

An analysis of the perception of stop consonants in bilinguals and monolinguals in different
phonetic contexts: a range-base language cueing approach

Adrián García-Sierra ^{a b}

Elizabeth Schifano^c

Gianna M. Duncan ^a

Melanie S. Fish ^d

^a Speech, Language and Hearing Sciences, University of Connecticut 2 Alethia Drive. Unit 1085
Storrs, CT 06269, USA

^b The Connecticut Institute for the Brain and Cognitive Sciences

^c Department of Statistics, University of Connecticut 415 Glenbrook Road. Unit 1085 Storrs, CT
06269, USA

^d Institute for Learning & Brain Sciences, University of Washington, 1715 NE Columbia Road.
Portage Bay Building Box 357988, Seattle, WA 98195, USA

Corresponding address:

Adrián García Sierra
University of Connecticut
2 Alethia Dr., Unit 1085
Storrs, Connecticut 06269
adrian.garcia-sierra@uconn.edu

Abstract

Bilinguals' observed perceptual shift across language contexts for shared acoustic properties between their languages supports the idea that bilinguals, but not monolinguals, develop two phonemic representations for the same acoustic property. This phenomenon is known as the double phonemic boundary. This investigation replicated previous findings of bilinguals' double phonemic boundary across a series of Go/No-Go tasks while controlling for known confounding effects in speech perception (i.e., contrast effects) and differences in resource allocation between bilinguals and monolinguals (i.e., left or right hand response). Using a range-base language cueing approach, we designed 2 experiments. The first experiment tested if a VOT range representative of either Spanish or English phonetic categories can cue bilinguals, but not monolinguals, to use language specific perceptual routines. The second experiment tested a VOT range with a mixture of Spanish and English phonetic categories to determine if directing attention to a specific phonetic category can disambiguate the competition of the non-attended category. The results for Experiment 1 showed that bilinguals can rely on the distributional patterns of their native phonetic categories to activate specific language modes. Experiment 2 showed that attention can change the weight given to a native phonetic distinction. However, this process is restricted by the internal phonetic composition of the native language(s).

An analysis of the perception of stop consonants in bilinguals and monolinguals in different phonetic contexts: a range-base language cueing approach

Introduction

The present investigation focuses on the perception of acoustically similar speech sounds that are categorized differently between languages. For example, the acoustic properties of Spanish and English stop consonants overlap in such a way that English speakers perceive Spanish *p*, *t*, and *k* as English *b*, *d*, and *g* (respectively). Relevantly, Spanish-English bilinguals' perception of these shared acoustic properties shift towards an English-like speech sound (/b/) in an English context, and a Spanish-like speech sound (/p/) in a Spanish context. This perceptual shift is referred to as *bilinguals' double phonemic boundary* (García-Sierra, Diehl, & Champlin, 2009), and is thought to mirror a psychological shift of language mode (Grosjean, 1998; Grosjean & Miller, 1994). Accordingly, bilinguals use the language context to engage either a bilingual mode where both languages are equally active, or a given monolingual mode where one language is more active than the other. Importantly, the switch between language modes is not an on/off switch; it is a gradient in which bilinguals can move between unbalanced monolingual mode(s) and a more balanced bilingual mode. Given this gradient nature of bilinguals' linguistic activation, it is unlikely that one language "turns off" completely during a monolingual mode (Marian, Bartolotti, Rochanavibhata, Bradley, & Hernandez, 2017; Marian & Spivey, 2003). However, linguistic information specific to one of bilinguals' native languages (e.g., prosody, phonetic, semantic, and/or syntactic structure) establishes a language context that ideally shifts bilinguals into a monolingual mode (e.g., Spanish mode or English mode).

Studies of bilinguals' double phonemic boundary have explored a variety of bottom-up mechanisms by which specific language representations can be accessed and maintained (i.e.,

monolingual mode) for a given moment in time. For example, studies that have used lexical information to engage bilinguals in a monolingual mode have reported significant shifts in the perception of stop consonants as a function of language context. These studies have tested bilingual speakers of English and Spanish (Elman, Diehl, & Buchwald, 1977; García-Sierra et al., 2009; García-Sierra, Ramírez-Esparza, Silva-Pereyra, Siard, & Champlin, 2012; Wig & García-Sierra, in press), English and French (Hazan & Boulakia, 1993), English and Dutch (Flege & Eefting, 1987), and English and Greek (Antoniou, Tyler, & Best, 2012).

Other studies have explored if pre-lexical information is sufficient to establish a language mode in bilinguals. This question has been approached in more than one way, and has subsequently yielded different results. For example, language-specific sounds in nonwords (i.e., Pafri/Bafri with Spanish ‘r’ vs. English ‘r’) were able to produce a shift in bilinguals’, but not monolinguals’, perception of acoustically-shared sounds (Gonzales & Lotto, 2013). Simply, language-specific sounds were sufficient to create a monolingual mode in bilinguals. This methodology also has shown that bilinguals shift across two given language modes rapidly (i.e., monolingual to monolingual, bilingual to monolingual, etc.), and that the degree of shift is mediated by language proficiency (Casillas & Simonet, 2018; Elman et al., 1977; García-Sierra et al., 2009).

However, other studies have used consonant-vowel stimuli (CV) to investigate pre-lexical influences in bilinguals’ language mode. Accordingly, different phonetic contexts were established by presenting syllables with stop consonants unique to Spanish (i.e., prevoiced + vowel) or unique to English (long-lag aspirated stop consonants + vowel), and a perceptual shift was measured by presenting syllables with acoustically shared stop consonants (i.e., short-lag + vowel) across both phonetic contexts. These studies report similar perceptual shifts in bilinguals

and monolinguals (Bohn & Flege, 1993), which suggests that at the phonetic level, stop consonants cannot establish a language context in bilinguals.

The present investigation introduces another approach to learn if bilinguals can use pre-lexical information to shift between language modes. This approach, which we will name range-base language cueing approach, will serve to better understand (1) if exposure to basic distributional properties of each language's phonetic categories, without additional linguistic information, is sufficient to engage bilinguals in a specific language mode, and (2) if directing attention to a specific phonetic category can disambiguate the competition between phonetic distinctions present in a VOT range with a mixture of Spanish and English phonetic categories. Accordingly, we test speech categorization along different ranges of a speech continuum that either present only one phonemic contrast appropriate for a given language (i.e., Spanish or English phonetic contexts; Experiment 1), or phonemic contrasts appropriate for both languages (i.e., Spanish-mix and English-mix phonetic contexts; Experiment 2). In other words, we use phonetic ranges (VOT continua) to establish monolingual (i.e., English-specific or Spanish-specific VOT continua) and bilingual (i.e., English-mix or Spanish-mix VOT continua) language modes.

Background

The study responsible for the motivation behind our investigation is Bohn & Flege (1993) who, too, studied if pre-lexical information can induce a language mode in bilinguals. Their goal was to learn if the perception of short-lag sounds, which are categorized differently in English and Spanish (i.e., /d/ vs. /t/, respectively), depended on the phonetic contexts in which they are presented. Accordingly, Spanish-English bilinguals, Spanish monolinguals, and English monolinguals were told to identify syllables (/do/ and /to/) produced by Spanish or English

monolingual speakers in a Spanish-like phonetic context (i.e., prevoiced sounds and short-lag unaspirated sounds) or in a mixed Spanish- and English-like phonetic context (i.e., prevoiced, short-lag unaspirated, and long-lag aspirated sounds). Importantly, the stimuli tested did not form a VOT continuum, but instead represented the “end points” of Spanish and English phonetic categories (i.e., Spanish: prevoiced and short-lag unaspirated; English: short-lag unaspirated and long-lag aspirated). It was anticipated that the Spanish phonetic context would engage bilinguals in a Spanish monolingual mode, whereas the mixed phonetic context would encourage a bilingual mode. Therefore, bilinguals’ perception of short-lags was not expected to change in the Spanish context (i.e., always ‘t’), but it was expected to shift between ‘t’ and ‘d’ in the mixed phonetic context. As expected, bilinguals perceived more ‘t’ sounds during the Spanish phonetic context than during the mixed phonetic context. However, Spanish monolinguals and English monolinguals, too, showed this perceptual trend.

The authors concluded that all participants created two sets of phonetic rules in accordance with the VOT stimuli used to define each phonetic context, rather than in accordance with the phonetic categories used to define their native language(s). Namely, the small VOT range presented in the Spanish context condition (i.e., prevoiced to short-lag) encouraged participants to place the phonetic boundary close to the short-lag region, while the large VOT range present in the mixed context condition (i.e., from prevoiced to long-lag) encouraged them to place the boundary toward the long-lag aspirated region. Hence, participants labeled more short-lags as ‘t’ sounds in the prevoiced to short-lag condition than in the prevoiced to long-lag condition. The authors explained this phenomenon as a postperceptual decision process that has little or nothing to do with bilinguals’ double phonemic boundary. In other words, under some

methodological scenarios, speech perception is driven more strongly by general auditory mechanisms than by linguistic processes.

Accordingly, the post-perceptual effect described by Bohn & Flege, (1993) fits the description of what is known as range-effects. Range effects refer to the fact that participants tend to identify the midpoint respective to a given range of stimuli. Therefore, the range of stimuli along a speech continuum may influence boundary placement. For instance, Brady & Darwin (1978) tested a group of English monolingual participants' perception along a VOT speech continuum that ranged from 5 to 55 ms of VOT in 5 ms steps. Participants' labeling performance was assessed along the full stimuli range and subranges of the VOT continuum (5-25 ms, 15-35 ms, 25-45 ms, and 35-55 ms). Accordingly, different phonetic boundaries were reported for each of the VOT subranges tested. Therefore, Bohn and Flege explained their results as participants finding the midpoint of each condition (range effects) rather than creating phonetic rules in accordance with native phonetic categories.

However, Bohn and Flege's results can also be explained by differences in acoustical salience. Namely, it has been shown that the distinction between short-lags vs. long-lags is more acoustically salient than the distinction between short-lags vs. prevoiced/lead VOT (Abramson & Lisker, 1972; Keating, Mikos, & Ganong, 1981; Pastore et al., 1977; Streeter, 1976; Williams, 1977, 1979). Importantly, the differences in acoustical salience encountered in a VOT continuum are attributed to the differences in the acoustic dimensions among these phonetic categories; hence, postulated as a bottom-up effect (Williams, 1979). This means, the phonetic boundary shifts observed in Bohn and Flege's study can also be explained by the contrastive difference between the endpoints tested.

Another study that is relevant to range effects and bilinguals' double phonemic boundary is that of Keating, Mikos, and Ganong (1981), which presented different subranges of a VOT continuum to monolingual speakers of Polish and monolingual speakers of English. Keating and colleagues' study is relevant to bilinguals' double phonemic boundary because results showed that speakers of some languages are more prone to range effects than others. Specifically, most languages show 2 distinct VOT distributions; one for voiced sounds (i.e., /b, d, g/) and another for voiceless sounds (i.e., /p, t, k/). Relevantly, the VOT productions of these consonants do not overlap in languages that contrast voicing amongst prevoicing and short-lags (e.g., Polish and Spanish), while the production of these consonants do overlap in languages that contrast voicing amongst short-lags and long-lags (i.e., English and German). Thus, the internal composition of phonemic categories in Spanish creates a larger gap between voiced and voiceless categories than in English (Caramazza, Yeni-Komshian, Zurif, & and Carbone, 1973; Fish, García-Sierra, Ramírez-Esparza, & Kuhl, 2017; Williams, 1977; Hay, 2005). Keating and colleagues' results showed that Polish speakers' phonetic boundary shifted more, by far, as a function of VOT ranges compared to that of English speakers. Accordingly, Keating and colleagues proposed that the wide gap between Polish speakers' phonetic categories allowed them to move their phonetic boundary more "freely" along a VOT continuum, and still maintain an appropriate phonemic distinction. Further, the wide gap in the Polish phonetic structure may result in Polish speakers using different strategies to detect VOT differences. That is, Polish speakers may rely on detecting the presence or absence of voicing, while English speakers may rely on detecting small VOT differences. Altogether, this suggests that shifts in the phonetic boundaries can, in part, be explained by the internal composition of each language's phonetic categories.

