MONOLINGUAL SPEECH PRODUCTION IN A BILINGUAL CONTEXT

Jessica A. Barlow[1] and Philip N. Combiths[1,2]

[1]San Diego State University and [2]University of California, San Diego
jbarlow@sdsu.edu, pcombiths@sdsu.edu

ABSTRACT

In language contact contexts, monolinguals may have knowledge of another language due to exposure. We tested the phonetic and phonological knowledge of 20 English monolinguals from Southern California (an English-Spanish contact community) on a bilingual task. We analysed their productions of /l/ and /p t k/, phonetically different phonemes that vary by context in English but not Spanish. Specifically, English /l/ varies in "darkness" due to an allophonic velarization rule, while /p t k/ vary in voice onset time due to an allophonic aspiration rule. Results suggest that so-called monolinguals in contact contexts may have phonetic but not phonological knowledge of the "unknown" contact language. Participants showed a qualitative difference between their productions of the same phonemes shared by the two languages, but transferred allophonic patterns of English to their Spanish productions. We consider these findings in light of future research that compares monolingual and bilingual speech production.

Keywords: monolingual, language contact, transfer, allophony, acoustic

1. INTRODUCTION

Research has shown that bilinguals show subtle phonetic differences in production of phonemes shared between their languages. These differences are also impacted by phonological context, particularly in instances of allophonic variation [1, 2]. As well, such differences are distinct from productions by monolinguals in the respective languages, due to interaction between the two languages [3].

By "shared phonemes", we refer to those sounds that have the same articulatory description, and perhaps the same orthographic representation, even though the sounds may differ along more subtle phonetic dimensions and phonological patterning. Despite such differences, these shared phonemes do appear to be recognized as such by bilinguals and second language learners, based on production and perception evidence [2, 4, 5]. Sensitivity to the shared properties of sound systems extends beyond influence of orthography, and is present in production patterns of preliterate children who do not necessarily have access to the shared orthography [6-8].

Studies of accent imitation have shown that second language learners exhibit tacit awareness of non-distinctive phonetic differences between the first and second language [9-11], but the distributional properties associated with allophones may influence the extent to which they invoke these changes [12]. In language contact contexts, even monolinguals may have some knowledge of another language due to exposure and may show similar such knowledge of phonetic differences between shared sounds of the two languages [13, 14]. Knowledge of this sort has been referred to by some as "incipient bilingualism" [15], more typically reflecting initial stages of contact between languages in a community.

The purpose of this study was to evaluate the phonetic and phonological knowledge of *Spanish* by self-identified *English* monolinguals residing in an English-Spanish language context. We conducted a bilingual task to compare production of phonemes that are shared between the two languages but differ along phonetic and phonological dimensions. We focused on the shared phonemes /l/ and /p t k/, which have subtle phonetic differences across the two languages that are detectable by novice second language learners [9-11], and also vary by phonological context in English, but not Spanish.

Specifically, English approximant /l/ is qualitatively "darker" than Spanish /l/, due to lower second formant (F2) values and a smaller difference between F2 and the first formant (F1). Further, English has an allophonic rule of postvocalic /l/ velarization, such that postvocalic /l/ is produced with an even darker quality than prevocalic /l/ (e.g., "lid" [lɪd] vs. "dill" [dɪɫ]). In Spanish, /l/ is produced as "clear" across contexts.

Similarly, foot-initial /p t k/ plosives in English show longer lags (based on measures of voice onset time [VOT]), as compared to non-foot-initial contexts, due to an allophonic aspiration rule (as with "pie" [p^haɪ] vs. "spy" [spaɪ]). In Spanish, all plosives are produced with short lags, and do not vary by phonological context.

2. METHODS

2.1. Participants

Participants were 20 college students (16 females), with an average age of 20.5 years (range: 18 to 27 years). All were English monolinguals per self-report and responses to a detailed language use questionnaire that allowed for the opportunity to indicate any and all experience with another language, in terms of both input and output (adapted from [16]).

All participants spent the majority of their childhoods in Southern California, where both English and Spanish are prevalent and part of everyday print, signage, and broadcast media. All participants spoke a Southern California dialect of English. In addition, based on self-report, all participants were determined to have normal or corrected-to-normal vision, normal hearing, and no history of developmental delays or disorders, acquired cognitive disabilities, or any speech or language difficulties requiring clinical intervention.

