



OPEN ACCESS

EDITED BY

Jennifer Cabrelli,
University of Illinois Chicago, United States

REVIEWED BY

Shannon Barrios,
The University of Utah, United States
John Matthews,
Chuo University, Japan

*CORRESPONDENCE

Fernanda Barrientos
fernanda.barrientos-contreras@
uni-konstanz.de

RECEIVED 15 September 2023

ACCEPTED 17 January 2024

PUBLISHED 09 February 2024

CITATION

Barrientos F (2024) Out with the old, in with the new: contrasts involving new features with acoustically salient cues are more likely to be acquired than those that redeploy L1 features. *Front. Lang. Sci.* 3:1295265. doi: 10.3389/flang.2024.1295265

COPYRIGHT

© 2024 Barrientos. This is an open-access article distributed under the terms of the [Creative Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.

Out with the old, in with the new: contrasts involving new features with acoustically salient cues are more likely to be acquired than those that redeploy L1 features

Fernanda Barrientos*

Department of Linguistics, University of Konstanz, Konstanz, Germany

Feature-based approaches to second language (L2) phonology conceptualize the acquisition of new segments as operations that involve either the addition of new phonological features, or the rebundling of existent ones. While the deficit hypothesis assumes that only features that are fully specified in the L1 can be redeployed to the L2 in order to create new segments, it has been shown that features which are completely absent in the L1 can also be learned. This article investigates whether a learning scenario in which features are only partially available (that is, they are present in the L1, but are redundant with other features) is less challenging than learning an entirely new feature, even when the new feature has acoustically salient cues. Since Spanish has a much smaller vowel system /i e a o u/, L2 learners of German with Spanish as L1 need to learn a system with front rounded vowels as well as tense/lax contrasts. We tested L1 Spanish speakers' perception of the German contrasts /i/ ~ /ɨ/ (e.g., *Miete/mitte*, where [+/- tense] is acquired) and /u/ ~ /y/ (e.g., *Spulen/spülen*, where L1 feature [+/-round] redeployes to a front vowel). The results showed that experienced L2 learners are more successful when discriminating between sounds in a feature acquisition scenario than in redeployment; however, neither of the non-native contrasts was easier to perceive than the other in the identification task. The differences in performance between tasks and in the acoustic saliency of the cues by contrast (F2 vs. duration and F1) suggests that L2 phonological acquisition is likely to take place at a surface level and favors learning through attunement to auditorily salient acoustic cues over internal rearrangement of abstract features, regardless of their presence in the L1.

KEYWORDS

speech perception, feature-based models, L2 phonology, underspecification, vowel contrast, perceptual cues, feature redeployment

1 Introduction

In L2 phonology, whether transfer and acquisition are surface-based or phonological phenomena (Major, 2008, p. 68) is an ongoing discussion; nevertheless, research following an acoustic-based approach vastly outnumbers feature-based accounts. The main claims of the Speech Learning Model (SLM) and its revised version (SLM-r) (Flege, 1995; Flege and Bohn, 2021), which propose an acoustic-based approach to the acquisition of L2 segments, or the Perceptual Assimilation Model (PAM/PAM-L2) (Best, 1995; Best and Tyler, 2007), with a direct-realist perspective, have been extensively tested and confirmed by a large number of empirical studies.

On the other hand, more abstract models have also been proposed, where topics such as the role of universals and feature prominence are discussed, e.g., the Ontogeny-Phylogeny Model (OPM) (Major, 2001) or the Feature Competition Model (FCM) (Hancin-Bhatt, 1994), respectively. Among such phonology-based accounts, of particular interest for this research is the Feature-Based Model (FBM) (Brown, 1998, 2000). The model builds upon FCM and is motivated by Feature Geometry (Clements, 1985; Clements and Hume, 1995), a phonological framework wherein segments are no longer defined as “bundles” of unorganized features (à la SPE) but rather as a hierarchy, with tiers to which features are associated. Further developments of Feature Geometry include Underspecification Theory (Archangeli, 1988; Avery and Rice, 1989), according to which abstract segmental representations only contain the features needed for contrast with regards to the rest of the segments in the inventory.

The key aspect of Brown’s FBM is that the L1 constrains acquisition whenever the L2 phonemic inventory has features that the L1 does not have, or rather, when such features are underspecified in the learner’s L1. In such cases, the L2 segment simply cannot be fully acquired due to the fact that the relevant features are underspecified in the L1 and thus are unavailable for transfer. Brown’s interpretation of the inability to learn new features is that the L1 hinders the L2 learner’s access to Universal Grammar. As a corollary to the above, cases in which the L1 has the same features as the L2 but in different natural classes (e.g., when length is available in the L1 only in vowels, but the L2 has it in consonants) or when the acoustic cue associated to a particular contrast is relevant in the L1 but to a lesser extent, should result in easier acquisition, a prediction that has been borne out, at least in the case of length contrasts (e.g., McAllister et al., 2002; Pajak and Levy, 2014).

However, other studies have suggested that L1 speakers/listeners of languages without a tense/lax contrast are indeed able to distinguish between tense and lax segments in an L2, e.g., speakers of L1 Catalan perceiving English /i/ ~ /ɪ/ and /e/ ~ /ɛ/ (also “long/short,” since the contrast is usually realized via differences in duration) when performing certain perceptual tasks, at least to a certain extent, regardless of their amount of experience in the L2 (Cebrian, 2006); this ability to “turn off” the attunement to the L1 acoustic cues in favor of new ones was also observed by Bohn (1995) in L1 speakers of Spanish learning English /i/ ~ /ɪ/, in a phenomenon known as the “Desensitization Hypothesis.”¹ Furthermore, it has also been found that the presence or absence of the relevant phonological features in the L1 does not seem to play a role when explaining difficulty in the acquisition of L2 contrasts (Barrios et al., 2016). In sum, the evidence regarding the acquisition of tense/lax contrasts disproves a strong interpretation of Brown’s theory, although the feature’s status in the grammar may differ across experiments: while Brown’s experiment looks into the

effect of underspecified features below the coronal node that may nevertheless allow for allophonic forms that match contrasting L2 sounds, the [tense] feature is completely absent in both Spanish and Catalan.

This research examines the effect of underspecification in the acquisition of new L2 sounds; that is, whether there is a difference in the perception of nonnative sounds when the feature responsible for a given contrast in the L2 is completely absent in the L1’s grammar, vs. when the feature in question is present but also redundant with other features (i.e. feature +A occurs always in segments with feature +B) in the L1. In a null hypothesis, both underspecification and absence of a feature in the grammar would have the same effect on L2 learning: that is, that new L2 contrasts involving either absent or redundant features should pose the same degree of difficulty in acquisition. Conversely, the alternative hypothesis would state that either type of feature (underspecified or absent) would be easier to acquire than the other. On the one hand, it could be hypothesized that a present feature is better than nothing at all; on the other hand, if the absent feature has an acoustically salient cue it may be easier for L2 learners to perceive and produce. Furthermore, this study also analyzes the role of the acoustic cues involved in contrasts, that is, whether a new feature with acoustically salient cues may have an advantage over an existing one with cues that are perceptually less salient, due to either intrinsic auditory non-saliency or because the cue is already in use for the perception of another phonological contrast. By examining the perception of both types of L2 contrasts through perception tasks that tap different levels of representation, we aim to find potential differences which could in turn shed light on the phonological status of L2 representations.

2 Theoretical background: features in L2 phonology

2.1 Feature-based experiments

A considerable part of what we know about L2 phonological acquisition is largely based upon empirical studies that investigate the acquisition of new segments in the L2 from an acoustic-based approach; in this regard, the SLM(r) has been extensively cited as a model for acquisition that focuses on acoustic similarity to explain difficulties in L2 acquisition. Thus, SLM (and also PAM-L2) predicts that difficulties in L2 acquisition arise whenever the input is perceived by the L2 learner as similar enough to an L1 category. On the other hand, Brown (1998) points out that these models do not attempt to explain the impact of the native grammar on the perception of L2 sounds (139–140). More importantly, Brown looks into a more abstract level of representation, whereby the crucial factor for predicting difficulty is the internal configuration of the segment (that is, the segment’s abstract phonological features as well as their dependencies within the segment’s geometry).

The evidence offered by Brown (1998) consisted of two different perceptual tasks (AX discrimination and picture identification) involving the /r/ ~ /l/ contrast, carried out by experienced L2 listeners whose L1 was either Mandarin or Japanese (neither of these have these sounds in contrastive distribution).

1 The results of this study were statistically reanalyzed in Flege et al. (1997) where it was concluded that the differences in cue weighting between native and L1 Spanish nonnative speakers of English were not significant. However, and as discussed in Escudero and Boersma (2004), the statistical power of the test performed cannot be high since relatively high differences yielded non-significant results.

The results showed that Mandarin speakers performed better than the Japanese participants in both tasks. Brown suggests that this advantage of Mandarin listeners over Japanese listeners is given by the transfer of the feature [COR] and terminal node [anterior], present in Mandarin's sibilants, to English liquids /r/ ~ /l/. On the other hand, Japanese does not contrast any of their coronal phonemes, since the [COR] feature is not further specified.² However, in order for this account to work, Brown suggests that English /r/ is a phoneme whose place node is specified as [COR], whereas English /l/ has an empty place node. This is an alternative to the usual assumption that the relevant feature in English for the /r/ ~ /l/ distinction is [lateral], along the lines of previous research (e.g., Spencer, 1984). This led to Brown's proposal that fully specified L1 features not only can be transferred to the L2 in order to convey phonemic contrasts between sounds that are not present in the L1, but also that the perception of new L2 contrasts depends *exclusively* on this transfer operation; that is, if the features in question are not present, then acquisition is not possible. Surface phenomena such as the acoustic similarity of L2 sounds in relation to L1 sounds, or the type of acoustic cue involved in the perception of new contrasts, do not seem to play a role in the transfer of phonological features to the L2. Follow-up experiments (Brown, 2000) have shown again that features that are contrastive in the L1 can be redeployed to the L2, but not those that are redundant in the L1.

