On segmental representations in second language phonology: A perceptual account

Fernanda Barrientos
University of Konstanz, Germany

Abstract
The extent to which exposure to new phonemic contrasts (i.e. contrasts that are present in the L2 but not in the L1) will lead to the creation of a new phonemic category in L2 speakers, as well as the phonological nature of these categories, remains an open question insofar as there is no consensus on whether acquiring a new contrast would result in abstract, phoneme-like categories, or if they belong to a less abstract level of representation. This work explores the perception of the /ɑ/–/ʌ/ contrast (cop – cup) in American English by Spanish speakers of L2 English through a discrimination task. The results show that while the interlanguage state of less experienced learners is best described as a case of single-category assimilation, the interlanguage state achieved by advanced learners is not a full phonemic split, despite the increased sensitivity to otherwise within-category perceptual cues; rather, it seems that while the ability to perceive differences is not affected, the ability to create a new phonemic representation is impaired.

Keywords
L2 phonology, representations, speech perception

I Introduction
There is no question that the native language (L1) has an influence over the way in which second language (L2) learners perceive the speech sounds of the target language (TL), in particular those who learned it at a later age. In this respect, a large body of research has attempted to determine the extent of the influence of the L1 phonological system on learning the sound system of the L2. Some concerns that this line of research has addressed include the effects that the sound system of a specific L1 has over the learning of the L2 (Flege, 1987; Iverson and Evans, 2007), the role of linguistic input across the lifespan (Flege, 1995; Iverson and Kuhl, 1995), the relationship between the L1 and Universal Grammar (UG) mechanisms when learning the sound inventory of a new language (Brown, 1998), the different types of learning situations when it comes to
acquisition of new phonemic contrasts (Eckman et al., 2003), and how learners develop the ability to change their perception of the acoustic cues when learning the L2 sound system (Escudero, 2005; Escudero and Boersma, 2003; Kim et al., 2018), just to name a few.

Likewise, and despite the interference of the L1 in learning the sound system of an L2, it has also been shown that late L2 learners are able to learn new phonological and specifically perceptual contrasts to a considerable extent (Bohn and Flege, 1997; Flege et al., 1997, 1999). In these studies, it has been found that L2 speakers can learn how to perceive nonnative contrasts after a relatively long period of exposure to the TL.

But how deep is that learning? Can it be comparable to the robustness of the perceptual mechanisms involved in L1 perception? Is the result of this learning process a long-term, phonemic representation, or is it a surface-like structure that can be perceived by L2 speakers only under very specific conditions? In fact, the terms used in the literature vary greatly and include abstract as well as surface-level terms. Hence, the question is whether the phonological nature of L2 representations that correspond to new, but similar sounds in the L2 are abstract enough to allow for the retrieval of different lexical items; that is, whether minimal pairs can be formed so that they can be robustly recognized in perception, and accurately articulated in production. In the alternative situation, speakers/listeners perceive the different sounds in a minimal pair as two instances of the same sound, unless confronted with a task that triggers a non-phonological mode of perceiving speech sounds.

This research deals with the perception of the /ɑː/~/ʌ/ contrast (cop – cup) in American English by Spanish speakers of L2 English. Since the Spanish inventory has only five monophthong vowels /i e a o u/, L2 Spanish speakers of English (at least at an early stage of their L2 learning process) are likely to perceive both /ɑː/ and /ʌ/ as members of the same category (Escudero and Chládková, 2010; García, 2014). The acquisition of this contrast poses an additional difficulty, since according to the Markedness Differential Hypothesis (Eckman, 1977), L2 contrasts are predicted to be more difficult to acquire when they are marked; in this regard, Flemming (2004: 233–234) suggests that vowel systems with back rounded and unrounded vowels are perceptually marked. Likewise, and following the predictions of the Perceptual Assimilation Model (PAM; Best, 1995) and PAM-L2 (Best and Tyler, 2007), with such a small number of L1 categories the options for nonnative speakers are: (1) to conflate both sounds as /a/ (i.e. single-category assimilation), (2) to map /a/ onto the Spanish vowel /a/, and /ʌ/ to the Spanish category /o/ (two-category assimilation); or (3) to perceive both /a/ and /ʌ/ as /a/, but one being a better exemplar of the L1 category than the other (category-goodness assimilation). A further question is whether these mappings change across the lifespan of an L2 learner, or if they are kept until very advanced learning stages. And finally, this research is also concerned with the nature of the perceptual categories that late learners of L2 English have developed over time until reaching a state of high proficiency in the TL.

1 Surface and abstract representations in L2

L1 effects in the L2 at the phonetics/phonology level, and more specifically in perception, have been acknowledged by phonological theory at least since the days of the
Prague School. Trubetzkoy (1969) [1939] conceptualized the sounds of the L1 as a “sieve” through which the L2 sounds are perceived and produced; likewise, Polivanov (1931: 231) referred to the difficulties in the perception and production of L2 sounds as “language habits attained by every given individual in the process of mastering his mother (native) tongue”. The question that follows is whether these “language habits” can be unlearned, or rather, whether L2 learners can create a narrower mesh for their sieves; and ultimately, how this learning process can be described by phonological theory in terms of the possible levels of segmental representation.

In this regard, different approaches to phonology have recognized two main levels of representation: an abstract one (phonemic), and a concrete one (a phonetic, or surface level). However, research in second language phonology does not seem to have reached an explicit agreement over the exact type of representation that late learners of a given L2 create throughout their learning process, or at least after they reach a high level of proficiency in the TL. The terminology used by researchers in the field include higher-level representations such as “phoneme” (e.g. Hayes-Harb, 2005; Winkler et al., 1999) and “phonological/phonemic category” (Antoniou et al., 2012), surface-level concepts such as “phonetic category” (Flege, 1995; Flege and Bohn, 2021), and less specific terms such as “perceptual category” (Escudero, 2005; Escudero and Boersma, 2004).

Such different conceptualizations raise the question about the nature of these newly created representations at the segment level: are they phonemic or not? It is worth noting here that the concept of phoneme, despite being a fundamental notion in phonology, has been defined in different ways throughout time, with views that acknowledge it as a physical phenomenon, a psychological reality, or simply fiction (Dresher, 2011). In the present study, the concept of phoneme will be understood in the classic structuralist way, that is, as an abstract representation of a segmental speech sound with psychological reality (Cole and Hualde, 2011).

While it could be argued that the phoneme as a unit of representation is an outdated notion in phonological theory, the concept is particularly relevant when addressing perception Sapir (1963) [1949] conceptualized the phoneme mostly as a psychological unit, since listeners (specifically native speakers) are unaware of the detail of phonetic realizations (Dresher, 2011; Martin and Peperkamp, 2011). When explaining the phonemic principle, Swadesh (1934: 118) stresses the perceptual nature of the phoneme, stating that “if they [the speakers of a given language] hear a foreign tongue spoken, they still tend to hear in terms of their native phonemes.”

The above motivates the hypothesis that learning in the phonetics/phonology domain in L2 at a later age does not lead to creation of new phoneme-like representations; rather, this learning belongs to the domain of phonetics. However, this is not to say that phonetic/phonological learning in L2 simply cannot take place after a certain age; instead, the question is whether these new representations are comparable to those of the L1 at the more abstract level, and more particularly in cases when the L2 learner deals with sounds in the target language (TL) that are non-contrastive in the L1. In this regard, the notion of interlanguage (Selinker, 1972) becomes crucial as it posits that the representations that L2 speakers create are a self-contained system with characteristics that are neither target-like nor a product of L1 transfer (Adjemian, 1976), both in terms of performance and competence (Lakshmanan and Selinker, 2001). While the concept has been
applied mostly to the domains of morphology and syntax, it has also been used to describe phenomena in the phonetics/phonology domain (Hancin-Bhatt, 2008; Major, 1998; Tarone, 1978)

This research intends to show that perceptual representations in L2 in cases where two L2 sounds are perceptually assimilated to one L1 category are better described by the concept of interlanguage, which in phonological terms can be interpreted as an in-between state where sounds that ought to be perceived as contrastive may be perceived as different, but not in a phoneme-like way. In empirical terms, we will assume that a robust, phoneme-like representation will produce clear-cut boundaries when perceiving continua between two target sounds and reach perfect sensitivity at the endpoints in a discrimination test designed to tap phonological information. This is the pattern expected to be seen in native speakers and will be the benchmark for nonnative speakers. In real-life terms, in this interlanguage state L2 listeners seem to be aware of acoustic differences in L2 contrasts, but the representations that they create are not robust enough to become phonemic categories that are independent from L1 sounds.

