Replication Study

CONCEPTUALLY CUED PERCEPTUAL CATEGORIZATION IN ADULT L2 LEARNERS

Cristina Lozano-Argüelles ⓘ

Rutgers University

Laura Fernández Arroyo ⓘ

Rutgers University

Nicole Rodríguez ⓘ

Rutgers University

Ezequiel M. Durand López ⓘ

Rutgers University

Juan J. Garrido Pozú ⓘ

Rutgers University

Jennifer Markovits

Rutgers University

Jessica P. Varela

Rutgers University

Núria de Rocafiguera [ID]

Universitat Pompeu Fabra

Joseph V. Casillas [ID] *

Rutgers University

Abstract

Previous studies attest that early bilinguals can modify their perceptual identification according to the fine-grained phonetic detail of the language they believe they are hearing. Following Gonzales et al. (2019), we replicate the double phonemic boundary effect in late learners (LBs) using conceptual-based cueing. We administered a forced choice identification task to 169 native English adult learners of Spanish in two sessions. In both sessions, participants identified the same /b/-/p/ voicing continuum, but language context was cued conceptually using the instructions. The data were analyzed using Bayesian multilevel regression. Learners categorized the continuum in a similar manner when they believed they were hearing English. However, when they believed they were hearing Spanish, "voiceless" responses increased as a function of L2 proficiency. This research demonstrates the double phonemic boundary effect can be conceptually cued in LBs and supports accounts positing selective activation of independent perception grammars in L2 learning.

INTRODUCTION

Bilinguals often navigate complex communicative situations in which they produce and perceive speech in both of their languages. This necessarily entails switching between language-specific phonetic targets (e.g., Flege & Eefting, 1987b), as well as phonological rules (e.g., Simon, 2010), often in real time. For instance, English and Spanish both contrast bilabial stops based on voicing, /b/-/p/, though the phonetic realizations differ in each language. English distinguishes between lag stops, whereas Spanish distinguishes between prevoiced and short-lag stops. Bilingual speakers regulate fine-grained phonetic differences such as this by seamlessly adjusting to the ambient languages. Manifold studies demonstrate that early and late bilinguals[1] develop language-specific phonetic categories in speech production (Flege & Eefting, 1987b, among many others). There is a growing body of research showing that early and late bilinguals also employ language-specific perceptual categorization routines (Antoniou et al., 2012; Elman et al., 1977; Hazan & Boulakia, 1993, among others). Specifically, perception of the incoming acoustic signal seems to be modulated by the language the bilingual listener believes they are hearing. The listener's beliefs can derive directly from the acoustic signal (e.g., Casillas & Simonet, 2018; Gonzales & Lotto, 2013), and recent studies suggest they can also be cued conceptually by tapping

* Correspondence concerning this article should be addressed to Joseph V. Casillas, Department of Spanish and Portuguese, Rutgers University, 15 Seminary Place, New Brunswick, New Jersey 08904. E-mail: joseph.casillas@rutgers.edu

into the listeners conceptual knowledge about which language is spoken in different language contexts (Gonzales et al., 2019; Yazawa et al., 2019). That is, bilinguals modify their perceptual categorization routines in accordance with the phonetic realizations of the language they believe they are hearing. It is unclear how dynamic perceptual categorization develops and whether or not it can be conceptually cued in late bilinguals. In the present study, a conceptual replication[2] of Gonzales et al. (2019), we tested the extent to which adult second language (L2) learners of Spanish modulated their identification of a resynthesized voice timing continuum based on the language they believed they were hearing.

BACKGROUND AND MOTIVATION

Previous studies have explored the notion that bilinguals develop independent phonetic (sub)systems between which they are able to switch as a function of the communicative setting (Bohn & Flege, 1993; Caramazza et al., 1973; Caramazza et al., 1974; Casillas & Simonet, 2018; Elman et al., 1977; Flege & Eefting, 1987a; Garcia-Sierra et al., 2009; Garcia-Sierra et al., 2012; Gonzales et al., 2019; Gonzales & Lotto, 2013; Hazan & Boulakia, 1993; Williams, 1977, 1979). These "language set" experiments have attempted to demonstrate language-specific bilingual speech production and perception behavior by manipulating language mode (Grosjean, 2001b). According to Grosjean (2001a, 2001b), bilingual language modes refer to the level of activation of an individual's languages in any given moment. This framework posits a continuum between a monolingual mode, in which only one of the languages is activated, and a bilingual mode, which supposes dual activation of the languages. It is uncontroversial that bilinguals produce language-specific phonetic targets in unilingual testing situations (e.g., Flege & Eefting, 1987b), and cross-linguistic interactions are attested when the communicative setting activates both languages (e.g., Simonet, 2014). There is an increasing amount of evidence indicating that bilinguals also adjust their perceptual categorization routines based on the language context in monolingual (Garcia-Sierra et al., 2009, 2012; Gonzales et al., 2019; Gonzales & Lotto, 2013, among others) and bilingual (Casillas & Simonet, 2018) modes. This has been referred to as the double phonemic boundary effect.