The hypothesis that it is the segment's feature geometry, and not the acoustic similarities of the input with regard to L1 sounds, that predicts different degrees of difficulty in L2 acquisition was also explored by LaCharité and Prévost (1999) who examined the perception of English /θ ~ t/, /ŋ ~ n/ in coda position, and /h/ ~ ∅ by native speakers of French. The tasks were all discrimination (AX with minimal pairs, ISI = 500 ms, and ABX, same minimal pairs, ISI = 1000 ms). The results showed an effect of contrast only in the task with the shorter ISI; the ABX task did not yield any significant differences. The conclusions suggest that creating a new articulator node, e.g., Pharyngeal for /h/, may be more difficult than adding a terminal feature to an existing node (such as [distributed] for /θ/). However, it is worth noting that the tasks in the study point to different results because they tap different types of knowledge, with the AX task being more likely to tap phonetic/acoustic perception than the ABX one; likewise, discrimination tasks with longer ISI are more likely to tap phonological knowledge (Werker and Tees, 1984; Werker and Logan, 1985).

The perception of Polish post-alveolar sibilants /ɕ, ʑ/ by naïve Croatian listeners was tested by Čavar and Hamann (2011). While these sounds do not have a phonemic status in Croatian, they do exist as allophones of different phonemes (of /s/ and /z/ respectively). The participants carried out an identification task

whose stimuli consisted of the target consonant with a prefixed vowel. While the Polish group performed at 99.7% accuracy, the Croatian group performed at 96%. Further language groups (Slovenian and German) showed poorer performance. The results suggest that Croatian speakers were able to transfer [continuant, voice, back] from their L1, unlike German and Slovenian speakers, since they lack this exact configuration of features within one sound. However, it is still possible that the Croatian listeners were able to map these allophonic representations onto their own phonemic categories, which are contrastive via different features.

All in all, experiments taking a feature-based approach may be leading to different results not just because of differences in the methodology, but also because the role of the features in the L1 grammar differs across experiments. However, generalizations regarding the validity of feature-based approaches are difficult to establish given the lack of empirical research on the acquisition of abstract phonological features in L2.

2.2 Experiments involving duration

Current approaches to phonological length propose that phonological length is not a property of the segment itself and therefore not a feature such as [long] proposed in SPE. Instead, and at least in the case of vowel length, a long vowel can be conceptualized as the result of two higher-level, prosodic representations linked to one lower-level representation (segment) (Odden, 2011). Therefore, and assuming that English /i/ ~ /i:/ is not a length-only contrast but a tense/lax one, the results observed in these experiments are not informative regarding feature-based approaches.

Experiments involving length-only contrasts have shown that length, when realized only via duration, is not difficult to perceive by nonnative speakers, regardless of their level of experience. The discrimination of consonant length by speakers of Korean, Cantonese, Vietnamese and Mandarin, tested on nonce words modeled after Polish phonology, yields higher d-prime scores for speakers of the first three languages, all of them with a *vowel* length contrast; Mandarin speakers, who do not have a length contrast at all in their L1, showed the lowest d-prime scores (Pajak and Levy, 2014). The perception of vowel length in experienced speakers of L2 Swedish with different native languages was tested by McAllister et al. (2002), where listeners had to listen to a real word and a nonce word, and determine which one was the real word based on vowel duration. The results showed that native speakers of L1 Estonian (with a three-way length distinction) performed best, followed by speakers of L1 English, and in last place, L1 Spanish. Length of residence (LOR) in Sweden did not play a role in the results of the L1 Spanish group, but there was a significant correlation between accuracy and LOR in the L1 English group; this suggests that speakers of languages that make use of a certain acoustic cue (if only partially) for establishing phonemic contrasts in the L1, may have an advantage over languages that do not use this cue for phonemic contrast.

² Even though Brown states that "Japanese, on the other hand, does not contrast any phonemes within the coronal place of articulation" (p. 153) and presents the Japanese vowel inventory without postalveolar consonants, one reviewer pointed out that Japanese does have alveolar/postalveolar minimal pairs, such as [sakai] "border" vs. [çakai] "society." The best account for this apparent contradiction is that (ç) is a merge of two underlying segments: /s/ and a palatal /j/ (Labruno, 2012, p. 67).

2.3 Experiments involving the tense-lax contrast

The [tense] feature has been traditionally used for the characterization of the difference between tense/lax segments such as /i/ ~ /ɪ/ in English (Giegerich, 1992) and German (Wiese, 2000). While the tense/lax contrast may be seen in phonological terms as simply another type of feature contrast, this is not necessarily the case. From a phonetic perspective, the contrast between tense and lax vowels has been shown to be based both on spectral and durational differences, whereby tense vowels are associated to longer duration as well as peripheral formant frequency values, in English (Hillenbrand et al., 1995; Leung et al., 2016) and German (Delattre and Hohenberg, 1968; Jessen et al., 1995). Since the inherent differences in duration in the tense/lax distinction are not encoded in the segmental representation as features, this leads to the question whether [tense] should instead be conceptualized as a suprasegmental only or replaced by the [ATR] feature, among other options (see Durand, 2005 for a detailed discussion).³

Experiments in the acquisition of vowel contrasts where the [tense] feature is involved show that L2 learners are able to attune their perception to the duration cue, even when this is not present in their L1. Bohn (1995) tested the perception of English vowels /i/ ~ /ɪ/ in L1 speakers of Spanish and Mandarin, who made use of the duration cue (not used in Spanish or Mandarin for phonemic contrasts) more than the spectral differences. Bohn called the phenomenon of “shutting down” the perception of an L1 cue in favor of a completely new one “the desensitization hypothesis,” which seems to take place in L2 listeners regardless of their language experience. Cebrian (2006) examined the perception of the tense-lax distinction in English vowel pairs /i/ ~ /ɪ/ and /e/ ~ /ɛ/ by native speakers of Catalan, whose native vowel system has /e/ ~ /ɛ/ but the realization differs only in terms of the spectral values. The results showed that L2 listeners over-rely on duration regardless of their level of experience in the L2. On the other hand, it also seems to be the case that L2 experience has an influence in the perception of acoustic cues: Escudero and Boersma (2004) looked into the perception of English /i/ ~ /ɪ/ by L1 Spanish speakers who were living in either southern England or Scotland and found out that L2 experience correlated to the listeners’ ability to match the acoustic cue weightings shown by the native speakers of the linguistic community in which they were immersed.

3 Phonological features, acoustic cues, and perceptual tasks

3.1 Acoustic cues = phonological features?

While all the experiments mentioned above touch upon the idea that non-native speakers can(not) learn L2 segments, either by transferring existing features or by learning completely new

ones, they are different in crucial ways. These experiments differ not only in the object to be measured (cues or features), but also in the assumptions made by the authors about them (for instance, whether they make the connection between cue and feature, or not). For instance, it is worth noting that some of these experiments (e.g., Bohn, 1995; Escudero and Boersma, 2004) do not straightforwardly refer to the idea of “feature” as an abstract phonological notion, since what the learners are being tested on is whether they are able to make a different use of the available acoustic cues in the input in order to acquire a new contrast. While this may count as a very concrete way to operationalize the abstract notion of feature into a specific acoustic cue, this is not *exactly* the same as testing for the acquisition of an abstract feature. Thus, if we assume that F1 values relate to [high] and [low], that F2 values relate to [back], and that both F1/F2 and duration relate somehow to [tense], then it could be hypothesized that L2 speakers who are learning how to process the acoustic cues in a target-like manner would in turn acquire an underspecified feature in the L1 that is relevant for a contrast in the L2.

However, while learning native-like cue weighting in L2 *may* lead to the creation of a new abstract feature, it does not necessarily entail the acquisition of the feature in question; that is, it might be a necessary condition, but not a sufficient one. For instance, it can be claimed that (a) learning new weights for the available cues could lead to a remapping of the cues onto different *existing* representations; that is, L2 sounds that were initially mapped onto the same L1 segment are now mapped onto two different ones. In the PAM/PAM-L2 framework, this would mean that learners would go from single-category assimilation to two-category assimilation. And (b), the learning of new cue-weightings could also lead to an in-between state where the difference is only noticeable at a pre-lexical level, but does not permeate to a more abstract, lexical representation.

Furthermore, do the specifics of the acoustic cue matter? In this regard, three points must be taken into consideration. Firstly, it is important that the acoustic cue associated to a contrast is relatively easy to perceive by the L2 listener. Archibald (2009) argues that the robustness of the acoustic cue is crucial, regardless of whether the learning operation is feature reassembly or feature acquisition. Experiments on the production of the /n/ ~ /ɲ/ contrast in Spanish by L1 English/L2 Spanish speakers suggest that both the weakness of the relevant acoustic cue as well as the functional load of the contrast are possible causes for a lack of L2 convergence (Stefanich and Cabrelli, 2021). By taking only this into account, a great deal of the variation in the results of different tests could be explained.

