

Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity

Juli Cebrian^{a)}

Departament de Filologia Anglesa i de Germanística, Facultat de Filosofia i Lletres, Universitat Autònoma de Barcelona, Campus UAB, Bellaterra 08193, Spain

ABSTRACT:

This study reports the results of two experiments aimed at assessing the perceived similarity between English and Catalan vowels and diphthongs. Perceived similarity judgements were elicited from speakers of both languages involved and were compared to first language identification data from the same speakers. In experiment 1, a group of 27 naive Catalan listeners performed a perceptual assimilation task in which they were asked to identify Catalan and English vowels in terms of native categories and provide a goodness of fit rating. In experiment 2, a group of native speakers of Southern British English performed an adaptation of the same task. The results showed that most non-native vowels were consistently perceived as instances of a given native category, with varying degrees of goodness of fit. In a few cases, assimilation scores were very high in both experiments, pointing to the possibility of near-identical or shared categories. A few asymmetrical mappings were found, which were linked to the influence of language-specific cues such as the role of vowel duration in English. These results emphasize the importance of contrasting native and non-native perception and the potential of reciprocal approaches for making predictions about non-native perception and second language development. © 2021 Acoustical Society of America.

<https://doi.org/10.1121/10.0004257>

(Received 13 September 2020; revised 15 March 2021; accepted 23 March 2021; published online 15 April 2021)

[Editor: Megha Sundara]

Pages: 2671–2685

I. INTRODUCTION

The idea that the first language (L1) phonological system influences the perception of non-native sounds is widely accepted. For example, the native language magnet model (Kuhl and Iverson, 1995) posits that through exposure to the ambient language, speakers develop L1 sound prototypes that act as perceptual magnets so that the L1 system functions as a filter through which native and non-native phones are perceived. Models of second language (L2) speech crucially draw on the notion of cross-linguistic similarity to explain the processes that take place in adult L2 speech acquisition and to make predictions about the difficulty of target language perception and production. Flege (1987) proposes that target language phones are perceived in terms of L1 categories to varying degrees through a process of equivalence classification. L2 phones that are perceived as identical or similar to L1 phones will be classified in terms of the closest L1 categories, while L2 phones that lack a clear counterpart in the L1 will pattern as new phones. According to the perceptual assimilation model (PAM; Best, 1995), non-native phones will be perceptually assimilated to (i.e., heard as instances of) the closest native category and will be categorized as better or worse exemplars of that category or be uncategorized as a L1 sound or even heard as non-speech. Based on the possible patterns of

assimilation of non-native phones to native categories, PAM makes a series of predictions for accuracy of discrimination of non-native segmental contrasts, which PAM-L2 (Best and Tyler, 2007) extends to L2 contrasts. Flege's (1995) speech learning model (SLM, recently revised as the SLM-r; Flege and Bohn, 2021) propounds that the closer a target language phone is perceived to be to a native phone, the more likely it will be identified with the native category and consequently the less likely it will be learned as a separate L2 category. By contrast, given enough authentic input from the target language, a more target-like category may be established for a more dissimilar L2 phone (Flege, 1995; Flege and Bohn, 2021). What these models have in common is their reliance on the construct of cross-linguistic similarity in order to make predictions about categorization of L2 sounds. The notion of phonetic similarity is in fact a key concept in the phonetics sciences, as discussed below. The current paper thus aims to contribute to our understanding of this concept by examining the perception of similarity between English and Catalan vowels by means of a novel method that evaluates the perceptual judgements of two parallel populations, namely, native English speakers and native Catalan speakers, and contrasts native and non-native perception.

A. Measuring perceptual similarity

Phonetics is concerned with the extent to which languages differ from one another in their inventories of sounds,

^{a)}Electronic mail: juli.cebrian@uab.cat, ORCID: 0000-0003-1547-3145.

and thus one of the principal questions it addresses is the degree to which different sounds resemble each other (Bohn, 2002; Ladefoged, 1990). This question is related to the second principle of the International Phonetic Association, which proposes that the same phonetic symbol should be used for any sound, or any similar shades of sounds, found in more than one language (e.g., Abramson, 1988; p. 66). Bohn (2002) underscores how the notion of phonetic similarity is crucial for a variety of areas of phonetics, including the description of sound inventories, the development of transcription systems and phonetic models, and the study of language typology and sound change, as well as for research on language acquisition, speech errors, and automatic speech recognition. Thus, finding a proper definition of phonetic similarity and establishing appropriate methods for measuring it is a basic issue not only for L2 speech research but for the phonetic sciences in general.

Cross-linguistic similarity measures typically include articulatory descriptions, acoustic comparisons, and perceptual measures [see Bohn (2002) and Strange (2007) for reviews of different approaches]. Strange (2007) notes that acoustic comparisons may be complicated by variability caused by phonetic context, speech rate, and individual variation. Discrepancies between acoustic and perceptual measures have been reported (Strange *et al.*, 2004; Strange, 2007). Given the difficulty of establishing appropriate acoustic parameters for cross-language comparisons, researchers have advocated for perceptual measures of cross-language similarity, particularly when intended to assess the extent to which attunement to L1 perception influences the categorization of non-native or L2 sounds (Bohn, 2002; Flege and Bohn, 2021; Strange, 2007; Tyler, 2021; Tyler *et al.*, 2014; among others).

A common perceptual technique for measuring cross-linguistic similarity is an interlingual identification task, also known as perceptual assimilation task or forced choice categorization task (Baigorri *et al.*, 2019; Best, 1995; Faris *et al.*, 2018; Strange *et al.*, 2009; among many others). In a perceptual assimilation task (PAT), listeners are asked to identify exemplars of non-native phones in terms of L1 categories, typically by means of labels representing the L1 sounds, and to provide goodness of fit ratings. Although PATs are probably the most widely used tasks in studies on cross-linguistic similarity, other approaches such as paired comparison techniques are also found, e.g., a rated dissimilarity task (RDT; Cebrian *et al.*, 2011; Flege *et al.*, 1994).¹ In a RDT, listeners are presented with two stimuli and are required to indicate on a (dis)similarity scale how similar the two stimuli are perceived to be.

The results of these tasks are then interpreted as the degree to which the non-native categories are similar to L1 categories. In fact, the interpretation of the results of perceptual tasks is not unproblematic. The classification of non-native (or L2) sounds as “new” phones or as “similar” or “identical” to L1 phones (Flege, 1987) is not straightforwardly defined in terms of specific assimilation scores. PAM (Best, 1995) and PAM-L2 (Best and Tyler, 2007)

distinguish between categorized and uncategorized phones, the latter being non-native phones that are not perceived to have a counterpart among the L1 phonological categories. The model explores whether pairs of contrasting non-native phones map onto two separate L1 categories or a single L1 category (with possible varying degrees of assimilation) or whether one of the two members is uncategorized. Based on the type of assimilation and the degree of cross-language assimilation (or perceived phonological) overlap between the members of each pair (i.e., overlap in the categorizations to native phones; Faris *et al.*, 2018; Levy, 2009b), different predictions are made for discrimination accuracy of non-native pairs. The threshold for categorization as a L1 phone, however, varies from one study to another: 50% assimilation as a given L1 category (Bundgaard-Nielsen *et al.*, 2011; Faris *et al.*, 2018), 70% (Tyler *et al.*, 2014), or even 90% (Harnsberger, 2001). The focus of this paper is not to evaluate the best way to quantify a categorization threshold, but rather to compare the perceived similarity of the sounds of two languages by speakers of those two languages. One of its goals is to investigate whether some non-native sounds can be perceived as good instances of L1 categories. The approach adopted is to evaluate cross-linguistic similarity by contrasting the perception of L1 and non-native sounds. In this regard, this paper follows Guion *et al.* (2000), who propose assessing assimilation scores in light of their similarity to L1 identification scores (see Sec. II C). The results of previous work analyzing cross-linguistic perception of Catalan and English are discussed next.

B. Previous work on Catalan and English

This study focuses on the perceived similarity between Standard Southern British English (SSBE) and Catalan vowels. The Central or Eastern variety of Catalan has an inventory of seven vowels (/i, e, ε, a, ɔ, o, u/), a reduced vowel ([ə]) in unstressed position, and a series of diphthongs resulting from the combination of vowels and high glides, e.g., /ai, ei, au, ou/. SSBE has a larger vowel inventory than Catalan, consisting of 12 vowels (/i, ɪ, e, ɜ, æ, ʌ, ɑ, ɒ, ɔ, ʊ, u, ə/) and eight diphthongs, e.g., /ai, aʊ, eɪ, əʊ/.² Figure 1 presents the vowel (monophthong) systems of Catalan and English based on the standard descriptions in Carbonell and Llisterra (1992) and Roach (2004), respectively.

Some previous studies have explored the perceived similarity between English and Catalan vowels. Cebrian (2006) examined the perceptual similarity between Canadian English /i, ɪ, eɪ, ε/ and the acoustically closest Catalan counterparts /i, e, ei, ε/. Vowel stimuli were presented in isolation. Catalan speakers with little knowledge of English perceived English /i/, /eɪ/, and /ε/ to be closest to Catalan /i/, /ei/, and /ε/, respectively, about 84%–99% percent of the time, with goodness of fit ratings (GRs) of 4.2–6.2/7, but English /ɪ/ obtained lower assimilation scores as Catalan /e/ (66%, GR: 3.5/7). These results were replicated in a later study involving a larger number of Catalan and Canadian English vowels and diphthongs presented in /bVs/ words

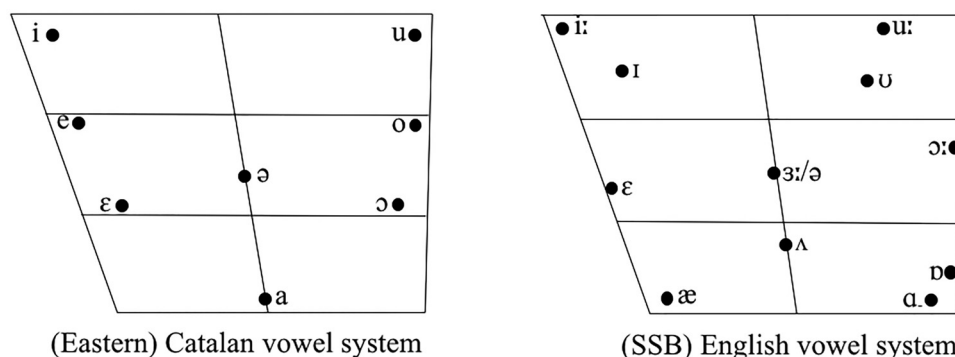


FIG. 1. Vowel systems of Catalan (Eastern variety) and English (SSBE) [based on Carbonell and Llisterrí (1992) and Roach (2004), respectively].