Recent investigations have bypassed the confound of range effects by presenting a restricted range of phonetic categories with respect to the relevant range of phonetic categories for bilinguals. For example, Spanish and English together use 3 phonetic categories to distinguish voiced and voiceless consonants (i.e., prevoiced, short-lag, long-lag), but studies have only presented 2 phonetic categories (i.e., prevoiced and short-lag). One study to do this was Gonzales and Lotto (2013). The voicing of the first consonant in nonwords (bafri-pafri) was manipulated to create a continuum ranging from a prevoiced category to a short-lag category; categories that are more representative of the Spanish language than the English language. To establish language contexts while controlling for range effects, only the prevoiced and short-lag categories within the nonwords were presented. While the experimental design is impressive, it should be noticed that a perceptual mismatch between English and Spanish phonology was created when presenting the English /r/ with Spanish VOT in the initial stop consonant. Thus, bilingual participants could have been perceptually aware of these confounding language-specific sounds and influenced bilinguals' perceptual shift across contexts (see Wig & Garcia-Sierra, in press). Further, it cannot be known how bilinguals' perception would be affected if all 3 phonetic categories were presented. The current study explores this by first investigating bilinguals' speech perception across two language-specific phonetic ranges, without confounding language-specific sounds. The intent is to learn if bilinguals are sensitive to the internal composition of the native languages during speech perception (Experiment 1). Then, in a second experiment, bilinguals are tested along a VOT range composed with all 3 relevant Spanish and English phonetic categories without any confounding language-specific sounds. The second experiment explores how bilinguals resolve competition across phonetic categories in speech perception (Experiment 2).

Another investigation that accounted for range effects is García-Sierra et al. (2012). Here, participants' brain activity was recorded to Spanish and English voicing contrasts in two language contexts established by reading a Spanish or English magazine. Importantly, this design allowed each language context to present a perceptual match (i.e., English contrast and English magazine) and mismatch (i.e., Spanish contrast and English magazine). Thus, unlike the above mentioned studies, bilinguals had equal opportunities to match phonetic categories across language contexts. The results showed changes in bilinguals' brain activity in accordance with the language context, again suggesting that bilinguals are sensitive to the internal phonetic composition of their native languages during speech perception. However, the experimenters spoke to the bilinguals in the language of the language context before testing. This is important to mention since it has been reported that conceptual knowledge (an expectation) of what language bilinguals expect to hear before a speech perception task can influence bilinguals' speech perception, similarly to perceptual cues (Gonzales, Byers-Heinlein, & Lotto, 2019). Namely, the only study that has not used linguistic information or phonetic information to establish language contexts is Gonzales and colleagues, (2019). In that study, bilinguals were asked to conceptually imagine they were hearing a native speaker from one of their native languages (English vs. Spanish or English vs. French) while being tested in a CV continuum ranging from prevoiced to short-lags sounds. The results showed a significant boundary shift in bilinguals. Though impressive, this result again shows a perceptual mismatch between phonemic categories that were mentally constructed (e.g., imagining English) and the categories presented in the VOT continuum (e.g., stop consonants representative of Spanish). In other words, similarly to Gonzales and Lotto (2013) bilingual participants could have been perceptually aware of these confounding information between conceptual and perceptual information and influenced

bilinguals' perceptual shift across contexts. For example, in a recent study, Wig and Garcia-Sierra (in press), designed an active speech perception task that concurrently collected event related potentials (ERPs) from Spanish-English bilinguals and English monolinguals. Participants categorized a VOT speech continuum ranging from prevoiced to short-lags in a condition where conceptual and perceptual cues matched or in a different condition where these cues mismatched. The results showed that bilinguals', but not monolinguals' brain responses (i.e., MMN-N2b), increased during the cue mismatch relative to the cue match. In other words, bilinguals maintain perceptual sensitivity to manage the interaction between their languages.

As mentioned previously, the present investigation avoids a mismatch between conceptual and perceptual cues by presenting a range of language-specific phonetic categories or a range with both languages phonetic categories, but the stimuli and the conceptual expectations remain the same. The present study uses a novel design, the range-base language cuing approach, to establish phonetic contexts as opposed to relying on linguistic information that requires proficiency in that given language to establish language contexts (i.e., reading, words, conceptual knowledge, etc.). This design allows monolinguals to be tested alongside bilinguals while giving both groups equal opportunities to label phonetic categories.

In summary, shifts in the phonetic boundary have been shown to extend beyond linguistic experience, and rather represent the consequence of non-linguistic phenomena, such as range effects and differences in acoustical salience. Therefore, the present study aims to investigate if bilinguals, but not monolinguals, shift their phonetic boundary in accordance with their native phonetic categories while controlling for these non-linguistic phenomena.

The present study

So far, the literature reviewed suggests that bilinguals have more than one phonetic criterion or perceptual routine when perceiving short-lag stop consonants. However, it has also been shown that some experimental designs tap more into general auditory mechanisms than into linguistic decisions (Bohn & Flege, 1993; Brady & Darwin, 1978; Pastore et al., 1977; Streeter, 1976; Williams, 1977, 1979). The present investigation builds on Bohn and Flege's and Keating et al.'s work by investigating if different subranges of a VOT continuum can cue bilinguals, but not monolinguals, to implement different phonetic criteria while controlling for auditory biases (i.e., range effects and differences in acoustical salience). In other words, we wanted to assess the degree to which bilinguals can rely on native language-internal phonetic structures to activate appropriate linguistic conceptual representation and facilitate the mapping of phonetic information. For this purpose, we used a range-base language cuing approach. Specifically, we created 4 phonetic contexts using different sections of a VOT continuum. The VOT ranges tested were either unique to Spanish (Spanish phonetic context), unique to English (English phonetic context), or consisted of a mix of both languages (Spanish-mix phonetic context and English-mix phonetic context).

Accordingly, two experiments were designed to test both Spanish-English bilinguals and English monolinguals across different phonetic contexts. A Go/No-Go task was implemented to assess perceptual shifts as a function of linguistic experience and auditory biases. The stimuli were presented using a quasi-normal frequency distribution in which participants were asked to press a button when they heard either /ba/ or /pa/. We used this procedure to direct participants' attention towards one endpoint (tail) of the distribution for a given continuum, and thus inform us if attention can mediate the acoustical differences between phonetic categories. We

additionally controlled for potential within-group differences in resource allocation by asking half the participants in a given group to respond with their dominant hand, while the other half was asked to respond with their non-dominant hand. Therefore, we could compare resource allocation across groups.

This is the first study to implement a Go/No-Go task to explore bilinguals' double phonemic boundary. One of the advantages in using a Go/No-Go task is that it allows assessing the interaction between bottom-up and top-down mechanisms during speech perception, such as the interplay between phonetic endpoints and attention as described above. Nonetheless, there are additional methodological challenges when Go/No-Go tasks are used: (1) bilinguals' cognitive advantage on No-Go responses and (2) non-linguistic perceptual mechanisms associated with perceptual labeling (i.e., contrast effects).

Bilinguals' cognitive advantage. Previous research has shown that bilinguals have greater control over attentional resources than monolinguals in three components of Executive Function (EF): updating, shifting and inhibiting (Bialystok, 2011; Bialystok, Craik, & Luk, 2012). Go/No-Go tasks require constant monitoring of the signal, but particularly require great effort for responses that require inhibition (i.e., No-Go trials). Consequently, it has been shown that bilinguals are better at inhibiting the No-Go trials in Go/No-Go tasks than monolinguals (Foy & Mann, 2014; Jiao, Liu, Wang, & Chen, 2017). This is relevant to our study because the expected group differences in speech perception can be the consequence of differences in resource allocation, as opposed to differences in speech perception. In order to explore this potential confound, we evaluated labeling performance between phonetic contexts across high and low task demand conditions. Specifically, the high task demand condition required participants to press the button with their non-dominant hand (i.e., left), whereas the low task demand condition

required participants to press the button with their dominant hand (i.e., right). If there are no differences between the low and high task demand conditions in both bilinguals and monolinguals, we can assume bilinguals' cognitive advantage on No-Go responses did not significantly contribute to their boundary shift.

Contrast effects. Go/No-Go tasks require a proportion of background and target sounds with different probabilities (.82 vs. .18; respectively in our case). It is known that the repetition of a background stimulus acts as a contrastive anchor in speech perception. Importantly, when a representative endpoint of the VOT continuum is used as a background, perception of stimuli near or at the phonetic boundary are affected the most (Diehl, Elman, & McCusker, 1978; Holt & Lotto, 2002; Holt, Lotto, & Kluender, 2001; Lotto & Kluender, 1998; Lotto, Kluender, & Holt, 1997). In other words, if an endpoint /ba/ in a VOT continuum is used as contrastive stimulus, then listeners report hearing more /pa/ sounds overall. Further, contrast effects are not the result of auditory adaptation since presenting a contrastive stimulus only once also creates similar results. Nonetheless, contrast effects are significantly reduced as the contrastive stimulus (i.e., background) gets closer to the phonetic boundary (Abbs & Sussman, 1971; Diehl et al., 1978; Diehl, Lang, & Parker, 1980; Eimas & Corbit, 1973). Therefore, in the present investigation we account for contrast effects by designing a sequence of stimuli in which target sounds (speech sounds near or at the boundary) were not preceded by endpoints of the continuum (please see contrast effects section in Supplementary Materials).

In summary, a range-base language cueing approach was designed to assess perceptual shifts as a function of linguistic experience and auditory biases in bilingual speakers of Spanish and English and monolingual speakers of English. The first experiment tests participants' ability to categorize two VOT ranges representing phonetic categories unique to the Spanish language

or unique to the English language. The second experiment tests participants' ability to use attentional mechanisms to disambiguate conflicting information from a VOT range consisting of a mix of both languages' phonetic categories.