Secondly, it might be tempting to assert that abstract features can be redeployed to the L2, without checking for the nature of the cue. However, (and this is something that needs to be stressed!), the presence of a feature in the L1 does not mean that both the L1 and the L2 will use the same cue for it. Consider for instance how [voice] is present in the plosives of both English and Spanish, but they are instantiated by different cues (voicing in Spanish vs. VOT in English); the effect of such differences in acoustic cues that implement the same contrast can be seen even in experienced bilinguals (Flege and Eefting, 1987). In a perhaps less extreme case, and if we assume that [round] is

³ ATR is identified as a pharyngeal feature and has been more widely used nowadays as it may account for an array of phonological phenomena in different languages (Vaux, 1996, 1999). However, the exact place for this feature in the geometry is immaterial for the purposes of this research; here we will use [tense], the traditional terminology used in L2 phonology.

present and contrastive in English,⁴ perceiving front rounded vowels should not be that difficult for L1 English speakers; however, English speakers' difficulty in perceiving and producing the /u/ ~ /y/ contrast in French has been widely studied (e.g., [Flege and Hillenbrand, 1984](#); [Gottfried, 1984](#); [Flege, 1987](#)) and increased experience does not seem to improve accuracy ([Levy and Strange, 2008](#)). Regarding German, studies in perceptual assimilation show that English speakers perceive /y/ mostly as /u/ ([Polka, 1995](#)) even after years of L2 experience ([Mayr and Escudero, 2010](#)). These studies suggest that even though the relationship between the [round] and the [+back, -low] in English is not the same as the one in Spanish, the F2 values of [+high] vowels that are not assigned to /i/ are nevertheless automatically parsed as [+back] and therefore also as [+round], which could be the case for Spanish as well.

Finally, one third issue is whether it is possible that a cue that is used in the L1 to parse a certain contrast [+/-back] may be reused in order to convey another contrast such as [+/-round]. It is worth noting that the relevant acoustic cue for the /u/ ~ /y/ contrast is F2 ([Delattre et al., 1952](#); [Fant, 1971](#)), which is also a cue used for vowel backness; therefore, for listeners of languages with no front rounded/back unrounded vowels in their inventory, /y/ may be perceived simply along the backness dimension ([Lisker and Rossi, 1992](#)), thus being categorized as exemplars of either /i/ or /u/. Then, the problem may not be the saliency of the cue, but the fact that L2 listeners make use of this cue in their L1 to break down the perceptual space into their L1 categories along a dimension that is already taken. Thus, and regardless of whether these L1 categories are far apart from each other (such as /i/ and /u/ in Spanish), the listener's L1 perception has already assigned boundary values across the cue dimension (i.e. F2 for backness). In this regard, listeners acquiring an L2 contrast that can be perceived through a new cue where no boundaries for L1 phonemes have been assigned would have less difficulty than when they learn a contrast where the relevant cue is already in use.

3.2 Perceptual tasks and their relationship to levels of representation in perception

The idea that perceptual tasks yield different results due to the level of representation to which they are aimed has been tested empirically in several works, for instance, in perception of L2 input which is phonotactically illegal in L1 ([Freeman et al., 2022](#)) and in L1 pre-lexical perception ([Gerrits and Schouten, 2004](#)). A model that considers different modes of perception which are prompted by different tasks is the Automatic Selective Perception model ([Strange, 2011](#)), which proposes a *phonological* mode and a *phonetic* mode of perception. Listeners engage in phonological

mode mostly when using the L1, and it recovers only the necessary information for word recognition. This is achieved by means of over-learned, automatic selective perception routines that aim to make perception efficient; tasks that trigger phonological mode involve lexical decisions or grammaticality judgments. On the other hand, the phonetic mode is characterized by a more selective focus on the acoustic cues, which involves a higher cognitive load; tasks that elicit information under phonetic mode are those where the attentional focus is on the acoustic input, for instance, discrimination or category goodness ([Strange, 2011](#), p. 460). Furthermore, the ASP model distinguishes between auditory and perceptual saliency. While auditory saliency is largely language-independent and refers to physiological responses to acoustic stimuli, perceptual saliency refers to the strength of the reaction to stimuli, which is modulated by linguistic experience ([Strange, 2011](#), p. 458). Ultimately, it can be argued that the saliency of a given cue cannot be detached from the linguistic system that hosts it; thus, the only potential tool for quantifying saliency in behavioral experiments is to either look into acoustic values and calculate raw differences, or to manipulate the cues in a given stimuli.

One further point made in the ASP model is the difference between automatic and attentional processing, where automatic processing takes place when processing L1 sounds, or in L2 processing that has been automatized enough to resemble that of the L1. Automatic processing is expected to yield shorter reaction times, while attentional processing requires more time. When taking L1 reaction times as a baseline, especially during a task that requires accessing lexical representations, it is possible to estimate how well-learned a given L2 perceptual routine is.

In sum, the question whether learning new L2 segments is licensed by the presence of a given redeployable feature in the L1, or if new features are also learnable, is a rather complex one. The complexity lies not just on the lack of homogeneity in the methodology, but also on the intricate relationship between acoustic cues, phonological features, and perceptual tasks. Thus, the usually stated view that feature-based approaches have little predictive power in L2 acquisition should be reconsidered on the basis of the points made above: is feature redeployment independent of acoustic cues, and therefore, is redeployment fully abstract or phonetic-driven? And if the latter, does the nature, saliency, and L1 use of the cue play a role in the potential redeployment of a feature? The present study attempts to shed light on these questions.

4 This study

The current study tests whether adult Spanish learners of German can combine [-back] with [+round] in order to acquire /y/ as a segment that contrasts with /u/ (e.g., *Blüten/bluten*; 'blossom'/to bleed') via [back], and whether acquiring the contrast between /u/ and /y/ is easier than learning the /i/ ~ /ɪ/ contrast, where the addition of a new feature (namely, [tense]) is required. The Spanish vowel system /i e a o u/ does not have front rounded vowels, but it does have front unrounded vowels /i e/ and back rounded vowels /o u/. This vowel system always presents [+round] with [+back] and never with [-back], for which acquiring /y/ requires restructuring the possibilities that the L1 gives in terms

⁴ Granted, this is a difficult assumption, since back, non-high vowels vary greatly across varieties of English. However, some of the many feature matrices that have been proposed for the English vowel system (e.g., [Giegerich, 1992](#), p.110) show that for instance /a/ and /ɔ/ are contrastive only by (round). Likewise, the existence of a mid, back unrounded vowel such as /ɘ/ shows that English can at least decouple the [+round] feature from [+back, -low], which is not the case of Spanish.

of which features are allowed to combine into one segment. The learning process referring to the reassembly of features to create a new segment has been referred to as feature redeployment (Archibald, 2005). However, it is worth noting that this term refers to features that bear contrast on their own; in Spanish, [+round] is predictable from the features [+back, -low], for which it does not count as a fully contrastive feature in the L1. Since a potential scenario wherein [+round] in Spanish rebundles with [+high, -back] is not covered by the original definition of redeployment, we will extend this term to include L1 features that are present but do not bear contrast.

On the other hand, when L1 Spanish learners of German acquire the /i/ ~ /ɪ/ contrast (e.g., *Miete/Mitte*; 'rent'/'middle') the learning task is of a different nature. The only difference between /ɪ/ and /i/ in German is given by the feature [tense], which is a feature that Spanish does not have. However, Spanish does use duration (one of the acoustic cues involved in production and perception of the tense feature) as one out of several (and equally relevant) cues for stress (Ortega-Llebaria, 2006), but not for signaling phonemic contrasts; thus, L1 Spanish learners of German could in theory transfer this cue for the perception of new phonemic contrasts. Nevertheless, and as stated in section 3.1., the use of a given cue in the L1 does not entail the existence (or the automatic acquisition) of an abstract feature such as [tense], although it can be assumed that this is a preliminary step in order to acquire the abstract feature. We will call this learning task feature acquisition.

Furthermore, the differences in acoustic cues in terms of their saliency could be privileging one scenario over the other; that is, whether learning the contrast between /u/ and /ʏ/ is easier than /i/ ~ /ɪ/, or vice versa. As mentioned in section 2, experiments suggest that L1 Spanish learners of English are more sensitive to duration, as it seems to be for them an auditorily salient cue; thus, this could mean that they would have less difficulty acquiring a new feature with an unused, salient cue, than redeploying an existing one with a cue that the L1 uses for the perception of the front/back contrast.

Thus, the following hypotheses can be considered:

- 1) *Feature redeployment is as difficult as feature acquisition*: In this case, the presence of the feature in question in the L1 [i.e., (round)] does not facilitate acquisition of a contrast involving this feature in L2; furthermore, a contrast involving a new feature is also difficult to learn. The results in Barrios et al. (2016) seem to support this hypothesis, which would also be supported by this study if the perception of the non-native contrasts /i/ ~ /ɪ/, and /u/ ~ /ʏ/ is significantly worse than their baseline L1-like performance (/u/ ~ /i/).
- 2) *Feature redeployment is easier than feature acquisition*: Brown's (1998, 2000) main claim is that features that are present/active and fully specified in the L1 should be able to rebundle in order to create a new segmental representation. On the other hand, Brown's hypothesis states that underspecified features will not be acquired. If we restrict ourselves to the definition of redeployment as the transfer of fully contrastive features only, then neither contrast should be easier to acquire than the other. However, by extending the term to cover the transfer of redundant features, we would see better performance with the /u/ ~ /ʏ/ contrast (eventually, comparable to /i/ ~ /ɪ/), assuming that redeployment is

licensed by the presence of the feature in the L1, and not by its ability to bear contrast on its own in the L1 grammar.

