BUCLD 38 Proceedings
To be published in 2014 by Cascadilla Press

Accessing Cross Language Categories in Learning a Third Language

Page Piccinini and Amalia Arvaniti

1. Introduction

It is well established that early bilinguals can accurately produce and perceive sounds in both of their languages (e.g. Flege and Eefting, 1987, Sundara, Polk, and Baum, 2006; but see also below). This applies both to sounds that occur in only one of the two languages and to sounds that are common to both but differ at the phonetic level. For example, the trilled /r/ in Spanish and the approximate /ɹ/ in English are appropriately produced from a young age by early Spanish-English bilinguals (Goldstein and Washington, 2001; Goldstein, Fabiano, and Washington, 2005). Similarly, English-Spanish bilinguals produce short-lag voice onset time (VOT) with initial /p/ in Spanish but long-lag VOT with initial /p/ in English (Flege and Eefting, 1987).

This suggests that bilinguals do indeed have two separate phonetic and phonological systems for each of their languages. However, much early research on bilinguals' phonological inventories focused on investigating if bilinguals behaved as two monolinguals, or if instead the languages in some way interact in speech production and perception. Generally, results have shown that bilinguals do not perfectly match monolinguals, however. This can result in convergence between the two languages (Sundara et al., 2006; Flege and Eefting, 1987 for English VOT of English-Spanish bilinguals) or even divergence, as a means to further differentiate the two languages (cf. Flege and Eefting, 1987, for Spanish VOT of English-Spanish bilinguals). Similar results have been found in perception, with language mode affecting phonological boundaries (Elman, Diehl, and Buchwald, 1977; Flege and Eefting, 1987). This is in contrast to theories positing that only one language is available at a time (Macnamar and Kushnir, 1971), or that bilinguals have access to both languages in production but only to their dominant language in perception (Cutler, Mehler, Norris, and Seguí, 1992).

---

* Page Piccinini, University of California, San Diego, ppiccinini@ucsd.edu (corresponding author), Amalia Arvaniti, University of Kent, a.arvaniti@kent.ac.uk.
Thank you to the members of the Phonetics Lab, Phon Company, Sarah Creel and the LASR Lab, and three anonymous BUCLD reviewers for providing comments on earlier versions of this work. Thank you to Zully Herrarte and Espi Gonzalez for help with subject running. This project was funded in part by the UCSD Center for Research in Language NIH Trainee Fellowship T32DC000041-20.

Psycholinguistic studies have also corroborated these results and led researchers to propose theories where both languages are always somewhat activated in speech production and perception, leading to interference effects (Green, 1986; Green, 1998; Roelofs and Verhoef, 2006). For example, eye-tracking experiments have found that bilinguals look longer at objects with names that are phonologically similar to the target word, both within the intended language of the study, and across both languages the speaker is fluent in (Marian and Spivey, 2003; Marian, Spivey, and Hirsch, 2003; Spivey and Marian, 2011). If the two systems were entirely separate and not regulated by an activation and suppression schema, then there should be no interference from the language not in use.

Interference from the language not in use need not have a negative effect. In picture naming studies, bilinguals are faster at naming pictures of cognates than non-cognates (Costa, Caramazza, and Sebastián-Gallés, 2000). This finding provides evidence in favor of the aforementioned theories since speakers constantly have access to the phonological make up of a word in both languages. ERP and other physiological studies have also found constant low level activation of the dominant language even when it is not in use (Abutalebi and Green, 2007, 2008; Indefrey, 2006; Leonard et al. 2010; Misra et al. 2012; Moreno, Rodríguez-Fornells, and Laine, 2008; Phillips et al. 2006; Rodriguez-Fornells et al. 2005; van Heuven and Dijkstra, 2010).

The present study tests cross language activation in bilinguals by examining their responses in exposure to a third language. To compare theories of bilingual sound system organization, early Spanish-English bilinguals were tested on a third language, Eastern Armenian, that required them to process a three-way contrast involving segments from both of their languages. If bilinguals are limited to the language currently in use then they should only hear a two-way contrast, incorporating into it categories from the language not in use. If both languages are available at all times, but controlled by a schema of activation and suppression, then bilinguals should be able to access the three-way contrast, but there may be interactions with language mode or language dominance, as either could affect the degree to which the "inactive" language is suppressed.