Method

Participants

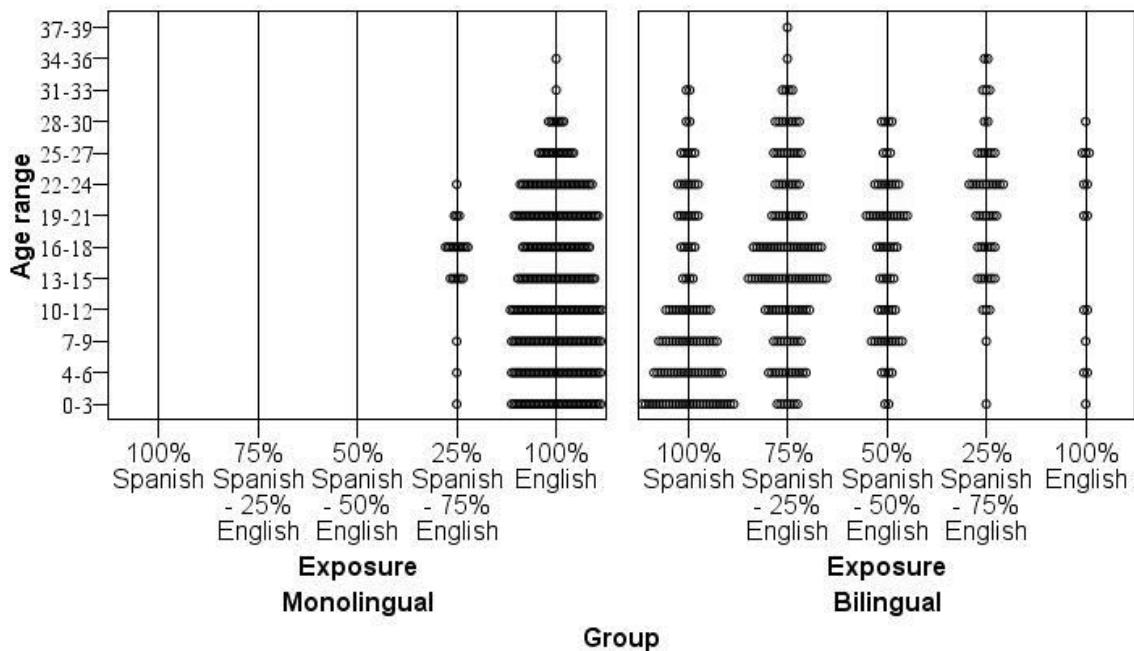
We recruited 36 participants (18 bilinguals) for experiment 1 and 40 new participants (20 bilinguals) for experiment 2. Our bilingual participants were speakers of Spanish and English and our monolingual participants were speakers of English. The mean age of our bilingual participants was 26.9 ($SD = 4.68$) and the age of our monolingual participants was 24.12 ($SD = 3.27$). Bilingual participants reported having been born in Chile (N=20), Mexico (N=1), Spain (N=4) US (N=9), and Latino America (N=3). All participants reported to be right-handed (using the right hand to write). Participants responded to questionnaires assessing their exposure to English and Spanish.

Questionnaires

The Language Questionnaire was completed by each participant to provide specific information about their *exposure* and level of *confidence* among both Spanish and English from childhood to adulthood. Questions for exposure were presented on a Likert scale from 1–5 (1= 100% Spanish; 2 = Spanish 75% - English 25%; 3 = Spanish 50% - English 50%; 4 = Spanish 25% - English 75%; and 5 = 100% English). Figure 1 shows a 2-D Dot Plot of participants' values for exposure (from birth; Panel A), and production (starting at 4 years old; Panel B) of both English and Spanish. Overall, monolingual participants' values show that they heard and spoke English most of the time, and across their lives. On the other hand, bilinguals' values

indicate that they heard and spoke both English and Spanish across their lives, with the most even distribution during ages 19-21.

PANEL A



PANEL B

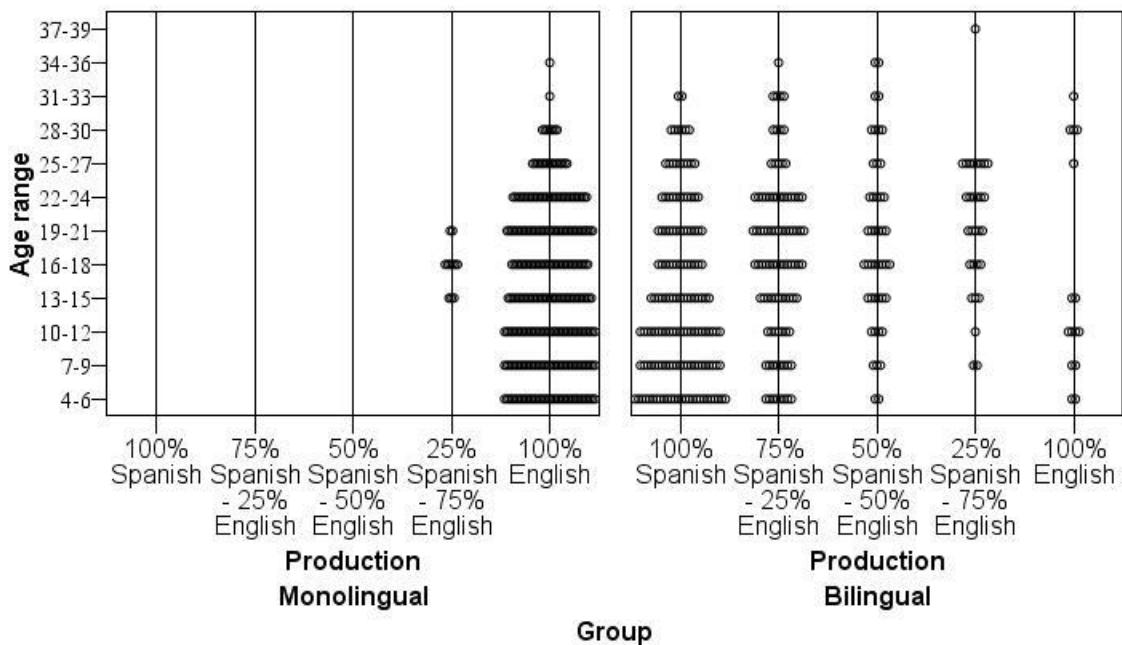
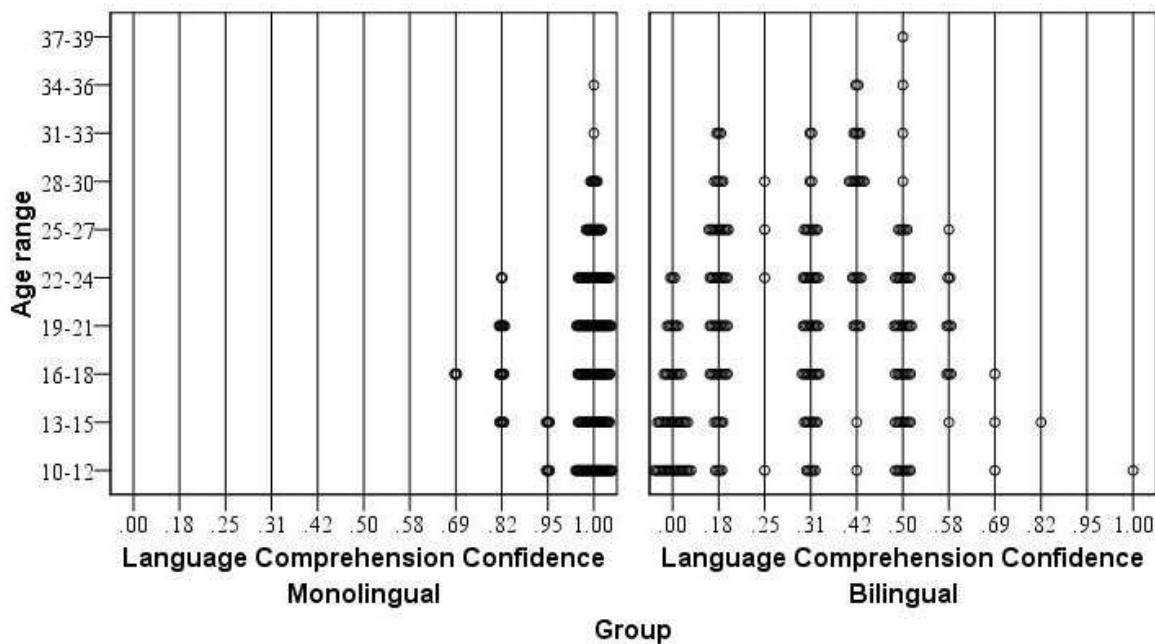


Figure 1. Panel A shows monolinguals' (left) and bilinguals' (right) exposure and production use of English and Spanish as a function of age. Panel B depicts monolingual (left) and bilingual (right) confidence in comprehending, speaking, and reading in English and Spanish as a function of age.

Questions for confidence were presented independently for English and Spanish on a Likert scale from 1–5 (1 = not confident; 2 = 25% confident; 3 = 50% confident; 4 = 75% confident; and 5 = 100% confident). Figure 2 shows participants' average confidence in comprehending (Panel A) and speaking (Panel B) both English and Spanish starting at 9 years of age. In general, monolingual participants' confidence scores are high across ages for English comprehension and production, but not for Spanish counterparts. On the other hand, bilingual participants' confidence scores are high across ages for both English and Spanish in comprehension and production, with the most even distribution across languages during ages 19–27 (See Figure 2 caption for specifics). Using the metric reported in the caption of Figure 2, we calculated the degree of change in the confidence in comprehending and producing Spanish and English from childhood (10-12 years-old) to adulthood (time of the experiment). Bilinguals

showed a significant transition from being more confident in comprehending Spanish during childhood (Mean = .272; $SD = .221$) to being more confident in comprehending both languages during adulthood (Mean = .414; $SD = .191$; $t(33.975) = 3.27$, $p = .002$). Similarly, bilinguals showed a significant transition from being more confident in producing Spanish during childhood (Mean = .264; $SD = .277$) to being more confident in producing both languages during adulthood (Mean = .409; $SD = .197$; $t(34.13) = 2.98$, $p = .005$).

PANEL A: Confidence in Comprehending English and Spanish



PANEL B: Confidence in the Production of English and Spanish

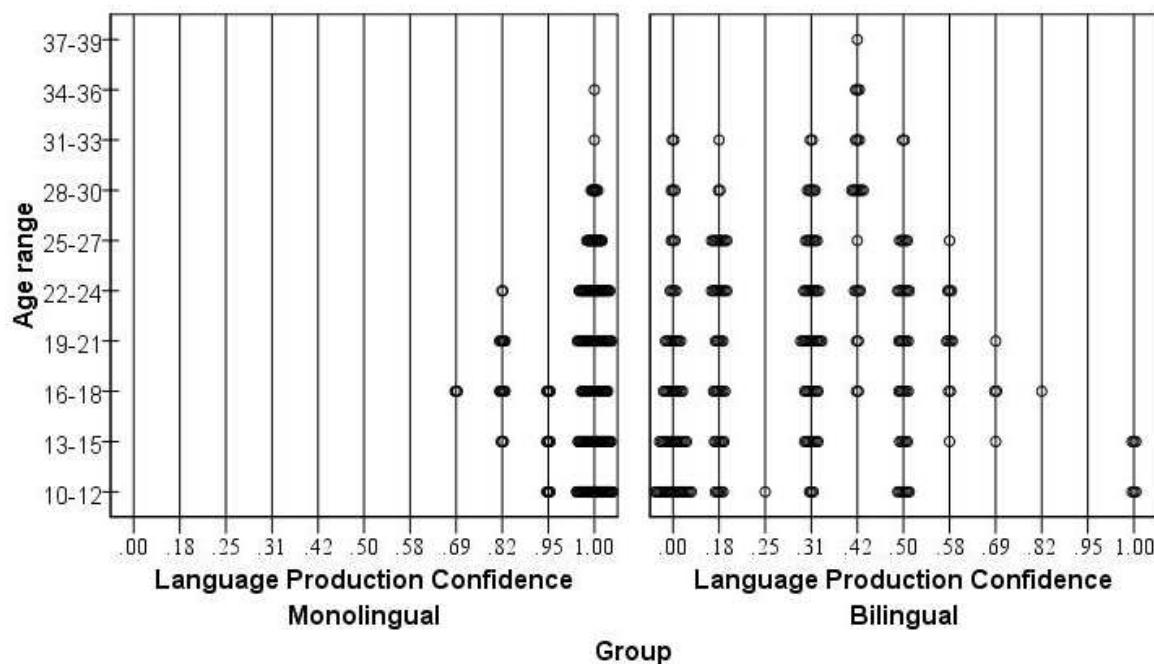


Figure 2. Confidence scores in comprehension and production were calculated with following formula: [scores in English / (scores English + scores in Spanish)] independently for each age range. The formula produced a continuum ranging from .167 as the lowest possible score to .833 as the highest possible score (i.e., $(5/(5+1))$ for maximum score). In order to create a continuum with 0 as the lowest score and 1 as the largest score the following formula was applied: $[(score - min) / (max-min)]$. In this way, higher values represent more confidence in English (i.e., values close to 1), lower values represent more confidence in Spanish (i.e. values close to 0), and equal confidence for both languages is represented with values close to .5.