- 3) *Feature acquisition is easier than feature redeployment*: In this scenario, the blank slate benefits learning, and/or it might be aided by the acoustic saliency of the cues involved in a given contrast. Such results would be in line with Bohn (1995) and the Desensitization Hypothesis. Better performance with the /i/ ~ /ɪ/ contrast (and comparable to /u/ ~ /ʏ/) would support this hypothesis.
- 4) *Feature redeployment is as easy as feature acquisition*: the most optimistic scenario is where native-like performance can be seen in both cases. While at least there is evidence that a fair amount of learning does take place (e.g., Escudero and Boersma, 2004), a scenario with ceiling effects for feature acquisition has, to the best of our knowledge, not been attested. Here, both nonnative contrasts /i/ ~ /ɪ/ and /u/ ~ /ʏ/ would be comparable to /u/ ~ /i/.

In order to test these hypotheses, the perceptual tasks need to tap different levels of representation. Since the /i/ ~ /ɪ/ contrast is aided by a new, acoustically salient cue, we set out to probe whether cue saliency facilitates perception at a more superficial level (that is, a phonetic mode of perception, without resorting to long-term representations). In this sense, an AX discrimination task would shed light on whether listeners are able to weight the acoustic cues involved in the contrasts in a target-like manner. However, potentially increased sensitivity to a cue does not necessarily entail the acquisition of a contrast at a phonemic level; therefore, a task that prompts the listener to make use of abstract representations, such as a categorization task, is needed. In this regard, a picture identification task is ideal for this purpose as it excludes potential orthographic effects in the responses as well as it probes the existence of L2 phonemic representations encoded at the lexical level.

Furthermore, reaction times (RT) are also measured, as it will provide a better understanding of the underlying cognitive process in the L2 learner, especially in terms of the automaticity of their perceptual routines. Differences in RTs by contrast, and particularly between native and nonnative contrasts, should indicate whether there is a substantial difference when processing the stimuli.

5 Methods

The experiment was carried out online. Most of the participants were recruited by word of mouth, while others were contacted via social networks. The experiment was set up on an online platform created specifically for this purpose in HTML and JavaScript using JsPsych (De Leeuw, 2015) in order to measure reaction times in a precise manner.⁵ Participants were asked to use their keyboard instead of the mouse when giving responses (i.e., to

5 As of 2020, JavaScript-based online experiments have been deemed less accurate than lab-based studies, but still within acceptable ranges in terms of variability (De Leeuw and Motz, 2016; Pronk et al., 2020; Anwyl-Irvine et al., 2021), where jsPsych shows a mean precision between 3 and 7 ms across operating systems and browsers (Bridges et al., 2020).

press Q for the option showing on the left side of the screen, and P for the one to the right). All instructions were given in German.

5.1 Participants

Participants ($N = 19$, mean age: 35.9 years) identified themselves as native speakers of Spanish, with a self-reported B2 CEFR level of German. Most of them learned German after age 18 (two participants started at age 9 and 5) and were living in Germany for at least a year at the time they were tested (one participant did not report, and another one had spent only 2 months). They also reported having used a variety of learning methods, of which the most frequent were language classes, both at their home country before their arrival and in Germany. Two of them reported two native languages: one Spanish and Catalan, and another Spanish and English; these participants were excluded.

5.2 Materials

Nineteen German words containing the vowels /y/, /i/, /u/, and /ɪ/ were recorded by a phonetically trained female native speaker of German. The selection of a small number of words is due mostly to the online nature of the experiment, where participants are more likely to get distracted or quit the experiment before finishing, and considering that repetition was needed in order to run statistical analyses. Furthermore, the picture identification task required that the words could be easy to represent in an image, which also increases the likelihood of being understood by B2-level speakers.

In order to avoid differences in intonational curves and obtain similar durations, the words were embedded in the carrier sentence *Ich sage ____ noch mal* ('I say ____ again'). Each word was recorded three times, and for each word one token was selected (ideally without clicks, vocalized consonants, or creaky voice). All tokens had the same intonation curve due to the embedding in the carrier sentence. For the AX task, trials with different words were manipulated with Praat so that the both words also had the same onset, coda, and schwa lengths while keeping the original pitch; this was to ensure that the only difference was the vowel itself and nothing else in the stimuli. While onsets were rather consistent in terms of duration, words ending with vocalized consonants or sonorants had different durations. In pairs where the differences in duration was small enough to not distort the pitch and both vowels were tense, the sound files were scaled to the average duration. In pairs with the /i/ ~ /ɪ/ contrast and/or with considerably different durations, the final sound in the longer word was shortened, avoiding abrupt changes in intensity and respecting zero-crossings. The list of words can be seen in [Table 1](#), arranged by minimal pairs per contrast (see [Table 1](#)). In order to rule out a potential assimilation of /y/ to /i/ (*contra* the evidence from L1 English learners of French and German), a fourth minimal pair /i/ ~ /y/ was added.

5.3 Procedure

After the consent and language background questionnaire, L1 Spanish learners of German were tested for perception of the /u/ ~ /y/, /i/ ~ /ɪ/, /i/ ~ /y/, and /u/ ~ /i/ contrasts with an AX (same-different) discrimination task with an ISI of 1500 ms (as in [Barrios et al., 2016](#)), followed by a two-way, forced-choice picture identification task. The online setup included a break screen between the tasks, which recommended the participant to take a 5-min break. The total experiment time was around 10 min.

5.3.1 AX discrimination task

The words were taken apart from the carrier sentence with Praat ([Boersma and Weenink, 2024](#)) and presented to the participants with an ISI of 1500 ms. Participants heard two German words consisting of a minimal pair from [Table 1](#) (e.g., *Blüten* and *bluten*); duplicates of the same word were also included. There was a total of 48 randomized trials. In trials where the words were different, the minimal pair was presented twice (once with each word in first place; that is, WordA – WordB, and then WordB – WordA). There was also one instance of WordA – WordA and one of WordB – WordB. During each trial, the screen showed the question: *Sind die Wörter gleich oder unterschiedlich?* ("Are the words the same or different?") The buttons provided were *Gleich* ("Same") on the left and *Unterschiedlich* ("Different") on the right. Participants had a maximum time of 5 s to provide a response.

5.3.2 Two-way forced-choice picture identification task

Participants carried out a picture identification task where only two alternatives were given. The same words in the previous task were here presented in isolation, while showing the participants two pictures (e.g., if the aural stimulus was the word *Blüten*, the pictures shown corresponded to *Blüten* and *bluten*) with the question *Welches Wort haben Sie gehört?* ("Which word did you hear?"). Participants had a maximum time of 5 s to make a decision. Each stimulus was shown three times; with 18 trials per contrast, there was a total of 72 randomized trials.

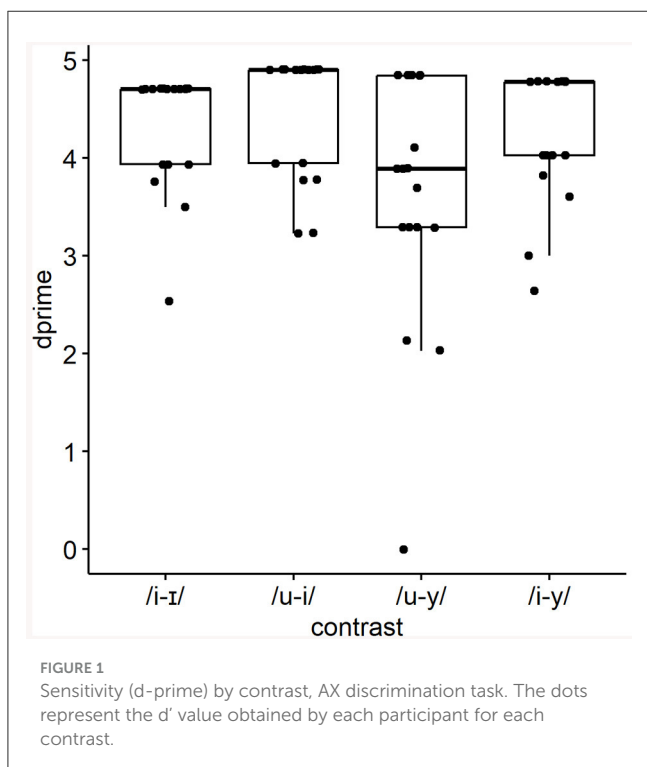
6 Results

6.1 AX discrimination task

Sensitivity (d') yields a single score per participant by vowel contrast, and is calculated by taking into account the z-scores of the total counts of hits minus false alarms ([MacMillan and Creelman, 1991](#)). In order to compute d' , all trials containing e.g., the /u/ ~ /i/ contrast (WordA-WordB, WordB-WordA, WordA-WordA and WordB-WordB for all three minimal pairs) were tabulated in terms of the participant's hits (answered "different" when the words were different), false alarms (answered "different" when the words were the same), misses (answered "same" when the words were different), and correct rejections (responded "same" when the words were the same). These d' scores were calculated with *R* (v.4.3.1; [R Core Team, 2023](#)) and the package *SensR* ([Christensen](#)

TABLE 1 List of German words used in the experiment.