2. Experiment 1: Production

2.1. Methods

2.1.1 Materials

A set of Eastern Armenian real words supplemented by nonce words was devised in consultation with a native speaker. Eastern Armenian was chosen as it has a three-way VOT contrast: pre-voiced, short-lag, and long-lag; cf. [bah] "spade", [pah] "moment", [p^hak] "closed". All three contrasts were included at two places of articulation: bilabial and velar. Alveolar was not included to reduce the chance of the alveolar stops sounding more "English-like" or "Spanish-like" due to the different place of articulation of English and Spanish coronal stops. In addition to test words, a set of filler words were used which began with one of [m], [n], [f], or [s]. All words were either C_1VC_2 or $C_1V_1C_2V_2$ so as to better conform with English or Spanish phonotactics respectively. In all words, the initial consonant was followed by one of three vowels [i], [ɛ], or [o]; [i] is present in all three languages, [ɛ] is found in Eastern Armenian and English, and [o] in Eastern Armenian and Spanish.[1] For all words C_2 was [m], while $C_1V_1C_2V_2$ words ended in either [a] or [o]. A total of 90 words were recorded as stimuli (10 initial consonants * 3 medial vowels * 3 endings). Two female native speakers of Eastern Armenian produced four tokens of each word in isolation. Speakers were instructed to say the words as if they were real words in Eastern Armenian even if they were not (i.e. with Eastern Armenian phonetics). Some of the constructed words were excluded because they exist in English or Spanish (e.g. *beam* [bim], *quemo* [kɛmo]), resulting in a final number of 48 unique words used as stimuli.

For the test words, VOT was measured to ensure that a three-way contrast was present. If a token did not have appropriate VOT duration, the VOT from another token of a word with the same initial CV sequence replaced the "incorrect" VOT. Splicing was done at zero-crossings at points of rising amplitude. For each speaker, stimuli were constructed so that the VOT distribution of one segment, e.g. [b], never overlapped with that of another segment, e.g. [p], with the same place of articulation. Unpaired t-tests showed that VOT for a given phoneme did not differ significantly between speakers. The final stimuli set included two tokens of each word for each speaker.

The production experiment included a total of 192 stimuli, the two tokens of the 48 words as produced by the two speakers. Of the 192 stimuli 96 were test words, of which 48 began with a bilabial stop and 48 began with a velar stop. The remaining 96 stimuli were fillers.

2.1.2. Speakers

Forty early Spanish-English bilinguals of Mexican-American heritage participated in the experiment. They were UCSD undergraduates who took part in the study in exchange for course credit. All participants had been exposed to both languages before age six. They were administered the Bilingual Dominance Scale (henceforth BDS, Dunn and Fox Tree, 2009) which assigns a score for how dominant a bilingual is in their two languages. A score of 0 is perfectly balanced, high positive numbers heavily English dominant, and high negative numbers heavily Spanish dominant. The participants in this study had an average score of 9.06, thus they tended to be English dominant. Half of the participants were tested with English as the language of instruction and half with Spanish. The two groups were fairly similar in regards to language dominance, with the English instruction group having an average BDS score of 9.85 (s.d. = 8.2), and the Spanish instruction group an average score of 8.60 (s.d. = 8.8) [t = 0.45, p = 0.65].

[1] Armenian and Spanish [o] are monophthongal, unlike English [o] which is diphthongal.

2.1.3. Procedures

Before the experiment started, participants were administered a language demographic questionnaire including the BDS. The language of the questionnaire matched the language of instruction. For the experiment, all words were presented over headphones with the aid of a computer and SuperLab Pro 4.5. Participants were told they would be hearing words in a new language and had to repeat the words back to the best of their abilities. They were instructed to focus on a fixation cross during which time they would hear a word. Participants repeated the word after the fixation cross went away. They then pressed the space bar to move to the next trial. The fixation cross reappeared and after 500 ms the next word would play. Each stimulus was presented twice across two blocks for a total of 384 productions per participant, 192 of which were test words. Stimuli were randomized within each block.