A recent line of research has empirically tested the double phonemic boundary effect by manipulating language context through the acoustic properties of the stimuli presented to the participants (Casillas & Simonet, 2018; Gonzales et al., 2019; Gonzales & Lotto, 2013), that is, language context was cued perceptually. For instance, Gonzales and Lotto (2013) investigated how bilinguals perceptually accommodate the fine-grained phonetic differences related to how stop consonants are pronounced in English versus Spanish. This language pair has the same voicing distinction at bilabial, coronal, and velar place; however, the phonetic realizations differ with regard to voice-onset time (VOT)—the duration between the release of the stop consonant and onset of voicing (Lisker & Abramson, 1964). While Spanish contrasts phonetically voiced /bdg/ with phonetically voiceless /ptk/, English differs in that the phonetic implementation of the same contrasts occur between lag stops. In Gonzales and Lotto (2013), early Spanish–English bilinguals identified resynthesized VOT continua of pseudowords ("bafri" and "pafri"). Language context was cued conceptually through the instructions by indicating which language they would hear and perceptually through the acoustic properties of the target word endings. The

English-specific final syllable [fɹi] was appended to the stimuli to create an "English-like" VOT continuum. A "Spanish-like" version was created by appending [fri] to the same continuum. Importantly, the continua were identical regarding the acoustic properties of the stop segment. Using a between-subjects design, Gonzales and Lotto (2013) found that early bilinguals' perceptual identification depended on the language-specific continua. That is, the bilinguals who heard the English-like continuum displayed identification functions typical of English speakers, and the bilinguals who heard the Spanish-like continuum produced more "voiceless" responses, essentially shifting the identification function to the left in a manner consistent with the phonetic realizations of Spanish stop voicing. Figure 1 depicts English and Spanish monolinguals' typical voicing identification functions consistent with English and Spanish phonemic boundaries.

The findings suggest that bilinguals adjust perceptual categorization across language contexts by switching between language-specific phonetic systems. Gonzales and Lotto (2013) questioned whether the same degree of phonetic (sub)system separation was also present in adult L2 learners. Models of L2 speech suggest that learner difficulties with nonnative segments/contrasts are explained using acoustic similarities and differences with L1 phonology. The Second Language Linguistic Perception (L2LP) model (Escudero, 2005; Van Leussen & Escudero, 2015) posits that during the initial stages of L2 learning a copy of the L1 perception grammar is made (Full Copying hypothesis), and then develops independently of the L1 grammar. With exposure to the new language, adjustments are made to the L2 perception grammar through a comparison module, the Gradual Learning Algorithm (GLA). The L2LP proposes three distinct learning situations. If the learner perceives the contrast as novel ("new scenario") the L2LP contends that a new phonetic category must be formed. If the learner perceives the contrast as familiar ("similar scenario"), (s)he must then reset the boundary between the acoustic characteristics of this contrast using the GLA. Finally, a "subset" scenario may occur
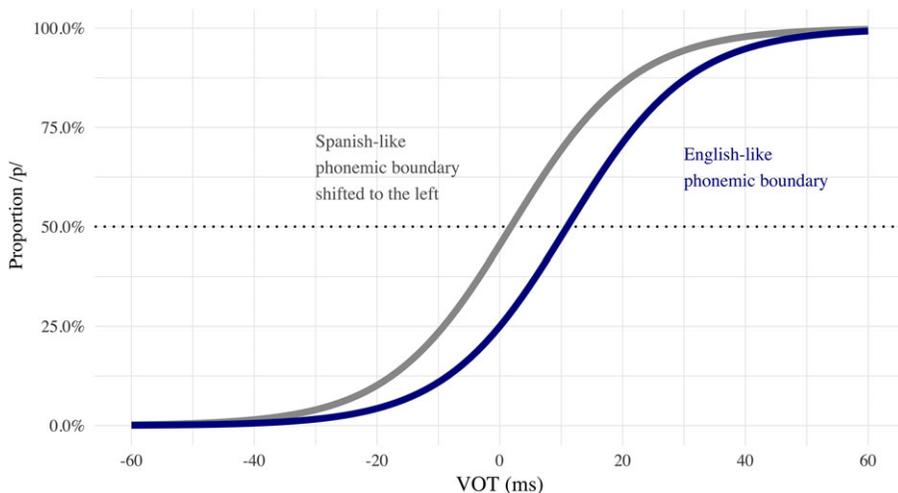

FIGURE 1.   Typical sigmoid identification functions consistent with English and Spanish phonemic boundaries. The Spanish sigmoid is representative of an overall higher proportion of /p/ responses such that the boundary is shifted to the left.

when the L2 has a smaller phonemic inventory and a single L2 category is perceptually assimilated to multiple L1 categories. The Spanish/English voicing contrasts are considered a "similar scenario," thus L2 learners of Spanish must adjust their voicing boundaries to accurately produce and perceive Spanish stops. Importantly, the model proposes that with exposure learners begin to selectively activate L1/L2 perception grammars during speech perception. Language activation can be triggered by a range of variables, linguistic and extralinguistic, such as the language of instruction or the language a given task requires (see Escudero, 2005; Yazawa et al., 2019).

In effect the L2LP provides a theoretical framework that incorporates bilingual language modes and makes the prediction that the double phonemic boundary effect should also occur in adult learners using perceptual and conceptual cueing. Casillas and Simonet (2018) found that both early and late bilinguals showed mode-specific perceptual normalization criteria in conditions of rapid, random mode switching. Importantly, the study showed that bilinguals can exploit language-specific perceptual processes in real time and that this ability appears to be modulated by language proficiency in late learners. Thus, there is evidence for perceptually cued phonemic boundary shifts using the acoustic signal in early and late bilinguals. However, it is still unclear how this ability develops in adult learners. Casillas and Simonet (2018) assessed self-reported proficiency based on a composite measure that suggested the sample was limited to low-intermediate/intermediate learners. It remains to be seen if L2 perceptual processes, such as the double phonemic boundary effect, develop in conjunction with standardized measures of L2 proficiency. There is evidence suggesting that L2 phonological development unfolds in a monotonic relationship with L2 vocabulary size (Bundgaard-Nielsen et al., 2012), though some studies suggest phonetic category development begins at an early stage of learning (Munro et al., 2013; Williams, 1979).