2 Models of speech perception in L2

Several models and frameworks for speech perception in L2 have been proposed, of which the most prominent ones are: The Speech Learning Model (SLM; Flege, 1995), the Revised Speech Learning Model (SLM-r; Flege and Bohn, 2021), the Perceptual Assimilation Model (PAM; Best, 1995), PAM-L2 (Best and Tyler, 2007), and the L2 Linguistic Perception model (L2LP model; Escudero, 2005). All of these models have dealt from different perspectives with the issue of learning a new phonemic contrast in L2. Of these models, SLM-r and PAM-L2 make specific claims regarding the nature of L2 representations, i.e. whether these are of a phonemic nature, or if they belong to a more surface level of representation, i.e. the phonetic domain. Likewise, the L2LP model also sheds some light on this matter, but is still not conclusive regarding a theory of interlanguage representations at the sound level of linguistic knowledge.

SLM-r states that their focus is on “how sequential bilinguals produce and perceive position-sensitive allophones of L2 vowels and consonants” (Flege and Bohn, 2021: 91), thus assuming that L2 learning occurs at the phonetic level. However, the reason why a phonemic level of representation is not taken into consideration is not entirely clear, since it also claims that the mechanisms for L2 learning are the same as those used in learning the L1.

While the PAM in its first version deals only with nonnative perception in naïve listeners (i.e. neither production nor advanced learning states in L2 speakers/listeners), it makes a series of predictions regarding the way in which the nonnative sounds will be perceptually assimilated to those of the L1, of which three are relevant for this research: single-category assimilation (henceforth SCA), where two L2 sound categories are perceived as one L1 category; two-category assimilation (henceforth TCA), where two L2 sound categories are perceived as two different L1 categories; and category-goodness assimilation (henceforth CGA), where one L2 category is perceived as a better exemplar of an L1 category than the other. We will use this terminology throughout this article in order to signal the L2-to-L1 relationships between sounds.
A later version of this model that extends to L2 learning, PAM-L2, assumes that “Mental representations of phonetic categories are not required for L2 perceptual learning” (Best and Tyler, 2007: 25); however, it does recognize two levels of representations: phonological categories, where “the linguistic-functional relationships within the sound system of a language are determined” and phonetic categories, which “refer to invariant gestural relationships that are sub-lexical yet still systematic and potentially perceptible to attuned listeners”. PAM-L2 acknowledges that both levels are involved in L2 speech perception, and assumes that SCa may evolve into learning a phonological contrast, provided that two conditions are met: (1) that the L2 learner moves onto a state of CGa, which could result in the creation of a new category for the less prototypical sound; and (2) that the minimally contrasting words are high frequency, and/or belong to dense phonological neighborhoods (2007: 30).

L2LP predicts that L2 learners create categories by noticing the differences between the underlying representations (URs) created during the initial state, and the actual acoustic properties of the input. The model operates under the premise that L2 speakers can learn new categories in the same way as L1 learners, provided that they receive the type, amount, and quality of input that would be expected in L1 acquisition (Escudero, 2005: 118). Crucially, the model distinguishes between two different types of learning tasks: a perceptual task (i.e. creating new mappings of the acoustic signal onto existing representations) and a representational task, that is, the creation of new perceptual categories (Escudero, 2005: 107). However, such division of tasks might allow for learning outcomes in L2 speakers wherein only one of the two tasks is performed, which might explain the nonnative-like behavior observed in participants with high proficiency when performing perceptual tasks.

In sum, while these models are by no means the only ones that deal with the acquisition of L2 sounds, they touch upon the problem of acquiring new perceptual contrasts and the best way to describe the cognitive state of the L2 learner. The following section will focus on how the L2LP model and PAM/PAM-L2 can offer further insight on the possible learning outcomes when acquiring a new contrast.

3 Different levels of representation

As mentioned above, L2LP further elaborates on the nature of perceptual learning by drawing a distinction between perceptual and representational tasks. Furthermore, the type of learning task will be defined by the L1 and L2 inventories: For instance, if both the L1 and L2 have the same elements in their corresponding inventories but different phonetic realizations that lead to different perceptual boundaries between these two elements, then only a perceptual task is needed: The learner needs to shift the perceptual boundary given by the L1 so that it matches that of the L2 (Escudero, 2005: 87). On the other hand, if the L2 has sounds that are not part of the L1 inventory, then the learner will need to additionally create a new abstract representation for the new sound; in other words, this learner will need to both create a new representation and learn which specific acoustic values should be mapped to this representation.

However, it is also possible that the acquisition process does not necessarily reach a final state that matches the competence of a native speaker in the TL. In this sense, the
learner might only perform one out of the two perceptual tasks, both, or neither. Hence, and in the case of perception of the /ɑ/–/ʌ/ contrast in English by L1 Spanish speakers, four possible outcomes can be outlined on the basis of achievement of these tasks (note that [–perception, +representation] is not a possible outcome, as it would create an empty perceptual category without any target acoustic values).3

- [–perception, –representation]: The L2 learner maps the relevant sounds (both /ɑ/ and /ʌ/) onto one L1 category, namely /a/. In PAM/PAM-L2 terms, this would entail SCa.
- [+perception, –representation]: The L2 learner maps the relevant sounds onto two different L1 representations (namely, /a/ and /o/), either from the beginning or after an initial SCa state. In PAM/PAM-L2 terms, this would result in TCa.
- [+perception, –representation]: The L2 learner recognizes the relevant sounds as subsets of one L1 category. However, this recognition can be retrieved only in certain specific listening tasks, such as discrimination (but not categorization). In this specific case, the categories /a/ and /ʌ/ are perceived as different, yet are still considered instances of the same phonemic representation /a/. Arguably, this might constitute a case of CGa in PAM/PAM-L2, provided that one of the L2 categories is perceived as a better exemplar of /a/.
- A full category split: [+perception, +representation]. The L2 learner creates two new perceptual categories for the relevant sounds, or at least one new perceptual category for the sound that is less likely to be recognized as a prototype of any L1 sound.

The first outcome, whereby the mappings of the initial state are kept throughout time, can be considered a type of fossilization, a term defined by Selinker (1972: 215) as “linguistic items, rules, and subsystems which speakers of a particular NL [native language] will tend to keep in their IL [interlanguage] relative to a particular TL [target language], no matter what the age of the learner or amount of explanation and instruction he receives in the TL.”4 The second and third outcomes, while not being states of native-like acquisition, also involve some degree of fossilization. Finally, the fourth outcome is the only one that entails complete acquisition.

While the first two outcomes involve only L1 category reuse, the fourth option assumes that an L2 speaker has managed to make the sound representations of her TL evolve into new phonemic-like categories. On the other hand, the third option shows an evolution from the L1-like representations towards a more target-like representation; however, the full category split is not achieved and thus both /a/ and /ʌ/ would be perceived as “two kinds of /a/”. This is a less deep level of representation than a phoneme, as it assumes that these sounds would not allow L2 learners to identify a minimal pair. In this matter, Altmann and Kabak (2011: 299) mention that the acquisition of contrast in L2 encompasses not only production and perception of an opposition but also an operation whereby the learner encodes this opposition in an abstract linguistic representation; hence, none of the previous three outcomes would involve the acquisition of a contrast (see Table 1).

Thus, the question is whether learners of L2 English at a very advanced stage of their acquisition process are able to perform both the perceptual and the representational task
thus reaching a full category split of /a/ into /ɑ/ and /ʌ/; if they rearrange their perceptual grammar so that what was initially perceived as members of the same category ends up being mapped onto two different ones (akin to TCa); if the perception of differences is possible but does not result in two different categories; or – worst case scenario – if everything remains the same and both English categories are still perceived as /a/.

II The current study

Spanish has a very simple vowel inventory, compared to English: While Spanish has only 5 monophthong vowels in all of its dialects (Hualde, 2005; Nadeu, 2014; Quilis, 1993),5 English shows a much more complex system that can at least double the number of monophthong phonemes: Bradlow (1995) considers 11 monophthongs in American English in dialects with no cot/caught merger. Thus, the acquisition of the /a/–/ʌ/ contrast would entail the creation of at least one vowel category (namely /ʌ/) which, when looking at the acoustic/perceptual space, will overlap with /a/ (Figure 1).