(Cebrian, 2009), which also found that English /æ, ʌ, u, ou/ were strongly assimilated to L1 vowel categories (namely, /a, a, u, ou/, respectively). Rallo Fabra (2005) tested two groups of Catalan learners of English varying in L2 experience, a group of naive listeners and a group of undergraduate students majoring in English, on their perception of American English vowels presented in /sVt/ words. Although generally the perceptual assimilation patterns were similar to previous findings, the assimilation rates obtained by Rallo Fabra were notably lower than those reported by Cebrian (2006, 2009). This may be due to methodological differences concerning the dialectal background of the native speakers who provided the stimuli (speakers were from different parts of the United States), the consonantal context of the vowels, and the fact that Rallo Fabra included “non-Catalan” as a possible response alternative.

Cebrian *et al.* (2011) collected perceptual similarity judgements from a group of 20 Catalan learners of English in an experiment that included Catalan and SSBE vowels presented in /bVt/ words. The results of a PAT showed that the most consistent English-to-Catalan assimilation patterns, with scores above 90% and GRs above 4.5/7, were /æ/-/a/ (100%, 4.7), /i:/-/i/ (96%, 4.6), and /ε/-/ε/ (91%, 4.7). These were followed by /u:/-/u/ (93%, 3.8), /ei/-/ej/ (90%, 3.5), /əu/-/ou/ (86%, 3.6), /ɔ:/-/o/ (86%, 3.6), /ʌ/-/a/ (85%, 3.7), and /ɪ/-/i/ (82%, 3.2) with very high scores but lower GRs. Finally, the assimilation patterns were least consistent with /ɑ:/-/a/ (70% and 2.5, while /ɑ:/-/ɔ/ reached 20% and 3.9), /ɒ/-/ɔ/ (70% and 4.1, with 28% and 4.4 for /ɒ/-/o/), and English /ɜ:/, which did not find a match in the L1 (as /e/ 30%, 2.0; as /ε/ 24%, 1.7; as /a/ 23%, 1.4, and as /o/ and /ɔ/ 9% and 2.1–2.7). Hence, the findings in Cebrian *et al.* (2011) showed that some English vowels were perceived as being very similar to native categories, a few more were consistently mapped onto a L1 category but judged as worse fitting, and a subset were perceived as not having a clear counterpart in the L1, particularly English /ɜ:/. However, the stimuli used in the experiments were unmodified /bVt/ words, and it is possible that differences in the prevoicing of the /b/ and the release (and aspiration) of the /t/ may have influenced the listeners’ similarity judgements. Consonantal context has been found to affect perception of vowel similarity (Bohn and Steinlen, 2003; Levy, 2009a, 2009b;

Schmidt, 2007; Strange *et al.*, 2001; but cf. Strange *et al.*, 2005). In addition, the listeners in that study were experienced L2 English speakers (undergraduate English majors). The current study differs from Cebrian *et al.* (2011) in the characteristics of the stimuli and in the population tested, made up of Catalan speakers with no experience or minimal experience with English. No previous study on Catalan and English has collected similarity judgements from SSBE native speakers. This study examines the perceived similarity of the same set of native and non-native vowels by native speakers of the two languages involved (hence its reciprocal nature), as discussed next.

C. Reciprocal study

Cross-linguistic similarity studies typically focus on a single population (a group of L2 learners or naive listeners), which is presented with a set of non-native or L2 sounds. A more complete picture of the similarity relationships between two sound systems may be achieved by eliciting similarity judgements from two parallel populations, that is, from speakers of the two languages being contrasted. This bidirectional approach may also help identify cases of phones that are possibly shared by the two languages, that is, cases of near-identity between native and non-native sounds. For example, Flege (1992) suggests that to know whether a non-native sound is perceived to be indistinguishable from a native category, this non-native phone should be undetectable when produced in a native context. Further, a reciprocal approach can help evaluate to what extent perceptual distance is a shared construct by speakers of the two languages involved, that is, an intrinsic characteristic of the sounds that are contrasted, or if it is influenced by the way sounds are organized in each of the languages involved. To our knowledge, no previous research has looked at cross-linguistic perceived similarity with reciprocal populations in the same study, particularly involving vowels.

The only prior research is a set of two studies on cross-linguistic perception of consonants by Korean (Schmidt, 1996) and English (Schmidt, 2007) speakers. Schmidt found evidence of parallel cross-linguistic mappings emphasizing the perceived similarity between the native and non-native sounds. However, she also found that cross-linguistic

similarity varied for the two language groups, depending on context-specific cues such as amount of labialization of the following vowel, variations in voice-onset time (VOT), and the presence of burst-like transitions in nasals. Schmidt concluded that “reciprocal studies...can delineate the relationship between sounds in two languages, as well as the cues used by speakers of each language so that more accurate predictions about cross-linguistic speech perception can be formulated” (Schmidt, 2007; pp. 199–200). With these arguments in mind, the current study assesses cross-linguistic perceived similarity by testing native speakers of the two languages examined.

D. Aims

This study aims to contribute to our general understanding of the concept of phonetic similarity by evaluating a novel method of measuring cross-linguistic similarity that involves (a) comparing the perceptual evaluation of native and non-native sounds and (b) contrasting the perceptual judgements of two parallel populations, naive listeners of the two languages involved. The study also investigates to what extent cross-linguistic perceived similarity is affected by the characteristics of the sound systems evaluated. The global objective is to assess the similarity of a near-complete set of Catalan and English vowels and diphthongs to be able to predict areas of difficulty in the perception and production of L2 vowels for adult L2 learners and to evaluate whether there are (perceptually) near-identical cross-linguistic categories, that is, if there are non-native phones that can be considered acceptable exemplars of native categories. The results of two experiments are presented. In the first experiment, a group of native Catalan speakers assessed the perceived similarity between Catalan and English vowels by means of a PAT. In the second experiment, an adaptation of the same task was administered to a group of English native speakers.

II. EXPERIMENT 1: PERCEIVED SIMILARITY OF SSBE VOWELS BY CATALAN SPEAKERS

A. Methodology

1. Participants

The participants were 27 Catalan-dominant Catalan-Spanish bilinguals (15 females) selected from a pool of 38 potential participants on the basis of their responses to a personal and linguistic background questionnaire. Selected participants were born to Catalan-speaking parents, lived in Catalan-speaking homes, and reported using Catalan 75%–100% of the time. Their average age was 24, ranging from 17 to 48. Most participants were undergraduate students. The ideal participants would be speakers with little or no knowledge of the non-native language in the experiment. While this was not a problem for the English subjects in experiment 2, it was more difficult to accomplish with the Catalan participants. Most had studied English as a foreign language (EFL) at school. However, the focus of EFL

instruction is typically grammar and vocabulary rather than pronunciation and perception. Thus, their familiarity with the English vowel system was considered minimal. None of the Catalan participants had spent more than 1 month in an English-speaking country. For these reasons, the participants can be considered naive listeners (Best and Tyler, 2007). All participants reported normal hearing and were compensated for their participation.

2. Stimuli

The stimuli consisted of 13 SSBE and 11 Catalan vowels and diphthongs. The Catalan vowels were the seven monophthongs /i e ε a ɔ o u/. The reduced vowel ([ə]) was not included, as it only appears in unstressed position in Eastern Catalan and is subject to large contextual variability (Recasens, 2014). In addition, four rising diphthongs, namely, /ej ou aj au/, were also included. To limit the complexity and duration of the task, only a subset of all possible Catalan rising diphthongs, i.e., combinations of a monophthong and a semivowel (/i/ or /u/), were used. The selected diphthongs were the ones that had been found to be perceptually closest to English vowels in previous studies involving Catalan and English (Cebrian, 2009; Cebrian *et al.*, 2011). The English stimuli contained the vowels and diphthongs /i: ɪ ε æ aɪ aʊ ʌ ɑ: ɒ ɔ u: eɪ əʊ/ (the English vowels /ɜ:/ and /ʊ/ were excluded from the PAT, as previous studies have shown that these vowels are not consistently assimilated to any single Catalan vowel and this paper explores the prevalence of reciprocal cross-linguistic mappings).³ The vowels were elicited in /b/ + vowel + /t/ sequences, which is a possible sequence in both languages. To elicit these sequences, the /bVt/ words were embedded in a carrier phrase that included a rhyming real word, as the /bVt/ context created some non-words particularly in Catalan. The English and Catalan carrier phrases were “It rhymes with hot. I say *bot*. I say *bot* again,” and “Rima amb dit. Ara dic *bit*. Ara dic *bit* un cop,” respectively. Each English speaker recorded six repetitions of the phrases with the words beat /i:/, bit /ɪ/, bet /ε/, bat /æ/, but /ʌ/, Bart /ɑ:/, bot /ɒ/, bought /ɔ:/, boot /u:/, bait /eɪ/, boat /əʊ/, bite /aɪ/, and bout /aʊ/, at a comfortable pace and with falling intonation. The order of the carrier phrases was randomized in each repetition. The Catalan words were elicited in the same fashion. The Catalan sequences were bat /a/, bét /e/, bèt /ε/, bit /i/, bót /o/, bòt /ɔ/, but /u/, bait /aj/, beit /ej/, baut /au/, and bout /ou/.

Stimuli were elicited from three male native speakers per language. The SSBE speakers (mean age: 35, range 29–44) had spent most of their lives in the South of England. They were recorded in a soundproof booth at the research laboratory at University College London. The Catalan speakers were Catalan-dominant Catalan-Spanish bilinguals living in Barcelona and neighbouring areas (mean age: 30, range: 21–35). They were selected on the grounds of a language use questionnaire and a short interview with a phonetically trained Catalan researcher. Catalan recordings were made in a sound-attenuated room at EUPMt technical

university (Escola Universitària Politècnica de Mataró, Mataró, Barcelona). Recordings were made with a digital recorder [Marantz (Kanagawa, Japan) PMD660] and a uni-directional dynamic microphone [Shure (Niles, IL) SM58]. The recordings were digitized at a 44 kHz sampling rate, and the stimuli were normalized for intensity (70 dB). The best two tokens per talker and language were selected, based on auditory judgements and spectrographic analyses. Figure 2 displays the distribution of the English and Catalan vowel stimuli in a $F1 \times F2$ (first and second formant) space, averaged across the three male speakers for each language [Lobanov's (1971) speaker normalization method was applied to the formant data; see Adank *et al.* (2004) for a discussion of normalization methods]. The $F1$ and $F2$ of every vowel stimulus were measured from a 50-ms window located manually at a steady-state portion between one-third and two-fourths into the vowel, using PRAAT (Boersma and Weenink, 2018). The resulting measurements were in accordance with reported values for male speakers of SSBE (Deterding, 1997; Ferragne and Pellegrino, 2010) and Eastern Catalan (Recasens and Espinosa, 2006; Recasens, 2014). Among the front vowels, some Catalan (C) and English (E) vowels appear to be acoustically very similar, namely, the low vowels C /a/ and E /æ/ and the mid front vowels C /e/ and E /ɛ/, followed by C /i/ and E /i:/ and C /ɪ/ and E /ɪ/. Regarding the back vowels, the plot indicates a less clear pattern of acoustic mappings, with C /ɔ/ and E /ɑ:/ having the greatest similarity. English /u/ appears to be notably fronted, in agreement with previous descriptions (Ferragne and Pellegrino, 2010; Harrington *et al.*, 2008). Diphthongs were measured at two different points, roughly 25% and 75% into the vowel, and presented expected values in relation to the location of the monophthongs. Finally, the duration of the Catalan vowels ranged from 129 ms (/i/) to 229 ms (/ou/), showing the expected increase in duration from high to low vowels (e.g., Recasens, 2014), and diphthongs were longer than monophthongs. English vowel duration ranged from 136 ms for /ɪ/ to 258 ms for /ɑ:/, with tense vowels being overall longer than lax vowels.