A related question deals with the type of cueing used to incite the double phonemic boundary effect in L2 learners. Gonzales and Lotto (2013) used perceptual and conceptual-based cueing and Casillas and Simonet (2018) used only perceptual cueing. Gonzales et al. (2019) tested for the double phonemic boundary effect using only conceptual cueing by presenting Spanish–English and French–English bilinguals a truncated version of the "bafri"/"pafri" VOT continuum. The stimuli were not language specific. Language context was manipulated conceptually through the instructions by informing the participants they were listening to incomplete utterances of two rare words constituting a /b-p/ minimal pair in English or their other language (e.g., for Spanish-English bilinguals, a Spanish or English bafri-pafri minimal pair). Bilinguals were asked to indicate, on each trial, which "word" the speaker was beginning to say. The continuum was exactly the same in both language contexts, ranging from the beginning portion of one member of the minimal pair (e.g., "ba") to that of the other member (e.g., "pa"). Thus, the contexts were in no way cued differentially by the acoustic properties of the continuum tokens. Nonetheless, both groups displayed different perceptual categorization routines corresponding with the language they believed they were hearing.

In sum, previous research suggests bilinguals can maintain some degree of separation between sound systems. Language-specific perceptual categorization can be perceptually cued in both early and late bilinguals (Casillas & Simonet, 2018; Gonzales & Lotto, 2013), and conceptually cued in early bilinguals (Gonzales et al., 2019). We build on this line of research by testing the conceptual-cueing hypothesis in adult L2 learners.

THE PRESENT STUDY

This research was a conceptual replication of Gonzales et al. (2019), who provided evidence for conceptually cued language-specific perceptual categorization routines in early Spanish/English and French/English bilinguals. Concretely, we extended the conceptual-cueing hypothesis to late bilinguals by following similar testing procedures and using the same auditory stimuli as Gonzales et al. (2019). Crucially, we test whether the double phonemic boundary effect can be conceptually cued in adult L2 learners. We operationalize conceptual cueing as the imparting of conceptual knowledge about which language will be spoken in different language contexts. Our language contexts, Spanish and English, differed from each other solely in terms of the conceptual content of the task instructions. Our study differed from Gonzales et al. (2019) in that we employed a within-subjects design and we assessed how L2 proficiency impacts perceptual categorization routines. Specifically, this work addressed the following research questions:

1. Can late bilinguals be conceptually cued to employ language-specific perceptual categorization routines?
2. If so, how is perceptual categorization modulated by L2 proficiency?

Following the L2LP, we hypothesize that target language boundary adjustments for the Spanish voicing contrast will develop such that conceptually cued language-specific perceptual categorization is possible. Specifically, the model predicts the L2 perception grammar will develop with increased L2 exposure, therefore, adult L2 learners of Spanish should display evidence for language-specific phonemic boundaries as proficiency in Spanish increases. This research fills a clear gap in the literature by adding to our knowledge about sound representation in the bilingual mind and how it develops in conjunction with target language proficiency in adult learners.

METHOD

PARTICIPANTS

A total of 169 individuals completed a two-alternative forced choice (2AFC) task in which the identity of a pseudoword VOT continuum was categorized in separate testing sessions. We recruited 139 undergraduate students enrolled in Spanish courses and graduate-level instructors from a university in the US Northeast. Students received course credit for their time. We also recruited 30 monolingual English speakers using the Prolific.ac online experimental platform. These participants received US$5 for completing the experiment. The pool of online-recruited participants was filtered using criteria set in Prolific.ac to insure participants self-reported as being monolingual English speakers born, raised, and currently living in the United States with no knowledge of any languages other than English. They reported no hearing difficulties and were required to use headphones on a desktop computer. Upon beginning the experiment, all participants responded to the following screening questions: 1) *Was English your first language?*, 2) *Did you start learning Spanish at the age of 13 or older?*, and 3) *Do you have significant knowledge of any languages besides English and Spanish?* We excluded data from any participant responding "no" to (1) or (2), or "yes" to (3). Participants responding categorically across

all trials were also excluded. Based on these criteria, the final dataset included 119 participants.

We administered the Lexical Test for Advanced Learners of Spanish (LexTALE-ESP) (Izura et al., 2014; Lemhöfer & Broersma, 2012) to provide a standardized assessment of the participants' proficiency/vocabulary size in Spanish. On this measure, scores can range from −20 to 60, with native speaker values generally above 50. Individuals with little or no knowledge of Spanish typically score from −20 to 0. For the purposes of the present study, proficiency is treated as a continuous variable, thus we consider a monolingual English speaker to have little to no proficiency in Spanish. Figure 2 plots the LexTALE data. Participants had a wide range of scores (Min. = −18, Max. = 56), suggesting all proficiency levels were likely represented in the sample. The mean value was 14.60 (95% HDI: [11.48, 17.89]). The horizontal lines indicate scores located at 1 and 2 standard deviations from the mean (SD = 17.75).

AUDITORY STIMULI

Instructions

Spoken instructions (duration = 35 seconds) were recorded in English by a 25-year-old female Spanish–English simultaneous bilingual. Following Gonzales et al. (2019), we created separate instructions for each language context. Prerecorded oral instructions conceptually cued the language context by informing participants they were going to hear fragments of rare words in the context-relevant language (Spanish or English), spoken by a native speaker of that language. Importantly, the audio contained no spoken instances of "bafri" or "pafri," though these forced choice alternatives were presented in orthographic form (see the following text). We used Praat (Boersma & Weenink, 2018) to interchange spoken instances of the words "English" and "Spanish" according to the language context, thus creating separate recordings of the instructions that were identical in content and acoustic information in all ways except for the words "English" and "Spanish." To
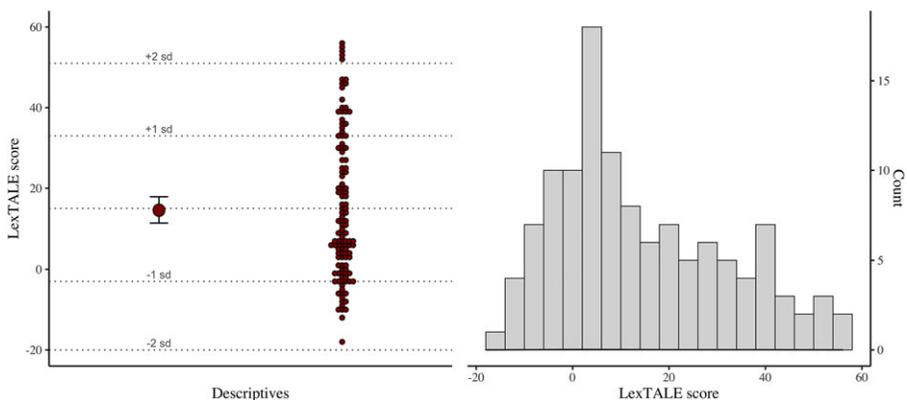