Here an important question arises: If /a/ overlaps with /ʌ/ in terms of acoustic values, then why do we claim that /ʌ/ will be the new category? Regarding previous findings,
Escudero and Chládková (2010) showed that native speakers of Spanish perceptually assimilate American English /ʌ/ to Spanish /a/, but also to Spanish /o/ to a similar extent (53% vs. 47%, respectively); on the other hand, /a/ was assimilated almost unequivocally to /a/ (99%). Thus it can be claimed that for Spanish listeners, /ʌ/ is a less prototypical exemplar of /a/. A similar behavior was found in Barrientos (2018) where native speakers of Spanish categorized /a/ as /a/ 91% of the times, whereas /ʌ/ was categorized as /a/ 69% of the times. Furthermore, in García (2014), the accuracy in identification is higher in /a/ than /ʌ/, thus suggesting that /ʌ/ is somewhat “problematic” (i.e. it does not map easily onto any L1 or L2 category).

Furthermore, from a featural perspective, /ʌ/ is described as [–tense] (a nonexistent feature in Spanish), but also its features are bundled in a way that cannot be found in the Spanish vowel system, namely, [–high, –low, –round, +back]. From this viewpoint, the question is whether Spanish speakers of L2 English can create new bundles of features, either by choosing a minimal solution whereby the existing features are recombined, or by adding an extra feature to their set.6

This study aims to address three questions. The first one is whether the early learning state in the perception of English /a/ and /ʌ/ by Spanish speakers corresponds to what PAM/PAM-L2 refers to as SCa; that is, if both /a/ and /ʌ/ are perceptually mapped onto the Spanish category /a/. Previous research has shown that this is the most likely scenario for native speakers of Spanish (Escudero and Chládková, 2010; García, 2014) and Catalan as well, which has a similar vowel system (Rallo Fabra and Romero, 2012; Safronova, 2016). While this is not the central research question, we consider important to make sure that, should there be an effect where Spanish learners of English do seem to be perceiving differences, it is not due to the fact that they are perceptually assimilating these sounds to two different L1 categories. The second question is whether this initial state changes as L2 speakers gain more experience in the TL; that is, whether learners at later stages of L2 acquisition are more likely to perceive differences between /a/ and /ʌ/ than learners who are at an earlier stage. Finally, the third question is how the end result (i.e. the segmental representations) can be described in phonological terms; that is, whether the perception of the nonnative contrast displays the same robust patterns as those found when perceiving native categories (thus being comparable to a phonemic representation) or not, and if not, then what the best way to describe such representations would be.

**III Empirical approach**

How can the nature of a given representation be determined from an empirical approach? According to Strange (2011), listeners may engage in two different modes of perception: phonological (which is mostly used in the L1) and phonetic. The phonological mode “enables the listener to detect sufficient phonologically relevant contrastive information for word-form identification” (Strange, 2011: 460). This mode of perception is over-learned, automatic, and allows for quick, robust lexical access through specific perceptual attunement to the relevant auditory cues. On the other hand, the phonetic mode of perception (where listeners can perceive acoustic detail that the phonological mode does not take into account) is less frequent and is more cognitively demanding than phonological mode
insofar as it requires a certain degree of “undoing” the perceptual routines that lead to optimal perception in the L1. These modes of perception can be elicited through different perceptual tasks. On the one hand, categorization tasks, where participants listen to a single stimulus and must answer the question “what sound did you hear?” tap phonological information (Boersma, 2009; Colantoni et al., 2015), insofar as their main goal is to make speakers recover a certain existing phonological category. However, this task may be particularly problematic for L2 speakers, either because they have not created a category for the sound in question, or because they do have it but they have trouble recognizing a specific label for it. This is particularly the case when the learner’s L1 has a transparent orthographic system that uses the same script as the L2; in these cases the learner has no reason to assume that the phonemic inventory in the L2 has sounds that do not map in the same way as in their L1. If the research question were simply whether the listener has a phonemic category or not, then this would probably be the simplest way to retrieve that information. However, if the question refers to the degree of abstraction of these representations, then this task would not give enough detail.

On the other hand, discrimination tasks, where listeners are presented more than one stimulus and respond whether these are the same or different, are more related to phonetic aspects of perception. However, the length of the interstimulus intervals (ISI), that is, the time elapsed between two stimuli within a trial, also plays a role in the type of information that the task can elicit: the longer the ISI, the more likely it is that responses tap phonological knowledge as the participants need to make use of long-term memory representations in order to provide an answer (Colantoni et al., 2015: 97).

Thus, since L2 learners cannot always correctly label the elements in a categorization task, it would not be particularly useful for the purposes of this research. Hence, the best approach is to use a discrimination task where the response is still same/different, but where the participant is forced to resort to higher levels of representation encoded in long-term memory. While ABX is indeed the most widely used method for retrieving phonological information, the long ISI and the nonce word format used in the AX task in the present study were there to provide a similar effect to that of ABX (that is, referencing abstract categories rather than acoustic information stored in short-term memory).

In this work we will take two approaches to the AX discrimination task. The first one will focus on the way in which L1 Spanish learners of L2 English (at earlier and later stages of learning) perceive the /æ/–/ʌ/ contrast in terms of sensitivity ($d'$): does sensitivity to these sounds change as experience in the L2 increases? The $d'$ statistic is obtained by taking into account hits (a “different” response to a pair of stimuli that are different), misses (a “same” response when the stimuli pair is different); a false alarm (a “different” response to same stimuli) and a correct rejection (a “same” response when stimuli are the same). The difference between the z-scores of hits and false alarms is fed to a formula that reduces the intrinsic bias of the task and yields a numeric value between zero (no sensitivity) and infinity (perfect sensitivity). However, infinity values can be corrected in order to obtain a reasonable value (Macmillan and Creelman, 1991).

The second approach is to test whether the nonnative participants change their responses in an abrupt way when presenting stimuli from a continuum: will they stop responding “same” if the trials start showing stimuli between which the acoustic distance
becomes increasingly longer, but are still within a native category? While this is not exactly a categorical perception test in the traditional way where the stimuli are always at a constant acoustic distance (Liberman et al., 1957) it still attempts to show that the nonnative discrimination patterns will not be the same as those of a native speaker. If we assume that the perception of the /a/-/ʌ/ contrast is a case of SCa or CGa, it can be expected that the changes in the “same” responses given by nonnative speakers will not take place in the same abrupt way as in the case of native speakers, since the latter will eventually reach a category boundary that the former will not. More specifically, an S-like curve will define the presence of a category boundary, where the bottom of the curve shows the proportion of “same” responses and the top of it shows the “different” responses. Conversely, a flat line would show that the participants are unable to perceive differences, as there would no category boundary within the given acoustic space. In other words, we would expect the S-curve in native speakers, and a flat line in nonnative speakers, if they don’t have a category boundary between /a/ and /ʌ/.

IV Hypotheses

- Hypothesis 1: L1 Spanish speakers of English map both /a/ and /ʌ/ onto the Spanish category /a/ in the earlier stages of their learning process; that is, a SCa scenario.

This will be tested by asking participants with less experience in the L2 to discriminate between an English and a Spanish vowel, namely, /a/ and /a/, /ʌ/ and /a/, /a/ and /o/, /ʌ/ and /o/. If the SCa scenario is true, then these participants will consistently respond “same” to the first two pairs, and “different” to the other two.

- Hypothesis 2: Advanced learners of English will be able to perceive acoustic differences between /a/ and /ʌ/, but not in the same consistent way as native speakers of English would do.

This will be tested by comparing $d'$ measurements at the endpoints of the /a/-/ʌ/ continuum across groups: while the less experienced group is expected to have low sensitivity, more experienced learners are expected to have a higher $d'$; native speakers of English are expected to have the highest.

- Hypothesis 3: Highly proficient L2 speakers of English whose L1 is Spanish will show a discrimination pattern along the /a/-/ʌ/ continuum that will be closer to that of native English speakers, but not a perceptual boundary as that of native speakers of English along the /a/-/ʌ/ continuum.

This will be tested by looking at the discrimination patterns along this continuum: it is expected that while native speakers of English show a clear S-shaped pattern with a significant change in the proportion of “different” responses around the mid point of the continuum, advanced learners show a more gradual pattern. Finally, less experienced learners are expected to show a flatter line with low counts of “different” responses all throughout the continuum.
**V Method**

1 Participants

The experiment considered 3 groups of participants: a first group of L1 American English (native speaker or NS, \( n = 8 \)) as a control group, a second group of native speakers of Spanish with some knowledge of English (NNSB, \( n = 13 \)), and a third group of native Spanish speakers with advanced knowledge of English (NNSA, \( n = 9 \)). The NS group were all native speakers of American English. The participants of the NNSA group were speakers of Chilean (8) and Mexican (1) Spanish who had lived in an English-speaking country (US and/or UK) for a minimum of 1 year (maximum 4 years) and had passed the required IELTS/TOEFL exams to begin studies at postgraduate level at the University of Manchester, UK (minimum language requirements set by the University are IELTS 6.5 or TOEFL100, which equals to B2–C1). The participants of the NNSB group were speakers of Chilean (12) and peninsular Spanish (1) who had been in Manchester (UK) for up to 3 months with purposes such as visiting relatives, learning English, and tourism. A further category of inexperienced speakers were spouses of workers and students who were neither engaging with the local community nor had further interest in learning English above the strictly necessary. For additional participant information, see Appendix 1.