/u/ ~ /i/	/i/ ~ /ɪ/	/u/ ~ /y/	/i/ ~ /y/
Tour – Tier (“tour” – “animal”)	Miete – Mitte (“rent” – “middle”)	Tour – Tür (“tour” – “door”)	Biene – Bühne (“bee” – “stage”)
Spulen – spielen (“spools” – “to play”)	Schief – Schiff (“crooked, not straight” – “ship”)	bluten – Blüten (“to bleed – blossom”)	spielen – spülen (“to play” – “to rinse”)
Stuhl – Stiel (“chair” – “stick”)	Stiel – still (“stick” – “quiet”)	Spulen – spülen (“spools” – “to rinse”)	Kiel – kühl (“Kiel” – “cool”)



and Brockhoff, 2017). In order to correct infinity values, a log-linear approach was used. Since the d' data violated normality assumptions, a series of non-parametric tests using the R package *rstatix* (Kassambara, 2023) were carried out. A Friedman test showed a significant effect of contrast on sensitivity expressed as d' values ($\chi^2 F(3) = 13.4, p < 0.01$). A Kendall's W test for effect size ($W = 0.26$) indicates a small effect. A *post-hoc*, two-sided Wilcoxon signed-rank test with a Bonferroni adjustment comparing performances in all three non-native contrasts against the native /u/ ~ /i/ showed that the only significant difference is that between /u/ ~ /i/ and /u/ ~ /y/ ($T = 136, p < 0.05$). Figure 1 shows the distribution of d' scores by contrast, where the differences in distribution by contrast can be seen: overall, the participants' d' was lower for the /u/ ~ /y/ contrast.

Regarding reaction times (RT), trials with null responses were discarded. Then, two analyses were carried out. First, differences in RT for all responses across stimuli pairs grouped by the vowels present in the stimuli (e.g., /u/ ~ /i/, /i/ ~ /ɪ/, /u/ ~ /y/) were probed with a mixed effects model using the R package *lme4* (Bates et al., 2014), with stimuli pair as a fixed effect, and subject and item as random intercepts. The model shows a significant effect of stimuli pair ($\chi^2 = 21.7, df = 7, p < 0.001$). Figure 2 shows the distribution of RT values for all stimuli pairs.

A comparison across all pairs (same and different) with least-squared means was calculated using the *emmeans* package (Lenth, 2022). The reaction times for stimuli pair /u/ ~ /y/ showed to be significantly higher than all other pairs, except /i/ ~ /ɪ/, /ɪ/ ~ /i/, and /y/ ~ /y/. Table 2 shows the statistical values for each significant comparison.

The values show that participants are considerably slower when discriminating between /y/ and /u/ than all same stimuli pairs, with the exception of the /ɪ/ ~ /ɪ/ and /y/ ~ /y/ pairs. On the other hand, participants were equally slow at reacting to the different pair /i/ ~ /ɪ/, which suggests that the non-native /ɪ/ triggers longer reaction times, even though the decisions made throughout the task with pairs involving this sound were mostly correct. It is also worth noting that trials with the /i/ ~ /y/ pair yielded faster reaction times than those with /u/ ~ /y/, which suggests that the participants perceive these sounds as clearly different and thus do not perceptually assimilate /y/ to their native category /i/.

One further mixed-effects model but only with the trials with different stimuli, with stimulus pair as predictor, and item and subject as random intercepts was fitted, and again there was a significant effect of stimulus pair ($\chi^2 = 14.39, df = 3, p < 0.01$). The least-squares mean analysis showed the same significant differences for /u/ ~ /y/ and /u/ ~ /i/ ($\beta = 354.67, SE = 107, t = 3.32, p < 0.05$) as well as for /u/ ~ /y/ and /i/ ~ /y/ ($\beta = 345.60, SE = 107, t = 3.24, p < 0.05$).

6.2 Picture identification task

Regarding the picture identification task, we carried out a mixed-effects model with correct/incorrect as the dependent variable, vowel contrast as predictor, and subject and item as random intercepts ($\chi^2 = 66.93, df = 3, p < 0.001$). The participants showed very high performance when choosing between /u/ and /i/ (95% correct); similar results were obtained for the non-native contrast /i/ ~ /y/ (99% correct). On the other hand, non-native contrasts /i/ ~ /ɪ/ and /u/ ~ /y/ yielded significantly lower correct counts: 76% and 73%, respectively. Figure 3 shows the categorizations by contrast.

A *post-hoc* comparison between all the contrasts with least-square means (with Tukey adjustment for p -values) showed significant differences across all contrasts, except between /i/ ~ /ɪ/ and /u/ ~ /y/, as well as /u/ ~ /i/ and /i/ ~ /y/. Table 3 presents the results of pairwise comparisons.

This task's RTs yielded different patterns by contrast, where the contrasts /u/ ~ /i/ and /i/ ~ /y/ displayed lower reaction times (Figure 4). However, a mixed-effects model including all

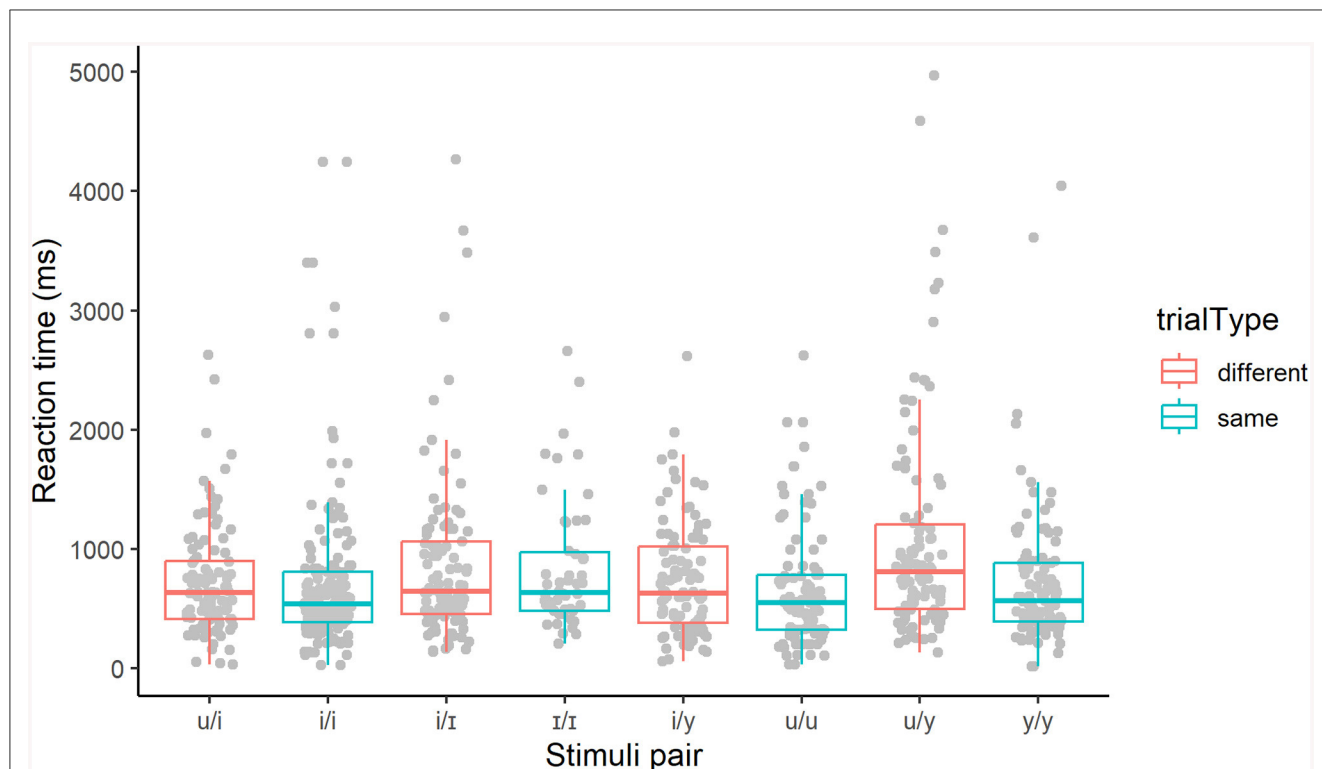


FIGURE 2 AX discrimination task: Distribution of reaction times (RT) by stimuli pair (all responses).

TABLE 2 Statistics for pairwise comparisons, reaction times.

Contrast comparison	Estimate	SE	t.ratio	p-value
/u/ ~ /y/ and /u/ ~ /i/	353.38	104.7	3.38	<0.05
/u/ ~ /y/ and /i/ ~ /y/	340.72	98.7	3.45	<0.05
/u/ ~ /y/ and /i/ ~ /i/	345.60	104.6	3.31	<0.05
/u/ ~ /y/ and /u/ ~ /u/	426.75	111.2	3.83	<0.01

responses except time-outs, with vowel contrast as fixed effect and subject and item as random intercepts showed that these differences in reaction times across vowel contrasts are not significant ($p = 0.11$).

7 Discussion

The present study has attempted to determine whether the /u/ ~ /y/ contrast found in German is more likely to be acquired by L1 Spanish learners of German than the /i/ ~ /I/ contrast, with four hypotheses in sight. Hypothesis 1 points to a no-learning scenario, with the native-like /u/ ~ /i/ contrast unmatched in perception. Hypothesis 2 follows feature-based approaches, wherein accurate perception of new L2 sounds can be facilitated by the presence of the same feature in in the L1. Hypothesis 3 favors the acquisition of the /i/ ~ /I/ contrast, as duration is an auditorily salient cue that offers learners a blank slate where new categories can be created. Finally, Hypothesis 4 predicts a full-learning scenario where all

nonnative contrasts are perceived in a native-like manner. To this end, two perceptual tasks were carried out: one tapping phonetic knowledge (AX discrimination), and another tapping phonological knowledge (picture identification).