2.1.4. Annotation and measurements

For all test words, VOT was measured from the release of the stop to the onset of voicing. In the case of negative VOT, measurements were taken from the onset of voicing to the release of the stop and marked as negative. All words were coded for whether the stimulus heard was pre-voiced, short-lag VOT, or long-lag VOT, e.g. [b], [p], or [p^h].

2.2. Results

The density curves for bilabial and velar stops in the three phonetic categories are presented in Figures 1(a) and 1(b). To test for significant effects linear mixed effects models (LMEM) were run with separate models for bilabial and velar stops. For both, model comparisons found that including language of instruction did not significantly improve the model, and thus this variable was not included in the final models. The dependent variable was VOT in milliseconds; the independent variables were stimulus VOT (pre-voiced, short-lag, long-lag) and BDS score. The baseline value for stimulus VOT was short-lag. Speaker was included as a random slope by stimulus VOT. Significance was set at a *p*-value of 0.05. For bilabials, short-lag tokens were significantly longer than pre-voiced tokens [$\beta = -45.20$, $t = -7.56$], and significantly shorter than long-lag tokens [$\beta = 63.02$, $t = 8.48$]. There was no significant effect of BDS score, however there was a significant interaction between BDS score and stimulus VOT that affected specifically [b] vs. [p] [$\beta = 1.54$, $t = 3.23$]. The intercept was not significant. For velars, short-lag tokens were significantly longer than pre-voiced tokens [$\beta = -54.46$, $t = -5.84$], and significantly shorter than long-lag tokens [$\beta = 55.59$, $t = 11.24$]. BDS score was not significant, however, and there were no significant interactions. The intercept was significant [$\beta = 34.69$, $t = 9.30$].

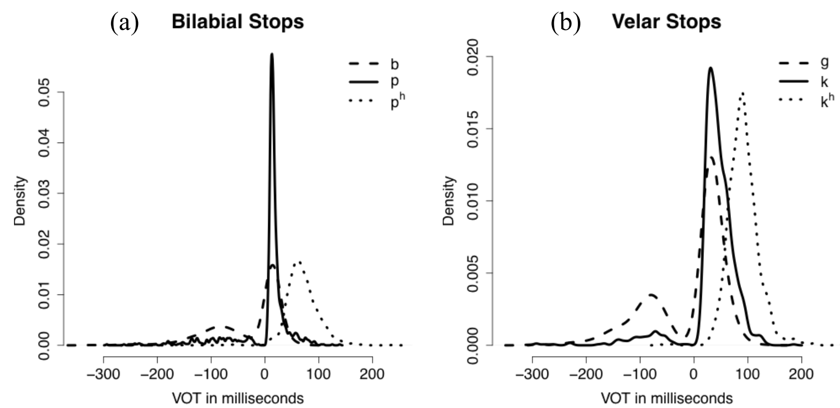

Figure 1: Density curves of stop productions for (a) bilabial and (b) velar stops in the three phonetic categories (pre-voiced, short-lag, and long-lag).

To further examine the interaction between stimulus VOT and BDS score found for bilabials, we calculated the difference between each speaker's mean VOT duration for the pre-voiced and short-lag categories. A large difference would suggest clear category differentiation, while a small difference would suggest the two categories had been collapsed. A linear regression was run with the difference between the two categories as the dependent variable and BDS score as the independent variable. The regression was significant [$r = -0.46$, $p < 0.01$], with the difference between [b] and [p] durations getting smaller as the BDS score increased. This result suggests that those who were more English dominant did not produce the two categories differently, while those who were more balanced most likely did. This accounts for the bimodal distribution of /b/ (see Figure 1a), though we note that a similarly bimodal distribution applies to /g/ but without BDS score being significant.