FIGURE 2.　LexTALE data. The left panel plots raw LexTALE scores, including the mean ± 95% HDI and the distribution spread (± standard deviations). The right panel provides a histogram of all scores (binwidth = 4).

illustrate, the instructions began with a sentence such as "In the following experiment you will listen to rare words in X" where "X" was either "English" or "Spanish" in accordance with the language context. Thus our language contexts, Spanish and English, differed from each other solely in terms of the conceptual content of the task instructions. We recorded the instructions in a quiet room using an AKG C520 head-mounted condenser microphone. The signal was digitized at 44.1 kHz and 16-bit quantization using a Sound Devices USBPre 2 audio interface.

Voicing Continuum

We used the same voicing continuum described in Gonzales et al. (2019), which was a slightly modified version of the stimuli described in Gonzales and Lotto (2013). For the sake of completeness, we provide an overview of how the stimuli were created. As the 2AFC task required that participants identify if the speaker was beginning to say "bafri" or "pafri," the voicing continuum consisted of bilabial stop tokens that served as plausible representations of Spanish and English segments. The continuum was created in Praat (Boersma & Weenink, 2018) and consisted of 14 stop + [af] sequences that varied in VOT from −35 to 35 ms in 5 ms increments (excluding a 0 ms step). An early Spanish–English bilingual produced a "pafri" token that was manipulated to create the other steps of the continuum. Specifically, the token was stripped of the [ɾi] segments and the voiceless interval of the stop, not including the release burst. Lead voicing was added in 5 ms increments to create seven prevoiced tokens. Voiceless intervals were added in 5 ms increments to create seven lag tokens.

Procedure

Participants completed the experiment in separate sessions that varied according to language context (Spanish, English). Session one consisted of the first iteration of the 2AFC task, followed by the LexTALE task. In session two participants completed the second iteration of the 2AFC task. The temporal order of the two language contexts was counterbalanced across participants. There was a minimum of 1 hour between testing sessions. In both sessions participants received general instructions before beginning. All interactions between participants and the experimenter took place in English, regardless of language context. PsychoPy3 (Peirce et al., 2019) presented the instructions and all stimuli. Participants received computer-based instructions in written and aural form simultaneously. In the 2AFC task the name of the language corresponding to the language context (Spanish, English) appeared at the top of the screen. Each trial began with the appearance of a fixation cross. After 500 ms the target words "bafri" and "pafri" appeared on either side of the screen. At the same time a randomly drawn continuum token was delivered binaurally through headphones. Participants were instructed to press the left or right arrow key to indicate whether the speaker was beginning to say the "rare word" located on the left or right side of the screen. A key press ended each trial. Participants were instructed to respond as fast and as accurately as possible. There were four initial practice trials pulling from the extremes of the continuum. During experimental trials, stimuli were drawn randomly in 10 separate blocks (14 steps × 10 blocks = 140 responses per participant). The same procedure was followed in both sessions. PsychoPy3 (Peirce

et al., 2019) presented the stimuli in the LexTALE lexical decision task. The Spanish target words appeared in the center of the screen and participants used the keyboard to indicate whether they believed the targets were fake (0) or real (1) words. The experiment lasted approximately 75 minutes (session one: 10 minutes, break: 60 minutes, session two: 5 minutes).

Statistical Analyses

We report two primary statistical analyses. First, we analyzed the 2AFC identification data using a Bayesian multilevel logistic regression model.[3] Second, we analyzed participants' perceptual boundaries using the random effects from the first model. The analyses were conducted in R (R Core Team, 2018, version 4.0.0). Both models were fit using stan (Stan Development Team, 2018) via the R package brms (Bürkner, 2017).

In the first analysis, the participants' responses, "bafri"/"pafri," were modeled as a function of the fixed effects *VOT*, *context* (English, Spanish), *LexTALE score* (henceforth z-LexTALE), *context order* (English first, Spanish first), and all higher order interactions (excluding *context order*). The likelihood of the outcome variable was binomially distributed and the probability of making a "pafri" versus "bafri" response (coded as 1 vs. 0) was mapped to the logistic space using a logit linking function. The continuous predictors were standardized and the categorical predictors were deviation coded (Spanish = −0.5, English = 0.5; Spanish first = −0.5, English first = 0.5).[4] The random effects structure included by-subject intercepts with random slopes for *VOT*, *context*, *z-LexTALE*, and the corresponding higher order interactions. The model included regularizing, weakly informative priors (Gelman et al., 2017), which were normally distributed and centered at 0 with a standard deviation of 5 for all population-level parameters. We established a region of practical equivalence (ROPE) of ±0.05 around a point null value of 0 (see Kruschke, 2018).

In the second analysis we used the posterior of the random effects estimates of the aforementioned model to calculate a distribution of plausible points at which the probability of responding "voiceless" was equal to 50%. The average 50% crossover boundary was calculated for each participant in each context. Next, we established a by-subject phonemic boundary effect by subtracting the Spanish boundary from the English boundary, which was then standardized to have a mean of 0 and a standard deviation of 0.5. We used Bayesian linear regression to analyze the phonemic boundary effect as a function of z-LexTALE. The model was fit using a Student's *t*-likelihood with a prior of *Gamma (4, 1)* for ν. The priors for β_0 and β_1 were normally distributed with a mean of 0 and a standard deviation of 1 (*Normal(0, 1)*), and the prior for σ was a Cauchy distribution set at 0 with scale 1. We established a ROPE of ±0.025 around a point null value of 0 (see Kruschke, 2018).

