

NOVEMBER 07 2022

Perception of English and Catalan vowels by English and Catalan listeners. Part II. Perceptual vs ephoric similarity

Juli Cebrian



J Acoust Soc Am 152, 2781–2793 (2022)

<https://doi.org/10.1121/10.0014902>



View
Online



Export
Citation

CrossMark

Related Content

Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs

J Acoust Soc Am (January 2019)

“If a perceptual problem exists, what are its roots?” Types of cross-language similarity, hypercorrection, and unexplored root canals

J Acoust Soc Am (March 2018)

Examining the success of three kinds of cross-language similarity in predicting English listeners’ discrimination of Danish vowels

J Acoust Soc Am (October 2021)



Advance your science and career
as a member of the

ACOUSTICAL SOCIETY OF AMERICA

LEARN MORE



Perception of English and Catalan vowels by English and Catalan listeners. Part II. Perceptual vs ephoric similarity

Juli Cebrian^{a)} 

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

ABSTRACT:

Although crosslinguistic similarity is a crucial concept for many disciplines in the speech sciences, there is no clear consensus as to the most appropriate method to measure it. This paper assessed the perceived similarity between English and Catalan vowels by means of an overt direct task evaluating perceptual similarity. The extent to which perceptual similarity is reciprocal is also explored by comparing perceptual judgements obtained by speakers of the two languages involved. Twenty-seven native Catalan speakers and 27 native English speakers rated the perceived dissimilarity between two aurally presented vowel stimuli. Trials included native–non-native pairs as well as native–native pairs to serve as baseline data. Some native–non-native pairs were perceived to be as similar as same-category native pairs, illustrating cases of very high crosslinguistic perceptual similarity. Further, in terms of reciprocity, the results showed a bidirectionality in similarity relationships that point to some cases of near-identical or shared categories and also illustrate the role of language-specific cue weighting in determining perceptual similarity. Finally, a comparison with the outcome of a previous study [Cebrian (2021). *J. Acoust. Soc. Am.* **149**(4), 2671–2685], involving the same participants and languages but exploring ephoric similarity, shows a generally high degree of agreement and a close relationship between the two types of similarity. © 2022 Acoustical Society of America. <https://doi.org/10.1121/10.0014902>

(Received 10 April 2022; revised 27 September 2022; accepted 29 September 2022; published online 7 November 2022)

[Editor: Ewa Jacewicz]

Pages: 2781–2793

I. INTRODUCTION

One of the main questions that phonetics addresses is the extent to which sounds from different languages resemble each other (e.g., Ladefoged and Ferrari Disner, 2012; Ladefoged and Johnson, 2014). Bohn (2002) stresses how this question, which we can refer to as the issue of crosslinguistic similarity, is crucial for different disciplines within the speech sciences. Still, there is no general consensus as to the most appropriate method of measuring crosslinguistic similarity (Bohn, 2002, 2017; Strange, 2007). The aim of this paper is to evaluate one method of assessing crosslinguistic similarity, namely, a paired comparison technique in which listeners rate the dissimilarity between a native and a non-native sound (i.e., a rated dissimilarity task or RDT). In addition, the extent to which the degree of perceived (dis)similarity between pairs of sounds from two different languages is determined by the specific characteristics of each language is also explored. This is performed by examining the perceived dissimilarity between the same pairs of Catalan and English vowels by native speakers of each of these two languages who performed the same RDT. Finally, the outcome of the current study is compared with the results of a previous study that involved the same languages and speakers as in this study but used an interlingual identification approach aimed at analyzing crosslinguistic

ephoric similarity (Cebrian, 2021). Sections IA–ID review the previous findings that motivate these goals.

A. Importance of crosslinguistic similarity in second language speech research

Most second language (L2) speech models aim to explain the processes that take place in adult L2 speech learning and make predictions about the difficulty of target language perception and production (see Bohn, 2002, 2017; Chang, 2019; Strange, 2007). For example, the speech learning model by Flege (1995) and its revised version, the SLM-r (Flege and Bohn, 2021), claims that adult L2 speech learners may be able to establish accurate or target-like categories for L2 sounds if they are able to detect differences between L1 and L2 phones. Given enough experience with the target language, this is more likely to happen in the case of dissimilar L1 and L2 phones than in the case of similar L1 and L2 phones (Flege, 1995). Further, according to the perceptual assimilation model (PAM) from Best (1995) on non-native perception and its adaptation to L2 perception (PAM-L2; Best and Tyler, 2007), target language phones will be perceptually assimilated to (i.e., identified with) the closest native category and categorized as better or worse exemplars of that category. On the other hand, a target phone with no counterpart in the L1 may be uncategorized or even heard as nonspeech. Based on the non-native/L2 to L1 perceptual mappings, PAM/PAM-L2 makes a series of predictions for accuracy of discrimination of non-native or

^{a)}Electronic mail: juli.cebrian@uab.cat

L2 segmental contrasts (see Tyler, 2021, for an overview of PAM's assumptions and predictions). Similarly, other speech perception and L2 speech learning models draw on the construct of crosslinguistic distance to explain how L2/non-native sounds are perceptually categorized. The native language magnet theory (NLM; e.g., Kuhl *et al.*, 2008) explains that the smaller the phonetic distance between a non-native sound and an L1 category prototype, the more likely it is that the non-native sound is perceived in terms of the native category. Finally, the second language linguistic perception model (L2LP; e.g., Elvin and Escudero, 2019) establishes different learning scenarios for non-native/L2 sounds that depend on the degree of acoustic and/or perceptual overlap between native and non-native sounds. Therefore, establishing a reliable and principled method of measuring crosslinguistic phonetic similarity is crucial for the evaluation of current theories of speech perception and L2 speech (Bohn, 2002; Strange, 2007).

B. Measuring perceived similarity: Ecphoric vs perceptual approaches

L2 speech researchers have advocated for the use of perceptual approaches over other types of methods of evaluating crosslinguistic phonetic similarity, such as articulatory descriptions and acoustic comparisons (Bohn, 2002; Strange, 2007). This preference is motivated by the fact that acoustic analyses are more likely to be affected by factors such as speech rate, phonetically conditioned variability, and individual variation. In addition, some studies have reported discrepancies between acoustic and perceptual measures (Strange *et al.*, 2004; Strange, 2007). Bohn (2002) and Strange (2007) conclude that perceptual tasks constitute a more direct method of evaluating crosslinguistic similarity, especially if the goal is to evaluate the way L1 perception influences the categorization of non-native or L2 sounds (Tyler, 2021; Tyler *et al.*, 2014; but cf. Thomson *et al.*, 2009 for a statistical approach based on acoustic measures).

The two most common perceptual approaches are paired comparisons and interlingual identification (Bohn, 2017). A paired comparison technique, such as a graded discrimination or RDT, involves using a Likert scale to indicate the degree of (dis)similarity between two stimuli presented aurally, typically a native and a non-native phone (e.g., Flege *et al.*, 1994; Fox *et al.*, 1995). RDT is a direct method of assessing crosslinguistic similarity as listeners are specifically asked to compare the two stimuli presented and indicate how close they are. An interlingual identification approach is exemplified by perceptual tasks known as a forced choice categorization task or perceptual assimilation task (PAT). In a PAT, listeners are presented with exemplars of non-native segments and asked to identify them in terms of L1 categories, typically using labels representing the L1 sounds, and provide a goodness of fit rating (GR; e.g., Cebrian, 2021). PAT illustrates an indirect method of assessing similarity as listeners are not directly instructed to judge the similarity between two sounds but rather to

identify a given stimulus as one from a set of possible responses. In addition, while RDT is an overt task in which the two stimuli to be compared are physically present, PAT is a covert task as the listener is indirectly required to compare a physical stimulus to a mental category or memory representation.

Bohn (citing Tulving, 1981) suggests that overt and covert tasks may be assessing different types of similarity, “what Tulving refers to as ‘perceptual similarity’ (in overt tasks) as opposed to ‘ecphoric similarity’ (in covert tasks)” (Bohn, 2002, p. 201). Despite the prevalence of PATs in crosslinguistic perception studies [see work by Strange and colleagues (e.g., Strange *et al.*, 1998) and those by Best and colleagues (e.g., Faris *et al.*, 2016), among many others], there is no clear consensus as to whether this is the most suitable technique. This paper examines the perceptual similarity between English and Catalan vowels using an overt direct approach (Bohn, 2002) in which speakers of both languages involved provide perceptual dissimilarity ratings (DRs) on a number of native–non-native vowel pairs. In the last part of the paper, these results will be compared to ecphoric similarity judgements obtained in Cebrian (2021) to assess the relationship between these two types of approaches (perceptual similarity vs ecphoric similarity), something which remains largely untested, at least in relation to L2 speech.¹

C. Previous findings involving perceptual similarity

Few previous studies have examined perceptual similarity between native and non-native vowels using a RDT or similar task (Cebrian *et al.*, 2011; Flege *et al.*, 1994). Flege *et al.* (1994) tested two groups of Spanish learners of English, differing in length of residence in the United States (1.8 vs 7 years), as well as native English-speaking controls. Using a nine-point scale, listeners rated the degree of dissimilarity between Spanish-Spanish, English-English, and Spanish-English vowel pairs. The results showed that DRs were influenced by auditory distance as the greater the distance between vowels in an F1-F2 bark space, the higher were the DRs obtained. However, L2 experience did not seem to play a role as the two groups of Spanish speakers did not differ in their DRs. Some native–non-native pairs received relatively low DRs by all of the groups, such as English /i/-Spanish /i/ and English /ʌ/-Spanish /a/, although the absence of same-category same-language pairs in the experiment (e.g., E /i/-E /i/) makes it difficult to assess how native-like the non-native vowels were perceived to be.