First, the results suggest that /y/ is unequivocally assimilated to /u/ (or rather, perceived as certainly not /i/), in accordance with studies on the perception of /y/ by L1 English learners of German and French. This can be seen in the results of the discrimination task, where d-primes and reaction times for /i/ ~ /y/ were not significantly different than those for the native pair /u/ ~ /i/. Furthermore, the picture identification task yielded excellent accuracy for /i/ ~ /y/. This suggests that, in the presence of a [+/-round] contrast such as /i/ ~ /y/, perception is driven by the acoustic cue F2 in such a way that a value between the usual ones for categories /i/ and /u/ is perceived as [+back], and not as [-back].

The fact that the same group of listeners showed lower accuracy for both non-native contrasts in the picture identification task (where more abstract representations are in use by the listener) but performed better with a contrast involving a new feature

(plus an additional cue) in the AX discrimination task (which focuses on prelexical perception) shows that the acquisition of the /u/ ~ /y/ contrast by native speakers of Spanish is not aided by the presence of the relevant phonological features [round, back, high] in the L1. Even if we interpret the more-than-above-chance performance in both tasks as a signal of facilitation, the even better performance with /i/ ~ /ɪ/ in the AX discrimination task shows that the active presence of the features in question in the L1 are not necessarily an advantage in relation to absent features such as [tense]. However, the following points should be taken into consideration.

7.1 Differences by task

The results of the AX discrimination task suggest that the /u/ ~ /y/ contrast is more difficult to acquire than /i/ ~ /ɪ/, with the latter being closer to their performance in the native contrast /u/ ~ /i/; based on this task alone, it may be concluded that L2 speakers are more likely to acquire a contrast when the feature in question is *not* present in the L1, therefore supporting Hypothesis 3. On the other hand, the results for the picture identification task show that

both non-native contrasts are significantly more difficult than the native one; that is, this task would support the idea that acquiring non-native contrasts is difficult regardless of whether the feature in question can be transferred or not (Hypothesis 1). Furthermore, the tasks yielded different patterns for reaction times: while the AX discrimination task was able to yield significant differences, the picture identification task did not.

The best approach for explaining the seemingly contradictory results would be to assume that discrimination is chiefly a phonetic task, regardless of the relatively long ISI and minimal pairs used. While shorter ISI might facilitate the retrieval of acoustic details stored in the short-term memory, the possibility that the participants were still able to remember the acoustic detail of the previous word after 1.5s is also possible, albeit to a lesser extent. Furthermore, the stimuli were recorded by one speaker, which does not require the listener to filter out interspeaker variation. Hence, learners may be less likely to rely on long-term representations when providing answers to the task. On the other hand, the picture identification task would not depend on acoustic detail and the listeners would need to rely on long-term representations; thus, changes in the perception of non-native contrasts due to L2 experience is more likely to be attested when

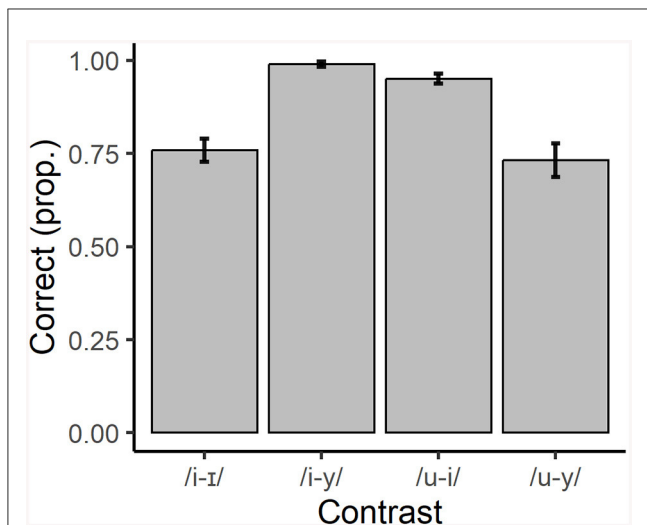


FIGURE 3 Mean proportion of correct responses by contrast in the picture identification task. Error bars show a 95% confidence interval.

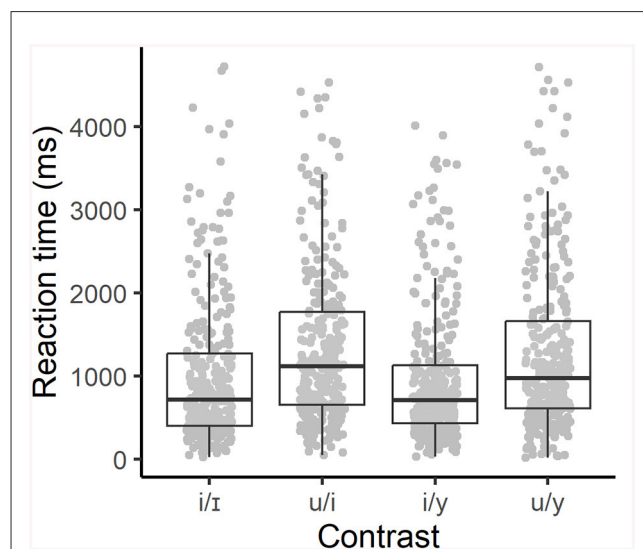


FIGURE 4 Reaction times by contrast (all responses), picture identification task. Dots show RTs per trial.

TABLE 3 Pairwise comparisons, identification task.

Contrast	Estimate (log odds)	SE	z.ratio	p-value
/u/ ~ /i/ and /i/ ~ /ɪ/	1.93	0.36	5.38	<0.0001
/u/ ~ /i/ and /i/ ~ /y/	-1.67	0.66	-2.53	0.056
/u/ ~ /i/ and /u/ ~ /y/	2.08	0.36	5.82	<0.0001
/i/ ~ /ɪ/ and /i/ ~ /y/	-3.59	0.62	-5.77	<0.0001
/i/ ~ /ɪ/ and /u/ ~ /y/	0.15	0.28	0.55	0.94
/i/ ~ /y/ and /u/ ~ /y/	3.73	0.62	6.01	<0.0001

tapping surface-like knowledge, and less so when looking into higher levels of representation.

The facilitatory effect of the AX discrimination task, with no inter-speaker variation, can be seen in the reaction times: not requiring a long-term representation to respond to the stimuli may have removed any potential differences due to lack of knowledge of a given stimulus. The significantly slower reaction times for /u/ ~ /y/ can be interpreted exclusively as a function of the difficulty of perceiving the difference between sounds, even when having the opportunity to direct their attention to the cue in question.

From a formal phonological perspective, the claim in [Hancin-Bhatt \(1994\)](#) about feature prominence, that is, how relevant the feature is in the grammar of the L1, may also account for the poorer performance for /u/ ~ /y/ in the AX discrimination task. In Spanish, the feature [+round] is redundant with [+back, -low], for which its prominence is not high. However, it is worth noting that this same contrast in L1 English learners of French is also difficult to perceive and produce, even when English does have [+back, -round], although only in low vowels. Here one could assume that since English has segments with the [+back, -round] and the [+back, +round] feature bundling, it could be easier for L1 English learners of L2 French to redeploy [+/-round] to [-low] vowels. It seems that this is not the case, and that in the end L1 speakers of both English and Spanish behave the same when it comes to the perception of /y/. This suggests that perhaps feature prominence needs to be coupled with the salience and functional load of the acoustic cues involved in the contrast.

One further point to consider is that the frequency effect from the words chosen for the stimuli could have affected the results for the picture identification task. In this regard, an item analysis shows a higher number of incorrect responses in the discrimination task for the stimuli pair *bluten/Blüten* (47%, while all other items show an incorrect rate between 0% and 16%), and a 33% of incorrect responses in the picture identification task. A post-experiment analysis of word frequency with German corpora ([Leipzig Corpora Collection, 2022](#)) was carried out, which showed that the frequency class of *bluten* and *Blüten* is 15 and 12, respectively (the higher the number, the less frequent the word); however, the analogous pair *Spulen/spülen* (frequency class 17 and 16, respectively) yielded higher correct response ratios (95% in discrimination and 79% in identification). It is worth noting that even though frequency in corpora may provide a range of probability as to whether an L2 learner has come across a given word, it may not be the best way to estimate the learner's familiarity with a certain lexical item, as L2 learners usually learn their vocabulary through modified input (e.g., textbooks, student materials) that does not contain the same lexical items and grammatical structures as the texts contained in corpora (newspapers, websites, etc. written and read by L1 speakers).

Furthermore, the reaction times for the /u/ ~ /y/ contrast in the picture identification task were not significantly slower than those obtained for the other contrasts, which suggests that the *bluten/Blüten* pair may have been more difficult to perceive not due to lexical frequency, but perhaps to the phonological context: together with *Miete/Mitte*, in this pair the vowel is preceded by a sonorant and followed by a plosive.

7.2 Differences by contrast

The differences in performance by contrast observed in the AX discrimination task suggest that learners could be more sensitive to auditorily salient acoustic cues when the attention is fully focused on perceiving differences between sounds, and not on lexical access. [Table 4](#) shows the differences in the acoustic cues present in the stimuli. Again, it is worth noting that these differences in acoustic cues do not play a significant role when the task prompts listeners to rely on long-term representations.

The values in [Table 5](#) show how the acoustic cues differ by contrast. Two large differences in F1 and duration can be seen in /i/ and /i/ (F1 and duration); on the other hand, one large difference between /u/ and /y/ is observed (F2). F3 values show relatively small differences across contrasts, especially in Bark. Regarding the perceptual assimilation of /y/, these values show, and especially in the Bark scale, that /y/ is much more distant from /u/ than /i/ in terms of F2; yet, the listeners seem to favor a perceptually and auditorily distant vowel over the closer one. Duration has also been deemed a more perceptually salient cue than frequency ([Bohn, 1995](#)), which seems to be relevant in a task where the focus is not on abstract representations but in picking up acoustic cues; likewise, most of the studies showing successful learning of new perceptual cue weightings and/or acquisition of new features focus on duration or quantity.

