Bilingualism: Language and Cognition

cambridge.org/bil

Consequences of mixing and switching languages for retrieval and articulation

Maria Fernanda Gavino and Matthew Goldrick

Department of Linguistics, Northwestern University, Evanston, IL, USA

Research Article

Supplementary material can be found online at https://doi.org/10.1017/S1366728922000682

Cite this article: Gavino MF, Goldrick M (2023). Consequences of mixing and switching languages for retrieval and articulation. *Bilingualism: Language and Cognition* **26**, 504–515. https://doi.org/10.1017/S1366728922000682

Received: 20 August 2021
Revised: 31 August 2022
Accepted: 20 September 2022
First published online: 28 November 2022

Keywords:
phonetics; lexical access; interaction; Spanish–English bilinguals

Address for correspondence:
Maria Fernanda Gavino
Department of Linguistics
Northwestern University
2016 Sheridan Rd.
Evanston, IL 60208 USA
Email: mariagavino@u.northwestern.edu

Abstract

A large literature has shown that language context –mixing and switching between languages – impacts lexical access processes during bilingual speech production. Recent work has suggested parallel contextual effects of language context on the phonetic realization of speech sounds, consistent with interactions between lexical access and phonetic processes. In this pre-registered study, we directly examine the link between lexical access and phonetic processes in Spanish–English bilinguals using picture naming. Using automated acoustic analysis, we simultaneously gather measures of reaction time (indexing lexical access) and acoustic properties of the initial consonant and vowel (indexing phonetic processes) for the same speakers on the same trials. Across measures, we find consistent, robust effects of mixing and language dominance. In contrast, while switching effects are robust in reaction time measures, they are not detected in phonetic measures. These inconsistent effects suggest there are constraints on the degree of interaction between lexical access and phonetic processes.

Introduction

Bilinguals have the amazing ability to flexibly control the language of production. They can limit speech to one language or mix languages within the same context, all while rarely producing errors (e.g., Gollan & Goldrick, 2018). However, this flexibility comes at a cost to processing speed. A large body of experimental work has shown bilinguals are slower to retrieve words from memory when participants are required to mix languages (e.g., name pictures in multiple languages as opposed to a single language); when mixing, retrieval slows further when switching from one language to another across naming trials (see Kleinman & Gollan, 2018, for a recent review). Recent studies have provided evidence that the articulation of speech sounds is also impacted by changing language contexts. Specifically, the phonetic distinction between similar speech sounds across languages is reduced when mixing multiple languages (see Amengual, 2021, for a recent review).

For example, voice onset time (VOT), the time between the release of the consonant's constriction and the onset of periodicity, is a primary cue to the distinction between voiced and voiceless sounds in both English and Spanish (e.g., "big" versus "pig", "beso" versus "peso"). In both cases, VOTs are shorter (or more negative) in Spanish than English. Spanish voiced stops ("beso") are pre-voiced (voicing before the release of the consonant constriction, a negative VOT), while English voiced stops ("big") are produced with a short positive gap between voicing and release of the stop (a small, ~0-30ms positive VOT). Spanish voiceless stops ("peso") are produced with short, positive VOTs (~0-30ms) relative to voiceless English stops ("pig;" ~30-90ms; Lisker & Abramson, 1964). As we review in more detail below, previous studies of cued language switching in picture naming have observed lengthening of Spanish VOTs and/or shortening English VOTs, such that each sound becomes more like its counterpart in the other language, reducing the contrast between them (e.g., Goldrick, Runnqvist, & Costa, 2014; Olson, 2013).

The effect of language context on both reaction times and the phonetic properties of speech in tasks such as cued picture naming has been interpreted through the lens of theories of speech production. These theories are integrating the processes of retrieving words from memory with the processes specifying the phonetic detail of words. As discussed in more detail below, both exemplar (Amengual, 2018) and cascading activation accounts (Goldrick et al., 2014) allow the processing of these two aspects of speech production to overlap, correctly predicting that both should be sensitive to the same variables.