Specifically involving Catalan (C) and English (E) vowels, Cebrian *et al.* (2011) collected perceptual similarity data from a group of Catalan learners of English by means of a RDT that involved pairs of L1-L2 vowels (e.g., CE /a/-æ/) as well as pairs of same-language same-category (e.g., EE /i/-/i:/, CC /u/-/u/) and different-category (CC /i/-/e/, EE /i/-/i/) vowels. The task used a 7-point scale and was completed by 47 undergraduate students majoring in English. The results showed that the two pairs of CC same-category

vowels obtained the lowest DRs (mean = 1.7 out of 7), followed closely by the EE same-category pairs (mean = 2, range = 1.7–2.3). CE pairs yielded DRs that ranged from 2.2 (CE /ɛ/-/ɛ/) to 6.4 (CE /ɛ/-/æ/) out of 7. Thus, some CE pairs yielded DRs within the range of those obtained for EE same-category pairs (CE /ɛ/-/ɛ/, mean = 2.2; /i/-/i/ and /ɔ/-/ɔ/, mean = 2.3; followed by /a/-/æ/, mean = 2.5). This shows evidence of a high degree of perceived crosslinguistic similarity. Still, the study had some shortcomings. First, the stimuli used in the experiments were unmodified /b/ + vowel + /t/ words (henceforth, /bVt/ words). 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 (see Sec. II A 2). In addition, the RDT task in Cebrian *et al.* (2011) contained only a subset of the possible vowel combinations, and the target vowels were not equally distributed across trials. Finally, the listeners in that study were experienced L2 English speakers (undergraduate English majors), and no native English speakers were tested. This study differs from Cebrian *et al.* (2011) in three main ways: (a) methodological limitations regarding the stimuli and task design have been addressed; (b) participants have no or minimal experience with the non-native language; and (c) the perceived similarity between the same set of Catalan and English vowels is elicited from native speakers of the two languages involved. This reciprocal or bidirectional approach to perceptual similarity is discussed next.

D. Reciprocal assessment of similarity

Most studies examining crosslinguistic perceived similarity obtain similarity judgements from speakers of one of the two languages involved. To our knowledge, only a study on crosslinguistic perception of consonants by Korean (Schmidt, 1996) and English (Schmidt, 2007) speakers and our previous study (Cebrian, 2021) have elicited similarity from two parallel populations. Cebrian (2021) investigated the extent to which speakers of the two languages examined would yield comparable similarity judgements for the same pairs of native and non-native vowels. This was done to obtain a more complete picture of the similarity relationships between two sound systems, examine if sounds from two different languages could be considered perceptually the same, and explore whether perceptual distance is a shared construct by speakers of the two languages involved. To this end, native speakers of Catalan and Standard Southern British English (SSBE) evaluated the ephoric similarity of the same set of 11 Catalan and 13 English vowels and diphthongs. Out of the 11 potential parallel matches, 8 Catalan-English pairs received reciprocal modal responses with assimilation scores equal to or higher than a 70% categorization threshold (Tyler *et al.*, 2014; see Tyler, 2021, for a discussion of categorization thresholds). These were CE /a₁/-/a₁/, /ɛ/-/ɛ/, /a/-/æ/, /o₁/-/ə₁/, /e₁/-/e₁/, /aj/-/ai/, /e/-/ɪ/, and /u/-/u₁/. For example, C /ɛ/ was assimilated to E /ɛ/ 92% of the time with an average GR of 4.2 out of 7, whereas E /ɛ/ was identified as C /ɛ/ 91% of the

time (GR, 5.6). Two more pairs obtained reciprocal but unequal mappings: E /i:/ and /ɒ/ were more consistently assimilated as C /i/ (87%) and /ɔ/ (82%) than the reverse (62% and 53%, respectively), and C /o/ and E /ɔ:/ did not receive reciprocal mappings. These discrepancies were linked to role of context-specific cues such as vowel duration in L1 perception. Catalan vowels are, on average, similar in duration to English lax vowels. Hence, C /i/ and /o/ are closer in duration to E /ɪ/ and /ɒ/ than to E /i:/ and /ɔ:/. Thus, the English tense vowels were not consistently selected as the L1 match for the Catalan stimuli despite being spectrally closer (Cebrian, 2021). Similarly, in the study by Schmidt (1996, 2007) on crosslinguistic perception of consonants, the English and Korean listeners generally provided reciprocal patterns of non-native to native mappings. Still, Schmidt also found that crosslinguistic similarity varied depending on context-specific cues such as variations in stop voice onset time (VOT), the presence of burst-like transitions in nasals, and the degree of labialization of the vowels following the target consonants. 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). This paper extends our previous study by testing the same group of participants and assessing perceptual (as opposed to ephoric) similarity.

E. The current study

The focus of this study is the perceptual similarity between Catalan and SSBE vowels. SSBE has an inventory of 12 vowels (/i: ɪ ɛ ɜ: æ ʌ ɑ: ɒ ɔ: ʊ u: ə/) and 8 diphthongs, e.g., /aɪ aʊ eɪ əʊ/ (Cruttenden, 2014). The Central or Eastern variety of Catalan has an inventory of seven vowels (/i, e, ɛ, a, ɔ, o, u/), a reduced vowel ([ə]) in unstressed position, and a number of diphthongs resulting from the combination of vowels and high glides, e.g., /a₁ e₁ a₁ o₁/, etc., (Recasens, 1993; see Cebrian, 2021, for an acoustic comparison of the two vowel systems). The results of two experiments examining the perceptual similarity between Catalan and English vowels are presented. In the first experiment, a group of native Catalan speakers rated the perceived dissimilarity between Catalan and English vowels by means of a RDT. In the second experiment, the same task was administered to a group of SSBE speakers.

II. EXPERIMENT 1. PERCEPTUAL SIMILARITY BETWEEN SSBE AND CATALAN VOWELS: CATALAN LISTENERS

A. Methodology

1. Participants

Twenty-seven Catalan-dominant Catalan-Spanish bilinguals (15 females; average age, 24 years old; age range, 17–48 years old) participated in the experiment and were mostly undergraduate students. Many had studied some

English at school, where the focus of English language instruction is mostly grammar and vocabulary. Participants reported minimal or no immersive experience in English. Hence, they were considered naive listeners (Best and Tyler, 2007). All of the 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 stimuli were the seven monophthongs /i e ε a o u/ and four rising diphthongs, namely, /eɪ ou aɪ au/. Diphthongs, i.e., combinations of a monophthong and a semivowel (/i/ or /u/), were included given their reported role in crosslinguistic perception (Cebrian, 2011, 2019; Escudero and Williams, 2011). Still, to limit the complexity and duration of the task, only those diphthongs that had been found to be perceptually close to English vowels in previous studies involving Catalan and English (Cebrian *et al.*, 2011; Cebrian, 2021) were selected. The English stimuli contained the vowels and diphthongs /i:ɪ ε ɜ: æ ʌ ɑ: ɒ ɔ u: eɪ aɪ aʊ əʊ/.³ The vowels were produced in a /b/ + vowel + /t/ context, a possible sequence in both languages, embedded in a carrier phrase. Phrases were produced by three male Catalan and three male English native speakers (average age, 32.5 years old, range 21–44 years old; see Cebrian, 2021, for further details). The production of stops in Catalan and English may differ in terms of the VOT of initial voiced stops (voice-lead vs short lag VOT) and the intensity of the release and possible aspiration in final voiceless stops (Docherty, 1992; Recasens 1993). Thus, stimuli were edited to prevent such crosslinguistic differences in consonant production from affecting perceptual judgements of vowel similarity. Consequently, each stimulus started from the release of the /b/ and ended with the beginning of the /t/ closure, therefore, maintaining intact the relevant cues to the vowel (i.e., vowel transitions; Strange, 1987).

3. RDT

In this task, listeners were presented with a pair of sounds and had to indicate the degree of perceived (dis)similarity on a continuous nine-point scale, where one meant very similar and nine meant very dissimilar (Flege *et al.*, 1994). Pairing all of the English vowels with all of the Catalan vowels with all of the possible talker and order combinations would have resulted in an extremely large and impractical number of stimuli. Hence, a subset of vowel combinations were selected, ensuring that all of the crucial L1-L2 vowel pairs that are adjacent in a perceptual space (following previous findings, e.g., Cebrian *et al.*, 2011; Cebrian, 2021; Rallo Fabra, 2005) were included. The task was designed so that all of the individual vowels appeared the same number of times throughout the task. Thus, each non-native vowel appeared in three vowel combinations in the Catalan-English pairs and every single Catalan and English vowel appeared in the same number of

combinations across the total number of same-language and different-language pairs. For example, E /æ/ was paired up with C /a/ and /ε/ (found to be perceptually close to the English vowel) and C /aɪ/ to ensure the same number of instances of these vowels throughout the experiment. This design resulted in 42 Catalan-English pairs (36 of these were used in experiments 1 and 2; see Sec. III A 3), and a set of same-language (L1-L1) pairs that were specific to each experiment. The L1-L1 pairs included some same-category and different-category pairs, which provided baseline ratings to compare to the native–non-native pairs. Each vowel combination appeared six times, that is, in three different talker combinations and two orders (e.g., E /æ/ - C /a/ and C /a/ - E /æ/). Given the resulting large number of stimuli, two versions of the task were created, grouping mostly front and mostly back vowels separately. Set A involved the English vowels /aɪ, æ, eɪ, ε, ɪ, i:, ɜ:/ and the Catalan vowels /i e ε eɪ aɪ aʊ/, whereas Set B included E /u: əʊ ɔ: ɒ ɑ: aʊ ʌ/ and C /u ou o ɔ aʊ a/. C /a/ appeared in both sets. Table I presents the list of the vowel pairs included in each of the subtasks (sets A and B) in experiment 1. The interstimulus interval was 1.2 ms to encourage the use of phonetic information stored in long term memory (Flege *et al.*, 1994; Pisoni, 1973; among others). The total number of trials per set was 180 (30 vowel pairs × 3 talker combinations × 2 orders).