3. Experiment 2: ABX Perception

3.1. Methods

3.1.1. Materials

The same set of words was used as in Experiment 1. The experiment included 288 trials. Each trial consisted of a series of three stimuli (three words) with an ISI of 500 ms. The first two stimuli were produced by one speaker, the third by the other, with the speaker order counterbalanced across trials. The third stimulus was always the same word as either the first or second stimulus, with matches between third and either first or second word counterbalanced across trials. Of the 288 trials, 48 were test trials in which the first two stimuli differed only in VOT duration (24 bilabial, 24 velar). These trials were constructed such that listeners heard each test contrast (e.g. [b] vs. [p], [b] vs. [p^h], etc.) eight times. The VOT category of the first stimulus in each trial and whether it was the correct answer were also counterbalanced. The other 240 trials were fillers. For 48 of these filler trials the first and second stimuli were test words that differed in more than just VOT. For 96 filler trials one of the first two stimuli was a test word and the other word a filler. For the remaining 96 filler trials, both of the first two stimuli were different fillers. The filler trials were constructed such that each word was heard an equal number of times (either four or five) through the course of the experiment.

3.1.2. Listeners

Half of the participants in the production experiment (N = 20) also participated in Experiment 2 (the other half participated in Experiment 3).

3.1.3. Procedures

All stimuli were presented over headphones with the aid of a computer and SuperLab Pro 4.5. Listeners were told they would be hearing words from the same language they had just heard for the production experiment. They were asked to focus on a fixation cross during which time they heard three words. Listeners then had to press a key to indicate whether the third word was the same as the first or the second word. The “z” key was always used for the former and the “m” key for the latter. A reminder of which key to use for each response was presented after each trial. After pressing “m” or “z” the fixation cross reappeared and 500 ms later the next trial began. Ten participants were tested with English instructions, the other ten with Spanish instructions. BDS score was similar for each group: the group who received instructions in English had a mean score of 12.2, and the group who received instructions in Spanish had a mean score of 9.8 [$t = 0.91$, $p = 0.38$].

3.2. Results

Results based on d′ scores were calculated for all participants and are presented in Figure 2. An ANOVA was conducted to test for significant effects. The dependent variable was d′ score, and the independent variables were test contrast (negative vs. long-lag, negative vs. short-lag, short-lag vs. long-lag) and language of instruction (English, Spanish). Listener was included as a within subject variable by test contrast. Bilabials and velars were analyzed together. There was a significant effect of contrast [$F(2, 36) = 22.11$, $p > 0.001$]. Paired t-tests with adjusted p-values using Holm-Bonferroni correction were run to further examine the effect of contrast. Listeners were significantly better at pre-voiced vs. long-lag than the other two contrasts [$p > 0.05$ for both comparisons]; short-lag vs. long-lag was also better discriminated than pre-voiced vs. short-lag [$p > 0.01$].

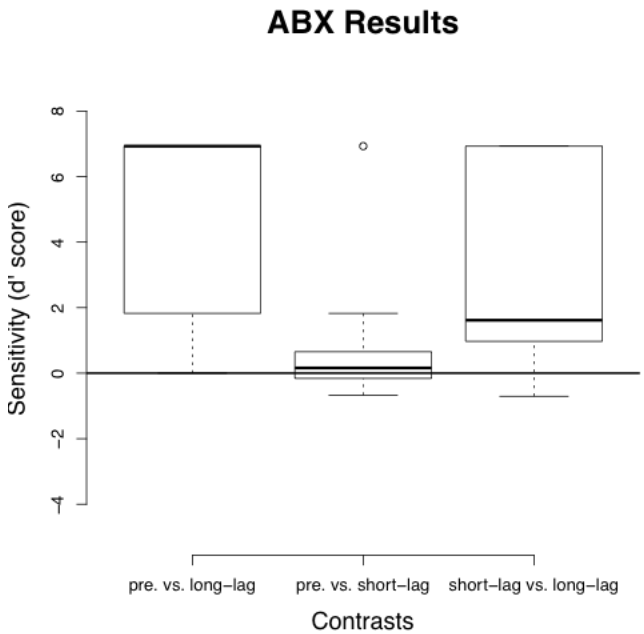

Figure 2: d′ scores for ABX experiment by contrast. Thick bands represent the second quartile (median) d′ scores, top and bottom bands of the box represent first and third quartiles of the data, whiskers represent the lowest and highest data points still within 1.5 interquartile range (IQR) of the lower and higher quartile; data points outside of the whiskers can be considered outliers.