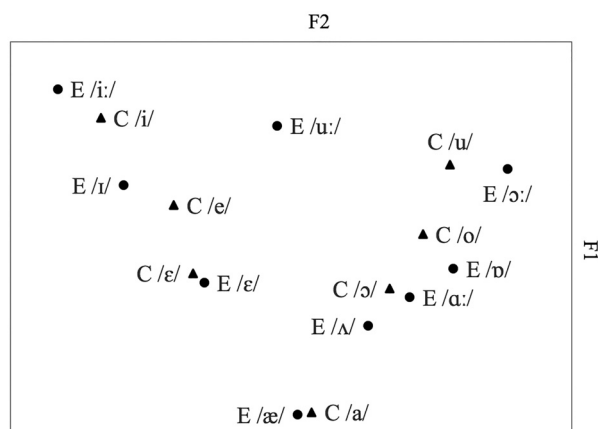


FIG. 2. Normalized $F1 \times F2$ values of each of the English (E) and Catalan (C) vowel stimuli, averaged across three male speakers of each language. The Lobanov transformation has been used (Lobanov, 1971).

Neighbouring sounds have been found to affect perceptual assimilation judgements (Bohn and Steinlen, 2003; Levy, 2009a, 2009b; Schmidt, 2007). In addition, English and Catalan differ in the amount of prevoicing of word-initial voiced stops (typically around 0 ms or short-lag VOT for English, voice lead for Catalan voiced stops). Consequently, stimuli were edited to include from the release of the /b/ until the end of the vowel portion and beginning of the /t/ closure. Thus, the portion corresponding to the /b/ closure and the release of the /t/ were excluded while maintaining intact the cues to the vowel.

3. PAT

In this task, listeners heard Catalan and English vowel stimuli presented in a random order and had to identify each stimulus as one of several possible L1 categories and then provide 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, example. Both L1 and non-native vowels were included in the PAT so that the tasks used with Catalan speakers and English speakers were as similar as possible. Crucially, the native perception data were used as a baseline for comparison (e.g., Cebrian, 2006; Escudero and Chládková, 2010; Guion *et al.*, 2000). Further, including L1 sounds facilitates the assessment of whether participants understood the instructions and used the response categories correctly.

Although some studies propose including all possible L1 categories as response options in PATs (e.g., Bundgaard-Nielsen *et al.*, 2011), a decision was made to divide the PAT into two smaller tasks, each with a subset of the possible L1 responses. This was motivated by the fact the inclusion of 11 Catalan vowels and 13 English vowels in a single task resulted in a very large number of stimuli, prone to fatigue effects. In addition, previous studies provide information concerning what L1 options are more often selected for each non-native vowel (e.g., Cebrian *et al.*, 2011; Rallo Fabra, 2005). Thus, the stimuli were divided into two sets: set A involved mostly front vowels (English /aɪ, æ, eɪ, ɛ, ɪ, i:, ʌ/ and Catalan /aj, a, ej, ɛ, e, i/), and set B consisted of mostly back and rounded vowels (English /aʊ, æ, ɒ, ɔ:, əʊ, u:, ɑ:/ and Catalan /au, a, ɔ, o, ou, u/). Two tokens of each vowel produced by each of the three talkers were included. The non-native tokens were repeated twice, while the native tokens were presented once. To have balanced sets, the English vowel /æ/ and the Catalan vowel /a/ appeared in both sets, and each token was presented only once in each set. The total number of trials in each task was 114, 78 involving non-native (English) vowels and 36 L1 Catalan vowels. The response alternatives consisted of high frequency keywords representing each of the L1 sounds. The response labels for set A were *xai* (for /aj/), *gat* (/a/), *rei* (/ej/), *set* (/ɛ/), *fet* (/e/), and *dit* (/i/), meaning sheep, cat, king, seven, fact, and finger, respectively. In the case of set B, the response options were *gat* (/a/), *dau* (/au/), *got* (/ɔ/),

mot (/o/), *pou* (/ou/), and *mut* (/u/), meaning cat, dice, drinking glass, word, water well, and mute.⁴

4. Procedure

The order of the two sets (A and B) was counterbalanced across all listeners. Instructions were given at the beginning of the experiment. Listeners were encouraged to use the whole range of the scale. The listeners performed the tasks individually in a soundproof room. They listened to the stimuli over headphones [Sennheiser (Wedemark, Germany) HD-25] and responded by clicking on one of the responses appearing on the computer screen. There was no time limit for the responses, and each stimulus could be replayed once if necessary, but participants were encouraged to respond based on their immediate impressions. There was a possibility for a short pause every 24 trials. PRAAT software (Boersma and Weenink, 2018) was used to conduct the experiments. The experimental task was preceded by a short practice phase to familiarize participants with the range of possible identification and rating responses. The practice session consisted of 12 trials (6 native and 6 non-native vowels). The participants took on average 20–30 min to complete both sets of the PAT. The participants in this study performed another perceptual task after the PAT, whose results are not reported in this paper.

B. Results

The percentage of times each L1 and non-native stimulus were identified as each of the L1 response categories and the mean GRs for each were calculated.⁵ This analysis provided the correct identification scores for Catalan vowels and modal assimilation scores (most frequent identification as a L1 vowel) for each non-native vowel. Table I shows the identification accuracy of each of the L1 vowels, the GR received, and the fit index score, a composite score that combines the two previous measures by multiplying the identification proportion by the goodness of fit rating (Guion *et al.*, 2000; Lengeris, 2009; among others).⁶ The fit index score will be used when comparing native and non-native

perception as well as when contrasting results across experiments.

The average correct identification of the Catalan vowels was 92% (range 80%–100%), with an average rating of 5.9 and a mean fit index score of 5.4. These identification percentages are within what has been found in comparable previous studies reporting L1 vowel identification [94.8% in Cebrian (2006) for Catalan vowels, with an average GR of 5.7/7; 95% (range 78%–99%) and 95.6% for Spanish vowels in Cebrian (2019) and Escudero and Chládková (2010), respectively]. The L1 vowels /e/ and /o/ obtained the lowest identification scores, 80% in both cases. The front vowel /e/ was misidentified as /i/ 13% and as /ɛ/ 7% of the time, while /o/ was misidentified as /ɔ/ 19% of the time. Other misidentification percentages were /ɛ/ as /e/ and /ɔ/ as /o/ (10%) and /u/ as /o/ (7%). These patterns of misidentification may reflect a possible lower degree of phonetic or perceptual stability of the mid vowel pairs in contrast with more peripheral vowels (e.g., Mora and Nadeu, 2012).

The crucial stimuli in the experiment were the non-native vowels. Recall that the order of presentation of the tasks, front vowels (set A) first or back vowels (set B) first, was counterbalanced among participants. Despite some variability across individuals, the modal responses were the same regardless of task order in all cases with one exception: for the subgroup who performed task B first, the English vowel /ɑ:/ (set B) was most frequently assimilated to Catalan /ɔ/ (69% of the time), followed by /a/ (16%), while for those who performed task B second, /ɑ:/ was more frequently assimilated to /a/ (47%), followed by assimilations to /ɔ/ (39%). An inspection of the individual data indicated that the perceptual assimilation of English /ɑ:/ was subject to both between-subject and within-subject variability: while some participants tended to chose one of the responses (/ɔ/, /a/, or /au/), roughly half of the participants (equally distributed between the two task orders) provided inconsistent responses. To assess whether task order (AB vs BA) could have influenced the results, Mann–Whitney *U*-tests were conducted on the percent identification/assimilation as the modal response and on the corresponding GRs. The results yielded no significant difference between the results of the participants who performed set A followed by set B and those who performed the tasks in the opposite order (% identification/assimilation: $U = 359$, $p = 0.145$; GR: $U = 310$, $p = 0.657$). The results were thus collapsed across task orders.

The variability in the responses for /ɑ:/ underscores the fact that individuals sharing the same L1 may differ in the categorization of non-native vowels (Faris *et al.*, 2018; Gilichinskaya and Strange, 2010; Strange *et al.*, 2009; Tyler *et al.*, 2014). Given this variability, for each non-native vowel, the number of participants who selected a given response at a consistent rate was tallied, following Strange *et al.* (2009) and Gilichinskaya and Strange (2010). This measure provides information about the consistency with which a given target vowel was assimilated to a L1 category. As discussed above, studies vary in the percent

TABLE I. Identification accuracy, GRs, and FIs for L1 Catalan vowels.

Catalan vowel	Correct identification (%)	Rating (out of 7)	FI (maximum 7)
/i/	100	6.3	6.3
/a/	99	5.9	5.9
/au/	99	5.9	5.9
/aj/	98	5.8	5.7
/ej/	93	5.5	5.1
/u/	93	6.1	5.7
/ou/	92	5.6	5.2
/ɛ/	90	6	5.3
/ɔ/	86	6	5.2
/o/	80	5.7	4.6
/e/	80	5.6	4.5
Means	92	5.9	5.4

assimilation considered as the categorization threshold. We are not concerned in this paper with establishing the most appropriate threshold, but rather with assessing the degree of consistency across listeners. For that purpose, the 70% (Tyler *et al.*, 2014) and 50% (Faris *et al.*, 2018; Bundgaard-Nielsen *et al.*, 2011) cutoffs are reported. The results are presented in Table II, following a decreasing order of assimilation scores. Only modal responses and assimilation scores above chance are provided. The full set of responses is provided in a confusion matrix in the Appendix (see Table V).