4. Procedure

The listeners performed two types of perceptual tasks, a perceptual assimilation task (reported in Cebrian 2021) and the RDT. The order of the two sets (A and B) was counter-balanced across all of the listeners. Instructions were given at the beginning of each task. The listeners were told to pay attention to the vowel sounds and disregard traces of consonants in the stimuli as well as possible voice quality or pitch differences. They were also 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 selected a rating on a nine-point scale, where one meant “very similar” and nine meant “very dissimilar” by clicking on a computer screen. Praat software (Boersma and Weenink, 2018) was used to conduct the experiments. Each task was preceded by a short practice phase to familiarize participants with the procedure. The practice session consisted of nine trials (four native–non-native and five native-native pairs). On average, participants took about 10 min to complete each task.

B. Results

The average DRs for each pair of vowels for each participant was calculated.⁴ The degree of inter-rater agreement among the 27 participants was evaluated first by obtaining the reliability coefficients (Cronbach’s alpha). A very high degree of agreement was observed (0.990; note that 0.700 is considered the lower limit for evidence of reliability; Cortina, 1993). This result indicates that the listeners were very consistent in

TABLE I. The vowel pairs included and number of trials in each RDT in experiment 1.

Number of vowel pairs per subtask	RDT A (mostly front vowels)	RDT B (back and rounded vowels)
5 L1 Catalan same-category vowel pairs	/a/-/a/, /ai/-/ai/, /ei/-/ei/, /e/-/e/, /i/-/i/	/a/-/a/, /au/-/au/, /ou/-/ou/, /ɔ/-/ɔ/, /u/-/u/
4 L1 Catalan different-category vowel pairs	/ai/-/a/, /ei/-/e/, /e/-/e/, /i/-/e/	/a/-/ɔ/, /o/-/ɔ/, /o/-/ou/, /u/-/au/
21 Catalan-English vowel pairs	/a/-/æ/, /e/-/æ/, /ai/-/æ/, /a/-/e/, /e/-/e/, /e/-/e/, /ai/-/ei/, /e/-/ei/, /ei/-/ei/, /e/-/i/, /ei/-/i/, /i/-/i/, /ei/-/i:/, /i/-/i:/, /ai/-/i:/, /ai/-/ai/, /i/-/ai/, /a/-/ai/, /a/-/ɜ:/, /e/-/ɜ:/, /e/-/ɜ:/	/a/-/ɑ:/, /au/-/ɑ:/, /ɔ/-/ɑ:/, /o/-/ɔ:/, /ɔ/-/ɔ:/, /ou/-/ɔ:/, /a/-/au/, /au/-/au/, /ou/-/au/, /ɔ/-/ɒ/, /o/-/ɒ/, /a/-/ɒ/, /u/-/u:/, /ou/-/u:/, /o/-/u:/, /ou/-/əʊ/, /o/-/əʊ/, /u/-/əʊ/, /a/-/ʌ/, /u/-/ʌ/, /au/-/ʌ/

their assessments of perceived dissimilarity. Further, the potential effect of task order, that is, front vowels (set A) first or back vowels (set B) first, was examined by comparing the average DRs for each vowel pair obtained by each order subgroup (AB, BA) and for each subset. The results of a generalized linear mixed model with vowel set (A or B), task order (AB or BA), their interaction as fixed factors, and a random intercept for participant indicated that the order in which the tasks were completed did not affect the average DRs for either subset of vowels [task order, $F(1,9554) = 0.039$, $p = 0.844$; vowel set, $F(1,9554) = 0.059$, $p = 0.809$; task order \times vowel set, $F(1,9554) = 3.002$, $p = 0.083$]. In addition, the results for the two order subgroups were very highly correlated ($r = 0.974$, $p < 0.001$, $N = 60$). SPSS software (IBM Corp, 2021) was used for the statistical analyses.

Figure 1 displays the average DRs per vowel pair, grouped by language combination, i.e., CC and CE, and vowel set, i.e., front vowels (set A) and back vowels (set B). The average DR for CC same-category pairs was two, ranging from 1.4 to 2.5, with a standard deviation of 1.2. The CC different-category pairs obtained ratings ranging from 4.4 (/ɔ/-/o/) to 8 (/u/-/au/). Regarding the CE pairs, the DRs ranged from 2.3 to 8, showing great variation among CE combinations in the degree of perceptual similarity. Three of the CE pairs received ratings within the range of those obtained for CC same-category pairs, namely, CE /e/-/e/, /e/-/i/, and /a/-/æ/, with ratings of 2.3, 2.3, and 2.4, respectively. Three more pairs obtained ratings that were equal or lower than 3.2, that is, within one standard deviation (SD) of the mean rating for CC same-category pairs (i.e., mean $2 + SD\ 1.2 = 3.2$): CE /i/-/i:/, /au/-/au/, and /a/-/ʌ/, with ratings of 2.8, 2.9, and 3.2, respectively, whereas /ɔ/-/ɒ/ received the following lowest rating (3.3).

These results follow the general trends reported by [Cebrian et al. \(2011\)](#), a previous study that also looked at SSBE and Catalan vowels but with some methodological differences (see Sec. IC). This study examined a total of 60 vowel combinations (18 CC pairs and 42 CE pairs), which included the 7 CC and 16 CE pairs tested in the previous study. The results of a Pearson correlation conducted on the DRs obtained for the common 23 pairs yielded a high level of agreement between the two studies ($r = 0.901$, $N = 23$,

$p < 0.001$). The cases of discrepancy involved pairings with vowel /ɪ/. Participants in both studies found E /ɪ/ to be perceptually more similar to C /e/ than to C /i/, but the difference was considerably larger in the current study (/e/-/ɪ/, 2.3 vs /i/-/ɪ/, 5.6) than in the previous study (/e/-/ɪ/, 3.7 vs /i/-/ɪ/, 4.3).⁵

C. Discussion

Experiment 1 examined the perceptual similarity between pairs of Catalan and SSBE vowels from the perspective of Catalan-speaking listeners. The results show that, generally, for each of the English vowels, there is a single native vowel that is perceived as least dissimilar, e.g., CE /e/-/e/, /e/-/ɪ/, /i/-/i:/, /au/-/au/, /o/-/ɔ:/, /ei/-/ei/, /ai/-/ai/, /ou/-/əʊ/, and /u/-/u:/, indicating a systematic pattern of non-native to native category mapping. However, on two occasions, the same native vowel was chosen as the most similar to two English vowels with varying DR values: C /a/ for E /æ/ (DR, 2.4) and /ʌ/ (3.2), and C /ɔ/ for E /ɒ/ (3.3) and /ɑ:/ (4.3). This outcome points to a potential difficulty on the part of Catalan speakers to distinguish these pairs of English vowels (/æ/-/ʌ/, /ɒ/-/ɑ:/). Finally, E /ɜ:/ appears to have no clear match in the L1 as it was consistently rated as rather dissimilar from all of the L1 vowels it was paired with (DRs, 6.3–6.9). The more dissimilar that the non-native vowels are from the closest L1 vowel (e.g., E /ɜ:/), the more likely it is that Catalan learners of English detect differences between them and ultimately establish a separate more target-like category for the non-native vowel given sufficient experience with the target language ([Flege, 1995](#); [Flege and Bohn, 2021](#)). By contrast, the fact that some CE pairs obtained DRs within the range of those received by same-category L1 vowels (e.g., CE /e/-/e/, /e/-/ɪ/, /a/-/æ/, /i/-/i:/, /au/-/au/) suggests that those English vowels may be readily assimilated to existing native categories. In cases of very close similarity (near-identity), the non-native vowels may be perceived and produced in terms of the corresponding native categories by Catalan learners of English, and the use of Catalan categories may go undetected by English listeners. Testing all of these possibilities lies beyond the scope of this paper. Still, analyzing the perception of Catalan vowels by native English speakers (experiment 2) can shed light