2 Stimuli

Five seven-step vowel continua were created: one between the English vowels /ɑ/ and /ʌ/, and four between each one of the former and the Spanish vowels /a/ and /o/. Stimuli were built with Praat (Boersma and Weenink, 2019) from recordings of real tokens by female native speakers of General American English (GAE)\(^7\) and Chilean Spanish (ChS). The GAE vowels were obtained by asking the participant to elicit CVC words (cup, cut, pot, top) 3 times each in the carrier sentence *I say ____ once*. The ChS vowels were obtained by asking the participant to elicit the Spanish CVC words pan, par, ron, por in the carrier sentence *Yo digo ___ de nuevo*.

The resulting continua were /ɑ–ʌ/ (the only English–English contrast, predicted to yield different levels of sensitivity across groups); /ɑ–a/ and /ʌ–a/ (both English–Spanish contrasts, predicted to be perceived as the same across groups, henceforth SCa continua), and /ɑ–o/ and /ʌ–o/ (also English–Spanish contrasts, predicted to be perceived as different across groups, henceforth TCa continua). These continua toward Spanish /o/ were included in order to rule out the possibility that the /ɑ/–/ʌ/ contrast is not a case of TCa, should the nonnative speakers show perceptual patterns in line with the presence of two categories (that is, that the participants are discriminating between L2 categories /a/ and /ʌ/, and not L1 categories /a/ and /o/). Formant values (F1 and F2) of the original vowels were modified in equal Hertz steps moving towards the vowel at the other endpoint, and were then re-synthesized using the same source; both F1 and F2 were modified orthogonally; that is, F1 was divided in 7 steps and so was F2, thus forming a straight line in the acoustic space where token 1 in a continuum has the F1 values of step 1 and the F2 values of step 1.

The final stimuli consisted of CVC nonce words (or low frequency real words) where C is a plosive sound and V is any of the resulting continuum tokens, followed by an ISI
of 1 second, and an isolated vowel corresponding to either one of the ends of the continuum (e.g. [tʌp] – (1s ISI) – [ɑ]). In a trial, continuum step 1 was embedded in the nonce word, followed by the ISI and then paired with step 1; in another trial, continuum step 2 was embedded in nonce word, followed by ISI and then by step 1; and so on. Likewise, continuum step 1 was embedded in a nonce word followed by the ISI and then paired with step 7; in another trial, we find step 2 in nonce word followed by ISI and then step 7, and so on; this was repeated within all the other continua. The nonce words were obtained by splicing and rearranging consonants and vowels from the recorded stimuli. During the resynthesis process, care was taken to minimize residual effects of coarticulation, such as traces of formants in the plosive bursts, differences in vowel length, etc. The validity of the English stimuli was tested by a native speaker of American English, and the Spanish stimuli were tested by a native speaker of Spanish.

Since the creation of the final stimuli also included pairs with identical stimuli at the endpoints (that is, 1–1 and 7–7 stimuli pairs), these stimuli were later analysed together with stimuli pairs 1–7 and 7–1 for calculating d’ at endpoints. Thus, 1–1 and 7–7 pairs provided false alarms and correct rejections, while 1–7 and 7–1 pairs provided hits and misses.

3 Procedure

The experiment was carried out in the Phonetics laboratory at the University of Manchester, UK. Both the American English and the Spanish speaker were recorded in a sound-attenuated booth using an AKG condenser microphone, model C520. The sounds were recorded on a HP ProBook 6570b with the Praat recording interface, at a sampling frequency of 44,100 Hz.

Before the experiment, participants were first asked to fill out a language background questionnaire. After completing the form, participants were asked to perform an AX task, discriminating between the vowel in the nonce word and the one after the ISI. The experiment was set up in E-Prime, and participants performed the task on computers in a sound attenuated booth using circumaural headphones. Stimuli were presented in random order and participants had a maximum time of 4 seconds to respond.

Participants were asked to respond by pressing a green button whenever they thought the vowel sounds in the stimuli were the same, and a red button when they were different (the screen on the computer showed the words “Same” and “Different” in the corresponding color and aligned with the buttons). For these purposes a Psychology Software Tools serial response box model 200a was used. The experiment had a total of 2 blocks with 35 trials each. All reaction times were recorded; timeouts were removed from the data.

VI Results

1 Endpoint-to-endpoint sensitivity (d’)

The trials consisting of endpoint-to-endpoint discrimination were analysed separately with R, using the log-linear approach for correction of extreme values suggested in Macmillan and Creelman (1991). Figure 2 shows the mean d’ values for each group.
Upon visual inspection, it seems that the differences in $d'$ by group is evident only in the /ɑ/–/ʌ/ pair and that all three groups behave similarly when discriminating against the other four endpoint pairs /ɑ/–/a/, /ʌ/–/a/, /ɑ/–/o/, and /ʌ/–/o/. In order to test the significance of these differences, a linear model for each contrast with $d'$ as response variable and group as predictor was fitted; Table 2 shows the relevant statistical values for each regression. As expected, the linear regression shows that only the /ɑ/–/ʌ/ pair gave significant results with group as predictor ($R^2 = 0.35$, $F = 6.89$, $df = 2, 26$; $p < .05$). In Table 3, the estimates for the $d'$ of the /ɑ/–/ʌ/ pair show that while the difference between the NNSA (reference group) and the NNSB group is significant ($\beta = –1.13$; SE = 0.48; $t = –2.35$; $p < .05$), the difference between the NNSA and the NS group is not ($p = .23$).

In order to test the SCa as the scenario for the inexperienced group, a linear model on their $d'$ values with stimuli pair as a predictor was fitted, where we check whether the stimuli pair /ɑ/–/ʌ/ yields significantly different $d'$ scores than /ʌ/–/a/ and /ɑ/–/ə/; the differences in $d'$ by stimuli pair were significant ($R^2 = 0.72$, $F(4, 56) = 36.59$, $p < .05$).
order to find out which pair differs from which, a post-hoc comparison via estimated marginal means was carried out; Table 4 shows the estimates. The non-significant differences for pairs /ɑ/–/a/ and /ʌ/–/a/ suggest that /ɑ/ and /ʌ/ are perceived as instances of the category /a/ to the same extent; however, the /ɑ/–/ʌ/ pair is significantly different from these two other pairs.

2 Discrimination along continua

These sets of stimuli tested discrimination between a vowel sound extracted from a continuum and both endpoints of this same continuum; e.g. token number 5 from the /ɑ–ʌ/ continuum was discriminated against /ɑ/ on one trial, and against /ʌ/ on a different trial. Since these stimuli did not include identical stimuli pairs for each step of the continuum, no sensitivity analysis was performed. Instead, a logistic regression was performed in order to model the effect of both the stimuli pair in terms of the acoustic distance between them (Euclidean distance in Hz), and the level or proficiency (independent variables) over the “same” and “different” responses (response variable). This, along with the confidence intervals (CI) will provide a quantitative comparison in terms of the distribution of the responses by group. Regarding the /ɑ–ʌ/ continuum against the /ɑ/ token (that is, each one of the continuum tokens as stimulus 1 and the first token, vowel /ɑ/, as stimulus 2), an ANOVA performed on a model considering Group and Acoustic distance as predictors showed that there is a significant effect of both the group of speakers ($\chi^2(2) = 29.71, p < .001$) and the acoustic distance between stimulus ($\chi^2(1) = 86.54, p < .001$) on discrimination, but no interactions were found. Figure 3 shows the changes in discrimination as a function of the acoustic distance between tokens.