The English vowels /æ aʊ ʌ aɪ ɛ i:/ were the most consistently assimilated to a Catalan category, both in terms of assimilation rates (91–99%) and agreement across participants (22–27). The GRs varied, however, ranging from /aɪ/’s comparatively lower rating as /aj/ (3.6) to the highest rating received by English /æ/ and /ɛ/ (5.6) as Catalan /a/ and /ɛ/, respectively. Both English /æ/ and /ʌ/ were consistently mapped onto the same L1 vowel, /a/, but the former received higher goodness ratings than the latter (5.5 vs 4.6). The vowels /ɒ u: ɪ əʊ ɔ: eɪ/ reached assimilation percentages between 76% and 82% with GRs ranging from 3.7 to 5.6, while the low vowel /ɑ:/ revealed the least consistent pattern, being identified mostly with Catalan /ɔ/ (54%), followed by /a/ (31%). This vowel displayed the greatest amount of variability in responses among participants.

C. Discussion

Experiment 1 examined the perceived similarity between Catalan and SSBE vowels from the perspective of Catalan-speaking listeners. The results for the native Catalan stimuli showed that the Catalan listeners were able to identify L1 vowels accurately and consistently, within the ranges of what is generally found in L1 perception studies. Regarding the non-native vowels, most English vowels show a consistently identified counterpart in the L1, with

assimilation scores of 76% or higher, including six cases reaching 87% and higher. Thus, most SSBE vowels were categorized in terms of L1 categories, assuming a categorization threshold of 70% (Tyler *et al.*, 2014). The exception was English /ɑ:/, perceived as Catalan /ɔ/ and /a/ 54% and 31% of the time, respectively, qualifying as the most dissimilar from a Catalan vowel (probably in addition to English /ɜ:/ and /ʊ/; see Footnote 3).

The outcome of the current experiment in fact follows the general trends reported by Cebrian *et al.* (2011), which also looked at British English and Catalan vowels. The three SSBE vowels that received the highest assimilation scores and goodness ratings in Cebrian *et al.* (2011), namely, /i: ɛ æ/ as Catalan /i ɛ a/, respectively, also received the highest scores in the current study. In general, the pattern of modal responses was highly consistent across both studies. This was not unexpected, as the stimuli in both studies involved bVt words, although in the current study, neighbouring consonants were edited to allow listeners to focus on the vowels (see Sec. II A 2). However, the two studies obtained different results for /ɑ:/ and /ɪ/. The modal response for E /ɑ:/ in Cebrian *et al.* (2011) was C /a/ (78%; GR: 2.5), followed by C /ɔ/ (20%; GR: 3.9), as opposed to C /ɔ/ (54%, GR: 4.8) and C /a/ (31, GR: 3.9) in the current experiment. The discrepancy may be in part related to the amount of variability across participants observed for this vowel, as can be observed by the comparatively small number of participants who consistently assimilated E /ɑ:/ to either C /ɔ/ or /a/, as mentioned above. The discrepancy is even greater with respect to E /ɪ/. Cebrian *et al.* (2011) report that /ɪ/ was assimilated to C /i/ 82% of the time and to C /e/ 15% of the time, both with similar GRs (3.4–3.5). In the current study, the modal response was C /e/ (80%, vs 12% as /i/), with higher GRs (5.6 as /e/, 3.8 as /i/). In addition to the difference between the two studies concerning the neighbouring consonants, the studies also differed in the population [naive

TABLE II. Perceived similarity between non-native (English) vowels and L1 (Catalan) vowels. No. of Ss assim. \geq 70%/50% = number of subjects who selected a given response 70%/50% of the time or more. The total number of participants was 27. Modal responses appear in boldface. Assimilations below chance level (16.7%) have been excluded.

English target V	L1 response	Assimilation (%)	GR	FI	No. of Ss assim. \geq 70%	No. of Ss assim. \geq 50%
/æ/	/a/	99	5.6	5.5	27	27
/aʊ/	/au/	99	4.9	4.9	27	27
/ʌ/	/a/	98	4.7	4.6	27	27
/aɪ/	/aj/	95	3.6	3.4	26	26
/ɛ/	/ɛ/	91	5.6	5.1	25	26
/i:/	/i/	87	5.7	5.0	22	27
/ɒ/	/ɔ/	82	5	4.1	20	26
/u:/	/u/	82	3.7	3.0	22	25
/ɪ/	/e/	80	5.6	4.5	20	24
/əʊ/	/ou/	79	3.7	2.9	22	24
	/au/	19	2.6	0.5	2	3
/ɔ:/	/o/	77	4.8	3.7	19	25
/eɪ/	/ej/	76	4.2	3.2	12	27
	/aj/	21	4.3	0.9		
/ɑ:/	/ɔ/	54	4.8	2.6	8	16
	/a/	31	3.9	1.2	3	9

listeners in the current study vs L2 English speakers in Cebrian *et al.* (2011)]. Interestingly, Rallo Fabra (2005) found that a group of inexperienced Catalan listeners identified American English /ɪ/ most often as C /e/ (63%), followed by /ɛ/ (19%) and /i/ (16%), while an experienced group (undergraduate students in an English Studies degree) was found to provide less consistent results (40% as /i/, 12% as /e/, 12% as /ɛ/, and 36% as “non-Catalan”). The lack of consistency across studies regarding the assimilation of E /ɪ/ to C /e/ or /i/ may be related to the fact that Catalan /e/ is perceptually less stable (e.g., Mora and Nadeu, 2012), as illustrated by its lower L1 identification scores (see Table I), and to cross-study methodological differences in population, variety of English, phonetic context of the stimuli, and response options used. Still, the results tend to indicate that English /ɪ/ is perceptually closest to Catalan /e/, particularly for naive listeners, and that the degree of assimilation observed with /ɪ/ is weaker than with other non-native to native assimilations. The results of the current study are generally in agreement with previous studies and probably present the most reliable account of Catalan-SSBE perceptual similarity, given the inclusion of a large number of vowels, the nature of the stimuli, and the choice of naive listeners.

One of the goals of this study was to examine how similar native and non-native vowels may be perceived to be. To that effect, the results for L1 identification and non-native to L1 assimilation can be compared by comparing cross-linguistic assimilation patterns with L1 vowel identification values. Guion *et al.* (2000) propose that non-native phones that fall within 1 standard deviation of the mean fit index (FI) obtained for the L1 phones can be considered “good” exemplars of that L1 category, while those non-native phones whose FIs fall within 1–2 standard deviations of the native mean value can be regarded as “fair” matches. The average FI score for L1 Catalan vowels was 5.4 (ranging from 4.5 to 6.3), and the standard deviation was 0.6. Thus, English vowels with FIs of 4.8 (i.e., $5.4 - 0.6 = 4.8$) or higher can be considered good matches to the L1 vowels they are assimilated to. A few mappings fall within this range, namely, E /æ/ as C /a/ (FI=5.5), E /ɛ/ as C /e/ (5.1), E /i:/ as C /i/ (5), and E /aʊ/ as C /au/ (4.9). English vowels that patterned as a fair match to L1 Catalan vowels had a FI between 4.3 and 4.7 (i.e., $5.4 - 2 \times 0.6 = 4.2$). These were E /ʌ/ as C /a/ (4.6) and E /ɪ/ as C /e/. All remaining modal assimilations patterned as poor matches, including E /aɪ/ as C /aj/, with a high assimilation score (95%) but low GR (3.6) and consequently a low FI (3.4). The rest of the poorly assimilated English vowels had identification scores of 82% or lower and FI of 4.1 (E /ɒ/ as C /ɔ/), 3.7 (E /ɔ:/ as C /o/), 3.2 (E /eɪ/ as C /ej/), 3 (E /u:/ as C /u/), 2.9 (E /əʊ/ as C /ou/), and the lowest being 2.6 obtained by E /ɑ:/ as C /ɔ/.

In brief, the results of experiment 1 show that most English vowels are readily assimilated to L1 Catalan categories, yielding a number of good (C-E /aʊ/-/au/, /ɛ/-/ɛ/, /a/-/æ/, /i/-/i:/) and fair matches (C-E /e/-/ɪ/ and /a/-/ʌ/). The remaining were poor matches (C-E /aj/-/aɪ/, /u/-/u:/, /o/-/ɔ:/,

/ou/-/əʊ/, /ej/-/eɪ/, /ɔ/-/ɒ/, /ɔ/-/ɑ:/), despite assimilation percentages above 76%, reaching 95% in the case of /aj/-/aɪ/. Differences in articulation such as the fronted nature of SSBE /u:/, particularly in alveolar context (e.g., Harrington *et al.*, 2008; Levy, 2009a), or central beginning of the diphthong /əʊ/ may account for the low ratings. One of the goals of this paper is to assess whether the Catalan-English similarity judgements will be replicated when evaluating the perception of English native speakers. This is explored in experiment 2.

III. EXPERIMENT 2: PERCEIVED SIMILARITY OF CATALAN VOWELS BY ENGLISH SPEAKERS

A. Methodology

1. Participants

Twenty-seven native speakers of SSBE participated in the experiment (23 females). They were undergraduate (19) and graduate (8) students in the Division of Psychology and Language Sciences at University College London. They had all acquired English as a first language and had been raised in Southern England or had been born to parents from Southern England. Most spoke no other language fluently except for one participant who was fluent in German. One participant reported having some knowledge of Spanish but was not fluent. Their average age was 22, ranging from 18 to 30. All participants reported normal hearing and were paid for their participation.

2. Stimuli

The stimuli were the same as in experiment 1.

3. PAT

The design of the PAT followed the same structure as in experiment 1 and involved the same stimuli. There was one task for front vowels (set A) and one for back vowels (set B), and both tasks included non-native (Catalan) and L1 (English) stimuli, in the same fashion as was done for experiment 1 (see Sec. II A 3). The total number of trials per task (A and B) was 108 (66 Catalan and 42 English stimuli). The response alternatives in this case were *bat/hat*, *bite/light*, *bait/hate*, *bet/set*, *bit/hit*, *beat/heat*, *but/hut*, and *Bart/art* for set A and *bat/hat*, *about/gout*, *bot/hot*, *bought/sort*, *boat/goat*, *boot/hoot*, *but/hut*, and *Bart/art* for set B.

4. Procedure

The procedure for experiment 2 was the same as the one described for experiment 1 in Sec. II A 4.

B. Results

The percentage of correct identification of the L1 English vowels and the corresponding average GRs and FIs are given in Table III. The average identification accuracy was 92% (range 76%–100%), with average GRs and FIs of 6.1 and 5.7 out of 7, respectively. These results are very

TABLE III. Identification accuracy, GRs, and FIs for L1 English vowels.