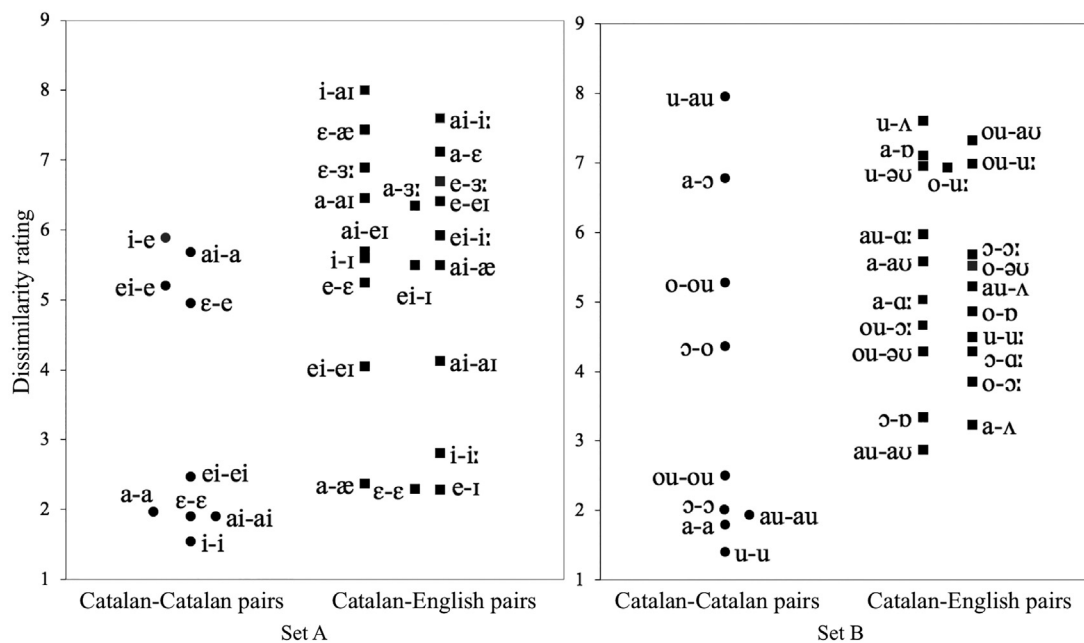


FIG. 1. The mean DRs for Catalan-Catalan (●) and Catalan-English (■) vowel pairs elicited from Catalan L1 speakers. Set A vowels are presented on the left, and set B vowels are presented on the right.

on the degree to which some Catalan and English vowels are perceptually similar or even near-identical.

Regarding the comparison to previous studies, the results of the current experiment are strongly correlated with the results reported by Cebrian *et al.* (2011), also involving SSBE and Catalan vowels. The two studies only showed some differences with pairs involving E /i/. Whereas in this study, E /i/ was heard as notably closer to C /e/ than to C /i/ (DRs of 2.3 vs 5.6), in Cebrian *et al.* (2011), the difference between the DRs for CE /e/-/i/ and CE /i/-/i/ was smaller (3.7 vs 4.3, respectively). One possible explanation for this may stem from the fact that unlike naive listeners, the L2 speakers in Cebrian *et al.* (2011) may have had orthographic and metalinguistic knowledge about English vowels. This knowledge may have resulted in crosslinguistic perceptual mappings that are more phonologically than phonetically motivated (Chang, 2019; Holliday, 2015). Specifically, the E /i/, which is acoustically and perceptually closest to C /e/ (Cebrian, 2021), may be perceived by Catalan L2 English learners as less dissimilar to C /i/ given the learners' metalinguistic knowledge [e.g., /i/ is commonly spelled with letter ⟨i⟩, which corresponds to C /i/, and is often learned in opposition to E /i:/, being often (incorrectly) described as a "short /i/"]. In any case, the remaining pairs obtained very consistent results across studies. This consistency was not unexpected as the stimuli in both studies had the same origin and both were elicited in bVt words produced by the same talkers even if in the current study, neighbouring consonants were edited to minimize consonant differences (see Sec. II A 2). The fact that the current participants were naive listeners while those in Cebrian *et al.* (2011) were L2 learners does not seem to have affected the overall results. This is in line with the finding by Flege *et al.* (1994) that two groups

of Spanish learners of English differing in amount of L2 experience (years of residence) did not differ in their DRs for pairs of Spanish and American English vowels. Rallo Fabra and Romero (2012), another study looking at Catalan and English vowels, did not specifically elicit perceptual dissimilarity but included seven Catalan-English pairs in a categorical oddity discrimination task that also included English-English pairs. Although the English stimuli represented American English, the results are, on the whole, in line with the current RDT results (DRs). The CE pair /i/-/i/ (current DR, 5.6) was generally better discriminated than /ε/-/ε/ (2.3), /a/-/æ/ (2.4), /i/-/i:/ (2.8), /a/-/Λ/ (3.2), and /a/-/ɑ:/ (5), and the latter three pairs tended to be better discriminated than /ε/-/ε/ and /a/-/æ/.

In sum, the results of experiment 1 have illustrated different patterns of perceptual similarity between English and Catalan vowels, from very high degrees of similarity to vowels with no clear match in the L1. The extent to which these patterns will be replicated in the parallel language situation, native English speakers' perception of Catalan, is explored next.

III. EXPERIMENT 2. PERCEPTUAL SIMILARITY BETWEEN SSBE AND CATALAN VOWELS: ENGLISH LISTENERS

A. Methodology

1. Participants

The participants were 27 native speakers of SSBE (23 females), who were undergraduate (19) and graduate (8) students in the Division of Psychology and Language Sciences at University College London (average age, 22 years old; range, 18–30 years old). They had been raised

in Southern England or born to parents from Southern England. Most spoke no other language fluently except for one subject who was fluent in German and another with passive knowledge of Spanish. All of the participants reported normal hearing and were paid for their participation.

2. Stimuli

The stimuli were the same as in experiment 1.

3. RDT

The design of the RDT used in experiment 2 followed the same criteria as in experiment 1. A total of 36 Catalan-English pairs were examined, which were also included in experiment 1 (the RDTs in experiment 1 had 3 more CE pairs given the greater number of vowels in English). Each Catalan vowel appeared in combination with an English vowel three times (e.g., C /e/-E /ɪ/, C /e/-E /eɪ/, C /e/-E /ɛ/). Also included were 22 English-English pairs, 10 same-category pairs, and 6 different-category pairs, yielding a total of 58 vowel pairs, which is almost equal to the number of pairs used with the Catalan listeners (60). Similar to experiment 1, two subtasks were created, one for (mostly) front vowels and one for back and rounded vowels. Table II lists all of the vowel pairs tested in experiment 2. The total number of trials per subtask was 174 (29 vowel pairs × 3 talker combinations × 2 orders).

4. Procedure

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

B. Results

As was done with the results for the Catalan listeners, the degree of agreement among the 27 listeners in their DRs was examined first. The reliability coefficient (Cronbach’s alpha) revealed a very high degree of inter-rater agreement (0.992), confirming the very high level of consistency in listeners’ responses, as was also observed in experiment 1. Next, the potential effect of task order (subtask A followed by subtask B or the reverse) was examined. As with experiment 1, the results of a generalized linear mixed model with vowel set (A or B), task order (AB or BA), and their

interaction as fixed factors and a random intercept for participant indicated no effect of task order on the average DRs for either subset of vowels [task order, $F(1,9566) = 2.7, p = 0.1$; vowel set, $F(1,9566) = 0.705, p = 0.401$; task order × vowel set, $F(1,9566) = 0.787, p = 0.375$]. Finally, the results for the two order subgroups were very highly correlated ($r = 0.988, N = 58, p < 0.001$).

Figure 2 presents the average DRs obtained for each pair of vowels, including the L1-L1 English pairs and native–non-native CE pairs, organized by vowel set. The English same-category pairs obtained an average DR of 1.5, ranging from 1.2 to 1.9, with a standard deviation of one, whereas the L1 different-category pairs obtained ratings ranging from 5.3 (/i:/-/ɪ/) to 8.1 (/ɜ:/-/ɪ/, /ɜ:/-/aɪ/, and /ɜ:/-/i:/). The ratings obtained for the CE pairs ranged from 1.8 to 7.7, showing that some native–non-native vowel pairs were perceived to be as close as some L1-L1 pairs. Specifically, four pairs were within the range of the values for same-category English pairs: CE /a/-/æ/, /ɛ/-/ɛ/, /aʊ/-/aʊ/ and /e/-/ɪ/, all four with a rating of 1.8. In addition, the DR for CE /eɪ/-/eɪ/ (2.4) was also within one standard deviation of the average rating for same-category L1 pairs (i.e., mean 1.5 + SD 1 = 2.5), followed closely by CE /oʊ/-/əʊ/ and /i/-/i:/ with 2.6 and 2.7, respectively. The remaining CE pairs received ratings that were above five.

C. Discussion

Unlike experiment 1, experiment 2 is the first study to report dissimilarity judgements between English and Catalan vowels elicited from L1 English speakers, hence, there are no previous results to compare the current results to. As was observed with the Catalan listeners, most non-native (Catalan) vowels were perceived as closest to a different L1 (English) vowel, namely, CE /a/-/æ/, /ɛ/-/ɛ/, /aʊ/-/aʊ/, /e/-/ɪ/, /eɪ/-/eɪ/, /oʊ/-/əʊ/, /i/-/i:/, /aɪ/-/aɪ/, and /u/-/u:/, some of which received DRs comparable to those given to same-category EE pairs. The exceptions were the Catalan high mid and low mid back rounded vowels (/o/ and /ɔ/, respectively). Given the height difference between the Catalan back rounded vowels, it could be expected that English listeners would find C /o/ to be least dissimilar to the English mid back rounded vowel /ɔ:/ and C /ɔ/ to the English low vowel /ɒ/.⁶ However, both Catalan vowels were found to be least dissimilar to E /ɒ/, followed by E /ɔ:/. In fact, this finding replicates the results in [Cebrian \(2021\)](#) as

TABLE II. The vowel pairs included and number of trials in each RDT in experiment 2.