Both models were fit with 2,000 iterations (1,000 warm-up). Hamiltonian Monte-Carlo sampling was carried out with 6 chains. For all models we report mean posterior point estimates for each parameter of interest, along with the 95% highest density interval (HDI), the percent of the region of the HDI contained within the ROPE, and the maximum probability of effect (MPE). We consider a posterior distribution for a parameter β in which 95% of the HDI falls outside the ROPE and a high MPE (i.e., values close to 1) as compelling evidence for a given effect. See the online supplementary materials for more information.

RESULTS

Figure 3 summarizes the posterior distribution of the omnibus model, illustrating point estimates ±70% and 95% HDI in graphical form. Table 1 of the supplementary materials provides a numeric summary of the posterior distribution. Holding all fixed effects constant, the log odds of responding "voiceless" were −0.85, or approximately 29.94% (β = −0.85, HDI = [−1.00, −0.68], ROPE = 0, MPE = 1), but, unsurprisingly, the likelihood of voiceless responses increased as VOT increased (β = 3.55, HDI = [3.19, 3.93], ROPE = 0, MPE = 1). At the mean proficiency level (x̄ LexTALE = 14.60, z-LexTALE = 0) participants' responses differed as a function of language context (β = −0.23, HDI = [−0.37, −0.07], ROPE = 0, MPE = 1). Specifically, holding VOT and proficiency constant at 0, participants were more likely to respond "pafri" in the Spanish context. We found no evidence that the order in which the language context was presented had an effect on participants' responses (β = −0.15, HDI = [−0.39, 0.07], ROPE = 0.16, MPE = 0.9). There was, however, evidence that the probability of responding "voiceless" increased in the Spanish context for participants with higher LexTALE scores above and beyond the effect of VOT. That is, conditional on the model, the data, and our prior assumptions, there was compelling evidence for a three-way interaction between VOT, language context, and z-LexTALE (β = −0.74, HDI = [−1.35, −0.06], ROPE = 0, MPE = 0.99). Concretely, "voiceless" responses increased as VOT increased and this effect was compounded in the "Spanish" condition, but only for participants with higher proficiency.

The triptych plot in Figure 4 provides draws from the posterior predictive distribution to illustrate the three-way interaction. Each panel depicts the proportion of voiceless responses as a function of VOT and context while holding proficiency constant at −2, 0, and 2 standard deviations, respectively (see Figure 2). One can see that independent of proficiency, voiceless responses increased as VOT increased. Importantly, there is also an observable shift to the left of the identification function in the "Spanish" context as
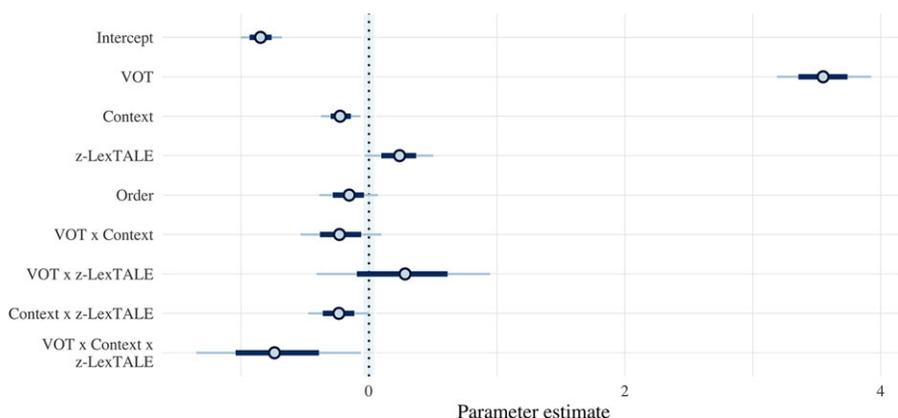

FIGURE 3. Summary of posterior model estimates. Points represent posterior means and horizontal bars indicate ±70% and 95% highest density credible intervals. The vertical discontinuous line represents the point null value (0) surrounded by a region of practical equivalence (ROPE) of ±0.05.
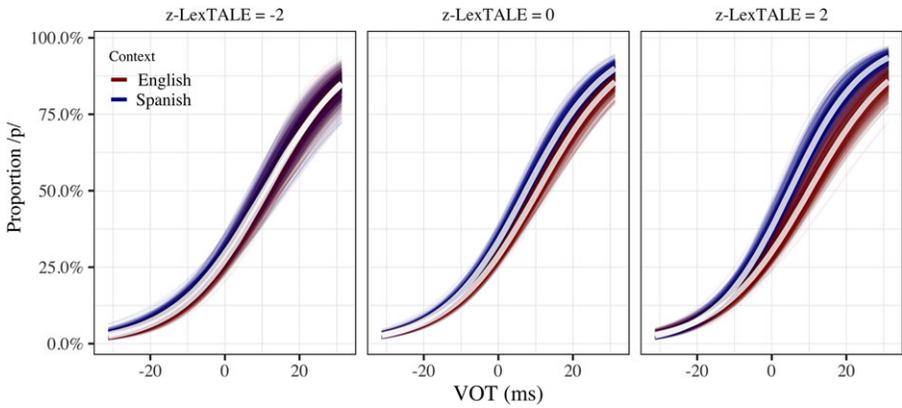
FIGURE 4.   Conditional effects of "voiceless" responses as a function of VOT and conceptually cued language context while holding proficiency (z-LexTALE) constant at −2, 0, and 2 standard deviations from the mean. Individual lines represent 500 samples from the posterior distribution and illustrate uncertainty around population averages (white lines).
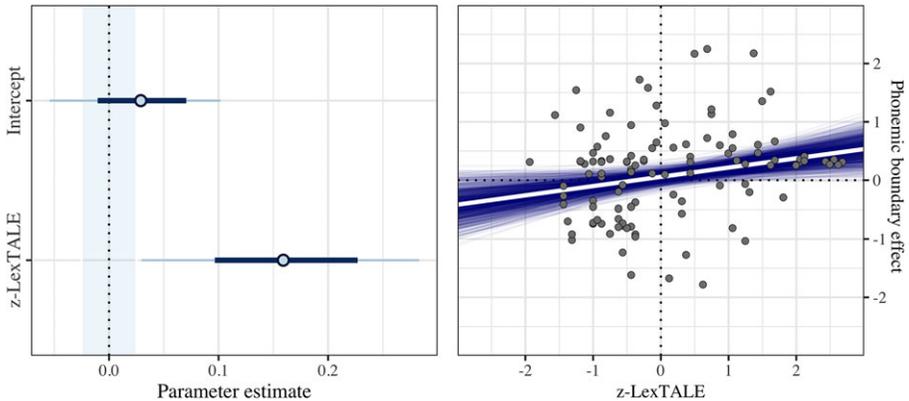