Furthermore, the regression shows that the discrimination pattern in advanced learners (NNSA; reference group) is significantly different from that of beginners, as well as that of the NS group. Table 5 shows the coefficients and p-values for each predictor; the estimates show the likelihood of “different” responses compared to the NNSA group.

| Table 3. Regression estimates for the $d'$ of the /ɑ/–/ʌ/ pair, with group as predictor. |
|---------------------------------|-------|--------|--------|-------|
|                                | Estimate | SE | t-value | p-value |
|-------------------------------|-----------|----|--------|---------|
| (Intercept)                   | 2.50      | 0.36 | 6.87   | < .05   |
| GroupNNSB                     | −1.13     | 0.48 | −2.35  | < .05   |
| GroupNS                       | 0.66      | 0.53 | 1.24   | .23     |

Notes. NS = L1 American English, i.e. native speakers. NNSB = native speakers of Spanish with some knowledge of English, i.e. non-native speakers.

| Table 4. Regression estimates for $d'$, inexperienced group. |
|---------------------------------|-------|--------|--------|-------|
|                                | Contrast | Estimate | SE | df | t-ratio | p-value |
|-------------------------------|-----------|-----------|----|----|---------|---------|
| /ɑ/–/ʌ/ and /ɑ/–/ʌ/           | −1.06243  | 0.35      | 56 | −3.06 | < 0.05  |
| /ɑ/–/a/ and /ʌ/–/a/           | 0.002     | 0.34      | 56 | 0.008 | 1.00    |
| /ɑ/–/ʌ/ and /ʌ/–/a/           | 1.07      | 0.33      | 56 | 3.20  | < 0.05  |
Figure 3. Changes in discrimination as acoustic distance between stimuli increases, /ɑ–ʌ/ continuum.
Notes. Shaded areas show a 95% CI.

Table 5. Logistic regression, discrimination (/ɑ–ʌ/ – /ʌ/).

|                    | Estimate | Std. error | z-value | Pr(>|z|) |
|--------------------|----------|------------|---------|----------|
| (Intercept)        | –6.10    | 0.99       | –6.13   | < 0.001  |
| GroupNNSB          | –1.34    | 0.57       | –2.34   | < 0.05   |
| GroupNS            | 1.75     | 0.61       | 2.90    | < 0.01   |
| AcDist             | 0.02     | 0.003      | 6.22    | < 0.001  |

Notes. NS = L1 American English, i.e. native speakers. NNSB = native speakers of Spanish with some knowledge of English, i.e. non-native speakers. AcDist = acoustic distance between stimuli.

Table 6. Logistic regression, discrimination (/ɑ–ʌ/ – /ʌ/).

|                    | Estimate | Std. error | z-value | Pr(>|z|) |
|--------------------|----------|------------|---------|----------|
| (Intercept)        | –4.25    | 0.68       | –6.28   | < 0.00   |
| GroupNNSB          | –0.86    | 0.48       | –1.77   | 0.08     |
| GroupNS            | 1.37     | 0.51       | 2.69    | < 0.01   |
| AcDist             | 0.01     | 0.002      | 6.27    | < 0.001  |

Notes. NS = L1 American English, i.e. native speakers. NNSB = native speakers of Spanish with some knowledge of English, i.e. non-native speakers. AcDist = acoustic distance between stimuli.

Similar results were found for discrimination of tokens corresponding to the /ɑ–ʌ/ continuum against the /ʌ/ token. The regression considering both predictors showed again that discrimination was significantly affected by group ($\chi^2(2) = 21.77, p < .001$) and acoustic distance between stimuli ($\chi^2(1) = 64.90, p < .001$); again, there was no interaction between the predictors. The coefficients and p-values can be found in Table 6, where
the estimates show the likelihood of “different” responses with the NNSA group as reference. However, in this case the responses of the advanced group do not seem to be particularly different from those of the beginners.

Regarding the SCa continua (i.e. /ɑ–a/ and /ʌ–a/), the regressions showed no effect of group on discrimination. However, it is worth pointing out that all three groups were in fact able to perceive differences between tokens when reaching maximal acoustic distance, albeit to a small extent; these results are in line with the PAM/PM-L2 predictions insofar they show poor discrimination. Figure 4 shows the fitted values along the continuum.

On the other hand, continua where TCa was expected (i.e. /ɑ–o/ and /ʌ–o/) showed excellent discrimination at the endpoints, with a clear S-curve depicting an increase in probabilities of perceiving the stimuli pair as different as the acoustic distance increases. Figure 5 shows the perceptual patterns of all three groups as a function of acoustic

Figure 4. Changes in discrimination as acoustic distances between stimuli increase, /ɑ–a/ and /ʌ–a/ continua. Notes. Shaded areas show a 95% CI.
distance between each stimuli in a given pair. However, the regressions show that, except for /ɑ–o/ – /ɑ/, responses to discrimination in TCa continua are also affected by Group. Post-hoc comparisons between groups showed that in /ʌ–o/ – /o/ and /ɑ–o/ – /o/ the discrimination in the NS group is significantly higher than the other two groups, but that both non-native groups do not differ from each other. The /ʌ–o/ – /ʌ/ continuum showed a similar pattern (only the NS group differs from the non-natives), but their discrimination is significantly lower than in the other two groups. Furthermore, and as seen in Figure 5, the differences are due to the shift of the boundary in NS, and not due to different slopes. Table 7 summarizes the values obtained by the ANOVAs performed on regressions corresponding to SCa and TCa continua.

In sum, the results show that all groups have a similar response to the SCa and TCa continua in terms of sensitivity at endpoints and discrimination along a given continuum.

Figure 5. Changes in discrimination as acoustic distance between stimuli increases, /ɑ–o/ and /ʌ–o/ continua.
Notes. Shaded areas show a 95% CI.
On the other hand, the /ɑ–ʌ/ continuum is the only one that triggered different responses across all three groups, which suggests that each group has a different degree of knowledge regarding the phonetic and phonological dimensions of the English vowel system. In this regard, the results suggest that while the beginners are still in a state where they are less likely to perceive differences between sounds that have been assimilated to the same Spanish vowel /a/ and lie within an acoustic distance of around 400 Hz, the advanced group is in an intermediate state between the behavior of beginners and the group of native speakers, insofar as they also assimilate both English vowels to /a/ but are more likely to perceive differences as a function of acoustic distance.

VII Discussion

Hypothesis 1 stated that L1 Spanish learners of English map both /ɑ/ and /ʌ/ onto the Spanish category /a/ in the earlier stages of their learning process, in what could be

| Expected assimilation vs. vowel | Continuum | Model | df | Dev | Res. df | Res. dev | p (>Chi) |
|---------------------------------|-----------|-------|----|-----|---------|----------|----------|
| SCa                             | /ɑ–a/ – /a/ | NULL | 206 | 80.11 |          |          |          |
|                                 | AcDist    | 1    | 12.17 | 205 | 67.94   | <.001    |          |
|                                 | Group     | 2    | 3.94  | 203 | 64.00   | .14      |          |
|                                 | /ɑ–ɑ/ – /ɑ/ | NULL | 209 | 122.85 |          |          |          |
|                                 | AcDist    | 1    | 8.45  | 208 | 114.41  | <.01     |          |
|                                 | Group     | 2    | 4.67  | 206 | 109.74  | .10      |          |
|                                 | /ʌ–a/ – /a/ | NULL | 208 | 117.89 |          |          |          |
|                                 | AcDist    | 1    | 4.23  | 207 | 113.66  | <.05     |          |
|                                 | Group     | 2    | 3.68  | 205 | 109.98  | .16      |          |
|                                 | /ʌ–ɑ/ – /ɑ/ | NULL | 209 | 80.41  |          |          |          |
|                                 | AcDist    | 1    | 7.27  | 208 | 73.14   | <.01     |          |
|                                 | Group     | 2    | 0.07  | 206 | 73.06   | .97      |          |
|                                 | /ɑ–o/ – /a/ | NULL | 208 | 288.93 |          |          |          |
|                                 | AcDist    | 1    | 137.05| 207 | 151.88  | <.001    |          |
|                                 | Group     | 2    | 1.18  | 205 | 150.69  | .55      |          |
|                                 | /ɑ–o/ – /o/ | NULL | 209 | 290.64 |          |          |          |
|                                 | AcDist    | 1    | 174.69| 208 | 115.96  | <.001    |          |
|                                 | Group     | 2    | 6.60  | 206 | 109.36  | <.05     |          |
|                                 | /ʌ–o/ – /a/ | NULL | 206 | 271.06 |          |          |          |
|                                 | AcDist    | 1    | 134.88| 205 | 136.18  | <.001    |          |
|                                 | Group     | 2    | 7.04  | 203 | 129.15  | <.05     |          |
|                                 | /ʌ–o/ – /o/ | NULL | 208 | 288.66 |          |          |          |
|                                 | AcDist    | 1    | 123.43| 207 | 165.23  | <.001    |          |
|                                 | Group     | 2    | 9.15  | 205 | 156.07  | <.05     |          |

Notes. SCa = single-category assimilation. TCa = two-category assimilation. NULL = null model. AcDist = acoustic distance between stimuli.
described as a SCa scenario. In order to test this, the sensitivity (d’) scores corresponding to the NNSB group were fitted in a linear regression model. The results showed that (1) there was no difference in sensitivity to the /a/-/a/ and /ʌ/-/a/ pairs, suggesting that both English vowels are being mapped onto Spanish /a/. However, we also see (2) a significant difference between these two pairs and /a/-/ʌ/, which suggests that when comparing these English vowels against each other, these learners can perceive a difference significantly better than the previous pairs. While (1) can be interpreted as evidence for what PAM/PAM-L2 would define as a SCa scenario, (2) seems to provide support for a CGa scenario. However, both /ʌ/ and /a/ are more or less equidistant from /a/, with the former at a distance of 182 Hz (higher and more fronted) and the latter at 254 Hz (lower and further back). It can be assumed then, that the difference perceived between /ʌ/ and /a/ is due to the higher acoustic distance (416 Hz) and not to the fact that one is a better exemplar of /a/ than the other.