As a third assessment of level of bilingualism, bilingual participants were asked to provide their *current* confidence in speaking and understanding English and Spanish. Questions were presented on a 1–5 Likert scale (1 = “I cannot speak the language, I have a few words or phrases, and I cannot produce sentences”; 5 = “I have a native-like proficiency with few grammatical errors and I have good vocabulary”). The overall mean for confidence in speaking English was 4.24 ($SD = 0.76$) and 4.89 ($SD = 0.31$) for Spanish. The overall mean for confidence in understanding English was 4.4 ($SD = 0.72$) and 4.92 ($SD = 0.27$) for Spanish. Accordingly, our bilingual group’s self-ratings indicate high confidence levels in both English and Spanish.

General Procedure

All interactions between participants and research assistants occurred in English. Before the experimental session, all participants received hearing screenings. Hearing was assessed in both ears at the frequencies of 250, 500, 1000, 2000, 4000, 6000, and 8000 Hertz. Participants were excluded from the experiment if more than two frequencies were detected at a level higher than 25 dB in the same ear.

Participants' perceptual judgments were assessed across two experimental sessions on separate days, such that only one phonetic context was presented per session. The actual task lasted approximately 60 min. One set of participants were tested across the English and Spanish phonetic contexts (i.e., Experiment 1), while another set of participants were tested across the English-mix and Spanish-mix phonetic contexts (i.e., Experiment 2) (see *Phonetic Contexts* below). Behavioral responses were recorded by asking participants to press a button with their thumb upon perceiving the sounds of interest. All participants were right-handed; however, only half of the participants were instructed to press the button with their right (i.e., dominant) hand, while the other half was instructed to use their left (i.e., non-dominant) hand in a given group.

Stimuli

The speech continuum was generated using the cascade method described by Klatt, (1980). All speech stimuli were 305 ms in duration with a 10 ms burst, and 40 ms formant transitions. Vowel length varied from 215 to 295 ms depending on VOT duration. Vowel F0 was kept constant at 130 Hz until the last 95 ms portion of the vowel, when it declined to 90 Hz. A turbulent noise source (Amplitude of Frication or AF) of 10 ms duration with 80 dB amplitude was applied to simulate the consonant release. Formants and formants' bandwidths (BW) were manipulated in the following way:

Formant transitions were linearly interpolated from values appropriate for a labial stop consonant ($F1 = 380$ Hz, $BW1 = 250$ Hz; $F2 = 950$ Hz, $BW2 = 160$ Hz; $F3 = 1875$ Hz, $BW3 = 330$ Hz; $F4 = 3200$ Hz, $BW4 = 500$ Hz; $F5 = 3500$ Hz, $BW5 = 500$ Hz) to values suitable for vowel /a/ ($F1 = 790$ Hz, $BW1 = 130$ Hz; $F2 = 1280$ Hz, $BW2 = 70$ Hz; $F3 = 2655$, $BW3 = 70$ Hz; $F4 = 3200$ Hz, $BW4 = 500$; $F5 = 3500$ Hz, $BW5 = 500$ Hz).

Stop consonant changes in aspiration (from voiced to voiceless) were accomplished by delaying the energy in $F1$ relative to the onset of higher formants, and by applying a noise source in $F2$ (amplitude of aspiration or $AH = 80$) during the $F1$ cutback period. $F1$ energy was delayed until the start of voicing. Pre-voicing was accomplished by manipulating three parameters; fundamental frequency ($F0 = 130$ Hz), amplitude of voicing ($AV = 55$ dB), and amplitude of voice exciting $F1$ ($A1V = 45$ dB) during the pre-voicing period (Flege & Eefting, 1987b).

An insert earphone (EAR Tone, model 3A 10 k Ω) was used to present the speech sounds. The peak sound intensity (dB SPL) of each stimulus was measured with a sound-level meter that was connected to a 2-cc coupler. All stimuli were delivered at 74 dB peak-equivalent SPL, which is considered a comfortable listening level.

Stimuli were presented using Neuroscan STIM2 software. The responses were recorded with a 4-button hardware latched response pad. The inter-stimulus interval (ISI) occurred in 5 possible lengths; 1210, 1275, 1350, 1425, and 1500 ms. Each of these occurred randomly while maintaining a normal distribution, in which 1350 ms was the most frequently presented. Each phonetic context condition lasted about 60 minutes and consisted of 40 blocks of 48 to 51 stimuli presentations. After each block, there was a 30 s resting period. Participants' responses with latencies shorter than 250 ms or longer than 1000 ms were rejected from the final average.

Reaction times were evaluated from the onset of stimulus presentation to the moment of button press.

Task

A Go / No-Go task was implemented to test participants' phonetic boundaries across two different phonetic contexts in each experiment. The stimuli were presented using a quasi-normal frequency distribution in which the left or right tail (end point of the VOT continuum) of a given distribution represented either a /b/ or a /p/ category. Specifically, participants were instructed to press a button with their thumb when detecting the sound of interest, /ba/ or /pa/ (see Figure 3). For simplicity, the sounds that participants were instructed to detect *if phonetic ranges do not cue language* will be named target sounds (the left or right tails of the distribution), the sounds that fall in the center of the distribution will be named immediate backgrounds, and the sounds in the opposite side of the targets sounds will be named endpoint backgrounds. These labels were chosen to aid the reader in better understand how contrast effects were accounted for (see below). Importantly, the label "target" should not be interpreted as correct responses (i.e., Hit), since the responses given to stimuli falling at the center of the distribution (i.e., immediate backgrounds) are also informative during the categorization process.

Contrastive mechanisms were reduced by controlling the sequence of stimulus presentation in such a way that endpoint backgrounds never followed target sounds (Diehl et al., 1978) (see Figure 3-B). Also, target sounds were separated by at least 3 background sounds. Target sounds were presented with a probability of .18 (360 sounds), while immediate background and end point background sounds were presented with a total probability of .82 (1620 sounds). Figure 3-A shows the frequency distribution of each of the sounds tested as well as the sequence used when presenting the stimuli.

Experiments

Four phonetic contexts were created by presenting VOT ranges that represented phonetic categories unique to Spanish, unique to English, or a combination of both (see Figure 3).

Experiment 1: specific language cueing

Participants: We recruited 36 participants for the first experiment (18 bilingual). See Table 1 for specifics. Each phonetic condition was tested in different days, and the presentation of phonetic contexts was counterbalanced across participants in a given group.

In Experiment 1, we created two phonetic contexts using two phonetic ranges intended to cue bilinguals, but not monolinguals, to use different phonetic criteria when categorizing speech sounds (range-base language cuing approach). Each phonetic range used the same target sounds, but different background sounds. Two conditions were tested:

Spanish Phonetic Context: The speech continuum consisted of 9 stimuli ranging from -60 to 20 ms of VOT in steps of 10 ms. The target stimuli were 0, 10, and 20 ms of VOT, which are representative of both Spanish and English sound categories that are categorized differently (i.e., /p/ and /b/; respectively). The background stimuli ranged from -60 to -10 ms of VOT, which are more representative of a Spanish sound category (i.e., /b/) than an English one. It is important to clarify that although monolingual English speakers also produce voiced consonants with prevoicing (i.e., negative VOT), it occurs less frequently than in Spanish and does not affect phonetic categorization. That is, native speakers of English perceive prevoiced stops as English voiced stop consonants (Flege, 1982; Hay, 2005; Keating, et al., 1981; Dmitrieva, Llanos, Shultz, & Francis, 2015; Fish, et al., 2017). In order to reduce auditory contrastive mechanisms, only the immediate backgrounds of -10, -20, and -30 ms of VOT preceded target sounds. Each

immediate background preceded each target sound 40 times (see Supplementary Materials section for contrast effect analyses). Participants were instructed to press a button when perceiving a /p/ sound, which directed their attention towards the endpoint of the continuum. To account for resource allocation, half of the participants were instructed to press a button with the dominant hand and the other half was instructed to use the non-dominant hand. To make results easier to interpret, the figures represent /ba/ ($/ba/ = 1 - \% \text{ perceived as } /pa/$).

English Phonetic Context: The same participants from the Spanish phonetic context participated. The speech continuum consisted of 9 stimuli ranging from 0 to +80 ms of VOT in steps of 10 ms. As mentioned, the target stimuli were the same as in the Spanish phonetic context (0, 10, 20 ms of VOT) to represent existing sound categories in both Spanish and English that are categorized differently (i.e., /p/ and /b/; respectively). However, the background stimuli now ranged from +30 to +80 ms of VOT to represent a sound category that exists in only English (i.e., /p/). Importantly, only immediate backgrounds of +30, +40, and +50 ms of VOT preceded target sounds. Each immediate background preceded each target sound 40 times (see Supplementary Materials section for contrast effect analyses). Participants' attention was directed towards the endpoint of the continuum by instructing them to press a button with the same hand as in the Spanish context when perceiving a /b/ sound.

Expectations: A group of bilinguals and monolinguals were tested in VOT ranges representative of either Spanish or English phonetic categories. This was done to learn if phonetic ranges acts as a linguistically relevant perceptual cue in speech perception. With this in mind, we predicted two outcomes: (1) if phonetic ranges cue language specific perceptual routines, only bilinguals would shift their phonetic boundary toward the voiced side of the continuum in the Spanish phonetic context and toward the voiceless side of the continuum in the

English phonetic context, and (2) if phonetic ranges do not cue language specific perceptual routines, then neither bilinguals nor monolinguals would show a perceptual shift across contexts. This latter scenario would suggest that exposure to basic properties of each language's phonetic structure, is not sufficient to engage bilinguals in a specific language mode. Figure 3 shows the visualization of the expected shifts in perception for bilinguals and monolinguals in the tested phonetic VOT ranges.

Experiment 2: non-specific language cueing

Participants: We recruited 40 new participants for the second experiment (20 bilingual). See Table 1 for specifics. Each phonetic condition was tested in different days, and the presentation of phonetic contexts was counterbalanced across participants in a given group.

Unlike Experiment 1, the same phonetic range was presented in both phonetic contexts. This phonetic range had a mixture of Spanish and English phonetic categories. Importantly, using the same phonetic range with both Spanish and English cues allowed us to explore if bilinguals would show a boundary shift along a VOT continuum with no language-specific cues. We created two different phonetic contexts by using different groups of target sounds that were preceded by the same background sounds. Two conditions were tested:

Spanish-mix Phonetic Context: The speech continuum consisted of 9 stimuli ranging from -30 to +50 ms of VOT in steps of 10 ms. Accordingly, the target stimuli were -30, -20, and -10 ms of VOT, while the background stimuli ranged from 0 to +50 ms of VOT. In this case, the target stimuli are more representative of the Spanish /b/ category than the English /b/ category, while the background stimuli are divided in two groups of sounds: 1) immediate backgrounds, which represent existing sound categories in both Spanish and English that are categorized differently (/p/ and /b/; respectively), and (2) the endpoint backgrounds which represent a sound

category that exists only in English (/p/). Importantly, only the immediate backgrounds of 0, +10, and +20 ms of VOT preceded target sounds (-10, -20, -30 ms of VOT). Each immediate background preceded each target sound 40 times (see Supplementary Materials section for contrast effect analyses). Half of the participants were instructed to press a button with the dominant hand and the other half was instructed to use the non-dominant hand when perceiving a /b/ sound.