To test for effects of dominance, for each listener, their d′ score for the English contrast (short-lag vs. long-lag) was subtracted from their d′ score for the Spanish contrast (pre-voiced vs. short-lag). A large difference would suggest listeners did much better at one contrast than the other, while a small difference would suggest they did equally well at both. A linear regression was conducted with difference of d′ scores as the dependent variable and BDS score as the independent variable. The regression was significant [$r = -0.66$, $p < 0.01$], with differences getting more negative (better at the English contrast) as the BDS score increased (more English dominant). Thus participants who were more balanced did equally well at both language specific contrasts, while participants who were more English dominant did much better at the English than the Spanish contrast.

4. Experiment 3: AX Perception

4.1. Methods

4.1.1. Materials

The same set of words was used as in Experiment 1. The experiment included 288 trials. Each trial consisted of two stimuli with 500 ms ISI. The two stimuli in each trial were produced by different speakers, with speaker order counterbalanced across trials. For half of the trials (N = 144) the two stimuli were the same word. In the "same" set, half the trials included test items (the initial consonant was a stop) and the other half fillers (see Section 2.1.1.). Listeners heard each "same" pair in two trials (for which different tokens were used). Of the 144 "different" trials, 36 were test items differing only in the VOT of the initial stop. For a given test triplet listeners heard all three possible pairings, e.g. [bɛma]-[pɛma], [bɛma]-[p^hɛma], [pɛma]-[p^hɛma]. The order in which stimuli appeared was counterbalanced so that each word only appeared first once per set of trials. Also, if Speaker A produced [bɛma] in the paring with [pɛma], then Speaker B produced [bɛma] in the pairing with [p^hɛma]. Of the "different" filler trials, 24 consisted of two test stimuli that differed in more than just VOT, 48 trials consisted of a test and a filler stimulus, and 48 of two filler stimuli. Fillers were chosen such that each word and token were heard an equal number of times through the course of the experiment.

4.1.2. Listeners

The listeners were those who had participated in Experiment 1 but had not taken part in Experiment 2.

4.1.3. Procedures

All stimuli were presented over headphones with the aid of a computer and SuperLab Pro 4.5. Listeners were told they would be hearing words from the same language they had just heard in the production experiment. They were told to focus on a fixation cross; while the cross was on screen they heard the two stimuli of a trial. Once the fixation cross went away listeners indicated whether the two words were the "same" or "different" using the keyboard and pressing either "z" or "m" with the meaning of "z" and "m" ("same" or "different") counterbalanced across listeners. Listeners were told that words would not sound exactly the same as they were produced by two different speakers. A reminder of which key to use for each response was presented after listeners heard the words. After pressing the selection button the fixation cross reappeared followed by the next trial 500 ms afterwards. As in Experiment 1, participants were tested with either English or Spanish instructions. BDS scores were similar across groups, with those who received instructions in English having a mean score of 7.5 and those who received instructions in Spanish having a mean score of 7.4 [$t = 0.02$, $p = 0.98$].

4.2. Results

The d′ scores for test contrasts are presented in Figure 3. An ANOVA was conducted to test for significant effects. The dependent variable was d′ score, and the independent variables were test contrast (negative vs. long-lag, negative vs. short-lag, short-lag vs. long-lag) and language of instruction (English, Spanish). Listener was included as a within subject variable by test contrast. There was a significant effect of contrast [$F(2, 36) = 12.45$, $p > 0.001$] which was further investigated using paired t-tests with Holm-Bonferroni correction. Listeners were significantly better at pre-voiced vs. long-lag as compared to both pre-voiced vs. short-lag [$p < 0.001$], and short-lag vs. long-lag [$p > 0.001$]; they performed similarly at short-lag vs. long-lag and pre-voiced vs. short-lag. There were no significant effects of language of instruction and no significant interaction. Again, to test for effects of dominance, for each listener, their d′ score for the English contrast (short-lag vs. long-lag) was subtracted from their d′ score for the Spanish contrast (pre-voiced vs. short-lag). The regression was not significant [$r = 0.17$, $p = 0.47$], suggesting that there was no effect of language dominance for the language specific contrasts.

