel in the case of the two high vowels than in the case of the two low vowels, suggesting a possible greater effect of vowel dominance with high vowels than with low vowels. This was not found to be the case, as statistical models evaluating the effect of contrast (/i:/-/ɪ/ vs. /æ/-/ʌ/) did not yield significant effects involving contrast. This can be explained by the fact that although the difference in perceived similarity to an L1 vowel was greater between the two high vowels than between the two low vowels, the dominant vowels in both cases obtained highly comparable results in Experiment 1 (English /i:/ was assimilated to L1 /i/ 96% of the time with a mean goodness rating of 5.5 out of 7 while English /æ/ was assimilated to L1 /a/ 100% of the time with a goodness rating of 5.4). Therefore, the lack of vowel contrast effect may follow from the comparable levels of cross-language assimilation of the two dominant target vowels.

On the whole, the results of the two experiments show a clear link between the two measures and highlight the appropriateness of the eye-gaze data as a measure of cross-linguistic similarity. As discussed in the introduction, perceptual approaches to cross-linguistic similarity, such as the perceptual assimilation task in Experiment 1, are commonly used in speech perception and L2 speech research. However, these offline tasks have a number of limitations, including the large number of stimuli and trials required, the consequently repetitive nature of the task, and the potential influence of the response labels used in the task, which typically represent L1 categories orthographically. An online task

like a word recognition task implemented in a visual world paradigm can help to overcome these shortcomings. Nevertheless, the fact that the results of the two approaches examined in the current paper are closely related, as evidenced by the role of vowel dominance in Experiment 2, also indicates that despite its limitations, an offline perceptual approach does, in fact, yield valid results. In addition, a perceptual task is easier to design and implement than an eye-tracking online task, as the latter requires specialized equipment and a very careful selection of auditory and visual stimuli. From a methodological viewpoint, thus, both the offline perceptual approach and the online word processing alternative represent appropriate methods to assess cross-linguistic perceived similarity.

The relationship between the two measures at an individual level was also examined. The results of a correlational analysis exploring the relationship between perceptual and online measures at an individual level for each vowel separately yielded no significant correlations. A second analysis including all vowels in a single analysis revealed that the fixation durations to the interlingual competitor were correlated with two measures of perceived similarity (percent assimilation and fit index). Still, the correlations were weak. One possible explanation for the lack of stronger correlations may be the fact that for three of the four vowels, the results of the PAT were very consistent and showed very little variability. Still, even in the case of English /ɪ/, which displayed more variability in assimilation to an L1 category, no correlations were found in the analysis per vowel. It is possible that the measures under study are not fully comparable at the individual level and that the relationship between cross-linguistic perceived similarity and degree of L1 lexical activation during L2 spoken word recognition is better captured by contrasting and comparing patterns of group performance on both tasks.

The results of this study must be interpreted in light of several potential limitations. First, the results of Experiment 2 provided some evidence that the presence of an interlingual competitor affected the performance not only of the learners but also of the native controls, even if to a lesser extent. This shows that the vowel dominance effect was also present with native speakers, which was attributed to the English speakers' knowledge and exposure to Catalan and Spanish. Therefore, native speakers' baseline measures in Experiment 2 could have been more reliable and consistent with our predicted native baseline measures if native participants had been monolingual English speakers recruited in an English-speaking context. Secondly, the relatively small number of critical items in the cross-language lexical activation task (12, 3 for each of the 4 target vowels), although comparable to previous cross-language lexical activation studies (e.g., [Spivey and Marian 1999](#); [Marian and Spivey 2003](#); [Chambers and Cooke 2009](#)), probably affected the consistency of the fixation proportion measures at an individual participant level when comparing the degree of cross-language activation between dominant and non-dominant vowels. This limitation is inherent to the design of a cross-language lexical activation task, which is severely restricted by the possibilities the L2 and L1 lexicons offer to find adequate L1 interlingual homophone competitors for L2 words containing target confusable vowels. Finally, although target and competitor objects were carefully selected, taking into account their lexical frequency, related word familiarity, and graphic representability (most items were concrete nouns with comparable word frequency indices across AOI and language conditions), we cannot discard the potential influence of differences in the visual salience of images, which could also affect inter-subject differences in speed of lexical access. The inclusion of a familiarization task prior to the spoken word recognition task and the inclusion of random intercepts for both item and subject in the statistical analysis were expected to control for the potential effect of differences in visual representability. Still, future studies could consider incorporating a measure of visual salience of the object pictures to fully address this issue.