While such accounts are consistent with existing data, most empirical studies fail to gather measures of retrieval AND phonetics from the same participants while they are performing the same task (for exceptions, see Goldrick, Shrem, Kilbourn-Ceron, Baus, & Keshet, 2021; Gustafson & Goldrick, 2018; Jacobs, Fricke, & Kroll, 2016). To our knowledge, no studies have examined this issue in the context of language mixing and switching. The current study aims to address this gap. English-dominant Spanish–English bilinguals performed a cued-language-switching picture naming task; we gathered reaction times (to examine word

© The Author(s), 2022. Published by Cambridge University Press

retrieval) and phonetic measures of consonants and vowels (to examine phonetic processing in speech production). All analyses were done within participants. These existing models predict that we should observe similar effects of mixing and switching in both reaction times and phonetic measures, and that variation in reaction times should be related to variation in phonetic measures.

The remainder of this paper is structured as follows. We begin by reviewing previous work examining the effect(s) of language context on bilinguals' lexical access and phonetic processing in speech production. This motivates the design and methods of our current study. The results show a mixture of results across lexical access and phonetic processing, presenting a challenge for existing theories. We conclude by discussing what aspects of these theories need to be revised and extended to account for these findings, including areas for future empirical research.

Effects of language context on lexical access and phonetic processing

Language contexts in speech production

We define a SINGLE LANGUAGE context as an experimental block where only one of their languages is used (e.g., a Spanish–English bilingual using exclusively English within a group of trials). MIXED LANGUAGE context refers to an experimental block where both languages are used (e.g., a Spanish–English bilingual using both English and Spanish within a group of trials). Mixed contexts include STAY (i.e., preceding trial is in same language) and SWITCH (i.e., preceding trial in different language) contexts (Gollan & Ferreira, 2009; Meuter & Allport, 1999).

Mechanisms of bilingual speech production

Psycholinguistic theories of speech production typically distinguish several stages of post-semantic processing (e.g., concepts, lexical items, phonological, and phonetic representations; Levelt, 1989). At each level of processing, there is co-activation of representations (see Melinger, Branigan, & Pickering, 2014, for a review), both within and across languages (e.g., the concept DOG activates both the target word <*perro*> and its translation equivalent <*dog*>). In bilinguals, the degree of co-activation varies across language contexts (e.g., when switching languages, there is increased co-activation of representations in the target and non-target language relative to stay trials; Bobb & Wodniecka, 2013). Within this framework, some proposals incorporate interaction between these stages of processing (e.g., Goldrick et al., 2014). Co-activation in lexical access yields co-activation of target and non-target phonetic properties, producing articulations that blend properties of both languages – reducing the contrast between speech sounds across languages (e.g., reducing the difference in length of VOTs for voiceless stops in Spanish versus English). To the extent that co-activation in retrieval varies as a function of language context, such an account predicts parallel effects of context on retrieval and phonetic measures.

Exemplar theories of speech production (e.g., Pierrehumbert, 2002) represent an alternative conceptualization of these processes but make similar predictions. In such accounts, experiences of speech perception/production (exemplars), encoded at multiple levels of linguistic structure (lexical, phonological, phonetic, social, contextual, etc.), are linked together in long-term memory. Production is guided by co-activated exemplars produced in similar contexts, leading to enhanced activation of cross-language exemplars in mixing and switching (Amengual, 2018). Similar to the cascading activation, this predicts reduction of the phonetic contrast of speech sounds across languages.

Effects of language context on lexical access

Under each of these accounts, lexical access should be easiest in single language contexts because of lower co-activation of representations across languages. This is consistent with the findings of studies which found faster reaction times in single language blocks in comparison to mixed language blocks (e.g., Christoffels, Firk, & Schiller, 2007; Gollan & Ferreira, 2009; Hernandez & Kohnert, 1999; Kleinman & Gollan, 2018; Prior & Gollan, 2013; Weissberger, Wierenga, Bondi, & Gollan, 2012). In mixed language contexts, bilinguals have been found to retrieve words even slower in switch contexts when compared to stay contexts; increased competition between the target word and its cross language competitor causes the target word to be retrieved more slowly (Declerck, 2020; Meuter & Allport, 1999; for a review, see Bobb & Wodniecka, 2013). Note that different types of control processes can contribute to variation in retrieval times in mixed language contexts. In other words, proactive control processes anticipate interference, which contribute to mixing costs, while reactive control processes engage at points where the non-target language interferes with word selection (see Declerck, 2020, for discussion). However, the key predictions of this study rely only on differences in the degree of coactivation and not these control process distinctions. Theories integrating lexical retrieval and phonetic detail of words predict that when there are greater differences in coactivation, larger phonetic effects should be observed.