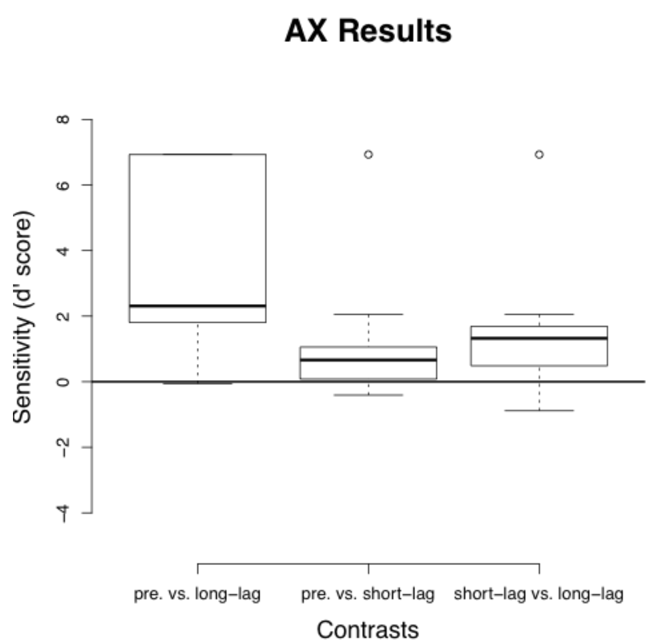

Figure 3: d′ scores for the AX experiment by contrast. Thick bands represent the second quartile (median) d′ scores, top and bottom bands of the box represent first and third quartiles of the data, whiskers represent the lowest and highest data points still within 1.5 interquartile range (IQR) of the lower and higher quartile; data points outside of the whiskers can be considered outliers.

This result was unexpected: it was anticipated that since the AX paradigm is meant to access *phonetic* categories, if bilinguals have access to both their languages they should be able to discriminate all contrasts equally well. Alternatively, if listeners operated in Spanish mode they should discriminate pre-voiced and short-lag stops, while if they operated in English they should discriminate short-lag and long-lag VOT; yet neither was the case. To further investigate why listeners performed poorly on both of these contrasts, results were analyzed separately according to whether the correct answer was same or different. This would allow us to see on which trials listeners were making errors, specifically, whether they were being too liberal in their responses (calling same trials different) or too conservative (calling different trials same). ANOVAs were conducted on each set of data with percent correct as the dependent variable, and contrast, and language of instruction as the independent variables. Listener was included as a within subject variable by test contrast. The results indicate a strong conservative bias. On same trials listeners performed well and performance was not affected by either contrast or language of instruction. A regression with percent correct as the dependent variable and BDS scores as the dependent variable was not significant either [$r = -0.18$, $p = 0.16$]. For different trials, on the other hand, there was a significant effect of contrast [$F(2, 36) = 27.54$, $p < 0.001$] that replicated the d′ scores: listeners were significantly better at pre-voiced vs. long-lag as compared to both pre-voiced vs. short-lag and short-lag vs. long-lag [$p < 0.001$ for both]. There was no difference between pre-voiced vs. short-lag and short-lag vs. long-lag; for both contrasts listeners were at guessing level. There was no effect of language of instruction in the ANOVA, and the regression with BDS scores was not significant either [$r = -0.07$, $p = 0.59$].

5. Discussion

Results from both the production and ABX perception experiments showed that bilinguals are able to access phonological categories from both languages at once in both production and perception, however this was not uniform across all bilinguals. Specifically, only relatively balanced bilinguals were able to access categories from both languages at once. Bilinguals who were English dominant were limited to the English contrast, and grouped pre-voiced with short-lag VOT as a monolingual English speaker would. Due to the nature of the pool of bilinguals available to us, we were not able to examine whether the same effect would hold for bilinguals who were Spanish dominant, i.e. whether such bilinguals would discriminate pre-voicing from short-lag VOT, as in Spanish, but have difficulty discriminating short-lag and long-lag VOT. Based on these results though, we would predict that Spanish dominant bilinguals would respond in this manner.