In conclusion, the results of the two experiments indicate that Catalan learners of English exhibit asymmetric levels of cross-language lexical competition between dominant and non-dominant L2 vowel categories. The amount of L1 activation during L2 speech processing is determined by the degree of similarity between the L1 and L2 vowels. This

is reflected in the fact that the degree of competition from interlingual homophones is greater (as indicated by greater proportions of looks to the interlingual competitor) when the target word contains a dominant vowel, that is, the L2 vowel that is perceptually closest to an L1 category (SSBE /æ/ and /i:/ in the current study). This outcome is in line with the asymmetries in lexical access observed in previous studies examining the L1 interference during L1 processing (Weber and Cutler 2004; Cutler et al. 2006). From a methodological point of view, the comparison between the two tasks shows that the two methodological approaches effectively capture differences in the degree of cross-linguistic similarity between native- and second-language sounds. Online tasks, thus, provide a potentially effective method of assessing cross-linguistic similarity without the shortcomings inherent to more traditional offline perceptual approaches. Still, perceptual tasks remain a valid and more easily implemented method of assessing cross-linguistic perceived similarity.

Author Contributions: J.C. and J.C.M. are responsible for the conceptualization, methodology, data collection, data analysis and storage, resources, writing, review and editing of the manuscript and funding acquisition. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by Research Grants No. FFI2017-88016-P and PID2021-122396NB-I00 to the first author and No. PID2019-107814GB-I00 and PID2022-138129NB-I00 to the second author from the Spanish Ministries of Economy and Competitiveness and Science and Innovation and by research grants from the Catalan Agency for Management of University and Research Grants (AGAUR) to the Experimental Phonetics research group [grant numbers 2017SGR34 and 2021SGR00544] and to the GRAL Research Group [grant numbers 2017SGR560201 and 2023SGR00303].

Institutional Review Board Statement: The study protocol adhered to the good practices of data collection, data anonymization, data processing, and data storage of the Institutional Review Board of the University of Barcelona (IRB0003099) and in accordance with the Declaration of Helsinki.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data presented in this study are available upon request from the corresponding author.

Acknowledgments: We would like to thank Joan Borràs-Comes and the Laboratori de Fonètica “Eugenio Martínez Celdrán” of the Universitat de Barcelona (Barcelona, Spain) for help with data processing and statistical analyses. We also thank Gisela Sosa for her help finding and selecting pictures.

Conflicts of Interest: The authors declare no conflicts of interest. The funders had no role in the design of the study; in the collection, analyses, or interpretation of data; in the writing of the manuscript; or in the decision to publish the results.

Appendix A

Code *orderNorm* in the package *bestNormalize* used to normalize the dependent variable used in the statistical analyses (fixation durations):

```
var_normalize = function(x, poly_n = 3, get_coefs = F) {
  require(tidyverse)
  x = x + abs(min(x, na.rm = T)) + 1
  ordernorm = x %>% bestNormalize::orderNorm(warn = F) %>% pluck("x.t")
  fit = lm(ordernorm ~ poly(log10(x), poly_n, raw = T), NULL)
  poly = poly(log10(x), poly_n, raw = T) %>% as_tibble() %>% pluck("1")
  normalized = predict(fit, as.data.frame(x))
  if (get_coefs == T) {result = list(norm = normalized, coef = enframe(coef(fit), name =
  NULL) %>% rename(coef = value), poly = poly)} else {result = normalized}
  return(result)
}
```

Appendix B

Table A1. Parameter estimates of the linear mixed-effects model with *group* (L2 learners, native English speakers), *dominance* (dominant vowels /i:/ and /æ/; non-dominant vowels /ɪ/ and /ʌ/), contrast (high vowels /i:/ and /ɪ/; low vowels /æ/ and /ʌ/) and *AOI* (target, competitor) and their interactions as fixed effects, and random intercepts for *subject* and *item*.

glmmTMB (fixdur_norm~group * dominance * contrast * aoi + (1 | subject) + (1 | item), data).