English-mix Phonetic Context: The same participants from the Spanish-mix phonetic context participated. The same speech continuum from the Spanish-mix context was used, but now sounds on the opposite side of the continuum served as the target sounds. Accordingly, target stimuli were +30, +40, and +50 ms of VOT (i.e., English /p/), while the background stimuli ranged from +20 to -30 ms of VOT (i.e., English /b/; Spanish /p/). Hence, the target sounds now represented a sound category that exists only in English (/p/). Each immediate background preceded each target sound 40 times (see Supplementary Materials section for contrast effect analyses). Participants were instructed to press a button with the same hand as in the Spanish-mix phonetic context when perceiving a /p/ sound. To make results easier to interpret, the figures represent /ba/ (/ba/ = 1 - % perceived as /pa/).

Expectations: A new group of bilinguals and monolinguals were tested in a VOT range that consisted of a mixture of phonetic cues representing Spanish and English categories. Namely, the VOT range tested did not cue for a specific language category. Instead, the task directed participants' attention towards one end of the continuum representing a language-specific phonetic category (negative VOT in Spanish-mix; long-lag VOT in English-mix) in attempt to engage language specific routines. This was done to better understand the interaction between language-specific perceptual routines and the general acoustical salience across given

VOT categories. Specifically, English uses the more acoustically salient distinction (short- vs. long- lag VOT) compared to Spanish (prevoiced/lead vs. short-lag). Hence, this condition taps into the interaction between bottom-up (differences in acoustic salience) and top-down (attention) processes during speech categorization. We predicted two possible outcomes: (1) if language-specific perceptual routines can override general acoustical salience between VOT categories, then only bilinguals would show a boundary shift toward the voiced side of the continuum in the Spanish-mix phonetic context and toward the voiceless side of the continuum in the English-mix phonetic context, and (2) if language-specific perceptual routines cannot override general acoustical salience between VOT categories, neither bilinguals nor monolinguals would show a boundary shift between phonetic contexts (see Figure 3).

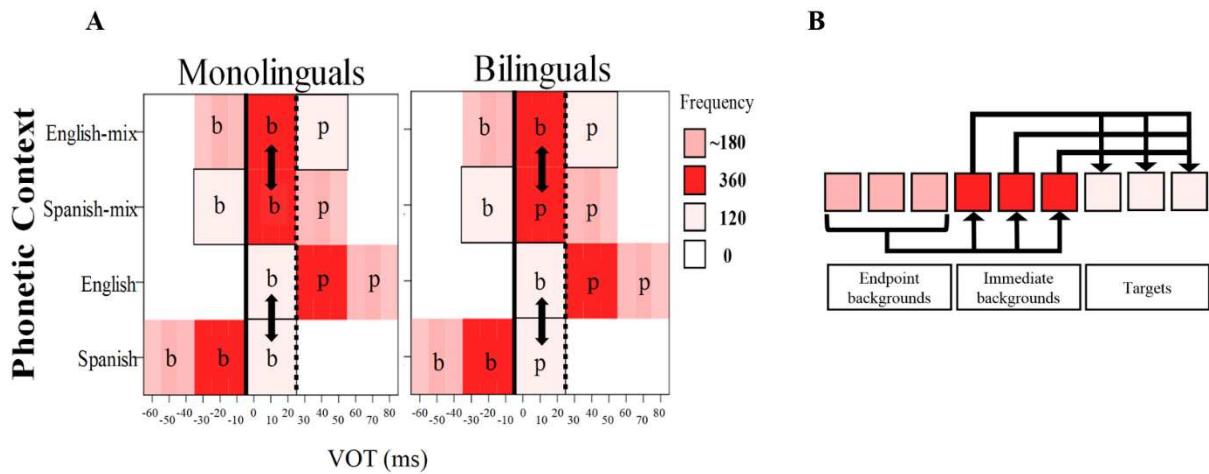


Figure 3. The left side of the figure (A) shows the schematic of the 3 VOT sub-ranges used to cue language modes per phonetic context. The frequency distribution of the stimuli used in each of the sub-ranges is depicted with different colors. Target sounds were presented 120 times during all conditions, immediate backgrounds were presented 360 times during all conditions, and backgrounds were presented between 170 to 190 times during all conditions. The thick dark line shows the less salient contrast in a VOT continuum (i.e., prevoiced/lead vs. short-lags) and the dotted line shows the more acoustically salient contrast (i.e., short-lags vs. long-lags). The letters 'b' and 'p' represent the expected pattern of responses as a function of the phonetic context. The arrows indicate the expected shifts in perception. The right side of the figure (B) shows the schematic of the stimulus sequence used to reduce contrast effects. Importantly, each target sound was preceded 40 times by each of the immediate background sounds.

Statistical Analyses

Only participants with clear phonetic boundaries were retained in the final sample (see Table 1). Importantly, the statistical mixed model approach used to calculate the 50% crossover used a maximum likelihood method for parameter estimation that does not require complete cases for within group comparisons. Table 1 depicts that mostly left-hand monolingual responders were excluded from the Spanish phonetic context, while a similar number of bilingual and monolingual participants were excluded from the Spanish- and English-mix phonetic contexts, regardless of the hand used to give a response.

Phonetic boundary shift: Participant-specific VOT values for which the probability of pressing the button equaled 0.5 (i.e., 50% crossover) were estimated in each context by fitting generalized linear mixed models using a logit link: the number of times participants pressed the button (count responses) were modeled as binomial, and as a function of VOT (continuous), Group (monolingual or bilingual) and Hand (right or left), including all main effects and including all possible higher order interactions among the fixed factors. The mixed model approach implemented uses maximum likelihood methods for parameter estimation and does not require complete cases when doing within-group comparisons. Each context was modeled separately including random intercepts and slopes for participants nested within the between-subject factors (Group, Hand). Following Barr (2013), the highest order significant interaction(s) with VOT (and corresponding lower order terms) were considered as random slopes in the subsequent model fit. We only included participants with non-monotonic functions with clear trajectories showing categorical perception of the speech continuum within the -20 to 40 ms of VOT range. Exclusion analyses were also conducted to test whether exclusion status was

associated with Hand or Group using logistic regression models; neither Hand nor Group was significantly associated with exclusion status in any of the contexts.

The participant-specific VOT values for which the probability of pressing the button equaled 0.5 were compared across contexts, adjusted for Group and Hand effects , using either a linear mixed model with participant-specific random effects and all possible higher-order interactions, or a linear regression model with all possible higher-order interactions, depending on whether the participants were the same (i.e., Spanish vs. English; Spanish-mix vs. English-mix) or different (i.e., Spanish vs. Spanish-mix; English vs. English-mix) across the compared contexts.

For all mixed model analyses reported in this work, covariances and various covariance structures were assessed using (pseudo) likelihood ratio tests and reported information criteria and goodness of fit statistics, denominator degrees of freedom were computed using the Kenward-Roger correction, and analyses were performed in SAS version 9.4 using PROC GLIMMIX and PROC MIXED (SAS Institute, Cary NC). All planned (a priori) comparisons were adjusted for multiple testing using a Bonferroni correction. (Adjusted) p-values less than 0.05 are considered statistically significant (see Casillas & Simonet, 2018 for alternative approaches).

Reaction Time: The averaged reaction times (RTs) across the number of times the button was pressed for target sounds were calculated for each target independently for the four contexts. RTs for each target were submitted to a linear mixed model, allowing again for random intercepts and coefficients for participants nested within Group and Hand (i.e., between-subject factors), and with Target (3 levels, dependent on context) as the within subject factor. The

number of times participants pressed the response button was used as a covariate. Participants' RTs with latencies shorter than 250 ms or longer than 1000 ms were excluded from the analysis.

Table 1. Number of participants retained in the final sample as a function of phonetic contexts

	Spanish		English		Total	Spanish-Mix		English-mix		Total
	Left	Right	Left	Right		Left	Right	Left	Right	
Bilinguals	8	9	9	9	18	7	9	7	10	17
Monolinguals	6	7	9	8	17	8	8	9	8	17
Total	14	16	18	17	35	15	17	16	18	34

Note: A group of participants was recruited for the Spanish and English phonetic contexts and a new group of participants was recruited for the Spanish-mix and English-mix phonetic contexts. Missing observations for within-group comparisons (Spanish *vs.* English and Spanish-mix *vs.* English-mix) were fitted with the maximum likelihood method. The total columns show the number of observations included in the within-group analyses. For between group comparisons, only participants with observations in both conditions were retained.

Results

Results for Experiment 1: Specific-language cueing

Spanish phonetic context vs. English phonetic context. In accordance with bilinguals' double phonemic boundary literature, we predicted two outcomes: (1) if phonetic ranges cue language specific perceptual routines, only bilinguals would shift their phonetic boundary toward the voiced side of the continuum in the Spanish phonetic context and toward the voiceless side of the continuum in the English phonetic context, and (2) if phonetic ranges do not cue language specific perceptual routines, both bilinguals and monolinguals would show no perceptual shift across contexts. The results showed a main effect for phonetic context $F(1, 32.4) = 4.81, p = .035$, main effect for group $F(1, 31.2) = 9.43, p = .004$, and no interaction between phonetic context and group $F(1, 32.4) = 1.57, p = .219$. The planned comparisons with Bonferroni correction showed a significant shift for bilinguals (Estimate = -7.84, $SE = 3.02$) in the expected direction ($t(32.3) = -2.60, p = .028$), and no significant shift for monolinguals ($t(32.5) = -.63, p = .52$).

.534; Estimate = -2.142, $SE = 3.408$). These results suggest that the range of the stimuli tested (English: from 0 to +80 ms of VOT; Spanish: from -60 to +20 ms of VOT) significantly influenced the VOT value at which bilinguals placed their phonetic boundary (See Figure 4).