FIGURE 5.   Phonemic boundary effect data. The left panel provides a summary of the posterior model estimates. Points represent posterior means and horizontal bars indicate ±70% and 95% highest density credible intervals. The vertical discontinuous line represents the point null value (0) surrounded by a region of practical equivalence (ROPE) of ±0.025. The right panel plots the phonemic boundary effect as a function of proficiency (z-LexTALE). Individual blue lines represent 1,000 samples from the posterior distribution and illustrate uncertainty around the population average (white line).

proficiency in Spanish increases (2nd and 3rd panels). The identification function representing the "English" context is unaffected by changes in proficiency.

The left panel of Figure 5 summarizes the posterior distribution of the bivariate regression model analyzing the phonemic boundary effect. Table 3 of the supplementary materials provides a numeric summary of the posterior distribution. The model intercept represents the phonemic boundary effect at the average proficiency level ($\bar{x}$ Lex-TALE = 14.60, z-LexTALE = 0), which equates to approximately 3.66 ms ($\beta = 0.03$, HDI = [−0.05, 0.10], ROPE = 0.39, MPE = 0.77). Importantly, the phonemic boundary

effect increased as proficiency increased ($\beta = 0.16$, HDI $= [0.03, 0.28]$, ROPE $= 0$, MPE $= 0.99$). The right panel of Figure 5 illustrates the positive correlation between the phonemic boundary effect and z-LexTALE. More detailed visualizations and table summaries are available in the online supplementary materials.

DISCUSSION

The present work investigated the double phonemic boundary effect in adult L2 learners of Spanish. We administered a two-alternative forced choice task in which the language context was conceptually cued in a within subjects design. This study replicated the double phonemic boundary effect and extended it to a different population: adult L2 learners. Specifically, we found that participants at the average proficiency level were more likely to identify stimuli drawn at random from a stop voicing continuum as "voiceless" when they were led to believe they were hearing rare Spanish words spoken by a native Spanish speaker. Furthermore, the results supported a theoretically motivated three-way interaction between VOT, the proficiency of the learners, and the language context. At low proficiency, voiceless responses only varied as a function of VOT; however, we show that learners with higher proficiency were more likely to respond "voiceless" when they believed they were hearing Spanish. In other words, the present work provides evidence that language-specific perceptual routines can be conceptually cued in adult L2 learners. Specifically, the double phonemic boundary effect increased as LexTALE scores increased.

Our findings corroborate those of a long line of "language set" experiments in which perceptual categorization has been shown to depend on language context in monolingual language mode in early bilinguals (Bohn & Flege, 1993; Elman et al., 1977; Flege & Eefting, 1987a; Garcia-Sierra et al., 2009, 2012; Gonzales et al., 2019; Gonzales & Lotto, 2013; Hazan & Boulakia, 1993; Williams, 1977, 1979) and in late learners (Casillas & Simonet, 2018). This research can be contextualized in light of recent studies that have explored the double phonemic boundary effect by manipulating context-cueing strategies. Gonzales and Lotto (2013) used explicit instructions and acoustic information to cue language context. Casillas and Simonet (2018) used only acoustic information. Gonzales et al. (2019) showed evidence for conceptual cueing in early bilinguals and we replicate this finding and extend it to late bilinguals.

The results provide evidence supporting the L2LP's interpretation of Grosjean's language mode hypothesis (Escudero, 2005; Van Leussen & Escudero, 2015). Concretely, the model proposes that the ability to selectively activate L1 and L2 perception grammars develops with increased exposure to the L2. Previous research on cue-weighting strategies has supported the selective activation account by showing that Japanese learners of American English adjust their weighting of spectra and duration cues when identifying the /iː/-/ɪ/ vowel contrast (Yazawa et al., 2019). Yazawa et al. (2019) conceive of language mode as a continuum ranging from L1 monolingual mode to L2 monolingual mode, with an intermediate L1–L2 bilingual mode existing in between. Our findings compliment this research and add to the existing evidence for selective activation of L1 and L2 monolingual modes using the double phonemic boundary effect. Yazawa et al. (2019) draw attention to the L1–L2 bilingual mode as an avenue for future research. Casillas and Simonet (2018) found that simultaneous bilinguals and adult L2

learners utilized different categorization criteria in both unilingual (i.e., one language at a time) and bilingual (i.e., English and Spanish concurrently) sessions. Future research could test for conceptually cued language-specific processing routines using cue-weighting and double phonemic boundary effects in L1–L2 bilingual mode.