Number of vowel pairs per subtask	RDT A (mostly front vowels)	RDT B (back and rounded vowels)
5 L1 English same-category vowel pairs	/æ/-/æ/, /aɪ/-/aɪ/, /eɪ/-/eɪ/, /ɛ/-/ɛ/, /i:/-/i:/	/ɑ:/-/ɑ:/, /ʌ/-/ʌ/, /ɔʊ/-/ɔʊ/, /aʊ/-/aʊ/, /u:/-/u:/
6 L1 English different-category vowel pairs	/aɪ/-/ɜ:/, /eɪ/-/aɪ/, /ɛ/-/æ/, /i:/-/ɜ:/, /ɪ/-/ɜ:/, /i:/-/ɪ/	/əʊ/-/ɔ:/, /aʊ/-/ɔ:/, /aʊ/-/u:/, /ɑ:/-/ɒ/, /ʌ/-/ɒ/, /u:/-/ɒ/
18 Catalan-English vowel pairs	/a/-/æ/, /a/-/ɜ:/, /a/-/ɛ/, /ɛ/-/æ/, /ɛ/-/ɛ/, /ɛ/-/ɜ:/, /aɪ/-/æ/, /aɪ/-/eɪ/, /aɪ/-/aɪ/, /e/-/ɛ/, /e/-/ɪ/, /e/-/eɪ/, /eɪ/-/eɪ/, /eɪ/-/ɪ/, /eɪ/-/i:/, /i/-/ɪ/, /i/-/i:/, /i/-/aɪ/	/a/-/ɑ:/, /a/-/ʌ/, /a/-/aʊ/, /ɔ/-/ɑ:/, /ɔ/-/ɔ:/, /ɔ/-/ɒ/, /o/-/ɔ:/, /o/-/ɒ/, /o/-/əʊ/, /oʊ/-/ɔ:/, /ou/-/u:/, /oʊ/-/əʊ/, /aʊ/-/ɑ:/, /aʊ/-/aʊ/, /aʊ/-/ʌ/ /u/-/u:/, /u/-/əʊ/, /u/-/ʌ/

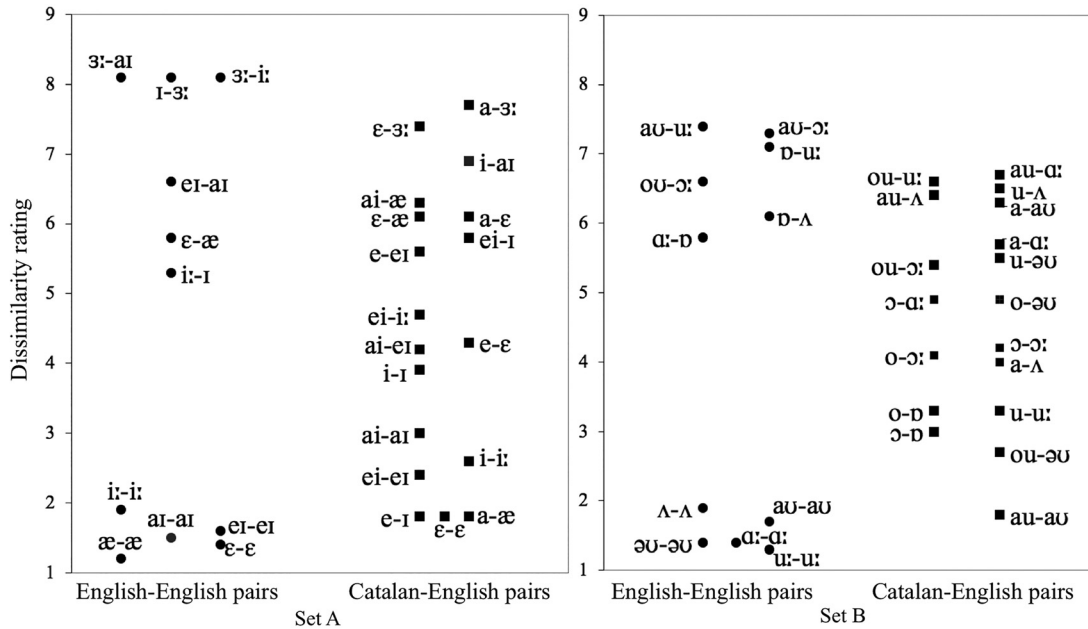


FIG. 2. The mean DRs for English-English (●) and Catalan-English (■) vowel pairs elicited from English L1 speakers. Set A vowels are presented on the left, and set B vowels are presented on the right.

discussed below. A possible explanation may lie in the role that duration may play as a phonetic cue to vowel identity for native English speakers. SSBE /ɔ:/ has a notably longer duration than either C /o/ or /ɔ/, both of which are close in duration to SSBE /ɒ/. English speakers' sensitivity to duration as a cue to vowel identity has been reported in previous studies (e.g., Fox *et al.*, 1995). Thus, despite C /o/'s greater spectral proximity to E /ɔ:/, the mismatch in duration may have rendered it perceptually closer to E /ɒ/ (Cebrian, 2021). The possible predictions for English learners of Catalan are along the lines of what was described above for Catalan learners of English. Some Catalan vowels, particularly those in CE pairs that received DRs within the range of EE same-category pairs, are likely to be readily assimilated to their English counterparts. The higher the DR between a native and a non-native vowel, the more likely it is that differences between them are detected and the non-native vowel is eventually categorized in a more target-like manner. Again, making finer predictions and evaluating them lies beyond the scope of this paper (but see Cebrian, 2021 for some general predictions). This study is concerned with the degree to which the findings for the two language situations are reciprocal. This is analyzed in more detail in Sec. IV.

IV. BIDIRECTIONAL PERCEPTUAL SIMILARITY

A. Comparison of the results of experiments 1 and 2

Experiments 1 and 2 shared the majority of the CE vowel pair stimuli (36 pairs; see Secs. II A and III A for details). Thus, the perceived (dis)similarity between Catalan and English vowels obtained from native speakers of each language can be compared. Table III presents the average DR per vowel and group. As can be observed, despite some isolated differences, the ratings across listener groups are,

overall, very consistent. This was confirmed by the results of a Pearson correlation ($r=0.847$, $N=36$, $p<0.001$) and is illustrated in Fig. 3. Despite the overall consistency, English listeners tended to provide lower DR (mean = 4.7, SD = 1.75) than Catalan listeners (mean = 5.1, SD = 1.57). This difference between the two groups reached significance on a paired samples *t*-test [$t(35) = 2.98$, $p = 0.005$].

TABLE III. The DRs obtained by the Catalan listeners in experiment 1 (Cat L) and English listeners in experiment 2 (Eng L) for each of the 36 Catalan-English vowel pairs common to both experiments (ordered from lowest to highest for the Cat L column).

CE pair	Cat L's DR	Eng L's DR	CE pair	Cat L's DR	Eng L's DR
/ɛ/-/ɛ/ ^a	2.3	1.8	/aɪ/-/æ/	5.5	6.3
/ɛ/-/ɪ/ ^a	2.3	1.8	/o/-/oʊ/	5.5	4.9
/a/-/æ/ ^a	2.4	1.8	/eɪ/-/ɪ/	5.5	5.8
/i/-/i:/ ^b	2.8	2.6	/a/-/aʊ/	5.6	6.3
/aʊ/-/aʊ/ ^b	2.9	1.8	/i/-/ɪ/	5.6	3.9
/a/-/ʌ/ ^b	3.2	4.0	/aɪ/-/eɪ/	5.7	4.2
/ɔ/-/ɒ/	3.3	3.0	/ɔ/-/ɔ:/	5.7	4.2
/o/-/ɔ:/	3.9	4.1	/eɪ/-/i:/	5.9	4.7
/eɪ/-/eɪ/ ^c	4.0	2.4	/aʊ/-/a:/	6.0	6.7
/aɪ/-/aɪ/	4.1	3.0	/a/-/ɔ:/	6.3	7.7
/oʊ/-/oʊ/	4.3	2.7	/ɛ/-/eɪ/	6.4	5.6
/ɔ/-/ɑ:/	4.3	4.9	/ɛ/-/ɜ:/	6.9	7.4
/u/-/u:/	4.5	3.3	/u/-/oʊ/	7.0	5.5
/oʊ/-/ɔ:/	4.7	5.4	/oʊ/-/u:/	7.0	6.6
/o/-/ɒ/	4.9	3.3	/a/-/ɛ/	7.1	6.1
/a/-/ɑ:/	5.0	5.7	/ɛ/-/æ/	7.4	6.1
/aʊ/-/ʌ/	5.2	6.4	/u/-/ʌ/	7.6	6.5
/ɛ/-/ɛ/	5.3	4.3	/i/-/aɪ/	8.0	6.9

^aDR within the range of EE and CC ratings.

^bDR within one SD of the mean for CC pairs.

^cDR within 1 SD of the mean for EE pairs.

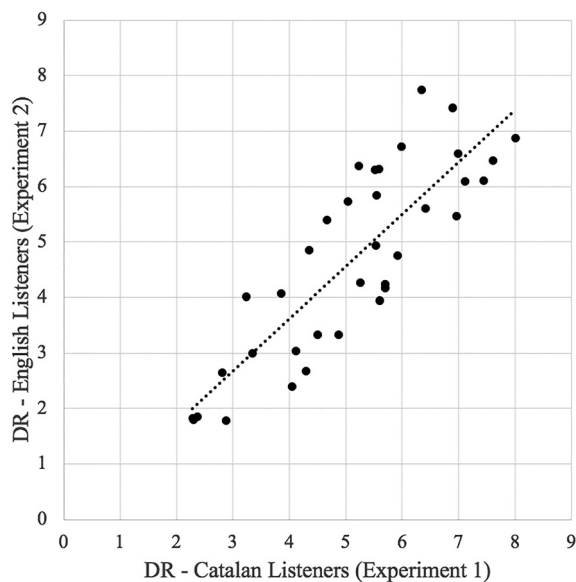


FIG. 3. The scatter plot of the average DRs obtained by the Catalan listeners (experiment 1) and English listeners (experiment 2) on 36 common Catalan-English vowel pairs.