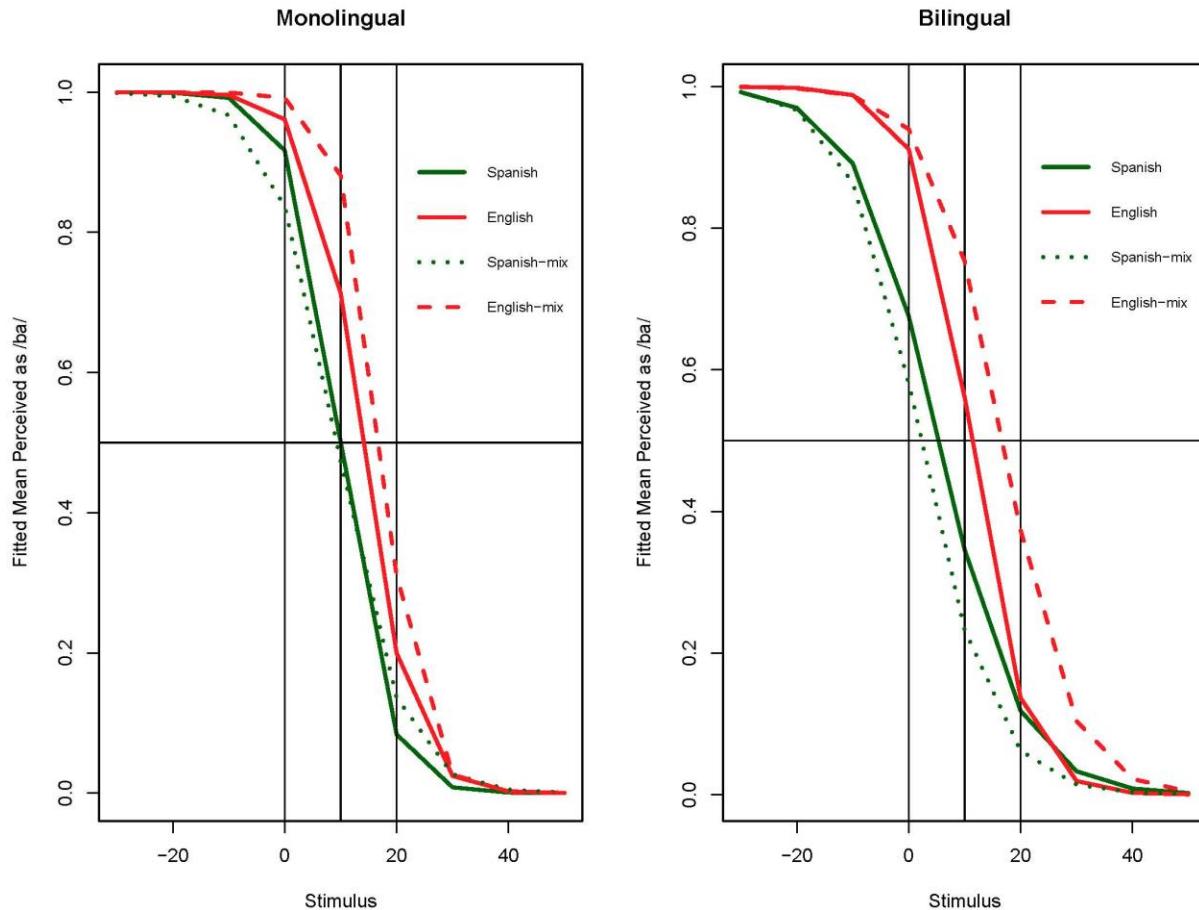


Figure 4. Logit probability values for fitted mean /ba/ responses for the 4 conditions.

Go/No-Go task. The next set of analyses were done to explore if bilinguals were better at inhibiting No-Go trials than monolinguals (i.e., bilinguals' cognitive advantage). Simply, we explored how differences in resource allocation, as opposed to phonetic knowledge, across groups may have influenced results in 2 ways: response hand (i.e., hand effects) and reaction times (i.e., RTs).

Hand effects. In order to explore this potential confound, we compared both groups' VOT values associated with the phonetic boundaries across low task demands (i.e., dominant hand response) and high task demands (i.e., non-dominant hand response) in both phonetic contexts. The results showed no interaction between group, condition and hand $F(1, 30.5) = .47, p = .496$. We report the pair-wise of interest (i.e., how left-hand responders performed against right-hand responders of a given group in a given context). The planned comparisons for the Spanish phonetic context showed no significant difference between hands for bilinguals ($t(30.7) = -.77, p = .45$) nor monolinguals ($t(30.7) = -1.33, p = .192$). Similarly, the planned comparisons for the English phonetic context showed no significant difference between hands for bilinguals ($t(27.8) = -.23, p = .820$) nor monolinguals ($t(30) = -.97, p = .338$). Altogether, the response hand did not seem to significantly influence bilinguals' nor monolinguals' VOT boundaries in any condition. This suggests that both bilinguals' and monolinguals' inhibitory control was similar between hands. Hence, a cognitive advantage unique to bilinguals is unlikely to explain our results.

Reaction Times. RTs allow us to explore an additional dimension of resource allocation (i.e., time) not provided by the VOT values. Specifically, we compared the Reaction Times (RTs) between groups as a function of phonetic context and hand. As mentioned in the Methods section, participants had 1 s to respond starting from the onset of stimulus presentation. We hypothesized that if bilinguals were better at allocating cognitive control demands, they would show faster RTs when using their non-dominant (left hand) when compared with monolinguals. The linear mixed model analysis for the Spanish phonetic context showed no significant interaction between group x hand ($F(1, 29.64) = .39, p = .534$) or group x target x hand ($F(2, 47.12) = 1.52, p = .228$) and no significant pair-wise comparisons were found for the comparisons of interest. The analysis for the English phonetic context showed no significant

interaction between group x hand ($F(1, 28.81) = 1.28, p = .267$) or group x target x hand ($F(2, 48.35) = 1.42, p = .252$) and no significant pair-wise comparisons were found for the comparisons of interest. Overall, the RTs observed across left-hand responders, and right-hand responders, did not differ between groups (see Additional Analysis section for figures showing RT means and standard errors for planned comparisons). This, again, suggests that bilinguals' cognitive advantage cannot adequately explain our results.

Discussion for Experiment 1: Specific Language cueing

The results obtained from our range-base language cuing approach are in accordance with bilinguals' double phonemic boundary (García-Sierra et al., 2009). Namely, during the Spanish and English phonetic contexts, bilinguals placed the phonetic boundaries in accordance with Spanish or English internal phonetic categories; a finding not observed in monolinguals. This means that bilinguals adopted language-specific phonetic routines based on the particular VOT categories presented within these ranges, whereas monolinguals implemented the same phonetic routine based on their single native internal phonetic distribution. Thus, our range-base language cuing approach shows that bilinguals also rely on the distributional patterns of their native phonetic categories to activate appropriate linguistic perceptual routines and facilitate the mapping of phonetic information. In other words, even without additional linguistic information (e.g., conceptual) during the perceptual task itself, the use of language-specific VOT ranges can help bilinguals access appropriate perceptual routines associated with a given language mode.

However, our results contradict Bohn and Flege (1993) who report that both bilinguals and monolinguals showed perceptual shifts across two different VOT ranges. Bohn and Flege explained their findings as a result of a "post-perceptual" decision process in which participants

identify the midpoint along a given range of stimuli regardless of its linguistic properties. This is known as range effects (Brady & Darwin, 1978). If range effects influenced our results as Bohn and Flege described, both bilinguals and monolinguals would have shown perceptual shifts irrespective of the Spanish or English phonetic categories. This was not the case. Instead, our results demonstrate that both groups identified phonetic boundaries in accordance with the internal composition of their native language(s), such that only bilinguals identified different phonetic boundaries across the ranges. This suggests bilinguals were sensitive to the Spanish and English phonetic distributions in each of the VOT ranges, while monolinguals were only sensitive to the English one.

In our investigation, we accounted for contrast effects, or how phonetic boundaries are displaced when stimuli near or at the boundary are contrasted with the endpoints of a given VOT continuum, but are less affected as the contrastive stimulus gets closer to the phonetic boundary (Diehl et al., 1978). Thus, we split the background stimuli into 2 categories: immediate backgrounds, or sounds closest to the targets, and endpoint backgrounds, or sounds furthest away from the targets (see contrast effects section in Supplementary Materials). Only immediate backgrounds preceded targets, and a variable ISI was used to reduce auditory strategies. This allows our results to better reflect participants' use of linguistic experience, as opposed to contrastive mechanisms in speech perception. Simply, bilinguals' observed boundary shift in this study is indicative of the double phonemic boundary.

It should be noted that the mixed model approach did not yield a significant interaction between group and phonetic context. This means that the observed shift in bilinguals (Mean difference = 8.12 ms VOT; Estimate = -7.84, SE = 3.02) is not significantly different from the observed shift in monolinguals (Mean difference = 1.42 ms of VOT; Estimate = -2.142, SE =

3.408). The lack of interaction between group and phonetic context can be explained by (1) the high variability encountered in bilingual populations and (2) by the fact that the perceptual shift is small. Namely, the dependent variable interacts with language proficiency and language usage which results in higher variability in bilingual compared to monolingual populations (Bialystok & Hakuta, 1994; Casillas & Simonet, 2018; Fishman & Cooper, 1969; Elman, Diehl, & Buchwald, 1977; García-Sierra, Diehl, & Champlin, 2009; Hakuta, Bialystok, & Wiley, 2003). In our study, this was not the exception. As depicted in Figure 5, bilinguals showed more variability than monolinguals, hence occluding the desired interaction. Regarding the perceptual shift, it should be noted that it is restricted by the internal composition of phonetic categories and hence it is small. In other words, the VOT range in which the boundary can shift is small (Antoniou, Tyler, & Best, 2012; Elman et al., 1977; García-Sierra et al., 2009) and hence occluded the desired statistical interaction. However, since our results showed a perceptual effect that is in accordance with theory and replicates previous research, we believe that it is valid to report the pair-wise interactions and to interpret them as evidence of bilinguals' double phonemic boundary (see Hsu, 1996; Wei, Carroll, Harden, & Wu, 2012 for a discussion in pair-wise comparisons).

Our results validate previous literature on bilinguals' double phonemic boundary that did not account for contrast effects. That is, our results replicated previous results even after minimizing contrastive auditory mechanisms.

Results for Experiment 2: Non-specific-language cueing

Spanish-mix phonetic context vs. English-mix phonetic context. These contexts consisted of a VOT range that did not cue for a specific language category. This was done to explore if attention can induce a perceptual shift and override the acoustical differences between the short- vs. long-lag distinction and the prevoiced/lead vs. short-lag distinction (Cutting & Rosner, 1974; Keating et al., 1981; Pastore et al., 1977; Williams, 1979). We assumed that if bilinguals' double phonemic boundary can be mediated by attention then bilinguals, but not monolinguals, would show a shift in accordance with the phonetic context. However, if attention cannot override the acoustical differences between the phonetic categories, then neither bilinguals nor monolinguals would show a boundary shift in accordance with the phonetic contexts. The results showed a main effect for phonetic context $F(1, 29) = 37.33, p < .0001$, main effect for group $F(1, 28.2) = 6.71, p = .015$, and no interaction between phonetic context and group $F(1, 29) = .53, p = .474$. The planned comparisons with Bonferroni correction showed a significant shift (Estimate = -13.01, $SE = 2.68$) in the predicted direction for bilinguals ($t(30.1) = -4.84, p = .0002$). Surprisingly, monolinguals also showed a significant shift (Estimate = -10.25, $SE = 2.70$) between contexts ($t(28) = -3.81, p = .0014$). Figure 5 shows the boxplots for the VOT values at .5 probability and Table 2 shows the means and standard deviations for each phonetic context with corresponding t-values (10,000 permutations) and Cohen's d values.