Regarding perceptual assimilation across groups, two observations can be made. First, all three groups assimilated /ʌ/ and /a/ to /a/ to a similar extent, which shows that tokens of neither /a/ nor /ʌ/ are considered good exemplars of /o/, the second closest vowel in the perceptual space. This departs from the findings in Escudero and Chládková (2010) and Barrientos (2018), where /ʌ/ is likely to assimilate to both /a/ and /o/. This might be due to the specific acoustic values in the reference stimuli provided by the native speaker of English (with a high F2 value), or to differences in the experimental design. Likewise, such similar patterns in both NNSA and NNSB groups suggest that perceptual assimilations of L2 sounds onto L1 sounds do not change throughout time. Second, and regarding the poor discrimination of native English listeners between /ʌ/-/a/ and /a/-/a/, it could be argued that if they show this pattern, then nonnative speakers also cannot change this assimilation. However, breaking such patterns of L2-to-L1 perceptual assimilation is neither a realistic expectation (especially considering the short acoustic distance) nor a proof of L2 category learning; what really matters is whether they can learn to perceive /ʌ/ and /a/ as different sounds despite the L2-to-L1 assimilation (and not due to it, in case they were mapped onto different L1 categories).

Hypothesis 2 claimed that the sensitivity to the /a/-/ʌ/ pair would be different across groups; this prediction was partially borne out. Native speakers showed the highest d’ scores, and the NNSA group showed significantly higher sensitivity than the NNSB group; however, the NNSA scores are not significantly different from those of the NS group. This suggests that the NNSA group is likely to have higher sensitivity to this vowel pair than previously expected and that their performance at the endpoints (that is, when discriminating between good exemplars of the L2 categories) is comparable to that of a native speaker.

Finally, Hypothesis 3 referred to the differences in the discrimination patterns along a continuum, and more specifically regarding the probabilities associated with perceiving differences as the acoustic distance increases in an abrupt pattern, due to the presence of a category boundary. In this regard, it was predicted that the discrimination along the /a–ʌ/ continuum in the experienced group would show an in-between state, whereby the probability of perceiving differences as the stimuli became increasingly different in terms of acoustic distance would be higher than that of the beginners’ group; however, it was also expected that this pattern would not be the same as that displayed by native
speakers on that continuum, or by all the three groups for continua where TCa was predicted. This prediction was borne out. Crucially, the lack of a clear S-curve pattern in advanced speakers contrasts with the clear display of it whenever the endpoints of a given continuum were predicted to undergo TCa. The previous finding suggests that advanced speakers have become more attuned to the acoustic differences in the L2 input, but that native-like behavior is not available to the advanced learners, as far as the “S-ness” of the curve is concerned. However, the NNSA behavior regarding Hypothesis 2 corresponds to that of a native speaker.

VIII Conclusions

The results have suggested that NNSA participants have not been able to create a full category split between /ɑ/ and /ʌ/ with a boundary akin to the one found in native speakers; that is, that this boundary, albeit existing, is not as clearly defined as the one in native speakers, or it is not at the same place. However, their perception differs from that of the inexperienced group in two ways: (1) their sensitivity (d′) to the acoustic difference between good exemplars of /ɑ/ and /ʌ/ is significantly higher, and (2) their discrimination improves as the acoustic distance between tokens lying within the /ɑ/–/ʌ/ range increases. Similar “in-between” results in L2 phonology where there seems to be a considerable degree of reliance in L1 representations have already been attested, both in perception and production (e.g. Freeman et al., 2016; Stefanich and Cabrelli, 2021). It can be concluded that this perceptual pattern is consistent with the subsetting outcome (partial perceptual categories in representation), whereby the learning is described as a perceptual task without a representational task; or rather, that the representational task is still incomplete. The main evidence for this claim is that the continua between tokens that are perceived as members of the same category will not yield an S-curve (e.g. the /a–ɑ/ continuum); however, the /a–ʌ/ continuum, whose endpoints were chiefly perceived as the same as Spanish /a/, displayed an in-between pattern closer (yet not entirely equal) to the perceptual behavior found in the NS group. Since the S-curve pattern is caused by the presence of two phonemic categories that warp the perception at the endpoints, it can be argued that the NNSA speakers have increased their sensitivity to the relevant acoustic cues, considering that the /a–ʌ/ continuum yielded something similar to an S-curve; nevertheless, they are not different categories. Furthermore, the ability of the NNSA group to resemble NS behavior may be aided by a potential CGa scenario, which is predicted by PAM-L2 to be more likely to evolve into a full category split.

The results also suggest that NNSA participants are not in the same state as the inexperienced group (NNSB) regarding the /a/– /ʌ/ pair, since the probability of perceiving these sounds as different along the continuum is not the same for each group. It can be hypothesized that since the /a/– /ʌ/ pair falls into the /a/ range, inexperienced speakers have lower probabilities of perceiving differences due to the fact that both English vowels are perceived as /a/ and the acoustic distance between the L2 categories is not taken into account. Second, the following closest L1 category is /o/, but the data does not support a hypothesis in which NNSA speakers would be remapping either /a/ or /ʌ/ onto it. Hence, and since inexperienced listeners are likely to perceive the endpoints of a given continuum as tokens of the same category, it can be hypothesized that the probability of perceiving these sounds as different is low even if the acoustic distance is perceivable by
a human ear due to the warping of the perceptual space that the existing L1 categories (/a/ in this case) create. On the other hand, experienced learners do become more attuned to such acoustic difference by means of input, and despite perceiving the L2 sounds as exemplars of the same L1 category just like inexperienced speakers do, they manage to break this “sand trap” effect.

As mentioned above, the lack of an S-curve pattern is given by the absence of a different phonemic category at one of the endpoints. But why are the NNSA participants able to nevertheless perceive these differences to at least some extent, and more importantly, to a significantly higher extent than NNSB? Two empirical facts must be taken into account. First, and as noted by Werker and Tees (1984), there is no evidence that the inability of late L2 learners to perceive certain contrasts is caused by sensorineural loss; rather, it is the L1 language experience that blocks the perception of differences at a later stage of the perceptual processing. And second, abundant L2 input does have an effect in increasing the L2 speaker’s sensitivity to nonnative contrasts, as L2 speakers learn different perceptual cue weightings for the L2 input (Bohn, 1995; Cebrian, 2006; Escudero et al., 2009; Escudero and Boersma, 2004), even by making use of perceptual cues that are not relevant in the L1 (Bohn, 1995). However, what does not follow from these two points is that this increased ability to perceive will imply the creation of a new phonemic category: the absence of these categories is then what causes the incomplete pattern. Whether these advanced speakers could eventually create a new category is, however, not in question: what is being suggested is that the ability to create a new phonemic category decreases in adult learners, and that it seems to be slower than acquiring linguistic knowledge in other domains, e.g. morphology and syntax. In this regard, theories of second language acquisition (SLA) have largely agreed that L2 learners rarely achieve native-like performance in the realm of phonetics and phonology (Ellis, 1994: 491–493). At least, and as far as the data analysed here is concerned, years of experience in the L2 are not a sufficient condition for showing perceptual patterns that are fully comparable to those of a native speaker and therefore their perceptual learning cannot be assumed to be phonemic.