The results of the AX perception experiment were less clear, as there was no effect of language dominance and listeners did not appear to do better at one language-specific contrast over another; rather, they did poorly in both. While the original intent of this experiment was to test for access of the phonetic categories, we believe that for a number of reasons (such as the use of fillers) phonological categories were accessed instead and consequently participants exhibited a preference for a two-way contrast as in both English and Spanish. However, since all three phonetic categories were present in the stimuli listeners were unsure how to treat short-lag VOT and as a result constantly switched between two possible categorizations (grouping [p] either with [b] or with [p^h]) and ultimately taking the conservative approach of discriminating only the two extremes of the continuum: negative vs. long-lag VOT. Note that this vacillation is also reflected in the bimodal distributions of /b/ and /g/ in the production experiment (see Figure 1) providing further evidence of a possible switch between languages during both production and perception. With respect to perception, one way to test if this interpretation stands would be to run the AX experiment with an initial training session with feedback. The feedback would in effect tell listeners from the start that there is indeed a three-way contrast in the new language they are listening to (though it could jeopardize the low-level processing one would hope to access in an AX experiment).

In contrast to previous studies, one variable that did not appear to have any effect on our results was language mode. There are several possible reasons why this manipulation did not have an effect. First, it is possible that the amount of exposure to the language (English or Spanish) before the experiment was not sufficient to truly put the participants in the "mode" for that language. This is particularly important for Spanish, since the participants habitually operate in a largely English-speaking environment. Second, due to the social evaluation of Spanish in the US context, participants may have actively tried to stay in English mode despite the use of Spanish before and during the experiment. For instance, participants in the Spanish-mode groups would often respond in English despite being spoken to in Spanish by the Spanish-English bilingual experimenter. Finally, language mode may simply be less of a factor than language dominance in bilingual processing when bilinguals are heavily dominant in one language. Perhaps with a more balanced sample of bilinguals language mode would have a larger effect.

Returning to bilingual language organization, the results go against theories that posit single language access based on language mode, as listeners were unaffected by language of instruction. However, there is support for a theory where language dominance is a key factor in processing, as shown by the different responses of the balanced participants and the English dominant group. Taken together their results suggest that bilinguals may indeed control their languages with a schema of activation and suppression, but with bilinguals dominant in one language being better able to suppress the language not in use, without, however, totally eliminating its effects.

As previously discussed though, in addition to questions of access the AX experiment may have also inadvertently tapped into metalinguistic awareness of language structure which could have biased the participants, given their experience with English and Spanish, towards a two-way contrast interpretation of the stimuli. While the original intent of the experiment was to see if bilinguals would access cross-language contrasts without being prompted, this may be harder to simulate in the context of exposure to a new language.

6. Conclusion

The present study examined early Spanish-English bilinguals' abilities to access phonetic categories from both of their languages when exposed to a third. The results show that language dominance but not language mode plays a role in access in both production and perception. Overall, the results suggest that balanced bilinguals have more ready access to both of their languages at a time, while bilinguals dominant in one of their languages operate mostly (but not entirely) in that language. These results argue in favor of a schema of suppression and activation and indicate that greater attention to the linguistic background of bilinguals is needed if the role of language dominance and mode in bilingual production and perception is to be fully understood.

References

Abutalebi, Jubin, Green, David. (2007). "Bilingual language production: The neurocognition of language representation and control". *Journal of Neurolinguistics*, 20(3), pp. 242-275.

Abutalebi, Jubin, Green, David. (2008). "Control mechanisms in bilingual language production: Neural evidence from language switching studies". *Language & Cognitive Processes*, 23(4), pp. 557-582.

Costa, Albert, Caramazza, Alfonso, Sebastián-Gallés, Nuria. (2000). "The cognate facilitation effect: Implications for models of lexical access". *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *26*(5), pp. 1283-1296.

Cutler, Anne, Mehler, Jacques, Norris, Dennis, Seguí, Juan. (1992). "The monolingual nature of speech by bilinguals". *Cognitive Psychology*, 24(3), pp. 381-410.

Goldstein, Brian A., Fabiano, Leah, Washington, Patricia S. (2005). "Phonological skills in predominantly English-speaking, predominantly Spanish-speaking, and Spanish-English bilingual children". *Language, Speech, and Hearing Services in Schools*, 36(3), pp. 201-218.