2.2. Task

A 59-item word list was created for each language, to include five mono- or di-syllabic /l/ words in word-initial ("initial") and -final ("final") contexts each, to total ten /l/ words. Also included were five mono- or di-syllabic words sampling /p t k/ each in initial position and following /s/ ("after /s/"), totaling 30 /p t k/ words. For each phoneme and context, words were balanced across the five corner vowels /i u e o a/. Twenty-nine additional foils that varied in syllable shape were included in the word list to total 59 words.

Each participant read the word lists three times in different order each time (to yield 177 productions) in the carrier phrase, "Say __ again" or *Di* __ *ahora* ("Say __ now"). They read the list three times in English and Spanish each, and the order of language was counterbalanced across participants. Because the participants were monolingual English speakers and did not know Spanish, it was necessary to explain the task. They were told to read each phrase casually and that some of the lists would be in English and some would be in Spanish. They were *not* given specific instructions to "sound like a Spanish speaker"; rather, they were simply told to not be concerned if they pronounced any words incorrectly. Because Spanish orthography is fairly transparent, it was not difficult for participants to decode the pronunciation of each word. The majority of phonemic "errors" (non-target forms) were on vowels (e.g., /a/ pronounced as [eɪ]), the insertion of a liquid (e.g., *pata* as [plɑtə]), or the deletion of a liquid (e.g., *clase* as [kɑsə]). Similar such errors also occurred for some of the English words (e.g., "cot" as [koʊt]). All such phonemic errors were omitted from planned acoustic analyses.

All utterances were digitally recorded directly onto a Roland Edirol R-09 digital recorder at a sampling rate of 44.1 kHz, via a Sony ECM-MS907 omnidirectional electret condenser microphone. Recording files were transferred to a computer server and stored in an uncompressed format.

2.3. Analyses

Of interest were the participants' productions of the 15 initial and 15 final /l/ productions and the 30 initial and 30 after-/s/ productions of /p t k/, which elicited 600 /l/ productions and 1,800 /p t k/ productions in each language across the 20 subjects. As stated, any phonemic errors described above were excluded from analysis. Additional items were excluded due to extraneous noise in the signal (e.g., the participant bumped against the table). In total, 1,186 /l/ forms and 3,498 /p t k/ forms were acoustically analyzed across the two languages. All acoustic analyses were done via Praat software [17] by trained undergraduate and graduate research assistants who followed a strict protocol for measurement of formant frequencies (for /l/ darkness) and VOT (for /p t k/ aspiration). These protocols are described next.

To evaluate relative /l/ darkness by language and context, we measured F1 and F2 values of each /l/ production following procedures of prior research [2, 6, 18]. We evaluated waveform and spectrogram displays of each /l/ utterance within Praat and identified the midpoint of each relevant production visually from the displays and perceptually via headphones. Raw F1 and F2 values were extracted from the midpoint of each /l/ production. From those values, the raw F2-F1 Difference was also calculated.

F2 values also were determined in the same manner for the vowels /i/ and /o/ in order to normalize /l/ F2 values, following the S-procedure of Watt and Fabricius [19], described by Simonet [18] and Barlow [2]. This procedure reduces interspeaker variation due to anatomical and physiological differences across participants, yet maintains the interspeaker variation attributable to language and dialect differences [20]. The /i/ and /o/ vowels were selected as front and back extremes of the vowel space, and served as reference points for determining the centroid F2 value of each speaker's vowel space. Specifically, mean F2 values were determined for /i/ and /o/ for each speaker (based on the above-mentioned word list), and then a grand mean of those two vowels served as the centroid F2 value for that speaker [2, 18]. (We used /o/ instead of /u/, because the latter is typically much "fronter" in the vowel space in the California English dialect [21-24]. In the Spanish spoken in this region, /o/ and /u/ have comparable backness [24].) From this value, the normalized F2 value for each /l/ production was calculated by dividing the raw F2 value for each /l/ by the centroid F2 value for each speaker.

To evaluate presence versus absence of aspiration in the /p t k/ productions in each language, we measured VOT of each plosive in initial position and after /s/ in Praat, following procedures described in prior research [25]. We evaluated the waveform and spectrogram displays of each /p t k/ within Praat and identified the closure period visually from displays and perceptually via headphones. The duration of this period from burst to onset of periodicity provided the VOT values for each plosive.