Regarding the existing speech perception models in L2, the PAM/PAM-L2 models make the correct predictions about the discrimination rate according to the type of perceptual assimilation: SCa scenarios yield poor discrimination and TCa trigger high discrimination. Finally, L2LP’s assumption of two different learning tasks (perception/representation) taking place seems to adequately describe the path towards developing a phonemic category; likewise, this division of tasks allows for a more accurate description of incomplete states of acquisition. Such incomplete patterns, which neither resemble native-like behavior nor the full transfer from the L1 that can be observed in initial states, are thus better explained by the concept of interlanguage category: in this case, we see how in some tasks the behavior observed in experienced speakers is closer to that of native speakers, but how they are either between native and inexperienced, or just closer to inexperienced speakers in others. More specifically, the fact that there is more attunement to acoustic cues but without showing clear boundaries suggests then that the perceptual task is being carried out, but that the representational task is not complete. Thus, the data supports the L2LP claim that learning can take place by means of input as L1 categories are learned, but whether input alone is enough for both learning tasks would have to be supported by evidence coming from very experienced speakers, e.g. with decades of extensive L2 experience.
A further implication regarding L2 phonology and SLA is that high proficiency in the L2 does not necessarily imply the development of native-like perception patterns. It is worth noting that in order to achieve a B2–C1 level of proficiency (as the NNSA group), the L2 listener needs to receive a considerable amount of input; hence, it is not necessarily a problem of lack of input. However, an actual measurement of input is at best controversial, considering that common measurements such as age of onset and length of residence in a foreign country conflate with psychological, social and cognitive factors, such as motivation, the time spent talking to native speakers, and educational experience, among others (Moyer, 2009). It might be worth exploring in the future whether the lack of native-like perceptual patterns (or rather, the lack of ability to create new phonemic categories) is actually caused by a later age of onset (thus confirming a weak critical period hypothesis), or if this later age of onset hides a qualitative difference in the input.

After taking the above in consideration, the data presented here does not support the hypothesis that experienced L2 learners create abstract phonemic representations, and thus, it suggests that the learning stays at a more surface level, at least for a considerable amount of time (i.e. at least years). Thus, these findings support SLM-r with regards to the focus on a phonetic level of representation. The data presented here suggest that looking at the issue of category acquisition in absolute learning/no learning terms is not the best approach; rather, the question should be shifted in first place to what is being learned and, second, to which extent this learning resembles native-like category learning. The data here suggests that creating interlanguage representations is largely an increased ability to perceive acoustic differences, but that the process in which a set of values is mapped onto new categories does not necessarily follow from the previous one. Thus, they are unlikely to look like categories in the way that L1 phonology understands them, that is, with a discrete nature and assuming clear perceptual boundaries.

Finally, a note on potential limitations of this study should be made. First – and while the majority of the participants acknowledged in their background survey that American English was the language variety to which they had most exposure when learning the language (see Note 7) – their presence in England could have played a role in the way they perceive the sounds in question. Second, the sample size is small, which might not be particularly representative of the population parameters as a whole. And, third, the inclusion of only a discrimination task (that is, neither a perceptual assimilation nor a lexical task) does not allow for proper conclusions on the question of the phonemic status of interlanguage representations in experienced learners, for which these results should be interpreted with caution. Nevertheless, we expect that the analysis and perspectives in this work open new paths of research regarding the nature of interlanguage representations at the phonological level, especially regarding the extent to which L2 structures can be comparable to those in the L1 and how these newly acquired representations fit within the framework of phonological theory.

Acknowledgements

I am very grateful to three anonymous reviewers for their thoroughness and patience. Likewise, I would like to thank Yuni Kim, George Walkden, and Henri Kauhanen for their feedback, comments, and support. All remaining errors are my own.
Declaration of conflicting interests

The author declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

Funding

The author received no financial support for the research, authorship, and/or publication of this article.

ORCID iD

Fernanda Barrientos https://orcid.org/0000-0001-6751-8136

Notes

1. However, production and perception do not necessarily match, even in highly proficient L2 speakers (de Leeuw et al., 2019).
2. While different levels of representations (ranging from more concrete to more abstract) have been proposed in phonological theory (e.g. McCarthy, 2007; Pierrehumbert, 1990), I will assume an abstract (phonemic) level and a concrete, non-phonemic level for the sake of simplicity.
3. This division of tasks and their possible combinations roughly equates the concepts of distinctiveness/contrastiveness in Stratal OT as presented by Kiparsky (2015).
4. Interestingly enough, this term has been much more used in SLA syntax and morphology rather than phonology, even though Selinker was also thinking about the sound level when defining the concept: “I have in mind such fossilizable structures as the well-known ‘errors’: French uvular /r/ in their English IL, American English retroflex /r/ in their French IL, [. . .] and so on” (Selinker, 1972: 215).
5. An exception can be made for Eastern Andalusian Spanish, which shows a tense–lax contrast linked to singular/plural distinctions (Hualde and Sanders, 1995).
6. Barrios et al. (2016) provide a detailed discussion on the mismatches between acoustic and feature-based approaches to similarity.
7. While the choice of an American English contrast and stimuli seems counterintuitive given that the participants were living in England, the reason is that in a previous survey the majority of the participants agreed that they were more familiar with the American accent, and that this was the pronunciation model that they first encountered while learning the language. This is usually the case of Latin American speakers of English as a second/foreign language. It is also worth noting that the University of Manchester has a high number of European Union and international students coming from all over the world, especially at postgraduate level, for which the influence of not only the local variety but also other accents, including nonnative ones, could affect their perceptual patterns.

References

Adjemian C (1976) On the nature of interlanguage systems. Language Learning 26: 297–320.
Altmann H and Kabak B (2011) Second language phonology. In: Kula N, Botma B, and Nasukawa K (eds) The continuum companion to phonology. London: Continuum, pp. 298–319.
Antoniou M, Tyler MD, and Best C (2012) Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? Journal of Phonetics 40: 582–94.
Barrientos F (2018) Perceptual representations in Interlanguage Phonology: subcategorial learning in late-learners with a smaller vowel inventory. Unpublished PhD thesis. University of Manchester, Manchester, UK.
Barrios S, Jiang N, and Idsardi WJ (2016) Similarity in L2 phonology: Evidence from L1 Spanish late-learners’ perception and lexical representation of English vowel contrasts. Second Language Research 32: 367–95.

Best C (1995) A direct realist view of cross-language speech perception. In: Strange W (ed.) Speech perception and linguistic experience: Issues in cross-language research. Timonium, MD: York Press, pp. 171–204.

Best C and Tyler M (2007) Nonnative and second-language speech perception: Commonalities and complementarities. In: Bohn O-S and Munro MJ (eds) Language experience in second language speech learning: In honor of James Emil Flege. Amsterdam: John Benjamins, pp. 13–34.

Boersma P (2009) Cue constraints and their interactions in phonological perception and production. In: Boersma P and Hamann S (eds) Phonology in perception. Berlin: Mouton de Gruyter, pp. 55–110.

Boersma P and Weenink D (2019) Praat: Doing phonetics by computer: 6.1 [computer program]. Available at: praat.org (accessed June 2021).

Bohn O-S (1995) Cross language speech production in adults: First language transfer doesn’t tell it all. In: Strange W (ed.) Speech perception and linguistic experience: Issues in cross-language research. Timonium, MD: York Press, pp. 279–304.

Bohn O-S and Flege JE (1997) Perception and production of a new vowel category by second-language learners. In: James A and Leather J (eds) Second-language speech: Structure and process. New York: Mouton de Gruyter, pp. 53–74.

Bradlow A (1995) A comparative acoustic study of English and Spanish vowels. Journal of the Acoustic Society of America 97: 1916–24.

Brown CA (1998) The role of the L1 grammar in the L2 acquisition of segmental structure. Second Language Research 14: 136–93.

Cebrian J (2006) Experience and the use of non-native duration in L2 vowel categorization. Journal of Phonetics 34: 372–87.

Colantoni L, Steele J, and Escudero P (2015) Second language speech. Cambridge: Cambridge University Press.

Cole J and Hualde JI (2011) Underlying representations. In: van Oostendorp M, Ewen CJ, Hume E et al. (eds) The Blackwell companion to phonology. Oxford: John Wiley, pp. 1–26.

de Leeuw E, Stockall L, Lazaridou-Chatzigoga D et al. (2019) Illusory vowels in Spanish–English sequential bilinguals: Evidence that accurate L2 perception is neither necessary nor sufficient for accurate L2 production. Second Language Research. Epub ahead of print 5 December 2019. DOI: 10.1177/0267658319886623.

Drescher BE (2011) The phoneme. In: van Oostendorp M, Ewen CJ, Hume E et al. (eds) The Blackwell companion to phonology: Volume 1. Oxford: Wiley-Blackwell, pp. 241–66.

Eckman F, Elreyes A, and Iverson G (2003) Some principles of second language phonology. Second Language Research 19: 169–208.

Eckman FR (1977) Markedness and the contrastive analysis hypothesis. Language Learning 27: 315–30.

Ellis R (1994) The study of second language acquisition. Oxford: Oxford University Press.