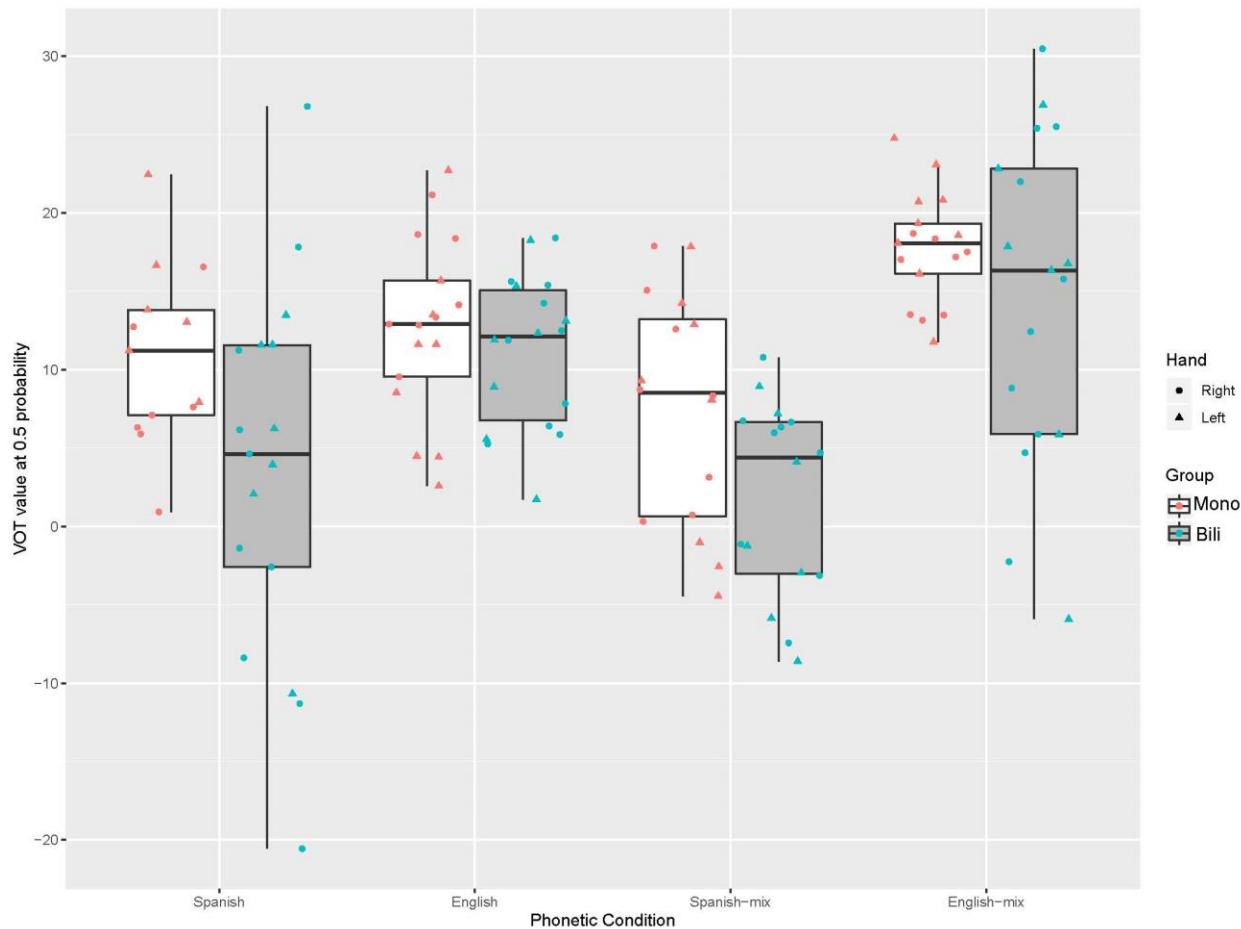


Figure 5. Box plots obtained from VOT values at .5 probability

Go/No-Go task. Similar to Experiment 1, we explored if bilinguals were better at inhibiting No-Go trials than monolinguals (i.e., bilinguals' cognitive advantage).

Hand effects. Similar analyses in Experiment 1 were performed (i.e., low task demand vs. high task demand). The results showed no interaction between group, condition and hand $F(1, 27.1) = .07, p = .793$. We report the pair-wise of interest (i.e., how left-hand responders performed against right-hand responders of a given group in a given context). The planned comparisons for the Spanish-mix phonetic context showed no significant difference between hands for bilinguals ($t(57.9) = .78, p = .438$) nor monolinguals ($t(57.8) = .41, p = .682$). Likewise, the English-mix phonetic context showed no significant difference between hands for

bilinguals ($t(57.8) = -.11$, $p = .916$) nor monolinguals ($t(57.8) = -.84$, $p = .405$). The results suggest that the response hand did not significantly affect bilinguals' nor monolinguals' VOT boundaries. Thus, a cognitive advantage unique to bilinguals is again unlikely to explain our results.

Reaction Times. Similar to Experiment 1, we compared the Reaction Times (RTs) between groups as a function of phonetic context and hand. The same linear mixed model was used from Experiment 1. The results for the Spanish-mix phonetic context showed no significant interaction between group x hand ($F(1, 36.88) = .67$, $p = .417$) or group x target x hand ($F(2, 683.4) = .42$ $p = .657$). The results for the English-mix phonetic context showed no significant interaction between group x hand ($F(1, 36.81) = .58$, $p = .449$) or group x target x hand ($F(2, 684.5) = .97$ $p = .378$). These results suggest that the RTs observed across left-hand responders, and right-hand responders, did not differ between groups (see Additional Analysis section for figures showing RT means and standard errors for planned comparisons), and hence bilinguals' cognitive advantage cannot adequately explain our results.

Table 2. Mean VOT values at .5 probability

Group	Phonetic Context	Mean	SD	Mean diff.	SD diff.	t-value	sig.	95% Confidence Int.		Cohen's d
								Lower	Upper	
Bilinguals	Spanish	3.57	11.87	8.12	14.77	-2.26	0.04	-14.91	-1.349	0.55
	English	11.69	4.36							
	Spanish-mix	2.36	5.82	9.53	11.9	-3.00	0.01	-16.34	-2.68	0.80
	English-mix	11.89	9.42							
Monolinguals	Spanish	10.94	5.72	1.42	8.14	-0.63	0.54	-5.32	3.03	0.17
	English	12.35	5.36							
	Spanish-mix	6.88	7.05	10.43	7.61	-5.31	0.0005	-14.25	-6.72	1.37
	English-mix	17.31	3.41							

Note: The means and p-values reported in the linear mixed model are different from the ones reported here because paired t-tests require complete cases while linear mixed models do not. Please refer to Table 1 for the number of participants included in each of the analyses. The t-test

statistic was performed with 10,000 permutations. Comparisons yielding significant p-values ($p < .05$), and medium to large effect sizes ($d > .5$) are bolded. Cohen's d values represent difference mean divided by pooled standard deviation.

Discussion for Experiment 2: Non-specific Language cueing

The Spanish-mix and English-mix phonetic contexts were implemented to better understand if bilinguals can access language specific routines when competition from multiple languages' categories is present. Specifically, our VOT range presented both, the less salient Spanish prevoiced/lead *vs.* short-lag distinction and the more salient English category short- *vs.* long-lag distinction. We investigated the degree in which attention can mediate the differences in acoustical salience between the Spanish and English distinctions. In order to accomplish this, the only difference between these contexts was the endpoint of the VOT continuum to which participants directed their attention (prevoiced in Spanish-mix or long-lags in English-mix).

Our results showed a significant boundary shift between the Spanish-mix and English-mix phonetic contexts in both our bilingual and monolingual participants. Monolinguals' shift was surprising given how the more acoustically salient distinction corresponds with the internal phonetic distinction in their native language. Specifically, despite where monolinguals' attention was directed along this continuum, a distinction based on perceptual routines or acoustical salience would occur at the same point (i.e., no shift). However, based on our results, both groups' phonetic boundaries were displaced in accordance with the side of the continuum to which their attention was directed. Relevantly, the shift observed in monolinguals is small and hence it is likely that they maintained a more English-like phonemic distinction (see Figure 4). In our opinion, this is not surprising since phonetic boundaries are susceptible to many experimental manipulations such as contrast effects (Diehl et al., 1978), range effects (Darwin, 1978), and stimuli distributions (Clayards et al., 2008). But relevantly, previously reported

phonetic boundary shifts in monolingual English speakers are small due to being restricted to the internal composition of the English phonetic categories (Keating et al., 1981). In contrast to monolinguals, our bilinguals showed a larger boundary displacement between both mix conditions, which suggests they accessed language specific phonetic distinctions (see Figure 4).

Differently from previous investigations, where the difference in acoustical salience is tested with different VOT ranges, we showed that acoustical salience can be mediated by attention in identical VOT ranges (Cutting & Rosner, 1974; Keating et al., 1981; Pastore et al., 1977; Williams, 1979). Namely, directing attention towards the less salient contrast in a VOT continuum (i.e., prevoiced/lead vs. short-lags) can assign a more appropriate weight to a given phonetic distinction, but this process is restricted by the internal phonetic composition of the native language(s).

General Discussion

The present investigation focused on the perception of acoustically similar speech sounds that are categorized differently between languages. Previous research has shown that by relying on linguistic information such as prosody, phonetic, semantic, and/or syntactic structure, bilinguals' perception of these sounds can shift. This shift has been referred as bilinguals' double phonemic boundary (García-Sierra et al., 2009), and is proposed to mirror a psychological shift of language mode (Grosjean, 2001; 1998). Here, we investigated how exposure to distributional properties of each's language phonetic structure, or a mix of both, affects bilinguals' speech perception. Specifically, we used a range-base language cueing approach to establish a monolingual mode or bilingual mode.

The results from our first experiment showed that when a single phonetic distinction is available in a VOT continuum, with no competition from other languages' phonetic categories, bilinguals and monolinguals are sensitive to the distributional patterns of their native language. This led bilinguals to shift their phonetic boundary across Spanish and English contexts, while monolinguals showed no shift. Thus, bilinguals use phonetic distributional patterns of their native languages to activate appropriate linguistic routines and facilitate the mapping of phonetic information.

The results from our second experiment showed that when more than one phonetic distinction is available in a VOT continuum, and thus presents competition from multiple languages' phonetic categories, then bilinguals and monolinguals show phonetic boundary displacements. We interpret these phonetic displacements as the result of top-down mechanisms (attention) during speech categorization. Namely, all participants displaced their phonetic boundary towards the phonetic category they attended to, despite their differences in psychoacoustic salience. However, the observed shift in monolinguals was restricted by the English phonetic categories (see Figure 4). In other words, the phonetic boundary of English monolinguals speakers has less space to move due to the internal composition of its phonetic categories (Keating et al., 1981). On the other hand, bilinguals' shift during the mix conditions aligned to the internal composition of the Spanish and English phonetic categories.

The small shift observed in monolinguals during Experiment 2 can be explained by the measures used to reduce general auditory contrastive mechanisms known to affect speech perception. Namely, endpoints of the continuum (endpoint backgrounds) were never presented directly before, or after, the sounds of interest (targets). Instead, only sounds close to the targets on the continuum (immediate backgrounds) surrounded the targets in the perceptual task. As a

result, the target sounds were embedded in a more phonetically similar sequence than if endpoint backgrounds were allowed to surround target sounds. Although this approach was done to reduce the acoustic contrast toward the target sounds, it may explain why monolinguals showed perceptual trends that did not align with the internal distribution of their native language. Specifically, directing attention towards sounds on a given end of the continuum, repeatedly surrounded by phonetically similar sounds, may have encouraged monolinguals to realign their phonetic categories. However, this shift seemed to be restricted by the internal phonetic composition of participants' native language(s). This possibility would need more research to better understand the interplay of attention and general auditory mechanisms.

It should also be noted that the measures used to reduce general auditory contrastive mechanisms resulted in an uneven distribution of stimuli (quasi-normal distribution). Namely, while the range of the continua used in the two mix conditions was the same, the tails of the distributions were not equal. The target sounds in one end of the continuum were presented fewer times (120) than the endpoint backgrounds at the other side of the continuum (between 170 to 190). Hence, the different distributions of stimuli in each of the conditions could have resulted in participants adapting to the distributions presented in each of the conditions. Further research is necessary to better understand the effects of uneven distributions in speech categorization.