For reliability purposes, 866 /l/ and vowel forms and 1,136 /p t k/ forms were reanalyzed for raw F1 and F2 values by a second researcher trained in the above-mentioned protocols. Correlation between the two judges for F1 was $r(854) = .682$; for F2, $r(854) = .903$; and for VOT, $r(1105) = .830$ (all p's < .01).

To test whether or not the participants produced a difference by language and by context, we used a mixed model to analyze mean normalized F2, F2-F1 difference, and VOT values by language and context, with subject as a random factor. For VOT, phoneme was also included in the model, because /p t k/ are known to differ with respect to VOT [26]. Interactions between language and context were included to determine how the variables interacted on normalized F2 values and F2-F1 differences during the production of /l/, and on VOT for the plosives.

3. RESULTS

3.1. /l/ darkness

Tables 1 and 2 display mean normalized F2 and F2-F1 differences, respectively, for /l/ by language and context. For normalized F2, analyses showed main effects of language, $F(1,1182) = 298.8$, $p < .01$, and context, $F(1,1182) = 109.2$, $p < .01$, but no Language x Context interaction ($p = .25$). Mean normalized F2 values were lower for English versus Spanish in both initial and final positions, indicating a darker /l/ for English in both contexts. In both languages, initial /l/ was produced with a higher mean normalized F2 value as compared to final /l/, indicative of a darker /l/ in final position in both languages.

Table 1: Means (and Standard Deviations) of Normalized F2 by Language and Context.

Context	Normalized F2	
	English	Spanish
Initial	.64 (.16)	.82 (.20)
Final	.55 (.11)	.70 (.19)

For F2-F1 differences, there were main effects of language, $F(1,1182) = 164.81$, $p < .01$, and context, $F(1,1182) = 312.4$, $p < .01$, but no interaction ($p = .69$). F2-F1 differences were smaller in English versus Spanish in both contexts, again consistent with a darker /l/ in English. As with normalized F2, the F2-F1 differences were greater in initial as opposed to final position in both languages, once again indicating a darker /l/ in final position.

Table 2: Means (and Standard Deviations) of F2-F1 Differences in Hz by Language and Context.

Context	F2-F1 Difference in Hz	
	English	Spanish
Initial	843.3 (297.5)	1069.4 (348.8)
Final	548.1 (181.4)	760.7 (322.9)

3.2. Voice onset time

Table 3 shows mean VOTs by language and context for /p t k/. Analyses revealed main effects of phoneme $F(2,3492) = 126.9$, language, $F(1,3492) = 117.4$, and context, $F(1,3492) = 3495.6$, as well as a Language x Context interaction, $F(1,3492) = 348.5$ (all p's < .01). Mean VOTs were overall longer in English versus Spanish; moreover, VOTs were longer in initial position versus after /s/ in both languages.

Table 3: Means (and Standard Deviations) of VOT in seconds by Phoneme, Language, and Context.

Phoneme	Context	VOT in seconds English	Spanish
/p/	Initial	.09 (.02)	.07 (.03)
	After /s/	.02 (.01)	.03 (.03)
/t/	Initial	.11 (.03)	.09 (.04)
	After /s/	.03 (.01)	.04 (.02)
/k/	Initial	.11 (.03)	.09 (.04)
	After /s/	.04 (.02)	.05 (.03)

Note that the Language x Context interaction revealed a greater difference by context for English than for Spanish. Further, following /s/, Spanish VOTs were slightly longer than English VOTs.

5. DISCUSSION

Participants exhibited a phonetic difference between English and Spanish productions in both allophonic contexts for /l/ and for /p t k/. For /l/, participants applied the allophonic velarization rule to both languages similarly, as evidenced by the lack of a Language x Context interaction. In contrast, for /p t k/, a Language x Context interaction revealed that the allophonic aspiration rule of English was present in the Spanish productions, but the lag differences by context were smaller, likely due to syllabification differences between after-/s/ forms in the two languages. That is, in English, /s/ + plosive sequences may occur, and were sampled herein, word-initially (e.g., "ski" [ski]). In Spanish, such clusters only occur word-internally, which also was the context that was sampled (e.g., *esquí* [eski]).