Escudero P (2005) Linguistic perception and second language acquisition: Explaining the attainment of optimal phonological categorization. Utrecht: Utrecht University and LOT.

Escudero P and Boersma P (2003) Modelling the perceptual development of phonological contrasts with Optimality Theory and the Gradual Learning Algorithm. Proceedings of the 25th Annual Penn Linguistics Colloquium: Penn Working Papers in Linguistics 8: 71–85.

Escudero P and Boersma P (2004) Bridging the gap between L2 speech perception research and phonological theory. Studies in Second Language Acquisition 26: 551–85.
Escudero P and Chládková K (2010) Spanish listeners’ perception of American and Southern British English vowels. The Journal of the Acoustical Society of America 128: EL254–60.
Escudero P, Benders T, and Lipski S (2009) Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. Journal of Phonetics 37: 452–65.
Flege JE (1987) The production of “new” and “similar” phones in a foreign language: Evidence for the effect of equivalent classification. Journal of Phonetics 15: 47–65.
Flege JE (1995) Second language speech learning: Theory, findings, and problems. In: Strange W (ed.) Speech perception and linguistic experience: Issues in cross-language research. Timonium, MD: York Press, pp. 233–77.
Flege JE and Bohn O-S (2021) The Revised Speech learning Model (SLM-r). In: Wayland R (ed.) Second language speech learning: Theoretical and empirical progress. Cambridge: Cambridge University Press.
Flege JE, Bohn O-S, and Jang S (1997) Effects of experience on non-native speakers’ production and perception of English vowels. Journal of Phonetics 25: 437–70.
Flege JE, MacKay I, and Meador D (1999) Native Italian speaker’s perception and production of English vowels. Journal of the Acoustical Society of America 106: 2973–88.
Flemming E (2004) Contrast and perceptual distinctiveness. In: Hayes B, Steriade D, and Kirchner R (eds) Phonetically based phonology. Cambridge: Cambridge University Press, pp. 232–76.
Freeman MR, Blumenfeld HK, and Marian V (2016) Phonotactic constraints are activated across languages in bilinguals. Frontiers in Psychology 7: 702.
García P (2014) Perception of American English vowels by adult Spanish–English bilingual listeners. Unpublished PhD thesis, Columbia University, New York, USA.
Hancin-Bhatt B (2008) Second language phonology in Optimality Theory. In: Hansen Edwards JG and Zampini ML (eds) Phonology and second language acquisition. Amsterdam: John Benjamins, pp. 117–46.
Hayes-Harb R (2005) Optimal L2 speech perception: Native speakers of English and Japanese consonant length contrasts. Journal of Language and Linguistics 4: 1–29.
Hualde JI (2005) The sounds of Spanish. Cambridge: Cambridge University Press.
Hualde JI and Sanders BP (1995) A new hypothesis on the origin of the Eastern Andalusian vowel system. In: Bilmes L, Liang AC, and Ostapirta W (eds) Proceedings of the twenty-first annual meeting of the Berkeley Linguistics Society: General session and parasession on historical issues in sociolinguistics/social issues in historical linguistics. Berkeley, CA: Berkeley Linguistics Society, pp. 426–37.
Iverson P and Evans B (2007) Learning English vowels with different first-language systems: Perception of formant targets, formant movement, and duration. Journal of the Acoustic Society of America 122: 2842–54.
Iverson P and Kuhl P (1995) Mapping the perceptual magnet effect for speech using signal detection theory and multidimensional scaling. Journal of the Acoustic Society of America 97: 553–62.
Kim D, Clayards M, and Goad H (2018) A longitudinal study of individual differences in the acquisition of new vowel contrasts. Journal of Phonetics 67: 1–20.
Kiparsky P (2015) Phonologization. In: Honeybone P and Salmons J (eds) The Oxford handbook of historical phonology. Oxford: Oxford University Press, pp. 563–82.
Lakshmanan U and Selinker L (2001) Analysing interlanguage: how do we know what learners know? Second Language Research 17: 393–420.
Liberman A, Harris K, Hoffman H et al. (1957) The discrimination of speech sounds within and across phoneme boundaries. Journal of Experimental Psychology 54: 358–68.
Macmillan NA and Creelman CD (1991) Detection theory: A user’s guide. Cambridge: Cambridge University Press Archive.
Major R (1998) Interlanguage phonetics and phonology: An introduction. Studies in Second Language Acquisition 20: 131–37.
Martin A and Peperkamp S (2011) Speech Perception and Phonology. In: van Oostendorp M, Ewen CJ, Hume E et al. (eds) The Blackwell companion to phonology. John Wiley, pp. 1–23.
McCarty JJ (2007) Derivations and levels of representation. In: de Lacy P (ed.) The Cambridge handbook of phonology. Cambridge: Cambridge University Press, pp. 99–118.
Moyer A (2009) Input as a critical means to an end: Quantity and quality of experience in L2 phonological attainment. In: Piske T and Young-Scholten M (eds) Input matters in SLA. Clevedon: Multilingual Matters, pp. 159–74.
Nadeu M (2014) Stress- and speech-rate-induced vowel quality variation in Catalan and Spanish. Journal of Phonetics 46: 1–22.
Pierrehumbert J (1990) Phonological and phonetic representation. Journal of Phonetics 18: 375–94.
Polivanov ED (1931) La perception des sons d’une langue étrangère [The subjective nature of the perception of language sounds]. In: Polivanov ED (ed.) Selected works: Articles on general linguistics. The Hague: Mouton, pp. 223–37.
Quilis A (1993) Tratado de fonología y fonética españolas [Treatise on Spanish phonology ad phonetics]. Madrid: Gredos.
Rallo Fabra L and Romero J (2012) Native Catalan learners’ perception and production of English vowels. Journal of Phonetics 40: 491–508.
Safronova E (2016) The role of cognitive ability in the acquisition of second language perceptual phonological competence. Unpublished PhD thesis, Universitat de Barcelona, Barcelona, Spain.
Sapir E (1963) The psychological reality of phonemes. In: David Mandelbaum (ed.) Selected writings of Edward Sapir in language, culture, and personality. Berkeley: University of California Press, pp. 46–60.
Selinker L (1972) Interlanguage. IRAL – International Review of Applied Linguistics in Language Teaching 10: 209–32.
Stefanich S and Cabrelli J (2021) The effects of L1 English constraints on the acquisition of the L2 Spanish alveopalatal nasal. Frontiers in Psychology 12. Epub 15 February 2021. DOI: 10.3389/fpsyg.2021.640354.
Strange W (2011) Automatic selective perception (ASP) of first and second language speech: A working model. Journal of Phonetics 39: 456–66.
Swadesh M (1934) The phonemic principle. Language 10: 117–29.
Tarone E (1978) The phonology of interlanguage. In: Richards JC (ed.) Understanding second and foreign language learning. Rowley, MA: Newbury House, pp. 15–33.
Trubetzkoy NS (1969) Principles of phonology. Foreword by Christiane AM Baltaxe. Translation. Berkeley, CA: University of California Press.
Werker J and Tees R (1984) Phonemic and phonetic factors in adult cross-language speech perception. Journal of the Acoustic Society of America 75: 1866–78.
Winkler I, Kujala T, Tiitinen H et al. (1999) Brain responses reveal the learning of foreign language phonemes. Psychophysiology 36: 638–42.
### Appendix 1. Additional participant information by group.

| Group   | Mean self-rated ability (1 = can't do it; 5 = very well) | First exposure after age 12 years? | Mean age (years) of previous immersions (if any) | Other languages spoken (if any) | Learning method |
|---------|----------------------------------------------------------|-----------------------------------|-----------------------------------------------|-------------------------------|-----------------|
|         | Listening American English | Listening British English | Reading | | |
| NNSB (n = 13) | 3.2 | 3.5 | 4.0 | yes (n = 8); no (n = 5) | 22.7 (n = 4) | French (n = 1); German and Catalan (n = 1) | a (n = 3); b (n = 4); c (n = 9) |
| NNSA (n = 9) | 4.3 | 4.6 | 5.0 | yes (n = 5); no (n = 4) | 23.6 (n = 6) | Portuguese (n = 2) | a (n = 1); b (n = 7); c (n = 6); d (n = 1); e (n = 2) |

Notes. The learning methods reported by the participants were: (a) attended an immersion school; (b) English classes; (c) self-taught; (d) spoken at home on a regular basis; (e) other. Most participants reported using more than one method. NNSA = native speakers of Spanish with advanced knowledge of English, i.e. non-native speakers group A. NNSB = native speakers of Spanish with some knowledge of English, i.e. non-native speakers group B.