Predictors	β	SE	CI	t	p
Intercept	0.161	0.097	−0.029–0.350	1.664	0.096
group [NES]	−0.257	0.144	−0.540–0.025	−1.788	0.074
dominance [non-dominant]	−0.25	0.137	−0.518–0.018	−1.828	0.068
height [low]	−0.096	0.136	−0.363–0.172	−0.702	0.483
aoi [target]	0.134	0.107	−0.075–0.343	1.258	0.208
group [NES] × dominance [non-dominant]	0.132	0.202	−0.264–0.528	0.653	0.514
group [NES] × height [low]	0.197	0.205	−0.205–0.599	0.959	0.338
dominance [non-dominant] × height [low]	0.18	0.193	−0.199–0.558	0.93	0.352
group [NES] × aoi [target]	0.272	0.204	−0.127–0.671	1.334	0.182
dominance [non-dominant] × aoi [target]	0.502	0.151	0.207–0.797	3.333	0.001
height [low] × aoi [target]	0.052	0.15	−0.242–0.346	0.344	0.731
group [NES] × dominance [non-dominant] × height [low]	−0.302	0.287	−0.864–0.260	−1.052	0.293
group [NES] × dominance [non-dominant] × aoi [target]	−0.128	0.286	−0.688–0.432	−0.449	0.654
group [NES] × height [low] × aoi [target]	0.159	0.29	−0.409–0.727	0.548	0.584
dominance [non-dominant] × height [low] × aoi [target]	−0.158	0.213	−0.576–0.259	−0.744	0.457
group [NES] × dominance [non-dominant] × height [low] × aoi [target]	0.161	0.406	−0.635–0.956	0.396	0.692
Random Effects					
σ^2	0.69				
τ_{00} subject	0				
τ_{00} item	0.01				
N subject	61				
N item	12				
Observations	1350				
Marginal R2/Conditional R2	0.096/NA				

Table A2. Parameter estimates of the linear mixed-effects model with *group* (L2 learners, native English speakers), *dominance* (dominant vowels /i:/ and /æ/; non-dominant vowels /ɪ/ and /ʌ/) and *AOI* (target, competitor) and their interactions as fixed effects, and random intercepts for *subject* and *item*.

glmmTMB (fixdur_norm~group * dominance * aoi + (1 | subject) + (1 | item), data).

Predictors	β	SE	CI	t	p
Intercept	0.113	0.068	−0.021–0.246	1.651	0.099
group [NES]	−0.16	0.103	−0.361–0.042	−1.556	0.120
dominance [non-dominant]	−0.16	0.097	−0.349–0.030	−1.654	0.098
aoi [target]	0.16	0.075	0.013–0.307	2.129	0.033
group [NES] × dominance [non-dominant]	−0.018	0.144	−0.300–0.264	−0.126	0.900
group [NES] × aoi [target]	0.349	0.145	0.064–0.633	2.399	0.016
dominance [non-dominant] × aoi [target]	0.423	0.107	0.214–0.632	3.966	<0.001
group [NES] × dominance [non-dominant] × aoi [target]	−0.046	0.203	−0.444–0.353	−0.224	0.823
Random Effects					
σ^2	0.7				
τ_{00} subject	0				
τ_{00} item	0.01				

Table A2. *Cont.*

glmmTMB (fixdur_norm~group * dominance * aoi + (1 | subject) + (1 | item), data).

Predictors	β	SE	CI	t	p
N subject	61				
N item	12				
Observations	1350				
Marginal R2/Conditional R2	0.091/NA				

Table A3. Parameter estimates of the linear mixed-effects model with *group* (L2 learners, native English speakers), *dominance* (dominant vowels /i:/ and /æ/; non-dominant vowels /ɪ/ and /ʌ/) and *AOI* (competitor, distractor) and their interactions as fixed effects, and random intercepts for *subject* and *item*.

glmmTMB (fixdur_norm~group * dominance * aoi + (1 | subject) + (1 | item), data).