The findings from this study indicate that even monolinguals in a language contact context have phonetic knowledge of subtle acoustic differences between phonemes that are shared between the two languages of the community. These findings are consistent with prior research on accent imitation by experienced and novice language learners alike, including research that has looked at the very same phenomena of VOT and /l/ darkness [9-11]. The monolinguals in the current study are not second language learners; yet, they may be demonstrating some level of incipient bilingualism observed in early language contact situations [15].

The study task may have tapped into participants' abilities to imitate accents, although they were not told to do so. Regardless, their differences in production between English and Spanish reflect some tacit knowledge of the phonetic differences between the two languages, likely due to input from the surrounding community; such information is rarely explicitly taught to learners, let alone non-learners. Through exposure to Spanish in a language contact context, the frequency with which short lag stops and clear /l/ occur in Spanish may be detectable to learners and non-learners alike, including the monolinguals in this study.

Although the participants showed knowledge of the phonetic differences between the two languages in their Spanish versus English productions of /p t k/ and /l/, the English phonological rules of aspiration and velarization were present in their productions in *both* languages. Allophonic rules such as these are known to transfer in second language acquisition and are difficult to suppress [27, 28]. Suppression of allophonic rules in second language acquisition is expected to occur once lexical representations in the target language input provide evidence of contrast between those sounds governed by the allophonic rule of the first language [12, 27, 28]. Yet, in the absence of any evidence of a phonemic contrast from the lexicon of the second language, there may not be sufficient evidence to motivate learners to suppress the allophonic rule. That is to say, the Spanish language does not contrast short- and long-lag stops, nor does it contrast clear and dark laterals. English learners of Spanish therefore are not exposed to lexical forms in the input that would disrupt the aspiration and velarization rules; thus, they would be expected to continue to apply these rules in their productions in Spanish. The monolinguals in the current study arguably had relatively few lexical representations and little to no knowledge of phonological rules in Spanish, despite their tacit phonetic knowledge of the language. Accordingly, they had no available evidence to motivate suppression of aspiration and velarization in their attempts at the Spanish words. It is interesting to note that allophonic rule suppression of this type is not widely studied in second language research and is a direction for future research.

There were larger standard deviations for Spanish productions as compared to English in all measures. This could be attributed to the likelihood that participants' articulatory patterns associated with the Spanish forms are less entrenched and therefore more variable than those of English [29]. The larger standard deviations may also be due to individual differences in the extent to which participants differentiated the two languages. As noted in 2.1, participants did not speak any language other than English; however, we did not document the amount of participants' incidental exposure to Spanish through their everyday activities. It may be that some participants lived in neighbourhoods or participated in activities that allowed for greater and/or more consistent exposure to Spanish. It is also possible that some participants had greater acoustic/phonetic awareness and were better able to detect subtle such acoustic differences in the language community. This is a direction for future inquiry.

If the findings of the current study are replicated in other language contact contexts, this would have important implications for bilingual research. Specifically, any differences that are observed between bilinguals and monolinguals in such communities are all the more robust [3], given that monolinguals show some knowledge of the second language. Future research also should compare productions of monolinguals from monolingual communities with those from (nearby) bilingual communities in order to further understand the role of interference on speech production in language contact communities [13].

6. ACKNOWLEDGEMENTS

Thanks to the participants of this study, to members of the San Diego State University Phonological Typologies Laboratory for assistance with data collection and analysis, and to three anonymous reviewers who provided input on an earlier draft of this paper.

7. REFERENCES

[1] Caramazza, A., Yeni-Komshian, G. H., Zurif, E. B. et al. 1973. The acquisition of a new phonological