A point of contention to keep in mind for future research relates to how the model conceives of L2 development. The L2LP specifically posits target language exposure as the driving force behind meaning-driven learning vis-à-vis the perception grammar (Van Leussen & Escudero, 2015, p. 4). In effect, perception improves if the current state of the grammar results in misunderstandings; however, there is no straightforward method for operationalizing nor quantifying exposure. There is evidence that phonological development is correlated with L2 vocabulary size (Bundgaard-Nielsen et al., 2012) and other lines of research show that learners' largest gains often occur at an early stage of development (Munro et al., 2013; Williams, 1979). The present work utilized a standardized assessment of vocabulary size as a proxy to L2 proficiency (see also Quam & Creel, 2017). The evidence provided herein cannot partial out the possible mediating effect of either. It may be the case that proficiency, vocabulary size, quantity and quality of input, or any combination thereof are the key to perceptual development. For instance, an individual could receive large amounts of L2 input and not improve in L2 proficiency. The converse is also true. On the surface it seems plausible that there are multiple paths to development of the L2 perceptual grammar. The directed acyclic graph in Figure 6 illustrates some possible causal relationships leading to perceptual development.

A case could be made, for instance, that input, vocabulary size, or some version of the construct "proficiency" could lead to perceptual development. However, at this time we cannot discount the possibility that vocabulary size or proficiency are also mediator variables associated with input. Future investigations should consider measuring and controlling for input and vocabulary size in conjunction with standardized measures of proficiency to shed light on how these variables interact during L2 perceptual development to better inform models of L2 speech learning. This also opens the door to new avenues for research on individual differences in the development of the L2 perception grammar.
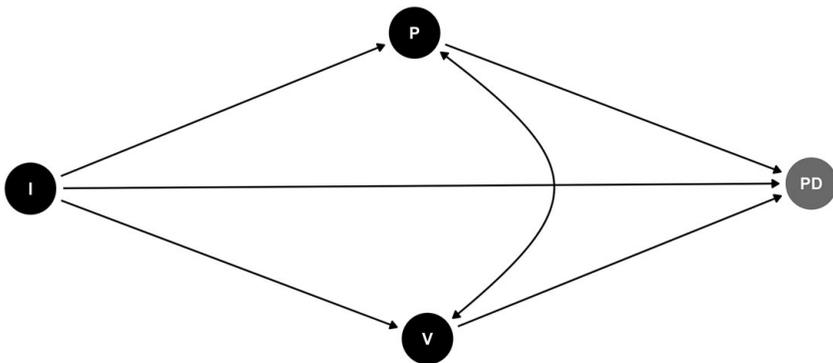

FIGURE 6.   A directed acyclic graph (DAG) illustrating possible causal paths from input (I), vocabulary size (V), and proficiency (P) to perceptual development (PD). The curved line indicates a bidirectional relationship.

CONCLUSION

The present study replicated the findings of Gonzales et al. (2019) and provided empirical evidence for conceptually cued language mode selection in late bilinguals. Specifically, we show that adult L2 learners of Spanish also display mode-specific perceptual normalization criteria in accordance with the fine-grained phonetic detail of the language they have been led to believe they are hearing. Additionally, we find that the double phonemic boundary effect develops as proficiency in the L2 increases. The results provide further evidence supporting the notion that there is some degree of separation between phonetic systems in the bilingual mind.

SUPPLEMENTARY MATERIALS

To view supplementary material for this article, please visit http://dx.doi.org/10.1017/S0272263120000273.

NOTES

[1]In the present work we operationalize early bilinguals as individuals who learn an L2 before puberty and late bilinguals as individuals who begin learning an L2 after this point.

[2]A conceptual replication differs from a direct replication in that a specific aspect of the methodology is changed (e.g., the population of interest) to test the same hypothesis. More information regarding types of replication studies and how they are particularly beneficial to the phonetic sciences is available in Roettger and Baer-Henney (2019).

[3]Bayesian data analysis (BDA) represents an alternative to frequentist data analysis. The interested reader is encouraged to consult the online supplementary materials for more information.

[4]Following Gelman and Hill (2012) we standardized the continuous predictors by subtracting the mean of the distribution from each individual value and dividing by two standard deviations. This transformation puts the parameter estimates of the continuous predictors on the same scale as the fixed factors *context* and *order*, and increases computational efficiency.

REFERENCES

Antoniou, M., Tyler, M. D., & Best, C. T. (2012). Two ways to listen: Do L2-dominant bilinguals perceive stop voicing according to language mode? *Journal of Phonetics*, *40*, 582–594.

Aust, F., & Barth, M. (2018). *papaja: Create APA manuscripts with R Markdown*. https://github.com/crsh/papaja

Boersma, P., & Weenink, D. (2018). *Praat: Doing phonetics by computer* [computer program]. http://www.praat.org/

Bohn, O.-S., & Flege, J. E. (1993). Perceptual switching in Spanish/English bilinguals: Evidence for universal factors in stop voicing judgments. *Journal of Phonetics*, *21*, 267–290.

Bundgaard-Nielsen, R. L., Best, C. T., Kroos, C., & Tyler, M. D. (2012). Second language learners' vocabulary expansion is associated with improved second language vowel intelligibility. *Applied Psycholinguistics*, *33*, 643–664.

Bürkner, P.-C. (2017). brms: An R package for Bayesian multilevel models using Stan. *Journal of Statistical Software*, *80*, 1–28. https://doi.org/10.18637/jss.v080.i01.

Caramazza, A., Yeni-Komshian, G., & Zurif, E. B. (1974). Bilingual switching: The phonological level. *Canadian Journal of Psychology/Revue Canadienne de Psychologie*, *28*, 310–318. https://doi.org/10.1037/h0081997.

Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B., & Carbone, E. (1973). The acquisition of a new phonological contrast: The case of stop consonants in French–English bilinguals. *The Journal of the Acoustical Society of America*, *54*, 421–428. https://doi.org/10.1121/1.1913594.

Casillas, J. V., & Simonet, M. (2018). Perceptual categorization and bilingual language modes: Assessing the double phonemic boundary in early and late bilinguals. *Journal of Phonetics*, *71*, 51–64. https://doi.org/10.1016/j.wocn.2018.07.002.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences*. Erlbaum.

Cohen, J. (2013). *Statistical power analysis for the behavioral sciences*. Routledge.