B. Discussion

The results of experiments 1 and 2 show a high degree of consistency, as evidenced by the strong correlation. The five CE pairs that obtained the lowest DRs were the same for both groups, namely, CE /a/-/æ/, /ɛ/-/ɛ/, /e/-/ɪ/, /aʊ/-/əʊ/, and /i/-/i:/, whose values were within the range or within one SD of the mean of same-category L1-L1 pairs. This indicates that some non-native vowels are perceived as representative exemplars of native categories by both populations, pointing to the possibility of shared or near-identical categories across languages. The two groups also tended to agree about the remaining vowel pairs, which were perceived to be gradually more dissimilar. Despite the overall consistency, English listeners tended to provide lower DR. For example, in the case of CE diphthongs (e.g., /ou/-/əʊ/, /aj/-/aɪ/, /ei/-/eɪ/), English speakers' DRs were over one point lower than those of Catalan listeners. CE /ei/-/eɪ/, in fact, received a DR which was as low as same-category EE values. It is possible that greater exposure to within-category variation in English (due to phonetic context or dialectal differences) may make English listeners more accepting of less canonical variants and, thus, provide lower ratings for non-native-native pairs. For example, more L1 variability in off-glide duration (e.g., diphthongs in closed vs open syllables) may explain the lower DR given to CE diphthong combinations by English listeners in general. Further, a less central more rounded beginning for E /əʊ/ (hence, closer to C /ou/) can be found in some L1 phonetic contexts (e.g., preceding /l/) and in some varieties of English. This may render C /ou/ perceptually more native-like to English ears than E /əʊ/ is to Catalan ears (Cebrian, 2021) and, therefore, CE /ou/-/əʊ/ may be judged to be closer by SSBE than by Catalan listeners. Along similar

lines, as discussed above, the fact that Catalan vowels are similar in duration to English lax vowels may render pairs of a lax English vowel and a Catalan vowel (e.g., CE /i/-/ɪ/, /o/-/ɒ/) less dissimilar to English listeners, who have been found to be sensitive to duration as a phonetic cue to vowel identity (e.g., Fox *et al.*, 1995), than to Catalan listeners. The latter, on the other hand, may be focusing on spectral differences only because duration is not a relevant cue in Catalan. These facts may underscore the role of L1 phonetic cue weighting in non-native vowel perception (Cebrian, 2021; Schmidt, 2007). For instance, Schmidt (1996, 2007) reported that Korean and English speakers tended to agree in their perception of similarity between Korean and English consonants with some exceptions. The exceptions involved cases where the two languages differ in the way some phonetic cues are weighted (e.g., variations in the amount of labialization of the following vowel, VOT, or presence of burst-like transitions in nasals). In brief, while the strong correlation of the results of experiments 1 and 2 illustrates that overall cross-language perceptual similarity may be independent of the listeners' L1, there is also evidence that the characteristics of the L1 may influence the degree of perceived similarity. Obtaining perceived similarity judgements from the two populations involved allows for a better understanding of the similarity between two sounds systems and helps evaluate the factors that affect the perception of similarity for each group of speakers.

V. PERCEPTUAL VS ECPHORIC SIMILARITY

A. Comparison of current results with Cebrian (2021)

The participants in this study also participated in a previous study examining ephoric similarity between Catalan and SSBE vowels (Cebrian, 2021). Thus, the DRs obtained in this study can be compared to the patterns of crosslinguistic mappings that were revealed in the previous study. Tables IV (Catalan listeners) and V (English listeners) present for each non-native vowel, the most similar native vowel in terms of the lowest DR in the current study, alongside the percent assimilation to a native category and the corresponding GRs observed in the PAT. The comparison is limited to assimilations above chance level in the PAT study (16.7% for Catalan listeners and 12.5% for English listeners).⁷ In fact, in all of the cases, the lowest DRs obtained in the RDT involve a pair made up of a native vowel and the non-native vowel that received the highest percent assimilation in the PAT (i.e., the modal response). For example, E /i:/ was perceived by Catalan listeners as the most similar to C /i/ in the RDT (with the lowest DR among all of the pairs involving C /i/ and an English vowel, namely, 2.8), and was very consistently assimilated to C /i/ in the PAT (87% of the time with a GR of 5.7 out of 7). Similarly, C /a/ was rated by English listeners as the least dissimilar to E /æ/ (DR = 1.8) and had E /æ/ as the modal response in the PAT (86% assimilation, GR = 5.3). In cases when a non-native vowel was assimilated to more than one native vowel above chance in the PAT, the difference in percent assimilation is

TABLE IV. A comparison of RDT (CE vowel pairs) and PAT (E to C assimilation) results obtained by the same 27 Catalan speakers in the current study and Cebrian (2021). For the latter, “>70%” indicates assimilations above a categorization threshold of 70% (Tyler *et al.*, 2014), and “<70%” and “<50%” indicate assimilations above chance level (16.7%) but below 70%/50%; GR = goodness of fit ratings.

Non-L1 vowel (English)	Native vowel (Catalan)	RDT results DR	PAT results (Cebrian, 2021)	
			Percent assimilation	GR (out of 7)
/ɛ/	/ɛ/	2.3	>70%	>5
/ɪ/	/ɛ/	2.3	>70%	>5
/æ/	/a/	2.4	>70%	>5
/i:/	/i/	2.8	>70%	>5
/aʊ/	/aʊ/	2.9	>70%	>4, <5
/ʌ/	/a/	3.2	>70%	>4, <5
/ɒ/	/ɔ/	3.3	>70%	=5
/ɔ:/	/o/	3.9	>70%	>4, <5
/eɪ/	/eɪ/	4	>70%	>4, <5
	/aɪ/	5.7	<50%	>4, <5
/aɪ/	/aɪ/	4.1	>70%	<4
/əʊ/	/oʊ/	4.3	>70%	<4
/u:/	/u/	4.5	>70%	<4
/ɑ:/	/ɔ/	4.3	<70%	>4, <5
	/a/	5	<50%	<4

TABLE V. A comparison of RDT (CE vowel pairs) and PAT (C to E assimilation) results obtained from the same 27 English speakers in the current study and Cebrian (2021). For the latter, “>70%” indicates assimilations above a categorization threshold of 70% (Tyler *et al.*, 2014), and “<70%” and “<50%” indicate assimilations above chance level (12.5%) but below 70%/50%; GR = goodness of fit ratings.

Non-L1 vowel (Catalan)	Native vowel (English)	RDT results DR	PAT results (Cebrian, 2021)	
			Percent assimilation	GR (out of 7)
/ɛ/	/ɛ/	1.8	>70%	>5
/aʊ/	/aʊ/	1.8	>70%	>4, <5
/a/	/æ/	1.8	>70%	>5
/e/	/ɪ/	1.8	>70%	=5
	/ɛ/	4.3	<50%	<4
/eɪ/	/eɪ/	2.4	>70%	=5
/i/	/i:/	2.6	<70%	>4, <5
	/ɪ/	3.9	<50%	>4, <5
/oʊ/	/əʊ/	2.7	>70%	=5
/aɪ/	/aɪ/	3	>70%	>5
/ɔ/	/ɒ/	3	<70%	=5
	/ɔ:/	4.2	<50%	>4, <5
/u/	/u:/	3.3	=70%	>4, <5
/o/	/ɒ/	3.3	<70%	>4, <5
	/ɔ:/	4.1	<50%	<4

reflected in the difference in DRs in the RDT. For instance, Catalan listeners rated CE /ɑ:/-/ɔ/ as less dissimilar than CE /ɑ:/-/a/ (DRs of 4.3 and 5, respectively); in the same fashion, E /ɑ:/ was assimilated to C /ɔ/ and C /a/ 54% and 31% of the time, respectively, in the PAT. Similarly, English listeners perceived C /ɔ/ to be more similar to E /ɒ/ than to E /ɔ:/ in the RDT (DRs, 3 and 4.2, respectively) and PAT (53% vs 33% assimilation, respectively).

In the case of the English listeners (Table V), the magnitude of the similarity was not always comparable across tasks and vowel pairs. For instance, CE /e/-/ɪ/ was among the five CE pairs that obtained the lowest DR (1.8), yet, its assimilation scores in the PAT were comparatively lower than those in the other four (73% vs 86%–94%), possibly because C /e/ was also perceived to be similar to another L1 vowel (E /ɛ/, 24%). Similarly, CE /i/-/i:/ obtained a relatively low DR (2.6) but the assimilation of C /i/ to E /i:/ did not reach the categorization threshold of 70% as it was split between E /i:/ and /ɪ/ (62% and 37%, respectively). Still, the patterns observed in the PAT are reflected by the DRs obtained in the RDT (CE /e/-/ɪ/, 1.8 vs /e/-/ɛ/, 4.3, and CE /i/-/i:/, 2.6 vs /i/-/ɪ/, 3.9). Therefore, there is a general pattern of consistency across the two tasks. This consistency was confirmed by the results of a series of Spearman’s correlation tests involving the DR obtained for each CE pair in the RDT and the assimilation percentages and GRs yielded for the corresponding crosslinguistic mapping obtained in the PAT (Catalan listeners, DR, and percent assimilation, $\rho = -0.860$, $N = 43$, $p < 0.001$; and DR and GR, $\rho = -0.733$, $N = 43$, $p < 0.001$; English listeners, DR, and percent assimilation, $\rho = -0.940$, $N = 42$, $p < 0.001$; and DR and GR, $\rho = -0.787$, $N = 42$, $p < 0.001$). The lower the DR in the RDT, the higher the assimilation rate and GR in the PAT.