English vowel	Correct identification (%)	Rating (out of 7)	FI (maximum 7)
/aɪ/	100	6.3	6.3
/uː/	100	6.6	6.6
/əʊ/	99	6.6	6.6
/æ/	98	6.4	6.3
/iː/	97	5.7	5.5
/aʊ/	96	6.2	6
/ɔː/	95	6.2	5.9
/eɪ/	93	6	5.6
/ɛ/	92	5.8	5.3
/ɑː/	89	6.3	5.6
/ɪ/	83	5.9	4.9
/ʌ/	80	6.2	4.9
/ɒ/	76	5.2	4
Means	92	6.1	5.7

similar to the L1 Catalan identification scores obtained in experiment 1 and are within the range of what has been reported in previous studies involving English vowels, e.g., 95% (range: 82%–99.6%) in Hillenbrand *et al.* (1995) and 89% (range: 75%–100%) in Rallo Fabra and Romero (2012) for American English vowels. Most English vowels obtained identification scores higher than 90% and GRs close to 6 or higher. The English vowels that obtained the lowest identification percentages were /ɪ, ʌ, ɒ/ (83, 80, and 76%, respectively), and the most common misidentifications were /ɪ/ as /ɛ/ (14%), /ʌ/ as /æ/ (13%) or /ɑː/ (7%), and /ɒ/ as /ɑː/ (11%), /ɔː/ (8%), or /ʌ/ (5%). It is possible that these vowels may have posed a greater challenge to native listeners as they are lax vowels, shorter in duration than the rest of the vowels, and typically followed by an audible consonant in the coda.

Table IV presents the modal responses for each of the Catalan target vowels in decreasing order of assimilation scores, as well as other assimilation patterns above chance

level, the average GR, FI, and the number of participants who selected the modal response 70% and 50% of the time or more (see Table VI in the Appendix for the complete range of responses). The modal responses were the same for the group who performed set A followed by set B and those who performed the tasks in the opposite order, and the results of Mann–Whitney *U*-tests conducted on the percent identification/assimilation as the modal response and on the corresponding GR confirmed there was no significant effect of task order (% identification/assimilation: $U = 340$, $p = 0.971$; GR: $U = 370$, $p = 0.552$).

The Catalan vowels and diphthongs /ou, ei, au, ɛ, ai/ were consistently assimilated to the English vowels /əʊ, eɪ, aʊ, ɛ, aɪ/, respectively, with assimilation scores equal to or higher than 90% (90%–97%), GR between 4.9 and 5.4 out of 7, and a high degree of agreement among English listeners (24–27 participants selected the modal response at least 70% of the time). Catalan /a/ reached 86% identification as English /æ/ with a GR of 5.3. Catalan /e/ and /u/ were mapped onto English /ɪ/ and /uː/ (73% and 70%, GR 5 and 4.1, respectively). Three Catalan vowels did not reach a 70% threshold for categorization: Catalan /i/, split between English /iː/ (62%, GR: 4.5) and /ɪ/ (37%, GR: 4.3), and Catalan /o/ and /ɔ/, which obtained the lowest assimilation rates, both being predominantly perceived as English /ɒ/ (59% and 53%, GR: 4.4 and 5, respectively) and then as /ɔː/ (18% and 33%) and /ʌ/ (12% and 9%).

C. Discussion

The perceptual similarity between Catalan and SSBE vowels from the perspective of native English speakers was examined in experiment 2 by means of a PAT using the same stimuli as in experiment 1, adapted to native English listeners. Regarding L1 perception, English listeners were overall successful identifying the native vowels, with a mean accuracy of 92%, within the average identification scores reported in the literature for native English vowels.

TABLE IV. Perceived similarity between non-native (Catalan) vowels and L1 (English) vowels. No. of Ss assim. $\geq 70\%/50\%$ = number of subjects who selected a given response 70%/50% of the time or more. The total number of participants was 27. Modal responses appear in boldface. Assimilations below chance level (12.5%) have been excluded.

Catalan V	L1 response	Assimilation (%)	GR	FI	No. of Ss 70% ($N = 27$)	No. of Ss 50% ($N = 27$)
/ou/	/əʊ/	97	5	4.8	27	27
/au/	/aʊ/	94	5.4	5.1	26	26
/ei/	/eɪ/	94	5	4.7	25	26
/ɛ/	/ɛ/	92	4.9	4.5	25	27
/ai/	/aɪ/	91	5.2	4.7	24	25
/a/	/æ/	86	5.3	4.5	19	26
/e/	/ɪ/	73	5	3.7	16	22
/u/	/uː/	70	4.1	2.9	16	21
/i/	/iː/	62	4.5	2.8	14	18
	/ɪ/	37	4.3	1.6	6	9
/o/	/ɒ/	59	4.4	2.6	11	16
	/ɔː/	18	3.8	0.7	1	5
/ɔ/	/ɒ/	53	5	2.6	7	14
	/ɔː/	33	4.3	1.4	1	9

The lower scores obtained by some lax vowels are possibly due to the added difficulty of identifying lax vowels in isolation.

As was observed with the Catalan listeners, a relatively consistent pattern of modal responses was found in the English listeners' perception of Catalan vowels. Six of the 11 vowels and diphthongs tested reached assimilation scores between 86% and 99%, two more vowels were just around the 70% categorization threshold, and three vowels were under 70% (53%–62%). As was done in experiment 1, to explore the extent to which the non-native vowels patterned as possible instances of L1 categories, the L1 identification scores and non-native vowel assimilation scores were compared in terms of fit index scores, following Guion *et al.* (2000). The average FI for L1 English vowels was 5.7 (ranging from 4 to 6.6), and the standard deviation was 0.75. Therefore, in this case, non-native vowels with FIs of 4.9 (i.e., $5.7 - 0.8 = 4.9$) or higher were classified as good instances of the L1 vowels. Only one Catalan phone qualified as such, namely, C /au/ as E /aʊ/ (5.1). Five Catalan vowels patterned as fair matches, i.e., within 2 standard deviations of the L1 identification mean, that is, between 4.1 and 4.9 ($5.7 - 2 \times 0.8 = 4.1$). These were C /ou/ as E /əʊ/ (4.8), C /ej/ as E /eɪ/ (4.7), C /aj/ as E /aɪ/ (4.7), C /ɛ/ as E /ɛ/ (4.5), and C /a/ as E /æ/ (4.5). The rest of the vowels obtained FI lower than 4.1 and, in fact, were also the vowels with the lowest assimilation scores (73%–53%): C /e/ as E /ɪ/ (3.7), C /u/ as E /u:/ (2.9), C /i/ as E /i:/ (2.8), and both C /o/ and C /ɔ/ as E /ɒ/ (2.6).

The results of the second experiment indicate that a number of Catalan vowels were readily assimilated to native English categories, particularly the C-E pair /au/-/aʊ/, qualifying as a good match, and the C-E pairs /ou/-/əʊ/, /ej/-/eɪ/, /aj/-/aɪ/, /ɛ/-/ɛ/, and /a/-/æ/ that patterned as fair matches. The assimilation scores for Catalan /e/ and /u/ as English /ɪ/ and /u:/, respectively, barely reached the categorization threshold of 70% and qualified as poor matches. Articulatory differences may account for /u/'s lower scores, as discussed above. In addition, English listeners identified Catalan /u/ as English /ʌ/ 12% of the time. It is possible that in these cases, the desired response was English /ʊ/, but the absence of a response alternative for this English vowel and the fact that letter ⟨u⟩ is a common spelling for both /ʌ/ and /ʊ/ (e.g., *putt* and *put*) may have resulted in some /ʌ/ responses for Catalan /u/. Finally, Catalan /i, o, ɔ/ obtained scores below 70%. Catalan /i/ was split between English /i:/ and /ɪ/ (62% and 37%, respectively). Although Catalan /i/ is acoustically closer to English /i:/ than to English /ɪ/ (Cebrian, 2006; see also Fig. 2), the Catalan /i/ stimuli were closer in duration to the latter than to the former (C /i/: 139 ms, E /ɪ/: 136 ms, E /i:/: 168 ms). It is possible then that the shorter duration of Catalan /i/ may have triggered some /ɪ/-responses. This is in line with Escudero and Boersma's (2004) finding that native Southern British English speakers may attend to temporal and spectral cues in their perception of the /i:/-/ɪ/ contrast. In addition to /i/, the two Catalan mid back vowels (/ɔ/ and /o/) obtained

assimilation scores under 70%, and in both cases, the modal response was English /ɒ/. Despite the presence of available native categories in the low and mid back vowel space (/ɑ:/, ɒ, ɔ:/), the two Catalan back rounded vowels were perceived as instances of the same L1 category /ɒ/. Vowel duration may have played a role in this case too. SSBE /ɔ:/ is considerably longer than /ɒ/, while the two Catalan vowels are similar in duration and closer to the English short vowel. The duration of the stimuli reflects this pattern: C /ɔ/: 174 ms, C /o/: 148 ms, E /ɒ/: 156 ms, E /ɔ:/: 245 ms. Hillenbrand *et al.* (2000) report that while native (American) English speakers predominantly rely on spectral differences, they may attend to duration as a cue to vowel identity, particularly in the case of mid and low vowels. The comparatively short duration of C /o/ may have made it an unlikely match for E /ɔ:/, prompting a preference for a lower, shorter, vowel, i.e., E /ɒ/. In any case, the Catalan /o/-/ɔ/ contrast emerges as the most difficult for English speakers to learn, in terms of their assimilation to the same native category and the relatively low assimilation scores and FIs. The comparison of the results of the two experiments is discussed in Sec. IV.

IV. GENERAL DISCUSSION AND CONCLUSIONS

This study has investigated the perceived similarity between Catalan and English vowels and diphthongs by eliciting similarity judgements from 27 naive native speakers of each language. The participants performed a PAT in which native and non-native exemplars were identified in terms of native categories and rated for goodness of fit to that category. The results for the native vowels were highly comparable across experiments, showing high levels of accuracy (92%) and high goodness ratings. Catalan mid vowels and English lax vowels obtained somewhat lower identification rates, which can be linked to their more unstable perceptual nature and difficulty of perceiving vowels in isolation. Still, the overall rates of identification were high and comparable to previous L1 perception studies.