However, it is worth stressing that (a) duration seems to be salient to L2 listeners even when this cue is not relevant for L1 perception; and (b) the level of perceptual saliency may be influenced not only by linguistic experience, but also by its role in the L1 regarding perception of other feature contrasts. Judging at least by the results obtained here, F2 seems to be a perceptually salient cue that is weighted by L1 Spanish speakers in such a way that non-peripheral F2 values are categorized as /u/ regardless of its acoustic proximity to /i/; on the other hand, F1 and duration seem to be more acoustically salient as the differences are larger. Furthermore, the studies in section 2 show how duration is a cue that Spanish speakers are reported to rely on, despite being irrelevant in their L1 for the perception of phonemic contrasts. It seems then, that acoustic cues to L1 vowel contrasts are more difficult to be recalibrated in order to perceive a category whose values lie between two L1 vowels; on the other hand, duration provides a blank slate that does not need to be shared with other L1 categories.

7.3 The role of the cue feature redeployment

While the discussion above seems to point to the conclusion that Brown's FBM has no predictive power, several aspects need to be taken into account. Apart from the limitation derived from the role of [round] in Spanish, this study showed a significant difference between pre-lexical and lexical levels of perception, which are conceptualized in terms of phonetic and phonological mode by [Strange \(2011\)](#) ASL model. In this regard, it is likely that tasks prompting a phonetic mode of perception show better results, as the listener's attentional focus is on the acoustic cues. But

TABLE 4 Mean acoustic values of vowels in stimuli and standard deviation (SD), in Hertz and Bark.

Vowel	F1 Hz (SD)	F1 Bark (SD)	F2 Hz (SD)	F2 Bark (SD)	F3 Hz (SD)	F3 Bark (SD)	Duration, ms (SD)
i	244 (23)	2.58 (0.23)	2428 (653)	13.94 (2.31)	3559 (356)	16.78 (0.73)	150 (50)
r	502 (49)	4.97 (0.42)	1901 (179)	12.54 (0.64)	3042 (293)	15.71 (0.68)	77 (2)
y	290 (39)	3.03 (0.39)	1726 (147)	11.91 (0.56)	2615 (153)	14.69 (0.39)	141 (48)
u	352 (72)	3.61 (0.67)	719 (224)	6.58 (1.52)	2956 (98)	15.53 (0.23)	156 (18)

TABLE 5 Differences in acoustic cues by contrast.

Contrast	$\Delta F1$ (Hz)	$\Delta F1$ (Bark)	$\Delta F2$ (Hz)	$\Delta F2$ (Bark)	$\Delta F3$ (Hz)	$\Delta F3$ (Bark)	Δ Duration (ms)
i ~ r	258	2.39	527	1.4	517	1.07	85
u ~ i	108	0.45	1709	7.36	603	1.25	32
u ~ y	62	0.59	1007	5.33	341	0.84	10
i ~ y	46	1.04	702	2.03	944	2.09	42

even though this type of task yields better performance, there is a limitation (and this is the second point): is the cue in question salient enough regardless of its relevance in the L1? What is the role of this cue in the L1? And how many cues are available for the learner?

A further point, also related to acoustic cues, is the relationship between cues and features. Is attunement to a certain cue a *conditio sine qua non* for the acquisition of a phonological feature? Our research has shown that even though duration is not relevant in the L1 for establishing phonemic contrasts, the listeners in this study are able to use this cue in the L2, but it still seems to be insufficient when the cognitive load in the task increases. However, the moderate-to-high performance in both tasks shows that learning does take place, so Hypothesis 1 (i.e. that both reactivating and rebundling features are equally difficult learning tasks) is valid only insofar as we compare this to L1 performance. One thing is clear: learning does take place, though the performance cannot be compared to native-language contrast perception.

8 Conclusion

Nonnative speakers with a reasonably high level of L2 experience are relatively more successful when discriminating between sounds where a new feature is acquired, than when the acquisition of a phoneme requires existing features (however underspecified) to be redeployed; however, neither contrast is easier than the other one when it comes to identification. These results support conflicting hypotheses: while the discrimination task supports a theory of acquisition along the lines of Bohn (1995) Desensitization Hypothesis, the identification task seems to replicate the results in Barrios et al. (2016). That is, this work found no supporting evidence of redeployment as posited by the FBM; however, further studies where the feature to be redeployed is fully contrastive would shed more light on the issue.

The evidence shown here suggests that the differences in performance between tasks and in the saliency of the acoustic cues involved (duration vs. F2) may be due to L2 phonological

acquisition taking place at a surface level, which could explain why the L2 speakers' sensitivity to acoustic cues is higher when the cue is acoustically salient (assuming that the larger the difference in values for the cue, the more auditorily salient the cue is), regardless of whether the learning task is redeployment or acquisition of a new feature. This may in turn suggest that new, auditorily salient cues would aid the learning of completely new features over internal rearrangement of abstract features, albeit at a surface, pre-lexical level.

Finally, among the limitations in this study is the inability to completely control for the type of acoustic cues to be tested, either in terms of its saliency or informativity in the L1; furthermore, the familiarity of the words used in the experiment may have differed by participant, although the minimum proficiency requirement, plus the instructions being given in German, made it less likely for the participants to be completely unfamiliar with the lexical items. One further point to address is that we did not control for exposure to English, which has an analogous vowel contrast /i~/ɪ/, though the realizations are not quite the same. Nevertheless, we hope that the present study motivates further research looking into the role of abstract phonological features in L2 speech perception, where comparisons between L2 feature acquisition and redeployment can be made on the basis of equally salient acoustic cues, and contrasts whose features in the L1 grammar are fully contrastive.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found at: <https://github.com/pandabar/features>.

Ethics statement

Approval for experiments within the Department of Linguistics was granted by the Ethics Committee at the University of

Konstanz on 04.02.21 (IRB statement 05/2021). The studies were conducted in accordance with the local legislation and institutional requirements. The participants provided their written informed consent to participate in this study.

Author contributions

FB: Conceptualization, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare no financial support was received for the research, authorship, and/or publication of this article.

References

- Anwyl-Irvine, A., Dalmaijer, E. S., Hodges, N., and Evershed, J. K. (2021). Realistic precision and accuracy of online experiment platforms, web browsers, and devices. *Behav. Res. Methods* 53, 1407–1425. doi: 10.3758/s13428-020-01501-5
- Archangeli, D. (1988). Aspects of underspecification theory. *Phonology* 5, 183–207. doi: 10.1017/S0952675700002268
- Archibald, J. (2005). Second language phonology as redeployment of L1 phonological knowledge. *Can. J. Ling.* 50, 285–314. doi: 10.1017/S0008413100003741
- Archibald, J. (2009). Phonological feature re-assembly and the importance of phonetic cues. *Second Lang. Res.* 25, 231–233. doi: 10.1177/02676583100284
- Avery, P., and Rice, K. (1989). Segment structure and coronal underspecification. *Phonology* 6, 179–200. doi: 10.1017/S0952675700001007
- Barrios, S., Jiang, N., and Idsardi, W. J. (2016). Similarity in L2 phonology: evidence from L1 Spanish late-learners' perception and lexical representation of English vowel contrasts. *Sec. Lang. Res.* 32, 367–395. doi: 10.1177/0267658316630784
- Bates, D., Mächler, M., Bolker, B., and Walker, S. (2014). Fitting linear mixed-effects models using lme4. *J. Statist. Softw.* 67, 1–48. doi: 10.18637/jss.v067.i01
- Best, C., and Tyler, M. (2007). “Nonnative and second-language speech perception: commonalities and complementarities,” in *Language Experience in Second Language Speech Learning: In Honor of James Emil Flege*, eds B. Ocke-Schwen and M. J. Munro (Amsterdam: John Benjamins), 13–34.
- Best, C. T. (1995). “A direct realist view of cross-language speech perception,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 171–204.
- Boersma, P., and Weenink, D. (2024). *Praat: Doing Phonetics by Computer [Computer Program]*. 6.1. Available online at: <http://www.praat.org> (accessed December 29, 2023).
- Bohn, O. S. (1995). “Cross language speech production in adults: first language transfer doesn't tell it all,” in *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, ed W. Strange (Timonium, MD: York Press), 279–304.
- Bridges, D., Pitiot, A., MacAskill, M. R., and Peirce, J. W. (2020). The timing mega-study: Comparing a range of experiment generators, both lab-based and online. *PeerJ* 8, e9414. doi: 10.7717/peerj.9414
- Brown, C. (2000). “The interrelation between speech perception and phonological acquisition from infant to adult,” in *Second Language Acquisition and Linguistic Theory*, ed J. Archibald (Wiley-Blackwell).
- Brown, C. A. (1998). The role of the L1 grammar in the L2 acquisition of segmental structure. *Second Lang. Res.* 14, 136–193. doi: 10.1191/02676589869508401
- Čavar, M., and Hamann, S. (2011). “Phonemes, features and allophones in L2 phonology,” in *Generative Investigations: Syntax, Morphology, and Phonology*, eds P. Bański, B. Łukaszewicz, M. Opalińska, and J. Zaleska, 308–320.
- Cebrian, J. (2006). Experience and the use of non-native duration in L2 vowel categorization. *J. Phonetic.* 34, 372–387. doi: 10.1016/j.wocn.2005.08.003
- Christensen, R. H. B. and Brockhoff, P. B. (2017). *sensR - An R-Package for Sensory Discrimination*. R package version 1.5-0. Available online at: <http://www.cran.r-project.org/package=sensR/>
- Clements, G. N. (1985). The geometry of phonological features. *Phonology* 2, 225–252. doi: 10.1017/S0952675700000440
- Clements, G. N., and Hume, E. (1995). “The internal organization of speech sounds,” in *The Handbook of Phonological Theory*, ed J. Goldsmith (Cambridge, MA: Blackwell), 245–306.
- De Leeuw, D., and Motz, J. R. (2016). Psychophysics in a Web browser? Comparing response times collected with JavaScript and Psychophysics Toolbox in a visual search task. *Behav. Res. Methods* 48, 1–12. doi: 10.3758/s13428-015-0567-2
- De Leeuw, J. R. (2015). jsPsych: a JavaScript library for creating behavioral experiments in a Web browser. *Behav. Res. Methods* 47, 1–12. doi: 10.3758/s13428-014-0458-y
- Delattre, P., and Hohenberg, M. (1968). (1968). Duration as a cue to the tense/lax distinction in German unstressed vowels. *J. Int. Rev. Appl. Ling. Lang. Teach.* 6, 367–90. doi: 10.1515/iral.1968.6.1-4.367
- Delattre, P., Liberman, A. M., Cooper, F. S., and Gerstman, L. J. (1952). An experimental study of the acoustic determinants of vowel color; observations on one- and two-formant vowels synthesized from spectrographic patterns. *Word* 8, 195–210. doi: 10.1080/00437956.1952.11659431
- Durand, J. (2005). “Tense/Lax, the vowel system of English and phonological theory,” in *Headhood, Elements, Specification and Contrastivity: Phonological Papers in Honor of John Anderson*, eds P. Carr, J. Durand, and C. Ewen (Amsterdam: John Benjamins), 77–98. doi: 10.1075/cilt.259.08dur
- Escudero, P., and Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Stu. Sec. Lang. Acquisition* 26, 551–585. doi: 10.1017/S0272263104040021
- Fant, G. (1971). *Acoustic Theory of Speech Production: With Calculations Based on X-ray Studies of Russian Articulations (No. 2)*. London: Walter de Gruyter.
- Flege, J. E. (1987). The production of new and similar phones in a foreign language: evidence for the effect of equivalence classification. *J. Phonetic.* 15, 47–65. doi: 10.1016/S0095-4470(19)30537-6
- Flege, J. E. (1995). Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research* 92, 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*, ed R. Wayland (Cambridge: Cambridge University Press), 3–83.
- Flege, J. E., Bohn, O. S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of English vowels. *J. Phonet.* 25, 437–470. doi: 10.1006/jpho.1997.0052
- Flege, J. E., and Eefting, W. (1987). Production and perception of English stops by native Spanish speakers. *J. Phonetic.* 15, 67–83. doi: 10.1016/S0095-4470(19)30538-8
- Flege, J. E., and Hillenbrand, J. (1984). Limits on phonetic accuracy in foreign language speech production. *The J. Acoustic. Soc. Am.* 76, 708–721. doi: 10.1121/1.391257
- Freeman, M. R., Blumenfeld, H. K., Carlson, M. T., and Marian, V. (2022). First-language influence on second language speech perception depends on task demands. *Lang. Speech* 65, 28–51. doi: 10.1177/0023830920983368