Furthermore, the reciprocity between English and Catalan listeners’ perceptual judgements reported in Sec. IV was also observed in Cebrian (2021). Regarding E /ɜ:/, this vowel was perceived to be highly dissimilar from any L1 Catalan vowel in the current study (/a/-/ɜ:/, 6.3; /e/-/ɜ:/, 6.7; /ɛ/-/ɜ:/, 6.9). This vowel was not tested in Cebrian (2021), but the current result is consistent with the results of the PAT in Cebrian *et al.* (2011), which revealed no clear pattern of assimilation for E /ɜ:/, with responses being split between C /e/ (30%), /ɛ/ (24%), and /a/ (23%) and low GRs (1.4–2 out of 7).

B. Discussion

The comparison of the results of this study and the outcome of Cebrian (2021) shows a close link between the two types of similarities, ephoric and perceptual. This is evident in the strong negative correlations between their respective measures: the higher the degree of assimilation of a non-native vowel to a native category and the higher the GR, the less dissimilar the two vowels are perceived to be. These results are in line with the outcome of Cebrian *et al.* (2011), who also found a high level of agreement between perceptual and ephoric similarity judgements obtained from a group of Catalan learners of English (20 participants completed a PAT, and 47—the previous 20 plus 27 more—completed a RDT). Although the RDT in that study did not include the same amount of vowel combinations as in the current study, the results did show that the English vowels that were perceived to have a more consistent match in the L1 (e.g., E /æ/ to C /a/, E /i:/ to C /i/, E /ɛ/ to C /ɛ/, and E /ɒ/ to C /ɔ/) corresponded to the L2-L1 pairs that obtained the lowest DRs.

Given the high level of cross-task consistency observed, the comparison of the current and previous results sheds little light on the question of which of the two types of methods constitutes a better approach to measuring crosslinguistic similarity, at least on the grounds of a potential difference in their results. L2 speech researchers have, however, argued for one or the other approach. Flege and Bohn (2021) support a perceptual similarity measure arguing that speakers may differ in the nature of their L1 categories and, thus, ecphoric comparisons may not be equivalent across individuals. They add that PATs may also accentuate differences among speakers because they involve two distinct actions (identifying the stimulus in terms of a native category and passing a judgement on how good an exemplar of that category the stimulus is). Flege and Bohn (2021) claim that these potential inter-speaker differences may be better controlled using a single-action task like RDT. Inter-speaker differences are also mentioned by Strange (2007) but as potentially problematic for a perceptual similarity measure. Strange points out that in a RDT, listeners directly compare non-native categories to L1 categories produced by L1 speakers other than themselves, and these L1 stimuli may differ from the listeners' own L1 categories. Our results do not show a greater likelihood of inter-speaker variability in one or the other task. As reported above, there was a very high degree of inter-rater agreement in experiments 1 and 2 (Cronbach's alpha values of 0.990 and 0.992, respectively). This was, in fact, also true of the PAT results in Cebrian (2021) with inter-rater reliability coefficients for assimilation percentages and GRs ranging from 0.920 to 0.956. This high degree of inter-rater agreement lends little support to a potential disadvantage of PATs (or RDTs) due to a possible greater inter-speaker variability.

Tyler (2021), on the other hand, argues that a PAT allows researchers to assess precisely how consistent perceptual mappings are across speakers and stimuli. In addition, Tyler favours a PAT approach as it provides more information about crosslinguistic perception than a RDT does (e.g., whether non-native sounds are perceived as speech or not or if they are consistently mapped onto a single native category or more). Moreover, PATs allow listeners to make their own choices about what non-native phones are closer to native phones. By contrast, in a RDT, it is the researcher who provides the listeners with the paired stimuli; thus, to represent all possible "choices," the experiment needs to include the whole list of possible combinations of stimuli. Strange (2007) refers to this as a disadvantage of RDTs, highlighting the prohibitive number of trials that a RDT would require to consider all of the possible combinations. Still, a practical approach is the one followed in this paper, where all native and non-native vowels were equally represented in the RDT while only including a subset of all of the possible vowel combinations. The selection of combinations can be based on the results of exploratory investigations using a PAT approach (as was performed in this study).

In brief, despite the advantages and disadvantages of each of the methods, the combined results of this and our

previous study indicate that both approaches are highly comparable and strongly related. However, it remains to be examined which type makes better predictions for L2 learning. Previous studies have shown that PATs and RDTs can be used to make predictions about L2/non-native sound categorization (e.g., Fox *et al.*, 1995; Strange *et al.*, 2011; Tyler *et al.*, 2014). However, to our knowledge, hardly any previous studies have actually contrasted the predictions of both methods. One such study is Bohn and Garibaldi (2021), who examined which of three methods (ecphoric similarity using a PAT, perceptual similarity from a RDT, or an acoustic comparison) best predicted non-native discrimination. Ten SSBE native speakers performed a PAT and RDT involving four SSBE and four Danish front unrounded vowels and were tested on their ability to discriminate the Danish vowels. The results indicated that ecphoric and perceptual similarity made different predictions for some vowels and none of the three methods fully predicted accuracy of non-native vowel discrimination. Still, that study was limited to four native and non-native vowels. Perhaps, a better approximation to the relationship between the two measures could be obtained if a greater representation of the native and non-native systems is assessed.

VI. GENERAL DISCUSSION AND CONCLUSIONS

This study has presented the results of two experiments specifically designed to assess the perceptual similarity (as opposed to ecphoric similarity) between Catalan and SSBE vowels elicited from two parallel populations (SSBE and Catalan native speakers) that performed a RDT. This task illustrates a direct overt method (Bohn, 2002) in which listeners are presented with a pair of sounds involving a native and non-native stimulus (overt method) and specifically asked (direct method) to assess the similarity between the two. Crucially, the RDT also included pairs of same-category and different-category native stimuli for baseline purposes. The results of the two experiments show that, generally, every non-native vowel has been found to be perceived as most similar to a given native vowel. One exception is E /ɜ:/, which seems to be dissimilar from all Catalan vowels. The other exceptions are pairs of non-native vowels perceived as most similar to the same native vowel, namely, E /æ/ and /ʌ/ as C /a/, E /ɑ:/ and /ɒ/ as C /ɔ/, and C /o/ and /ɔ/ as E /ɒ/. In the first two cases, this is related to the smaller size of the Catalan inventory and consequent unavailability of L1 categories to assimilate all of the non-native vowels to. By contrast, C /o/ and /ɔ/ were found to be most similar to E /ɒ/ despite the availability of additional L1 categories in the inventory, for instance, E /ɔ:/.

As discussed above, the role of duration as a phonetic cue to the identity of E /ɔ:/ may have rendered C /o/ as a poor match despite being spectrally closer (Cebrian, 2021). This shows that specific characteristics of the native language determine the way perceptual similarity relations are established. Finally, the results also show that a number of non-native vowels are perceived to be as similar to native

categories as two exemplars from the same L1 category, pointing to the possibility of near-identical or shared categories between the two languages.

The results of this study on perceptual similarity have been compared to earlier results on ephoric similarity (indirect covert method) involving the same stimuli and participants (Cebrian, 2021). This constitutes the first direct comparison of ephoric and perceptual similarity between a large number of native and non-native stimuli. The comparison indicates a very close relationship between the two types of measures: for both experiments, the RDT results (DRs) were highly correlated with the PAT measures (percent assimilation, GRs) obtained by Cebrian (2021). This suggests that individuals' judgements of perceptual similarity and ephoric similarity were highly comparable, at least regarding the similarity between native and non-native vowels. The use of the two tasks may help to obtain a more reliable measure of crosslinguistic similarity. For instance, researchers can design a RDT that includes all of the theoretically relevant native–non-native combinations based on the modal responses previously obtained in a PAT, thus, avoiding the need to include a prohibitive number of possible native–non-native vowel combinations. Further, it is possible that each type of task can serve a different purpose. RDTs may be more suitable in cases where the use of category labels should be discouraged to avoid problems of inconsistency and interpretability of responses (Flege and Bohn, 2021; Strange, 2007), e.g., if listeners' literacy or level of L2 proficiency is a concern. On a similar note, Cebrian *et al.* (2011) argued that PATs may be a good method to evaluate the perceived similarity between native and non-native sounds at a given point in time (e.g., naive vs inexperienced L2 learners) while RDT can be suitable for exploring changes in perceived similarity over time. This issue is left for future research.

This study is not without limitations. First, the RDTs in this study did not include all of the possible combinations of native and non-native vowels that would ideally be included to assess the full range of similarity relationships between native and non-native vowels (Strange, 2007). Still, the selection of vowel pairs was guided by the results of previous research using a PAT, allowing the inclusion of all of the crucial pairs. In addition, the phonetic context in which vowels were presented was limited to /bVt/. A more complete assessment of crosslinguistic perceptual similarity would need to include a greater number of contexts, given that consonantal context may affect vowel perception (e.g., Strange *et al.*, 2001). Finally, this study has focused on a specific variety of each language (Eastern Catalan and SSBE). The results may not be applicable if other varieties of English and Catalan are considered, something which would be worth pursuing.