Predictors	β	SE	CI	t	p
Intercept	0.112	0.045	0.023–0.200	2.472	0.013
group [NES]	−0.161	0.078	−0.313–−0.008	−2.065	0.039
dominance [non-dominant]	−0.158	0.064	−0.284–−0.033	−2.471	0.013
aoi [dist]	−0.138	0.057	−0.250–−0.026	−2.423	0.015
group [NES] × dominance [non-dominant]	−0.017	0.109	−0.231–0.196	−0.159	0.873
group [NES] × aoi [dist]	0.091	0.11	−0.125–0.307	0.824	0.410
dominance [non-dominant] × aoi [dist]	0.066	0.081	−0.093–0.224	0.816	0.415
group [NES] × dominance [non-dominant] × aoi [dist]	0.071	0.154	−0.231–0.373	0.461	0.645
Random Effects					
σ^2	0.40				
τ_{00} subject	0.00				
τ_{00} item	0.00				
N subject	61				
N item	12				
Observations	1350				
Marginal R2/Conditional R2	0.021/NA				

Notes

- ¹ The same pattern of cross-linguistic perceived assimilation is observed for English /æ ʌ α:/ with respect to Catalan /a/ (Cebrian 2021).
- ² The results for the remaining vowels were the following: SSBE /e/ 76% as Catalan /ε/ (GR: 5.6 out of 7) and 24% as /e/ (GR: 4.5); SSBE /ei/ 92% as Catalan /ei/ (GR: 4.8) and 7% as /ai/ (GR: 3.4); SSBE /ai/ 98% as Catalan /ai/ (GR: 4.5). These results are in line with previous findings (e.g., Cebrian 2021).
- ³ The lexical frequency indices were obtained from the SUBTLEXus resource for English (Brybaert and New 2009) and the SUBTLEXcat resource for Catalan (NIM, Guasch et al. 2013), which indicate the relative frequency of a word per one million words.

References

Alario, F-Xavier, and Ludovic Ferrand. 1999. A set of 400 pictures standardized for French: Norms for name agreement, image agreement, familiarity, visual complexity, image variability, and age of acquisition. *Behavior Research Methods, Instruments, & Computers* 31: 531–52. [CrossRef]

Alloppenna, Paul D., James S. Magnuson, and Michael K. Tanenhaus. 1998. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* 38: 419–39. [CrossRef]

Best, Catherine T. 1995. A direct realist view of cross-language speech perception. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Edited by Winifred Strange. Timonium: York Press, pp. 171–204.

Best, Catherine T., and Michael D. Tyler. 2007. Non-native and second language speech perception. In *Language Experience in Second Language Speech Learning. In Honor of James Emil Flege*. Edited by Ocke-Schewn Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 15–34. [CrossRef]

Birdsong, David, Libby M. Gertken, and Mark Amengual. 2012. Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism. COERLL, University of Texas at Austin. Web. Available online: <https://sites.la.utexas.edu/bilingual/> (accessed on 20 January 2012).