The present investigation is the first one using a Go/No-Go task to assess bilinguals' double phonemic boundary. We accounted for bilinguals' proposed cognitive advantage in Go/No-Go tasks, relative to monolinguals, by creating different task demands within a group (i.e., dominant hand response = low task demands; non-dominant hand response = high task demands) (Bialystok, 2011). Since bilinguals have been previously reported to demonstrate

better inhibition in No-Go trials (Foy & Mann, 2014; Jiao et al., 2017) and faster reaction times (RTs), we would expect such a bilingual cognitive advantage to be represented in our study as different VOT boundary values and RTs across bilinguals that used different response hands in a given context, as well as across bilinguals and monolinguals that used the same response hand in a given context. In contrast, we assumed that if a bilingual cognitive advantage did not play a role in our study, our results would show no differences with respect to response hand between groups. The results confirmed that bilinguals' and monolinguals' phonetic boundary placements (50% crossover) were not affected by the hand used to give a response. In other words, both groups were equally good at inhibiting their responses during No-Go trials using either hand. However, it should be noted that more left-hand monolingual responders than left-hand bilingual responders were excluded from the Spanish phonetic context for not showing clear phonetic boundaries. This result could be interpreted as left-hand responders not pressing the button within the allowed 1-second time interval as often as right-hand responders, and thus response hand could have influenced our results. Yet, our results showed that the RTs from monolingual left-hand responders did not differ from the RTs observed in monolingual right-hand responders, and the RTs from left-hand monolingual responders did not differ from the RTs observed in left-hand bilingual responders. Hence, the lack of clear phonetic boundaries in the left-hand monolingual responders cannot be attributed to increased processing time due to responding with the non-dominant hand. Another explanation could be that left-hand monolingual responders were more cautious in creating a phonetic criterion in accordance with their native language (i.e., English), and thus labeled the Spanish VOT range as one single category (i.e., /b/). Unfortunately, to our knowledge there is no specific research exploring speech categorization as

a function of hand-dominance (i.e., hemispheric lateralization), so more research is needed on this topic.

Relevantly, during the Spanish-mix and English-mix phonetic contexts there was a similar drop in performance in both groups regardless of the hand used to respond (see Table 1). We explain this finding as the Spanish- and English-mix phonetic contexts being more demanding than the Spanish and English phonetic contexts. Namely, participants were required to attend one of two phonetic distinctions, and perhaps for some participants, realigning their phonetic categories in accordance with this task was more demanding than for others. Altogether, since left- and right-hand responders varied their labeling performance patterns across the mix contexts, but yet show similar RTs, we interpret our task as demanding, but not affecting our results. Yet, it should be considered that (1) using the dominant and non-dominant hand as a way to create different levels of task demands did not serve its purpose, or (2) bilinguals' cognitive advantage is too small to have an impact in our task (Nichols, Wild, Stojanoski, Battista, & Owen, 2020). Nevertheless, it is unlikely that our results were affected by differences in cognitive control.

Conclusion

Our findings contribute to the literature in bilinguals' double phonemic boundary and speech perception in the following ways. (1) Bilinguals, but not monolinguals, can be cued by VOT ranges to use the most appropriate phonetic criteria (i.e., language mode) when categorizing speech sounds. Importantly, our findings were not the consequence of bilinguals' cognitive advantage or contrast effects. (2) Categorization of stop consonants can be mediated by attention. That is, when multiple acoustic features with different degrees of acoustic saliency

compete with each other, attention acts as a mediator to give appropriate weight to a given phonetic distinction. However, this process is restricted by the internal phonetic composition of the native language(s).

Acknowledgments

The research reported here was supported by a National Science Foundation Science of Learning Program grant to the LIFE Center (SBE-0354453). The authors thank Dr. Ramírez-Esparza and Noelle Wig for their valuable comments throughout the document.

Open Practice Statement: The data and materials for all experiments are available by request

References

Abbs, J. H., & Sussman, H. M. (1971). Neurophysiological Feature Detectors and Speech Perception: A Discussion of Theoretical Implications. *Journal of Speech, Language, and Hearing Research, 14*, 23-36.

Abramson, A. S., & Lisker, L. (1972). Voice-timing perception in Spanish word-initial stops. *Journal of Phonetics, 1*, 1-8.

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.

Barr, D. (2013). Random effects structure for testing interactions in linear mixed-effects models. *Frontiers in Psychology, 4*(328).

Bialystok, E. (2011). Reshaping the mind: The benefits of bilingualism. *Canadian Journal of Experimental Psychology/Revue canadienne de psychologie expérimentale, 65*, 229-235.

Bialystok, E., Craik, F. I. M., & Luk, G. (2012). Bilingualism: consequences for mind and brain. *Trends in Cognitive Sciences, 16*, 240-250.

Bialystok, E., & Hakuta, K. (1994). In other words: The science and psychology of second-language acquisition. New York, NY: Basic Books.

Bohn, O. S., & Flege, J. E. (1993). Perceptual Switching in Spanish-English Bilinguals. *Journal of Phonetics, 21*(3), 267-290.

Brady, S. A., & Darwin, C. J. (1978). Range effect in the perception of voicing. *J Acoust Soc Am, 63*, 1556-1558.

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

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.

Cutting, J. E., & Rosner, B. S. (1974). Categories and boundaries in speech and music*. *Perception & Psychophysics, 16*, 564-570.

Diehl, R. L., Elman, J. L., & McCusker, S. B. (1978). Contrast effects on stop consonant identification. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 599-609.

Diehl, R. L., Lang, M., & Parker, E. M. (1980). Further Parallel between Selective Adaptation and Contrast. *Journal of Experimental Psychology-Human Perception and Performance, 6*, 24-44.

Dmitrieva, O., Llanos, F., Shultz, A. A., & Francis, A. L. (2015). Phonological status, not voice onset time, determines the acoustic realization of onset f0 as a secondary voicing cue in Spanish and English. *Journal of Phonetics*, 49, 77-95.

Eimas, P. D., & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology*, 4, 99-109.

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

Fish, M. S., García-Sierra, A., Ramírez-Esparza, N., & Kuhl, P. K. (2017). Infant-directed speech in English and Spanish: Assessments of monolingual and bilingual caregiver VOT. *Journal of Phonetics*, 63, 19-34.

Fishman, J. A., & Cooper, R. L. (1969). Alternative measures of bilingualism. *Journal of Verbal Learning and Verbal Behavior*, 8(2), 276-282.

Flege, J. E., & Eefting, W. (1987). Cross-Language Switching in Stop Consonant Perception and Production by Dutch Speakers of English. *Speech Communication*, 6, 185-202.

Flege, J. E., & Eefting, W. (1987b). Production and Perception of English Stops by Native Spanish Speakers. *Journal of Phonetics*, 15, 67-83.

Foy, J. G., & Mann, V. A. (2014). Bilingual children show advantages in nonverbal auditory executive function task. *International Journal of Bilingualism*, 18, 717-729.

García-Sierra, A., Diehl, R. L., & Champlin, C. (2009). Testing the double phonemic boundary in bilinguals. *Speech Communication*, 51, 369-378.

García-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.

Gonzales, K., & Lotto, A. J. (2013). A bafri, un pafri: Bilinguals' Pseudoword Identifications Support Language-Specific Phonetic Systems. *Psychological Science*.

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.

Grosjean, F. (1998). Studying bilinguals: Methodological and conceptual issues. *Bilingualism: Language and Cognition*, 1, 131-149.

Grosjean, F. (2001). The bilingual's language modes. In J. Nicol (Ed.), *One mind, two languages: Bilingual language processing* (pp. 1-22). Oxford, England: Blackwell.

Grosjean, F., & Miller, J. L. (1994). Going in and out of Languages: An Example of Bilingual Flexibility. 5, 201-206.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical Evidence: A Test of the Critical-Period Hypothesis for Second-Language Acquisition. *Psychological Science*, 14(1), 31-38.
doi:10.1111/1467-9280.01415

Hay, J. F. (2005). How auditory discontinuities and linguistic experience affect the perception of speech and non-speech in English- and Spanish-speaking listeners (Doctoral dissertation). Available from Proquest Dissertations and Theses database. (UMI No. 3203519).

Hazan, V. L., & Boulakia, G. (1993). Perception and Production of a Voicing Contrast by French-English Bilinguals. *Language and Speech*, 36, 17-38.

Holt, L. L., & Lotto, A. J. (2002). Behavioral examinations of the level of auditory processing of speech context effects. *Hearing Research*, 167, 156-169.

Holt, L. L., Lotto, A. J., & Kluender, K. R. (2001). Influence of fundamental frequency on stop-consonant voicing perception: A case of learned covariation or auditory enhancement? *The Journal of the Acoustical Society of America, 109*, 764-774.

Hsu, J. C. (1996). *Multiple Comparisons: Theory and Methods*. New York, NY. Chapman & Hall/CRC.

Jiao, L., Liu, C., Wang, R., & Chen, B. (2017). Working memory demand of a task modulates bilingual advantage in executive functions. *International Journal of Bilingualism, 0*, 1-16.

Keating, P. A., Mikos, M. J., & Ganong, W. F. (1981). A Cross-Language Study of Range of Voice Onset Time in the Perception of Initial Stop Voicing. *Journal of the Acoustical Society of America, 70*, 1261-1271.

Klatt, D. H. (1980). Software for a cascade/parallel formant synthesizer. *Journal of the Acoustical Society of America, 67*, 971-990.

Lotto, A. J., & Kluender, K. R. (1998). General contrast effects in speech perception: Effect of preceding liquid on stop consonant identification. *60*, 602-619.

Lotto, A. J., Kluender, K. R., & Holt, L. L. (1997). Perceptual compensation for coarticulation by Japanese quail (Coturnix coturnix japonica). *J Acoust Soc Am, 102*, 1134-1140.

Marian, V., & Spivey, M. (2003). Competing activation in bilingual language processing: Within- and between-language competition. *Bilingualism: Language and Cognition, 6*(2), 97-115.

Marian, V., Bartolotti, J., Rochanavibhata, S., Bradley, K., & Hernandez, A. E. (2017). Bilingual Cortical Control of Between- and Within-Language Competition. *Scientific Reports, 7*(1), 11763.

Nichols, E. S., Wild, C. J., Stojanoski, B., Battista, M. E., & Owen, A. M. (2020). Bilingualism Affords No General Cognitive Advantages: A Population Study of Executive Function in 11,000 People. *Psychological Science*, 0956797620903113.

Pastore, R. E., Ahroon, W. A., Baffuto, K. J., Friedman, C., Puleo, J. S., & Fink, E. A. (1977). Common-factor model of categorical perception. *Journal of Experimental Psychology: Human Perception and Performance*, 3, 686-696.

Streeter, L. A. (1976). Language perception of 2-month-old infants shows effects of both innate mechanisms and experience. *Nature*, 259, 39-41.

Wei, J., Carroll, R. J., Harden, K. K., & Wu, G. (2012). Comparisons of treatment means when factors do not interact in two-factorial studies. *Amino acids*, 42, 2031-2035.

Wig, N., & Garcia-Sierra, A. (in press). Matching the Mismatch: The interaction between perceptual and conceptual cues in bilinguals'. *Bilingualism: Language and Cognition*.

Williams, L. (1977). The perception of stop consonant voicing by Spanish-English bilinguals. *Perception & Psychophysics*, 21, 289-297.

Williams, L. (1979). The modification of speech perception and production in second-language learning. *Perception & Psychophysics*, 26, 95-104.