Elman, J. L., Diehl, R. L., & Buchwald, S. E. (1977). Perceptual switching in bilinguals. *Journal of the Acoustical Society of America*, *62*, 971–974.

Escudero, P. (2005). *Linguistic perception and second language acquisition* (PhD thesis). Utrecht University, Utrecht, Holland; LOT Dissertation series 113.

Flege, J. E., & Eefting, W. (1987a). Cross-language switching in stop consonant perception and production by Dutch speakers of English. *Speech Communication*, *6*, 185–202. https://doi.org/10.1016/0167-6393(87)90025-2.

Flege, J. E., & Eefting, W. (1987b). Production and perception of English stops by native Spanish speakers. *Journal of Phonetics*, *15*, 67–83.

Garcia-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, *51*, 369–378. https://doi.org/10.1016/j.specom.2008.11.005.

Garcia-Sierra, A., Ramírez-Esparza, N., Silva-Pereyra, J., Siard, J., & Champlin, C. A. (2012). Assessing the double phonemic representation in bilingual speakers of Spanish and English: An electrophysiological study. *Brain and Language*, *121*, 194–205. https://doi.org/10.1016/j.bandl.2012.03.008.

Gelman, A., & Hill, J. (2012). *Data analysis using regression and multilevel hierarchical models.* Cambridge University Press.

Gelman, A., & Park, D. K. (2009). Splitting a predictor at the upper quarter or third and the lower quarter or third. *The American Statistician*, *63*, 1–8.

Gelman, A., Simpson, D., & Betancourt, M. (2017). The prior can often only be understood in the context of the likelihood. *Entropy*, *19*, 1–13. https://doi.org/10.3390/e19100555.

Gonzales, K., & Lotto, A. J. (2013). A Bafri, un Pafri: Bilinguals' Pseudoword identifications support language-specific phonetic systems. *Psychological Science*, *24*, 2135–2142. https://doi.org/10.1177/0956797613486485.

Gonzales, K., Byers-Heinlein, K., & Lotto, A. J. (2019). How bilinguals perceive speech depends on which language they think they're hearing. *Cognition*, *182*, 318–330. https://doi.org/10.1016/j.cognition.2018.08.021.

Grosjean, F. (2001a). *One mind, two languages bilingual language processing*. Blackwell Publishing.

Grosjean, F. (2001b). The bilingual's language modes. In *One mind, two languages bilingual language processing* (pp. 1–22). Blackwell Publishing.

Hazan, V. L., & Boulakia, G. (1993). Perception and production of a voicing contrast by French–English bilinguals. *Language and Speech*, *36*, 17–38. https://doi.org/10.1177/002383099303600102.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicologica*, *35*, 49–66.

Kimball, A., Shantz, K., Eager, C., & Roy, J. (2016). Beyond maximal random effects for logistic regression: Moving past convergence problems. *ArXiv E-Prints*.

Kruschke, J. K. (2018). Rejecting or accepting parameter values in Bayesian estimation. *Advances in Methods and Practices in Psychological Science*, *1*, 270–280.

Kruschke, J. K., & Liddell, T. M. (2018). The Bayesian new statistics: Hypothesis testing, estimation, meta-analysis, and power analysis from a Bayesian perspective. *Psychonomic Bulletin & Review*, *25*, 178–206.

Lemhöfer, K., & Broersma, M. (2012). Introducing LexTALE: A quick and valid lexical test for advanced learners of English. *Behavior Research Methods*, *44*, 325–343.

Lisker, L., & Abramson, A. S. (1964). A cross-language study of voicing in initial stops: Acoustical measurements. *Word*, *20*, 384–422. https://doi.org/10.1080/00437956.1964.11659830.

Munro, M. J., Derwing, T. M., & Saito, K. (2013). English l2 vowel acquisition over seven years. In J. Levis & K. LeVelle (Eds.), *Proceedings of the 4th Pronunciation in Second Language Learning and Teaching Conference, Aug. 2012* (pp. 112–119). Iowa State University.

Nicenboim, B., & Vasishth, S. (2016). Statistical methods for linguistic research: Foundational ideas—Part II. *Language and Linguistics Compass*, *10*, 591–613.

Peirce, J. W., Gray, J. R., Simpson, S., MacAskill, M. R., Höchenberger, R., Sogo, H., Kastman, E., & Lindelv, J. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior Research Methods*, *51*, 195–203.

Quam, C., & Creel, S. C. (2017). Mandarin–English bilinguals process lexical tones in newly learned words in accordance with the language context. *PloS One*, *12*, 1–27.

R Core Team. (2018). *R: A language and environment for statistical computing.* R Foundation for Statistical Computing. https://www.R-project.org/

Roettger, T. B., & Baer-Henney, D. (2019). Toward a replication culture: Speech production research in the classroom. *Phonological Data and Analysis*, *1*, 1–23.

Schoot, R. van de, & Depaoli, S. (2014). Bayesian analyses: Where to start and what to report. *European Health Psychologist*, *16*, 75–84.

Simon, E. (2010). Phonological transfer of voicing and devoicing rules: Evidence from L1 Dutch and L2 English conversational speech. *Language Sciences*, *32*, 63–86.

Simonet, M. (2014). Phonetic consequences of dynamic cross-linguistic interference in proficient bilinguals. *Journal of Phonetics*, *43*, 26–37. https://doi.org/10.1016/j.wocn.2014.01.004.

Stan Development Team. (2018). *Stan modeling language users guide and reference manual* (Version 2.18.0). http://mc-stan.org

Van Leussen, J.-W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, *6*, 6–12. https://doi.org/10.3389/fpsyg.2015.01000.

Williams, L. (1977). The perception of stop consonant voicing by Spanish–English bilinguals. *Perception & Psychophysics*, *21*, 289–297. https://doi.org/10.3758/BF03199477.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics*, *26*, 95–104. https://doi.org/10.3758/BF03208301.

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2019). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, 1–25. Advance online publication. https://doi.org/10.1177/0267658319832645.