The L1 identification results served as the basis for interpreting the assimilation scores obtained by the non-native vowels. The results of both experiments showed that most non-native vowels were predominantly assimilated to a single L1 category, receiving assimilation scores above a 70% categorization threshold (Tyler *et al.*, 2014). The perceptual assimilation patterns did not always follow from the acoustic proximity observable on an F1 × F2 plot (see Fig. 2), particularly in the case of the back vowels. This is in line with previous works showing discrepancies between acoustic and perceptual measures (see Strange, 2007). An examination of such discrepancies lies beyond the scope of this paper, whose focus is on perceptual data. Among the single-category assimilations, a few patterned as good and fair matches (Guion *et al.*, 2000). Both Catalan and English listeners predominantly linked Catalan and English diphthongs, indicating that diphthongs (often excluded from vowel studies) play a relevant role in cross-linguistic perception, in line with previous findings [e.g., Cebrian (2019)

and Escudero and Williams (2011) for Spanish and English diphthongs]. English /ɑ:/ (as well as /u/ and /ɜ:/ according to previous studies; see Footnote 3) and Catalan /i, o, ɔ/ were the only cases of uncategorized vowels. The extent to which English and Catalan listeners agreed in the judgements of cross-linguistic similarity is discussed next.

A. Reciprocity in perceived similarity

One of the goals of this study was to examine to what extent there is reciprocity in perceptual judgements of similarity between native and non-native vowels as perceived by native speakers of the two languages involved. The intention is to obtain a more complete picture of the similarity relationships between Catalan and English vowels and to evaluate whether some phones can be considered near-identical or shared by both languages. In terms of modal responses, there was a very high level of agreement between the two populations. Considering the number of Catalan stimuli was smaller (11 vowels and diphthongs tested, as opposed to 13 in the case of the English stimuli), there were 11 potentially reciprocal matches, that is, cases where the Catalan and English listeners could have provided the same native-non-native mappings. Figure 3 shows the pairs that obtained reciprocity. Of the 11 possible matches, 8 received reciprocal modal responses with assimilation scores equal to or higher than 70% (C-E /au/-/aʊ/, /ɛ/-/ɛ/, /a/-/æ/, /ou/-/əʊ/, /eɪ/-/eɪ/, /ai/-/aɪ/, /e/-/ɪ/, and /u/-/u:/).

The very consistent results for C-E /au/-/aʊ/, /ɛ/-/ɛ/, and /a/-/æ/ across experiments (good and fair matches) point to cases of near-identity and suggest that for both populations, the non-native vowels would be readily perceived and produced in terms of native categories. Contrasting the perceptual judgements of two parallel populations can shed light on a basic question in phonetics, namely, when two sounds from different languages can be considered the same (Ladefoged, 1990). We can speculate that these vowels may not be distinguishable from native categories and that, when used in L2 speech, they may go unnoticed by target language speakers. In addition, C-E mappings /ou/-/əʊ/, /eɪ/-/eɪ/, and /ai/-/aɪ/ were considered fair matches by English listeners but poor matches

by Catalan listeners, while the reverse was true for C-E /e/-/ɪ/. Speakers exposed to non-native or L2 sounds perceive them, at least initially, as variations on L1 sounds (Schmidt, 2007). Thus, differences between Catalan and English speakers in the GRs provided for the same C-E pairs may be related to differences in the amount of within-category variation and exposure to that variation in L1 speech. For instance, SSBE speakers are exposed to /əʊ/ productions with a less central beginning (thus closer to C /ou/) as a result of both allophonic variation (e.g., preceding consonants like /l/) and dialectal variation. By contrast, Catalan listeners do not encounter centralized versions of stressed /ou/ in their native input. Consequently, C /ou/ emerged as a better match for E /əʊ/ than the reverse.

Two more pairs received parallel mappings in terms of the modal responses, but with different categorization levels. English /i:/ was categorized as C /i/ (87% assimilation scores), but the reverse mapping was below 70% (62%). The difference was even greater with E /ɒ/ and C /ɔ/, which showed a consistent assimilation pattern with Catalan listeners (82%) but only reached a 53% assimilation score with English listeners. Finally, C /o/ and E /ɔ:/ did not receive reciprocal mappings. English listeners linked C /o/ with E /ɒ/ (59% assimilation scores), followed by E /ɔ:/ (18%), while Catalan listeners perceived E /ɔ:/ predominantly as C /o/ (77%). The discrepancy in the results for Catalan /i, o, ɔ/ may be linked to cross-linguistic differences in vowel duration, as discussed above, reflecting the way the specific characteristics of the listener's L1 affect cross-linguistic perception. Unlike English vowels, Catalan vowels tend to vary little in duration, and their average duration is notably closer to the average duration of English lax vowels than of English tense vowels. Thus, the English listeners selected the lax vowel /ɪ/ as the preferred match for the relatively short high front Catalan vowel more than a third of the time. This resulted in a rather asymmetric pattern, with C-E /i/-/i:/ being judged as a good match by Catalan listeners and a poor match by English listeners. Similarly, despite the quality difference between C /ɔ/ and /o/ and the availability of different L1 back vowels, the Catalan vowels' relative short duration rendered them unsuitable as instances of the comparatively long SSBE vowel /ɔ:/ and triggered a two to one assimilation of C /ɔ/ and /o/ to E /ɒ/. The only previous study that looked at reciprocal similarity, Schmidt (1996, 2007) on Japanese and English consonants, found a number of parallel or reciprocal mappings but also reported that cross-linguistic similarity varied for the two language groups depending on context-specific cues (e.g., the rounding of adjacent vowels, differences in VOT, and the presence of burst-like transitions in nasals). Schmidt (2007) suggests that adult L2 listeners attend to the phonetic cues that are relevant in the organization of sound contrasts in their L1. Although native English speakers rely mainly on spectral cues to identify L1 vowels, temporal differences may also play a role (Escudero and Boersma, 2004; Hillenbrand *et al.*, 2000). In the current case, the lack of duration distinctions among the Catalan stimuli may have

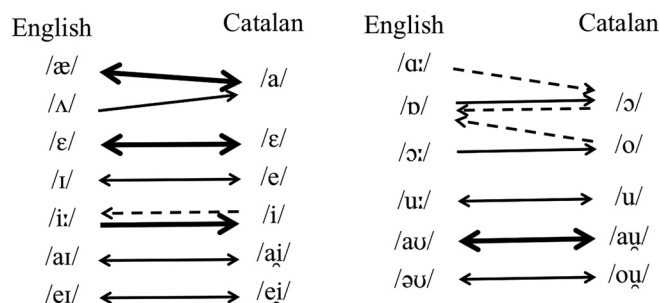


FIG. 3. Perceptual assimilation results from experiments 1 and 2. Double-headed arrows indicate reciprocal modal responses. Single-headed arrows show unidirectional modal responses (or differences in amount of assimilation). Darker lines indicate good/fair matches, all solid lines indicate assimilation scores $\geq 70\%$, dashed lines indicate assimilation scores between 50% and 70%.

rendered some non-native spectral differences like C /o/-/ɔ/ less interpretable in terms of available L1 categories. On the other hand, duration is not relevant for Catalan listeners who predominantly mapped each non-native back vowel with a different L1 vowel. The extent to which the two populations provide comparable similarity judgements is thus modulated by the role of language-specific acoustic cues. Interestingly, having a larger inventory in the case of English does not result in more single-category assimilations. The relative weighting of relevant L1 cues, such as duration, seems to play a greater role than the possibility of finding a different anchor for each non-native vowel.

Of the two remaining English vowels, /ʌ/ was consistently perceived as C /a/, while English listeners selected /ʌ/ as a match for non-native Catalan vowels on few occasions (12% for /o/ and /u/, 9% for /a/ and /ɔ/). Similarly, Catalan listeners perceived E /a:/ as /ɔ/ and /a/ (54% and 31% assimilation scores, respectively), but English listeners selected /a:/ as a response for C /ɔ/ and /a/ only 4% and 5% of the time. Given that English has a larger number of vowels, it is logical that some English vowels were not selected as modal responses. This may indicate that these L1 vowels were considered by English listeners to be the most different from Catalan vowels. In that sense, the results for E /a:/ are consistent across language groups, as it is the vowel that received the lowest assimilation scores by Catalan listeners and the L1 vowel that was used the least as a response by the English listeners. Regarding E /ʌ/ and /æ/, the former was hardly selected as a good match for C /a/ by English listeners (9%, cf. 86% for /æ/), while Catalan listeners linked E /ʌ/ predominantly with C /a/ (98%) but with lower GRs and FI (FI of 5.5 and 4.6 for /æ/ and /ʌ/, respectively). Hence, the results for both groups of listeners point to E /æ/ as being a better match for the C /a/ than E /ʌ/ and illustrate how the judgements from two parallel populations support the same findings.

Finally, although not the focus of the current paper, a few predictions are briefly presented in relation to the models discussed in the Introduction. In terms of the SLM/SLM-r (Flege, 1995; Flege and Bohn, 2021), differences between the L1 and the non-native vowels will be more likely to be detected in the case of poor matches, given enough exposure to authentic target language input, than in the case of fair matches. Those vowels that qualify as good matches may be perceived as nearly identical to the native categories and be readily accepted as instances of L1 vowels, and the resulting single L1-L2 category may yield accurate perception and production, particularly in cases of reciprocal assimilation (C-E /aʊ/-/aʊ/, /ɛ/-/ɛ/, /a/-/æ/). With respect to PAM's assimilation types and the corresponding predictions, E /a:/ and C /i, o, ɔ/ were assimilated above chance to more than one L1 phone and thus pattern as uncategorized-clustered (Faris *et al.*, 2016). For Catalan speakers, E /æ/-/ʌ/ constitute a category-goodness (CG) assimilation,⁷ while E /a:/-/ɒ/, /a:/-/æ/, and /a:/-/ʌ/ illustrate uncategorized-categorized assimilation types with partial assimilation overlap (UC-P; Faris *et al.*, 2016) as C /ɔ/, /a/, and /a/, respectively. These four pairs are predicted to pose the greatest discrimination difficulty for Catalan speakers.

Regarding English speakers, C /o/-/ɔ/ constitutes an uncategorized-uncategorized pair with complete assimilation overlap as E /ɒ/ and /ɔ:/ (UU-C), while C /i/ and /e/ pattern as a UC-P case, with partial overlap (as E /i/). According to PAM, the UU-C contrasts will be the most difficult to discriminate for non-native speakers, followed by the UC-P and CG cases, which will vary in accuracy depending on the degree of assimilation overlap. The remaining vowel combinations constitute two-category assimilations (e.g., C /a/-/ɛ/, E /æ/-/ɛ/) or non-overlapping categorized-uncategorized pairs (e.g., C /a/-/ɔ/), which are expected to be more accurately discriminated [see Faris *et al.* (2018) and Tyler (2021) for more discussion]. Testing these predictions is beyond the scope of the current paper and is left for further research.