- Bohn, Ocke-Schewn. 2002. On phonetic similarity. In *An Integrated View of Language Development: Papers in Honor of Henning Wode*. Edited by Petra Burmeister, Thorsten Piske and Andreas Rohde. Trier: Wissenschaftlicher Verlag, pp. 191–216.
- Bohn, Ocke-Schewn. 2017. Cross-language and second language speech perception. In *The Handbook of Psycholinguistics*. Edited by Eva M. Fernández and Helen Smith Cairns. Chichester: Wiley-Blackwell, pp. 213–39.
- Bonin, Patrick, Ronald Peereman, Nathalie Malardier, Alain Méot, and Marylène Chalard. 2003. A new set of 299 pictures for psycholinguistic studies: French norms for name agreement, image agreement, conceptual familiarity, visual complexity, image variability, age of acquisition, and naming latencies. *Behavior Research Methods, Instruments, & Computers* 35: 158–67. [\[CrossRef\]](#)
- Brooks, Mollie E., Kasper Kristensen, Koen J. Van Benthem, Arni Magnusson, Casper W. Berg, Anders Nielsen, Hans J. Skaug, Martin Machler, and Benjamin M. Bolker. 2017. GlmmTMB balances speed and flexibility among packages for zero-inflated generalized linear mixed modelling. *The R Journal* 9: 378–400. [\[CrossRef\]](#)
- Brysbaert, Marc, and Boris New. 2009. Moving beyond Kucera and Francis: A Critical Evaluation of Current Word Frequency Norms and the Introduction of a New and Improved Word Frequency Measure for American English. *Behavior Research Methods* 41: 977–90. [\[CrossRef\]](#)
- Cebrian, Juli. 2006. Experience and the use of non-native duration in L2 vowel categorization. *Journal of Phonetics* 34: 372–87. [\[CrossRef\]](#)
- Cebrian, Juli. 2019. Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs. *Journal of the Acoustical Society of America* 145: EL52–EL58. [\[CrossRef\]](#)
- Cebrian, Juli. 2021. Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity. *Journal of the Acoustical Society of America* 149: 2671–85. [\[CrossRef\]](#)
- Cebrian, Juli. 2022. Perception of English and Catalan vowels by English and Catalan listeners. Part II. Perceptual vs ephoric similarity. *Journal of the Acoustical Society of America* 152: 2781–93. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cebrian, Juli, Joan C. Mora, and Cristina Aliaga García. 2011. Assessing cross-linguistic similarity by means of rated discrimination and perceptual assimilation tasks. In *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*. Edited by Magdalena Wrembel, Małgorzata Kul and Katarzyna Dziubalska-Kołaczyk. Frankfurt am Main: Peter Lang, vol. I, pp. 41–52.
- Chambers, Craig G., and Hilary Cooke. 2009. Lexical competition during second-language listening: Sentence context, but not proficiency, constrains interference from the native lexicon. *Journal of Experimental Psychology: Learning Memory and Cognition* 35: 1029–40. [\[CrossRef\]](#) [\[PubMed\]](#)
- Cutler, Anne, Andrea Weber, and Takashi Otake. 2006. Asymmetric mapping from phonetic to lexical representations in second-language listening. *Journal of Phonetics* 34: 269–84. [\[CrossRef\]](#)
- Dahan, Delphine, and Michael K. Tanenhaus. 2005. Looking at the rope when looking for the snake: Conceptually mediated eye movements during spoken-word recognition. *Psychonomic Bulletin & Review* 12: 453–59. [\[CrossRef\]](#)
- Dahan, Delphine, James S. Magnuson, and Michael K. Tanenhaus. 2001. Time course of frequency effects in spoken-word recognition in French. *Cognitive Psychology* 42: 317–67. [\[CrossRef\]](#) [\[PubMed\]](#)
- Escudero, Paola, Rachel Hayes-Harb, and Holger Mitterer. 2008. Novel second-language words and asymmetric lexical access. *Journal of Phonetics* 36: 345–60. [\[CrossRef\]](#)
- Faris, Mona M., Catherine T. Best, and Michael D. Tyler. 2018. Discrimination of uncategorised non-native vowel contrasts is modulated by perceived overlap with native phonological categories. *Journal of Phonetics* 70: 1–19. [\[CrossRef\]](#)
- Flege, James Emil. 1995. Second-language Speech Learning: Theory, Findings, and Problems. In *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*. Edited by Winifred Strange. Timonium: York Press, pp. 229–73.
- Flege, James Emil, and Ocke-Schwen Bohn. 2021. The revised speech learning model (SLM-r). In *Second Language Speech Learning: Theoretical and Empirical Progress*. Edited by R. Wayland. Cambridge: Cambridge University Press, pp. 3–83. [\[CrossRef\]](#)
- Flege, James Emil, Murray J. Munro, and Robert Allen Fox. 1994. Auditory and categorial effects on cross-language vowel perception. *The Journal of the Acoustical Society of America* 95: 3623–41. [\[CrossRef\]](#) [\[PubMed\]](#)
- Guasch, Marc, Roger Boada, Pilar Ferré, and Rosa Sánchez-Casas. 2013. NIM: A Web-based Swiss Army knife to select stimuli for psycholinguistic studies. *Behavior Research Methods* 45: 765–71. [\[CrossRef\]](#) [\[PubMed\]](#)
- Guion, Susan G., James Emile Flege, Reiko Akahane-Yamada, and Jessica C. Pruitt. 2000. An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants. *Journal of the Acoustical Society of America* 107: 2711–24. [\[CrossRef\]](#)
- Iverson, Paul, Patricia K. Kuhl, Reiko Akahane-Yamada, Eugen Diesch, Andreas Kettermann, and Claudia Siebert. 2003. A perceptual interference account of acquisition difficulties for non-native phonemes. *Cognition* 87: B47–B57. [\[CrossRef\]](#)
- Ju, Min, and Paul A. Luce. 2004. Falling on sensitive ears: Constraints on bilingual lexical activation. *Psychological Science* 15: 314–18. [\[CrossRef\]](#)
- Kuhl, Patricia K., Barbara T. Conboy, Sharon Coffey-Corina, Denise Padden, Maritza Rivera-Gaxiola, and Tobey Nelson. 2008. Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e). *Philosophical Transactions of the Royal Society B* 363: 979–1000. [\[CrossRef\]](#)
- Levy, Erika S. 2009. Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French. *Journal of the Acoustical Society of America* 125: 1138–52. [\[CrossRef\]](#)
- Marian, Viorica, and Michael Spivey. 2003. Competing activation in bilingual language processing. *Bilingualism: Language and Cognition* 6: 97–115. [\[CrossRef\]](#)
- Meara, Paul M., and James Milton. 2003. *X-lex: The Swansea Levels Test*. Berkshire: Express Publishing.