Goldstein, Brian, Washington, Patricia S. (2001). "An initial investigation of phonological patterns in typically developing 4-year-old Spanish-English bilingual children". *Language, Speech, and Hearing Services in Schools*, 32(3), pp. 153-164.

Dunn, Alexandra L., Fox Tree, Jean E. (2009). "A quick, gradient Bilingual Dominance Scale". *Bilingualism: Language and Cognition*, 12(3), pp. 273-289.

Elman, Jeffrey L., Diehl, Randy L., Buchwald, Susan E. (1977). "Perceptual switching in bilinguals". *Journal of the Acoustical Society of America*, 62(4), pp. 971-974.

Flege, James E., Eefting, Wieke. (1987). "Production and perception of English stops by native Spanish speakers". *Journal of Phonetics*, 15, pp. 67-83.

Green, David W. (1986). "Control, activation, and resource: A framework and a model for the control of speech in bilinguals". *Brain and language*, 27(2), pp. 210–223.

Green, David W. (1998). “Mental control of the bilingual lexico-semantic system”. *Bilingualism: Language and Cognition*, 1(2), pp. 67–81.

Indefrey, Peter. (2006). “A meta-analysis of hemodynamic studies on first and second language processing: Which suggested differences can we trust and what do they mean?” *Language Learning*, 56(s1), pp. 279-304.

Leonard, Mattew K., Brown, Timothy T., Travis, Katherine E., Gharapetian, Lusineh, Hagler, Donald J. Jr., Dale, Anaders M., Elman, Jeffrey L., Halgren, Eric. (2010). “Spatiotemporal dynamics of bilingual word processing”. *NeuroImage*, 49(4), pp. 3286-3294.

Macnamara, John, Kushnir, Seymour L. (1971). “Linguistic independence of bilinguals: The input switch”. *Journal of Verbal Learning and Verbal Behavior*, 10(5), pp. 480-487.

Marian, Viorica, Spivey, Michael. (2003). “Competing activation in bilingual language processing: Within- and between-language competition”. *Bilingualism: Language and Cognition*, 6(2), pp. 97-115.

Marian, Viorica, Spivey, Michael, Hirsch, Joy. (2003). “Shared and separate systems in bilingual language processing: Converging evidence from eyetracking and brain imaging”. *Brain and Language*, 86(1), pp. 70-82.

Misra, Maya, Guo, Taomei, Bobb, Susan C., Kroll, Judith F. (2012). “When bilinguals choose a single word to speak: Electrophysiological evidence for inhibition of the native language”. *Journal of Memory and Language*, 67(1), pp. 224-237.

Moreno, Eva M., Rodríguez-Fornells, Antoni, Laine, Matti. (2008). “Event-related potentials (ERPs) in the study of bilingual language processing”. *Journal of Neurolinguistics*, 21(6), pp. 477-508.

Phillips, Natalie A., Klein, Denise, Mercier, Julie, de Boysson, Chloé. (2006). “ERP measures of auditory word repetition and translation priming in bilinguals”. *Brain Research*, 1125(1), pp. 116-131.

Rodriguez-Fornells, Antoni, van der Lugt, Arie, Rotte, Michael, Britti, Belinda, Heinze, Hans-Jochen, Münte, Thomas F. (2005). “Second language interferes with word production in fluent bilinguals: Brain potential and functional imaging evidence”. *Journal of Cognitive Neuroscience*, 17(3), pp. 422-433.

Roelofs, Ardi, Verhoef, Kim. (2006). “Modeling the control of phonological encoding in bilingual speakers”. *Bilingualism: Language and Cognition*, 9(2), pp. 167-176.

Spivey, Michael J., Marian, Viorica. (2011). “Cross talk between native and second languages: Partial activation of an irrelevant lexicon”. *Psychological Science*, 10(3), pp. 281-284.

Sundara, Megha, Polka, Linda, Baum, Shari. (2006). “Production of coronal stops by simultaneous bilingual adults”. *Bilingualism: Language and Cognition*, 9(1), pp. 97-114.

van Heuven, Walter J. B., Dijkstra, Ton. (2010). “Language comprehension in the bilingual brain: fMRI and ERP support for psycholinguistic models”. *Brain Research Reviews*, 64(1), pp. 104-122.