Studies have also found a reversed dominance effect in mixed contexts. In single language contexts, bilinguals are faster at retrieving words in their dominant language than in their non-dominant language; in mixed contexts, the asymmetry is weakened or reversed (e.g., Branzi, Martin, Abutalebi, & Costa, 2014; Declerck, Kleinman, & Gollan, 2020; Gollan & Ferreira, 2009; Kleinman & Gollan, 2018). This effect has been attributed to proactive control processes that select representation in the target language by inhibiting representations in the non-target language (Declerck et al., 2020). When bilinguals are using both their dominant and non-dominant language, they aim to equalize accessibility of the two languages, leading them to strongly inhibit the dominant language. Speakers sometimes 'overshoot' this equal accessibility target, applying greater inhibition than strictly required, yielding a reversal of dominance effects. As theories integrating lexical retrieval and phonetic detail of words are sensitive to relative levels of activation, they predict that when the dominant language is inhibited, it should be more susceptible to phonetic effects.

Effects of language context on phonetic processing

Cascading activation and exemplar accounts predict that variation in the co-activation of representations across languages should simultaneously influence reaction times and the phonetic contrasts between languages. Previous work has not directly examined the question. Instead, it has focused on how language context influences measures of retrieval only (studies reviewed above) or how context influences phonetic contrasts (studies reviewed below). We consider three previous studies that are

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

most similar to our work; these serve to illustrate the diversity of findings that have been reported in the literature.

Goldrick et al. (2014) tested Spanish(L1)–English(L2/3) speakers residing in Barcelona, Spain (all speakers had some knowledge of Catalan). They used a cued picture naming task in mixed language contexts, contrasting how the phonetic property of VOT varied across stay and switch trials. As introduced above, VOT is utilized in contrasting ways across these two languages. For example, English words beginning with voiceless stops like /p/ ("pig") have longer VOTs than Spanish words with initial voiceless stops ("peso"; note Catalan utilizes VOT similar to Spanish). The contrast between languages for voiceless stops could therefore be reduced by increasing Spanish VOTs and/or decreasing English VOTs. Consistent with this, Goldrick et al. found decreased VOT in English, the non-dominant language, for switch contexts in comparison with stay contexts. Similar results were found for voiced stops ("big" versus "beso"). English realizes voiced stops with short positive VOTs, while Spanish produces similar sounds with negative VOTs (*pre*voicing stops before closure). Goldrick et al. found that speakers reduced the contrast between Spanish and English voiced stops by altering productions in the non-dominant language; they were more likely to produce voiced stops in English words with Spanish-like negative VOTs on switch versus stay trials.

Olson (2013) tested Spanish(L1)–English(L2) and 10 English (L1)–Spanish(L2) bilinguals residing in Austin, Texas. He used a cued picture naming task in English and Spanish monolingual contexts (95% of the block was in one language and 5% was in the other language) and a bilingual language context (50% of trials were in English and 50% in Spanish), contrasting stay and switch trials. The results showed a decrease in sound contrast in switch contexts in comparison with stay contexts, with the reduction in contrast driven by the dominant language (English for English-dominant bilinguals and Spanish for Spanish-dominant bilinguals). This effect is larger with switch tokens in the monolingual contexts than in the bilingual context, which may reflect greater inhibition of the non-target language in the monolingual context.

Tsui, Tong, and Chan (2019) tested Cantonese(L1)–English(L2) bilinguals (of varying degrees of dominance) residing in Hong Kong using a cued picture naming task. They contrasted the production of voiceless stops in stay and switch trials. The results showed a decrease in sound contrast in switch contexts in comparison with stay contexts, with the reduction in contrast driven by the dominant language for unbalanced bilinguals (English for English-dominant bilinguals and Cantonese for Cantonese-dominant bilinguals). However, balanced bilinguals showed no variation in sound contrasts in switch contexts when compared to stay contexts.