- Mora, Joan C., and Ingrid Mora-Plaza. 2019. Contributions of cognitive attention control to L2 speech learning. In *A Sound Approach to Language Matters—In Honor of Ocke-Schwen Bohn*. Edited by Anne Mette Nyvad, Michaela Hejná, Anders Højen, Anna Bothe Jespersen and Mette Hjortshøj Sørensen. Aarhus: Department of English, School of Communication & Culture, Aarhus University, pp. 477–99.
- Peterson, Ryan A. 2021. Finding optimal normalizing transformations via bestNormalize. *The R Journal* 13: 310–29. [[CrossRef](#)]
- Peterson, Ryan A., and Joseph E. Cavanaugh. 2020. Ordered quantile normalization: A semiparametric transformation built for the cross-validation era. *Journal of Applied Statistics* 47: 2312–27. [[CrossRef](#)]
- Rallo Fabra, Lucrècia. 2005. Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test. *Vigo International Journal of Applied Linguistics* 2: 75–92.
- Snodgrass, Joan G., and Mary Vanderwart. 1980. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory* 6: 174–215. [[CrossRef](#)]
- Spivey, Michael J., and Viorica Marian. 1999. Cross talk between native and second languages: Partial activation of an irrelevant lexicon. *Psychological Science* 10: 281–84. [[CrossRef](#)]
- Strange, Winifred. 2007. Cross-language similarity of vowels. Theoretical and methodological issues. In *Language Experience in Second Language Speech Learning*. Edited by Ocke-Schewen Bohn and Murray J. Munro. Amsterdam: John Benjamins, pp. 35–55. [[CrossRef](#)]
- Strange, Winifred, Erika S. Levy, and Franzo F. Law. 2009. Cross-language categorization of French and German vowels by naïve American listeners. *The Journal of the Acoustical Society of America* 126: 1461–76. [[CrossRef](#)]
- Tanenhaus, Michael K., Michael J. Spivey-Knowlton, Kathleen M. Eberhard, and Julie C. Sedivy. 1995. Integration of visual and linguistic information in spoken language comprehension. *Science* 268: 1632–34. [[CrossRef](#)]
- Tsang, Cara, and Craig G. Chambers. 2011. Appearances aren't everything: Shape classifiers and referential processing in Cantonese. *Journal of Experimental Psychology: Learning, Memory, and Cognition* 37: 1065–80. [[CrossRef](#)] [[PubMed](#)]
- Tyler, Michael D. 2021. Phonetic and phonological influences on the discrimination of non-native phones. In *Second Language Speech Learning: Theoretical and Empirical Progress..* Edited by Ratreë Wayland. Cambridge: Cambridge University Press, pp. 157–74. [[CrossRef](#)]
- Tyler, Michael D., Catherine T. Best, Alice Faber, and Andrea G. Levitt. 2014. Perceptual assimilation and discrimination of non-native vowel contrasts. *Phonetica* 71: 4–21. [[CrossRef](#)] [[PubMed](#)]
- Weber, Andrea, and Anne Cutler. 2004. Lexical competition in non-native spoken-word recognition. *Journal of Memory and Language* 50: 1–25. [[CrossRef](#)]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.