To conclude, the use of a direct method of assessing perceptual similarity has been found to be a suitable approach to measuring crosslinguistic similarity. The inclusion of native-native pairs in the RDT, particularly same-category native-native pairs, makes it possible to assess the

degree to which high levels of perceptual similarity between native and non-native vowels may be an indication of near-identity or shared categories across the two languages. Further, the general reciprocity of the results of the two experiments confirms the language-independent nature and the bidirectionality of perceived similarity. However, the characteristics of each language, such as the degree of within-category variability or weighting of specific acoustic cues (e.g., vowel duration) also influence crosslinguistic perception of similarity and may account for differences between the two populations in the perceptual similarity of the same pairs of sounds. Future studies will need to evaluate the predictions that can be derived from these measures of perceptual and ephoric similarity for L2 speech learning, e.g., in terms of the ease of discrimination of non-native/L2 pairs or production of L2 sounds found to be perceptually very close or near-identical to native sounds (see Cebrian, 2021, for some predictions for Catalan learners of English and English learners of Catalan). Still, given the relevance of the notion of similarity for a number of disciplines and L2 speech research in particular, this paper hopes to have contributed to a better understanding of the nature of (crosslinguistic) perceived similarity, underscoring the importance of obtaining a reliable, yet practical method of assessment.

ACKNOWLEDGMENTS

This work was supported by Research Grant Nos. FFI2017-88016-P and PID2021-122396NB-I00 from the Spanish Ministries of Economy and Competitiveness and Science and Innovation and Research Group Grant No. 2017SGR34 from the Catalan Government.

¹Although it seems incoherent to refer to a task that measures “ephoric” similarity as a PAT—it should probably be renamed “ephoric assimilation task”—the term PAT will continue to be used here for the sake of comparison with previous studies that use this term. Thus, “perceptual” in PAT has a general meaning, i.e., “perceived.”

²The Catalan vowel inventory includes a reduced vowel ([ə]). This vowel was not included as in most varieties, including Eastern Catalan, it only appears in unstressed position and is subject to large dialectal variability (Recasens, 1993).

³The English vowel /ʊ/ was excluded as previous studies (Cebrian, 2009; Rallo Fabra, 2005) have shown that this vowel is perceived to be most dissimilar and not consistently assimilated to any single Catalan vowel, and often poorly identified by native English speakers [77% accuracy by native Canadian English speakers in Cebrian (2009) as opposed to 88%–100% for the rest of the vowels].

⁴Median values were also calculated as some studies report median rather than mean values for rating data (e.g., Strange *et al.*, 2001). The two measures did not differ notably and thus mean values are reported following other studies (e.g., Faris *et al.*, 2016).

⁵The DR scale used in Cebrian *et al.* (2011) was a seven-point scale. The actual scores were converted to a nine-point scale to facilitate the comparison of the two studies.

⁶Following the conventions for each language, the phonetic symbol /ɔ/ in Catalan represents the lowest of the two mid back rounded vowels, whereas the symbol /ɔ:/ in SSBE designates a vowel that is higher than /ɔ/.

⁷The chance levels of 16.7% (Table IV) and 12.5% (Table V) result from dividing 100% by the number of response options present per PAT (e.g., Faris *et al.*, 2016), that is, six for the Catalan listeners and eight for the English listeners [see Cebrian (2021) for details].

- 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 (Benjamins, Amsterdam), pp. 13–34.
- Boersma, P., and Weenink, D. (2018). "Praat: Doing phonetics by computer (version 6.0.43) [computer program]," available at <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, Trier, Germany), pp. 191–216.
- Bohn, O.-S. (2017). "Cross-language and second language speech perception," in *The Handbook of Psycholinguistics*, edited by E. M. Fernández and H. Smith Cairns (Wiley-Blackwell, Chichester, UK), pp. 213–239.
- Bohn, O. S., and Garibaldi, C. L. (2021). "Examining the success of three kinds of cross-language similarity in predicting English listeners' discrimination of Danish vowels," *J. Acoust. Soc. Am.* **150**(4), A44.
- Cebrian, J. (2009). "Effects of native language and amount of experience on crosslinguistic perception," *J. Acoust. Soc. Am.* **125**(4), 2775.
- Cebrian, J. (2011). "Cross-linguistic perception of Catalan and English diphthongs," in *Achievements and Perspectives in the Acquisition of Second Language Speech: New Sounds 2010*, edited by M. Wrembel, M. Kul, and K. Dziubalska-Kolaczyk (Lang, Frankfurt am Main, Germany), Vol. II, pp. 43–51.
- Cebrian, J. (2019). "Perceptual assimilation of British English vowels to Spanish monophthongs and diphthongs," *J. Acoust. Soc. Am.* **145**(1), EL52–EL58.
- Cebrian, J. (2021). "Perception of English and Catalan vowels by English and Catalan listeners: A study of reciprocal cross-linguistic similarity," *J. Acoust. Soc. Am.* **149**(4), 2671–2685.
- 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 (Lang, Frankfurt am Main, Germany), Vol. I, pp. 41–52.
- Chang, C. B. (2019). "The phonetics of second language learning and bilingualism," in *The Routledge Handbook of Phonetics*, edited by W. F. Katz and P. F. Assmann (Routledge, Abingdon, UK), pp. 427–447.
- Cortina, J. M. (1993). "What is coefficient alpha? An examination of theory and applications," *J. Appl. Psychol.* **78**, 98–104.
- Cruttenden, A. (2014). *Gimson's Pronunciation of English*, 8th ed. (Routledge, New York).
- Docherty, G. J. (1992). *The Timing of Voicing in British English Obstruents* (de Gruyter, Berlin, Germany).
- Elvin, J., and Escudero, P. (2019). "Cross-linguistic influence in second language speech: Implications for learning and teaching," in *Cross-Linguistic Influence: From Empirical Evidence to Classroom Practice*, edited by M. J. Gutierrez-Mangado, M. Martinez-Adrian, and F. Gallardo-del-Puerto (Springer, Cham, Switzerland), pp. 1–20.
- 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. T., and Tyler, M. D. (2016). "An examination of the different ways that non-native phones may be perceptually assimilated as uncategorized," *J. Acoust. Soc. Am.* **139**, EL1–EL5.
- Flege, 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.
- Flege, 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 Press, Cambridge, UK), pp. 3–83.
- Flege, 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.
- Fox, R. A., Flege, J. E., and Munro, M. J. (1995). "The perception of English and Spanish vowels by native English and Spanish listeners: A multidimensional scaling analysis," *J. Acoust. Soc. Am.* **97**(4), 2540–2551.
- Holliday, J. J. (2015). "A longitudinal study of the second language acquisition of a three-way stop contrast," *J. Phonetics* **50**, 1–14.
- IBM Corp. (2021). *IBM SPSS Statistics for Macintosh* (IBM Corp., Armonk, NY), version 28.
- Kuhl, P. K., Conboy, B. T., Coffey-Corina, S., Padden, D., Rivera-Gaxiola, M., and Nelson, T. (2008). "Phonetic learning as a pathway to language: New data and native language magnet theory expanded (NLM-e)," *Philos. Trans. R Soc. B* **363**, 979–1000.
- Ladefoged, P., and Ferrari Disner, S. (2012). *Vowels and Consonants*, 3rd ed. (Wiley-Blackwell, Chichester, UK).
- Ladefoged, P., and Johnson, K. (2014). *A Course in Phonetics*, 6th ed. (Wadsworth, Boston, MA).
- Pisoni, D. (1973). "Auditory and phonetic memory codes in the discrimination of consonants and vowels," *Percept. Psychophys.* **13**, 253–260.
- Rallo Fabra, L. (2005). "Predicting ease of acquisition of L2 speech sounds: A perceived dissimilarity test," *Vigo Int. J. Appl. Linguist.* **2**, 75–92, available at <https://revistas.uvigo.es/index.php/vial/article/view/10/10>.
- Rallo Fabra, L., and Romero, J. (2012). "Native Catalan learners' perception and production of English vowels," *J. Phonetics* **40**(3), 491–508.
- Recasens, D. (1993). *Fonètica i Fonologia (Phonetics and Phonology)* (Enciclopèdia Catalana, Barcelona, Spain).
- 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 Flege*, edited by M. Munro and O.-S. Bohn (Benjamins, Amsterdam, the Netherlands), pp. 185–200.
- Strange, W. (1987). "Information for vowels in formant transitions," *J. Mem. Lang.* **26**, 550–557.
- Strange, W. (2007). "Cross-language similarity of vowels," in *Language Experience in Second Language Speech Learning: In Honor of James Flege*, edited by M. Munro and O.-S. Bohn (Benjamins, Amsterdam, the Netherlands), 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., Akahane-Yamada, R., Kubo, R., Trent, S., Nishi, K., and Jenkins, J. (1998). "Perceptual assimilation of American English vowels by Japanese listeners," *J. Phonetics* **26**, 311–344.
- 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., Hisagi, M., Akahane-Yamada, R., and Kubo, R. (2011). "Cross-language perceptual similarity predicts categorical discrimination of American vowels by naïve Japanese listeners," *J. Acoust. Soc. Am.* **130**(4), EL226–EL231.
- Thomson, R. I., Neary, M. N., and Derwing, T. M. (2009). "A modified statistical pattern recognition approach to measuring the crosslinguistic similarity of Mandarin and English vowels," *J. Acoust. Soc. Am.* **126**, 1447–1460.
- Tulving, E. (1981). "Similarity relations in recognition," *J. Verbal Learn. Verbal Behav.* **20**, 479–496.
- 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 Press, 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.