B. Limitations, final conclusions, and issues for further study

This study has some limitations. First, the study falls short of a whole-system approach, where listeners identify all non-native monophthongs and diphthongs in terms of the entire vowel inventory of their L1 (Bundgaard-Nielsen *et al.*, 2011; Faris *et al.*, 2016). The current study excluded the English vowels /ɜ:/ and /ʊ/ and did not include all the possible Catalan diphthongs (combinations of the monophthongs and the semi-vowels /j/ and /ɥ/). The aim of this paper was to investigate the reciprocity of cross-linguistic perceptual assimilation, and thus the focus was on vowels that tended to be consistently categorized in previous studies involving Catalan listeners. Thus, English vowels /ɜ:/ and /ʊ/ were excluded on the grounds of previous results showing low degrees of assimilation for these vowels (see Footnote 3). Further, in an attempt to create a more practical less fatigue-prone perceptual task, front and back vowels were tested in two separate blocks (tasks A and B), with a subset of the response options each. This division may have prevented the detection of some similarity relationships. For instance, the English central vowel /ʌ/, was included with the front vowels (task A), and thus listeners could not choose one of the Catalan back vowels as responses (task B). Still, the division into subtasks was guided by the results of previous studies, and stimuli and responses were distributed in such a way that all likely responses were available for each task. For instance, in a previous study in which all possible response options were presented in a single task (Cebrian *et al.*, 2011), E /ʌ/ was perceived as C /a/ much more often than as a back vowel like C /ɔ/ (85% and 11%, respectively). Finally, it could be argued that the methodology could be improved with the addition of a “non-L1” response option so that listeners were not forced to choose one of the L1 categories for uncategorizable non-native stimuli. However, the consistency of the assimilation patterns and the use of goodness of fit data provide adequate evidence as to what non-native phones cannot be categorized in terms of L1 categories. A “non-L1” response option may be more informative for studies evaluating whether cross-linguistic perceived similarity varies as a function of L2 experience (Flege, 1991; Rallo Fabra and Romero, 2012).

Another limitation concerns the nature of the stimuli, which consisted of vowels produced in a single phonetic context, between /b/ and /t/. It is known that perceived similarity may vary depending on the phonetic and the prosodic context (Strange *et al.*, 2002). For instance, Bohn and Steinlen (2003) found that Danish speakers with English language experience classified English /i/ as instances of Danish /e/ in glottal and alveolar contexts but as /i/ in velar context. Levy (2009a) reported that assimilation patterns of French vowels to English vowels were more consistent in a bilabial context than in an alveolar context. Clearly, a more complete description of cross-linguistic similarity relations needs to include a wider variety of contexts. In addition, the current study has focused on two specific varieties of each language (Eastern Catalan and SSBE). It remains to be seen how cross-linguistic assimilation patterns would differ for speakers of other varieties, e.g., for speakers of Majorcan Catalan, a variety that has a mid central vowel (/ə/) as part of the stressed vowel inventory.

To sum up, this study has examined the perceived similarity between a near-complete set of SSBE and Catalan vowels and diphthongs and has contrasted the perceptual judgements of Catalan and English listeners. The cross-language assimilation scores have been compared to L1 identification results to evaluate the degree to which non-native vowels assimilate to L1 vowels. The results show that most non-native vowels received very consistent degrees of assimilation in terms of L1 vowels, with varying degrees of goodness of fit. The inclusion of L1 stimuli in the perceptual task has made it possible to assess that some high assimilation rates were comparable to the identification scores received by native vowels. In addition, by testing speakers of the two languages being compared, it has been found that the majority of the vowels were involved in reciprocal mappings, including some of the most strongly assimilated pairs (e.g., C-E /aʊ/-/aʊ/, /ɛ/-/ɛ/, /a/-/æ/). The interpretation of cross-linguistic assimilation patterns in relation to L1 identification data and the comparison of perceptual judgements from speakers of the two languages involved are thus proposed as a suitable novel approach to assess phonetic similarity and to answer questions about cross-linguistic identity (Ladefoged, 1990).

Further research evaluating L2 learning is necessary to investigate the extent to which these C-E pairs behave as near-identical or shared categories and are unnoticed when used in the non-native context, as well as examining cases of mismatched degree of assimilation and lower degrees of reciprocity. An interesting issue in relation to cases of asymmetric reciprocity (such as C-E /i/-/i:/) is what constitutes a better predictor of L2 learning difficulty, the perceived similarity judgements of the L1 speakers or of the target language speakers. A few cases of lack of reciprocity were observed, which underscore the role of language-specific cues such as duration, as illustrated by the two to one assimilation of the Catalan mid back vowels /o, ɔ/ to the English vowel /ɒ/. This study has not examined the relationship between perceptual and acoustic similarity or investigated other perceptual measures such as a paired comparison technique, i.e., dissimilarity ratings for pairs of native and non-native stimuli (Flege *et al.*, 1994). These issues, together with the examination of a wider range of phonetic and prosodic contexts and other varieties of Catalan and English, are left for further study. Still, we can conclude that reciprocal studies that include both L1 identification and non-native assimilation data can achieve a more complete assessment of cross-linguistic similarity and can help develop hypotheses about non-native perception and L2 learning, in addition to bringing us closer to a better understanding of the concept of phonetic similarity.

ACKNOWLEDGMENTS

The author is grateful to Joan Carles Mora for his help with the elicitation of the Catalan production data for the Catalan vowel stimuli. This work was supported by Research Grant No. FFI2017-88016-P from the Spanish Ministry of Economy and Competitiveness and by Research Group Grant No. 2017SGR34 from the Catalan Government.

APPENDIX

Confusion matrices displaying the percentage of responses for each stimulus vowel in experiment 1 and experiment 2 are shown in Tables V and VI.

TABLE V. Experiment 1: Percentage assimilation of SSBE vowels to Catalan vowels (Cat. resp.). GRs are given in parentheses. Modal responses appear in boldface. Assimilations below 3% have been excluded. Note: See Footnote 3 for results for English /ɜ:/ and /ʊ/ reported in previous studies.

	English vowel stimuli												
Cat. resp.	/i:/	/ɪ/	/eɪ/	/ɛ/	/æ/	/ʌ/	/aɪ/	/aʊ/	/əʊ/	/ɑ:/	/ɒ/	/ɔ:/	/u:/
/i/	87 (5.7)	12 (3.8)					3 (1.9)						
/e/		80 (5.6)		8 (5.1)									
/eɪ/	12 (4.2)		76 (4.2)										
/ɛ/		7 (4.8)		91 (5.6)									
/a/					99 (5.6)	98 (4.7)				31 (3.9)	3 (3.7)		
/aɪ/			21 (4.3)				95 (3.6)						
/aʊ/								99 (4.9)	19 (2.6)	11 (2.3)	3 (1.9)		8 (1.2)
/oʊ/									79 (3.7)			8 (3.4)	10 (2.4)
/ɔ/										54 (4.8)	82 (5)	7 (3.5)	
/o/										4 (4.5)	11 (4.6)	77 (4.8)	
/u/												6 (3.9)	82 (3.7)

TABLE VI. Experiment 2: Percentage assimilation of Catalan vowels to SSBE vowels (Eng. resp.). Goodness ratings are given in parentheses. Modal responses appear in boldface. Assimilations below 3% have been excluded.

	Catalan vowel stimuli										
Eng. resp.	/i/	/e/	/eᵢ/	/ɛ/	/a/	/aᵢ/	/aᵤ/	/oᵤ/	/ɔ/	/o/	/u/
/i:/	62 (4.5)		6 (4.3)								
/ɪ/	37 (4.3)	73 (5)									
/eɪ/			94 (5)			8 (4.5)					
/ɛ/		24 (3.7)		92 (4.9)							
/æ/				6 (2.9)	86 (5.3)						
/ʌ/					9 (4.5)				9 (4.2)	12 (4)	12 (3.4)
/aɪ/						91 (5.2)					
/aʊ/							94 (5.4)				
/ɑ:/					5 (3.6)						
/əʊ/								97 (5)		11 (3.1)	12 (3.1)
/ɒ/									53 (5)	59 (4.4)	
/ɔ:/							5 (3.2)		33 (4.3)	18 (3.8)	
/u:/											70 (4.1)

¹See also Daidone *et al.* (2015) for another technique referred to as the free classification task in which participants group stimuli in terms of perceptual closeness by moving visual representations of the stimuli on a computer screen.

²Note that in Catalan, the symbol /ɔ/ is used to represent the lowest of the mid back rounded vowels (/o, ɔ/), while in SSBE, /ɔ:/ represents a vowel that is higher than the low back rounded vowel /ɒ/.

³For example, Cebrian *et al.* (2011) found that English /ɜ:/ was assimilated to Catalan /e/ 30% of the time, followed by /ɛ/ (24%), /a/ (23%), /o/ (9%), and /ɔ/ (9%). Rallo Fabra (2005) reported that English /ʊ/ was identified as Catalan /u/ (31%), /o/ (16%), /ɛ/ (12%), and /ɔ/ (7%) and was perceived as “not Catalan” 30% of the time.

⁴An anonymous reviewer pointed out possible dialectal variation regarding the pronunciation of the word *mot* (with /o/ or /ɔ/). While this has been attested for Valencian Catalan, most sources indicate /o/ as the only or predominant pronunciation in Central or Eastern Catalan (e.g., Alcover and Moll, 1993). In addition, the participants in the current study were asked at the beginning of the PAT to consider the response options and indicate if any differed from their own pronunciation. No disagreement was mentioned.

⁵Some studies report median values rather than mean values for rating data (e.g., Baigorri *et al.*, 2019; Strange *et al.*, 2009). Median values were calculated and did not differ much from mean data. Mean values thus are reported, following other studies (e.g., Faris *et al.*, 2018; Guion *et al.*, 2000; Lengeris, 2009).

⁶A composite score like FI may not always appear appropriate (Faris *et al.*, 2018; Tyler, 2021). For example, a fit index score of 2.4 may be the result of different calculations, e.g., $80\% \times 3$, $60\% \times 4$, or $40\% \times 6$, making it difficult to distinguish among different assimilation patterns. However, this is not a problem in the current data, given that for all cases of multiple responses for non-native sounds, higher percentages of assimilation correspond to higher goodness of fit ratings.

⁷Following previous proposals (Tyler 2021, Tyler *et al.*, 2014), English /æ/-/ʌ/ was classified as a CG assimilation and not a single-category assimilation because the difference between the two vowels in the goodness of fit to Catalan /a/ reached significance ($t(26) = 3.902$, $p = 0.001$).

Abramson, A. S. (1988). “The principles on which the IPA should be based,” *J. Int. Phon. Assoc.* **18**(2), 66–68.

Adank, P., Smits, R., and van Hout, R. (2004). “A comparison of vowel normalization procedures for language variation research,” *J. Acoust. Soc. Am.* **116**, 3099–3107.