Conflict of interest

The author declares that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

- Gerrits, E., and Schouten, M. E. (2004). Categorical perception depends on the discrimination task. *Percept. Psychophys.* 66, 363–376. doi: 10.3758/BF03194885
- Giegerich, H. (1992). *English Phonology: An Introduction*. Cambridge: Cambridge University Press.
- Gottfried, T. (1984). Effects of consonant context on the perception of French vowels. *J. Phonetic.* 12, 91–114. doi: 10.1016/S0095-4470(19)30858-7
- Hancin-Bhatt, B. (1994). Segment transfer: a consequence of a dynamic system. *Sec. Lang. Res.* 10, 241–69. doi: 10.1177/026765839401000304
- Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. (1995). Acoustic characteristics of American English vowels. *The J. Acoustic. Soc. Am.* 97, 3099–3111. doi: 10.1121/1.411872
- Jessen, M., Marasek, K., Schneider, K., and Classen, K. (1995). Acoustic correlates of word stress and the tense/lax opposition in the vowel system of German. *Proc. Int. Congr. Phonetic Sci.* 13, 428–431. doi: 10.21437/ICSLP.1996-404
- Kassambara, A. (2023). *Rstatix: Pipe-Friendly Framework for Basic Statistical Tests*. R package version 0.7.2. Available online at: <https://CRAN.R-project.org/package=rstatix/> (accessed December 30, 2023).
- Labrune, L. (2012). *The Phonology of Japanese*. Oxford: Oxford University Press.
- LaCharité, D., and Prévost, P. (1999). The role of L1 and teaching in the acquisition of English sounds by francophones. *Proc. BUCLD* 23, 373–385.
- Leipzig Corpora Collection (2022). *German News Corpus Based on Material From 2022*. Available online at: https://corpora.uni-leipzig.de?corpusId=deu_news_2022 (accessed December 11, 2023).
- Lenth, R. V. (2022). *Emmeans: Estimated Marginal Means, Aka Least-Squares Means*. R Package Version 1.7.2. Available online at: <https://CRAN.R-project.org/package=emmeans> (accessed December 30, 2023).
- Leung, K. K., Jongman, A., Wang, Y., and Sereno, J. A. (2016). Acoustic characteristics of clearly spoken English tense and lax vowels. *The J. Acoust. Soc. Am.* 140, 45–58. doi: 10.1121/1.4954737
- Levy, E. S., and Strange, W. (2008). Perception of French vowels by American English adults with and without French language experience. *J. Phonetic.* 36, 141–157. doi: 10.1016/j.wocn.2007.03.001
- Lisker, L., and Rossi, M. (1992). Auditory and visual cueing of the [±rounded] feature of vowels. *Lang. Speech* 35, 391–417. doi: 10.1177/002383099203500402
- MacMillan, N. A., and Creelman, D. (1991). *Detection Theory: A User's Guide*. Cambridge: CUP Archive.
- Major, R. C. (2001). *Foreign Accent: The Ontogeny and Phylogeny of Second Language Phonology*. London: Routledge.
- Major, R. C. (2008). “Transfer in second language phonology: a review,” in *Phonology and Second Language Acquisition*, eds J. G. Hansen Edwards and M. L. Zampini (Amsterdam: John Benjamins Publishing Company).
- Mayr, R., and Escudero, P. (2010). Explaining individual variation in L2 perception: rounded vowels in English learners of German. *Biling. Lang. Cognit.* 13, 279–297. doi: 10.1017/S1366728909990022
- McAllister, R., Flege, J. E., and Piske, T. (2002). The influence of L1 on the acquisition of Swedish quantity by native speakers of Spanish, English and Estonian. *J. Phonetic.* 30, 229–258. doi: 10.1006/jpho.2002.0174
- Odden, D. (2011). “The representation of vowel length,” in *The Blackwell Companion to Phonology*, eds M. Oostendorp, C. J. Ewen, E. Hume, and K. Rice (New York, NY: John Wiley and Sons), 1–26. doi: 10.1002/9781444335262.wbctp0020
- Ortega-Llebaria, M. (2006). Phonetic Cues to Stress and Accent in Spanish. In *Selected Proceedings of the 2nd Conference on Laboratory Approaches to Spanish Phonetics and Phonology*, edited by Manuel Díaz-Campos, 104–18. Somerville, MA: Cascadilla Proceedings Project.
- Pajak, B., and Levy, R. (2014). The role of abstraction in non-native speech perception. *J. Phonetics* 46, 147–160. doi: 10.1016/j.wocn.2014.07.001
- Polka, L. (1995). Linguistic influences in adult perception of non-native vowel contrasts. *J. Acoustic. Soc. Am.* 97, 1286–1296. doi: 10.1121/1.412170
- Pronk, T., Wiers, R. W., Molenkamp, B., and Murre, J. (2020). Mental chronometry in the pocket? Timing accuracy of web applications on touchscreen and keyboard devices. *Behav. Res. Methods* 52, 1371–1382. doi: 10.3758/s13428-019-01321-2
- R Core Team (2023). *R: A Language and Environment for Statistical Computing*. Vienna, Austria: R Foundation for Statistical Computing.
- Spencer, A. (1984). Eliminating the feature [lateral]. *J. Ling.* 20, 23–43. doi: 10.1017/S002222670000983X
- Stefanich, S., and Cabrelli, J. (2021). The effects of L1 English constraints on the acquisition of the L2 Spanish alveopalatal nasal. *Front. Psychol.* 12:640354. doi: 10.3389/fpsyg.2021.640354
- Strange, W. (2011). Automatic selective perception (ASP) of first and second language speech: a working model. *J. Phonetics* 39, 456–466. doi: 10.1016/j.wocn.2010.09.001
- Vaux, B. (1996). The status of ATR in feature geometry. *Ling. Inquiry* 27, 175–182.
- Vaux, B. (1999). A note on pharyngeal features. *Harvard Working Papers Ling.* 7, 39–63.
- Werker, J. F., and Logan, J. S. (1985). Cross-language evidence for three factors in speech perception. *Percept. Psychophys.* 37, 35–44. doi: 10.3758/BF03207136
- Werker, J. F., and Tees, R. C. (1984). Phonemic and phonetic factors in adult cross-language speech perception. *J. Acoust. Soc. Am.* 75, 1866–1878. doi: 10.1121/1.390988
- Wiese, R. (2000). *The Phonology of German*. Oxford: Oxford University Press.