contrast: The case of stop consonants in French-English bilinguals. *J. Acoust. Soc. Am.* 54, 421-428.
[2] Barlow, J. A. 2014. Age of acquisition and allophony in Spanish-English bilinguals. *Front. Psychol.* 5, 1-14.
[3] Flege, J. E. 1991. Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *J. Acoust. Soc. Am.* 89, 395-411.
[4] Flege, J. E. 1981. The phonological basis of foreign accent: A hypothesis. *TESOL Quart.* 15, 443-455.
[5] Flege, J. E. 1987. The production of 'new' and 'similar' phones in a foreign language: Evidence for the effect of equivalence classification. *J. Phon.* 15, 47-65.
[6] Barlow, J. A., Branson, P. E., Nip, I. S. B. 2013. Phonetic equivalence in the acquisition of /l/ by Spanish-English bilingual children. *Biling.-Lang. Cogn.* 16, 68-85.
[7] Fabiano-Smith, L., Bunta, F. 2012. Voice onset time of voiceless bilabial and velar stops in 3-year-old bilingual children and their age-matched monolingual peers. *Clin. Linguist. Phonet.* 26, 148-163.
[8] Fabiano-Smith, L., Goldstein, B. A. 2010. Phonological acquisition in bilingual Spanish-English speaking children. *J. Speech Lang. Hear. R.* 53, 160-178.
[9] Flege, J. E., Hammond, R. M. 1982. Mimicry of non-distinctive phonetic differences between language varieties. *Stud. Second Lang. Acq.* 5, 1-17.
[10] Mora, J., Rochdi, Y., Kivisto-de Souza, H. 2014. Mimicking accented speech as L2 phonological awareness. *Lang. Aware.* 23, 57-75.
[11] Zuengler, J. 1988. Identity markers and L2 pronunciation. *Stud. Second Lang. Acq.* 10, 33-49.
[12] Whalen, D., Best, C., Irwin, J. 1997. Lexical effects in the perception and production of American English /p/ allophones. *J. Phon.* 25, 501-528.
[13] Mayr, R., Morris, J., Mennen, I. et al. 2017. Disentangling the effects of long-term language contact and individual bilingualism: The case of monophthongs in Welsh and English. *Int. J. Billing.* 21, 245-267.
[14] Spreafico, L., Vietti, A. 2013. On rhotics in a bilingual community: A preliminary UTI research. In: Spreafico, L, Vietti, A (eds), *Rhotics. New data and perspectives*. Bolzano, Italy: Bozen-Bolzano University Press, 57-80
[15] Diebold, A. R. 1964. Incipient bilingualism. In: Hymes, D (ed), *Language in culture and society*. New York: Harper and Row
[16] Gutiérrez-Clellen, V. F., Kreiter, J. 2003. Understanding child bilingual acquisition using parent and teacher reports. *Appl. Psycholinguist.* 24, 267-288.
[17] Boersma, P., Weenink, D. 2008. *Praat: Doing phonetics by computer*. In. v. 5.0.26 ed: Retrieved June 16, 2008, from http://www.praat.org
[18] Simonet, M. 2010. Dark and clear laterals in Catalan and Spanish: Interaction of phonetic categories in early bilinguals. *J. Phon.* 38, 663-678.
[19] Watt, D., Fabricius, A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1–F2 plane. *Leeds Working Pap. Linguist.* 9, 159-163.
[20] Flynn, N., Foulkes, P. 2011. Comparing vowel formant normalization methods. In: Lee, W-S, Zee, E (eds), *Proc. 17th ICPhS*, 683-686
[21] Hagiwara, R. 1995. *Acoustic realizations of American /r/ as produced by women and men* [Doctoral dissertation]. University of California, Los Angeles,
[22] Hagiwara, R. 1997. Dialect variation and formant frequency: The American English vowels revisited. *J. Acoust. Soc. Am.* 102, 655-658.
[23] Labov, W., Ash, S., Boberg, C. 2006. *Atlas of North American English: Phonetics, phonology, and sound change*. New York: Mouton de Gruyter.
[24] Grijalva, C., Piccinini, P. E., Arvaniti, A. 2013. The vowel spaces of Southern Californian English and Mexican Spanish as produced by monolinguals and bilinguals. *J. Acoust. Soc. Am.* 33, 3340.
[25] Torre III, P., Barlow, J. A. 2009. Age-related changes in acoustic characteristics of adult speech. *J. Commun. Disord.* 42, 324-333.
[26] Lisker, L., Abramson, A. S. 1964. A cross-language study of voicing in initial stops: Acoustical measurements. *Word* 20, 384-422.
[27] Simon, E. 2010. Phonological transfer of voicing and devoicing rules: evidence from L1 Dutch and L2 English conversational speech. *Lang. Sci.* 32, 63-86.
[28] Curtin, S., Goad, H., Pater, J. V. 1998. Phonological transfer and levels of representation: the perceptual acquisition of Thai voice and aspiration by English and French speakers. *Second Lang. Res.* 14, 389-405.
[29] Nip, I. S. B., Blumenfeld, H. K. 2015. Proficiency and Linguistic Complexity Influence Speech Motor Control and Performance in Spanish Language Learners. *J. Speech Lang. Hear. R.* 58, 653-668.