Alcover, A. M., and Moll, F. d B. (1993). *Diccionari Català-Valencià-Balear (Catalan-Valencia-Balearic Dictionary)* (Moll, Palma de Mallorca, Spain).

Baigorri, M., Campanelli, L., and Levy, E. S. (2019). “Perception of American-English vowels by early and late Spanish-English bilinguals,” *Lang. Speech* **62**(4), 681–700.

Best, C. (1995). “A direct realist view of cross-language speech perception,” in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 171–204.

Best, C. T., and Tyler, M. D. (2007). “Nonnative and second-language speech perception: Commonalities and complementarities,” in *Language Experience in Second Language Speech Learning: In Honor of James Flege*, edited by M. Munro, and O.-S. Bohn (John Benjamins, Amsterdam), pp. 13–34.

Boersma, P., and Weenink, D. (2018). “Praat: Doing phonetics by computer (version 6.0.43) [Computer program],” <http://www.praat.org/> (Last viewed September 8, 2018).

Bohn, O.-S. (2002). “On phonetic similarity,” in *An Integrated View of Language Development: Papers in Honor of Henning Wode*, edited by P. Burmeister, T. Piske, and A. Rohde (Wissenschaftlicher Verlag, Trier), pp. 191–216.

Bohn, O.-S., and Steinlen, A. K. (2003). “Consonantal context affects cross-language perception of vowels,” in *Proceedings of the XV International Congress of Phonetic Sciences*, edited by D. Recasens, M. J. Solé, and J. Romero (Causal Productions, Barcelona), pp. 2289–2292.

Bundgaard-Nielsen, R., Best, C. T., and Tyler, M. D. (2011). “Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories,” *J. Appl. Psycholinguist.* **32**, 51–67.

Carbonell, J. F., and Llisterrri, J. (1992). “Catalan,” *J. Int. Phon. Assoc.* **22**(1), 53–56.

Cebrian, J. (2006). “Experience and the use of non-native duration in L2 vowel categorization,” *J. Phon.* **34**(3), 372–387.

Cebrian, J. (2009). “Effects of native language and amount of experience on crosslinguistic perception,” *J. Acoust. Soc. Am.* **125**(4), 2775.

Cebrian, J. (2019). “Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs,” *J. Acoust. Soc. Am.* **145**(1), EL52–EL58.

Cebrian, J., Mora, J. C., and Aliaga García, C. (2011). “Assessing crosslinguistic 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 M. Wrembel, M. Kul, and K. Dziubalska-Kolaczyk (Peter Lang, Frankfurt am Main, Germany), Vol. I, pp. 41–52.

Daidone, D., Kruger, F., and Lidster, R. (2015). “Perceptual assimilation and free classification of German vowels by American English listeners,” in *Proceedings of the 18th International Congress of Phonetic Sciences*, edited by The Scottish Consortium for ICPhS 2015 (University of Glasgow, Glasgow, UK).

Deterding, D. (1997). “The formants of monophthong vowels in Standard Southern British English pronunciation,” *J. Int. Phon. Assoc.* **27**(1), 47–55.

- Escudero, P., and Boersma, P. (2004). "Bridging the gap between L2 speech perception research and phonological theory," *Stud. Second Lang. Acquis.* **26**(4), 551–585.
- Escudero, P., and Chládková, K. (2010). "Spanish listeners' perception of American and Southern British English vowels," *J. Acoust. Soc. Am.* **128**(5), EL254–EL260.
- Escudero, P., and Williams, D. (2011). "Perceptual assimilation of Dutch vowels by Peruvian Spanish listeners," *J. Acoust. Soc. Am.* **129**(1), EL1–EL7.
- Faris, M., Best, C., and Tyler, M. (2016). "An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized," *J. Acoust. Soc. Am.* **139**(1), EL1–EL5.
- Faris, M., Best, C., and Tyler, M. (2018). "Discrimination of uncategorized non-native vowel contrasts is modulated by perceived overlap with native phonological categories," *J. Phon.* **70**, 1–19.
- Ferragne, E., and Pellegrino, F. (2010). "Formant frequencies of vowels in 13 accents of the British Isles," *J. Int. Phon. Assoc.* **40**(1), 1–34.
- Fllege, J. E. (1987). "The production of new and similar phones in a foreign language: Evidence for the effect of equivalence classification," *J. Phon.* **15**, 47–65.
- Fllege, J. E. (1991). "The interlingual identification of Spanish and English vowels: Orthographic evidence," *Q. J. Exp. Psychol.* **43**(3), 701–731.
- Fllege, J. E. (1992). "The intelligibility of English vowels spoken by British and Dutch talkers," in *Intelligibility in Speech Disorders, Theory, Measurement, and Management*, edited by R. D. Kent (John Benjamins, Amsterdam), pp. 157–232.
- Fllege, J. E. (1995). "Second language speech learning: Theory, findings and problems," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 233–277.
- Fllege, J. E., and Bohn, O.-S. (2021). "The revised speech learning model (SLM-r)," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by R. Wayland (Cambridge University, Cambridge, UK), pp. 3–83.
- Fllege, J. E., Munro, M. J., and Fox, R. A. (1994). "Auditory and categorical effects on cross language vowel perception," *J. Acoust. Soc. Am.* **95**, 3623–3641.
- Gilichinskaya, Y. D., and Strange, W. (2010). "Perceptual assimilation of American English vowels by inexperienced Russian listeners," *J. Acoust. Soc. Am.* **128**(2), EL80–EL85.
- Guion, S. G., Fllege, J. E., Akahane-Yamada, R., and Pruitt, J. C. (2000). "An investigation of current models of second language speech perception: The case of Japanese adults' perception of English consonants," *J. Acoust. Soc. Am.* **107**(5), 2711–2724.
- Harnsberger, J. D. (2001). "On the relationship between identification and discrimination of non-native nasal consonants," *J. Acoust. Soc. Am.* **110**, 489–503.
- Harrington, J., Kleber, F., and Reubold, U. (2008). "Compensation for coarticulation, /u/-fronting, and sound change in standard southern British: An acoustic and perceptual study," *J. Acoust. Soc. Am.* **123**, 2825–2835.
- Hillenbrand, J. M., Clark, M. J., and Houde, R. A. (2000). "Some effect of duration on vowel recognition," *J. Acoust. Soc. Am.* **108**, 3013–3022.
- Hillenbrand, J. M., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *J. Acoust. Soc. Am.* **97**, 3099–3111.
- Kuhl, P. K., and Iverson, P. (1995). "Linguistic experience and the perceptual magnet effect," in *Speech Perception and Linguistic Experience: Theoretical and Methodological Issues*, edited by W. Strange (York, Timonium, MD), pp. 121–154.
- Ladefoged, P. (1990). "Some reflections on the IPA," *J. Phon.* **18**, 335–346.
- Lengeris, A. (2009). "Perceptual assimilation and L2 learning: Evidence from the perception of Southern British English vowels by native speakers of Greek and Japanese," *Phonetica* **66**, 169–187.
- Levy, E. S. (2009a). "Language experience and consonantal context effects on perceptual assimilation of French vowels by American-English learners of French," *J. Acoust. Soc. Am.* **125**, 1138–1152.
- Levy, E. S. (2009b). "On the assimilation-discrimination relationship in American English adults' French vowel learning," *J. Acoust. Soc. Am.* **126**(5), 2670–2682.
- Lobanov, B. M. (1971). "Classification of Russian vowels spoken by different listeners," *J. Acoust. Soc. Am.* **49**, 606–608.
- Mora, J. C., and Nadeu, M. (2012). "L2 effects on the perception and production of a native vowel contrast in early bilinguals," *Int. J. Biling.* **16**(4), 484–500.
- Rallo Fabra, L. (2005). "Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test," *Vigo Int. J. Appl. Linguist.* **2**, 75–92.
- Rallo Fabra, L., and Romero, J. (2012). "Native Catalan learners' perception and production of English vowels," *J. Phon.* **40**(3), 491–508.
- Recasens, D. (2014). *Fonètica i fonologia experimentals del català. Vowels i consonants (Catalan Experimental Phonetics and Phonology. Vowels and Consonants)* (Institut d'Estudis Catalans, Barcelona, Spain).
- Recasens, D., and Espinosa, A. (2006). "Dispersion and variability of Catalan vowels," *Speech Commun.* **48**, 645–666.
- Roach, P. (2004). "British English: Received pronunciation," *J. Int. Phon. Assoc.* **34**(2), 239–245.
- Schmidt, A. M. (1996). "Cross-language identification of consonants. Part 1. Korean perception of English," *J. Acoust. Soc. Am.* **99**(5), 3201–3211.
- Schmidt, A. M. (2007). "Cross-language consonant identification," in *Language Experience in Second Language Speech Learning: In Honor of James Fllege*, edited by M. Munro and O.-S. Bohn (John Benjamins, Amsterdam), pp. 185–200.
- Strange, W. (2007). "Cross-language similarity of vowels," in *Language Experience in Second Language Speech Learning: In Honor of James Fllege*, edited by M. Munro, and O.-S. Bohn (John Benjamins, Amsterdam), pp. 35–55.
- Strange, W., Akahane-Yamada, R., Kubo, R., Trent, S. A., and Nishi, K. (2001). "Effects of consonantal context on perceptual assimilation of American English vowels by Japanese listeners," *J. Acoust. Soc. Am.* **109**, 1691–1704.
- Strange, W., Bohn, O.-S., Nishi, K., and Trent, S. A. (2005). "Contextual variation in the acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **118**, 1751–1762.
- Strange, W., Bohn, O.-S., Trent, S. A., and Nishi, K. (2004). "Acoustic and perceptual similarity of North German and American English vowels," *J. Acoust. Soc. Am.* **115**, 1791–1807.
- Strange, W., Levy, E., and Law, F. (2009). "Cross-language categorization of French and German vowels by naive American listeners," *J. Acoust. Soc. Am.* **126**, 1461–1476.
- Strange, W., Weber, A., Levy, E., Shafiro, V., and Nishi, K. (2002). "Within- and across-language acoustic variability of vowels spoken in different phonetic and prosodic contexts, American English, North German, Parisian French," *J. Acoust. Soc. Am.* **112**(5), 2384.
- Tyler, M. D. (2021). "Phonetic and phonological influences on the discrimination of non-native phones," in *Second Language Speech Learning: Theoretical and Empirical Progress*, edited by R. Wayland (Cambridge University, Cambridge, UK), pp. 157–174.
- Tyler, M. D., Best, C., Faber, A., and Levitt, A. (2014). "Perceptual assimilation and discrimination of non-native vowel contrasts," *Phonetica* **71**, 4–21.