These studies suggest that language contexts can influence the realization of phonetic contrasts. In each case, such effects are limited to one language, although the particular language that is impacted (dominant vs. non-dominant) varies, and some participants appear to show no effects (balanced bilinguals in Tsui et al., 2019). These mixed results may stem from several limitations of this work. These phonetic studies have not distinguished between single, stay and switch contexts, making it unclear to what extent they are measuring switching and/or mixing effects. Most critically, previous work has not studied bilingual lexical access and phonetic processing together. Previous phonetic studies have not collected data on reaction times – making it unclear if the same participants, tested on these materials, would show effects in retrieval that parallel what's been shown in phonetic measures. To be clear, this same limitation impacts studies focused on retrieval – none have collected phonetic measurements. It is therefore important for research to encompass methods used in both lexical access and phonetic processing, allowing us to strongly test theoretical approaches that aim to join these two domains together.

The current study

Theories integrating retrieval and phonetic processing predict that increased difficulty in language selection will result in both slower reaction times and more accented productions – i.e., more English-like phonetic properties in Spanish targets and/or more Spanish-like properties in English targets. However, no study has tested this prediction. We aimed to address this empirical gap. Spanish–English bilinguals named pictures in a cued language-switching task, allowing us to examine performance in single language and mixed contexts, as well as stay vs. switch trials. For each trial, reaction times (indexing retrieval) and phonetic measures were analyzed. Extending previous studies of cued language-switching, we analyzed both initial consonant VOT and properties of the following vowel. This provided a fuller picture of phonetic processing, examining two distinct types of speech sound properties on two different segments. Importantly, sample sizes were set by a pre-registered power analysis based on a previous phonetic study (see "Supplemental Materials" for details). This required a large sample size; to address this, we utilized automated methods for analysis of phonetic measures, a key technical advance over previous work.

Based on the research reviewed above, we expected to observe slower reaction times in mixed vs. single language blocks (a mixing cost), and a smaller increase in reaction times on switch vs. stay trials within mixed language blocks (a switch cost). Reversed dominance effects may be observed, particularly in mixing costs. Theories integrating retrieval and phonetic processing predict parallel costs in phonetics: a decrease in the contrast between languages in mixed vs. single language blocks (a phonetic mixing costs), along with a smaller decrease in the contrast between languages on switch vs. stay trials (a phonetic switch cost). If reversed dominance effects are observed in reaction times, these theories predict stronger phonetic costs for the dominant vs. non-dominant language. Finally, because such theories claim a common source for effects in retrieval and phonetics, the theories predict that such relationships will hold at the level of individual trials; when retrieval is difficult (i.e., trials with longer RTs), there will be corresponding difficulties in phonetic processing (i.e., a decreased contrast between languages).

To preview the results, we find parallel as well as divergent effects across these measures. The mixture of results challenges theories that assume very strong integration of lexical and phonetic processing, suggesting that these two processes must be separated to some degree.

Methods

This study was pre-registered (Mertzen, Lago, & Vasishth, 2021) with the Open Science Foundation (https://osf.io/f8gd4). This means that the sample size and analyses were planned out before collecting data and uploaded to the OSF website. Note that due to issues in controlling for frequency and length effects, we deviated from our pre-registered analysis plan by omitting cognate status (see below for details). However, results for mixing and switching costs were qualitatively similar when cognate status was included in the model. As noted below, we also deviated from the pre-registration in the phonetic analysis of vowels.

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

Participants

A Monte Carlo power analysis, based on the results of Goldrick et al. (2014), suggested that, for 48 target items, 18 participants were required to reach power exceeding 0.8 (see Supplemental Materials for details). Nineteen English-dominant Spanish–English bilinguals from the metro Chicago area participated in the study for financial compensation. One participant was excluded from the analysis due to an equipment error. All participants acquired Spanish from birth and learned English in childhood (mean age of onset of learning = 3.2 years old, range 0-9; see Appendix A for more self-report data on language background). Participants were informally screened for ability to produce the distinctions between the English target vowels (/i/ - /ɪ/ and /e/ - /ɛ/; see below) by the experimenter, a native Spanish–English bilingual. Dominance was measured by language proficiency, assessed by the Multilingual Naming Test (MINT; Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012). Based on the size of their productive vocabularies, participants were English dominant[1]; each participant had a higher MINT score in English (mean, 64.9, range, 58-68) in comparison to Spanish (mean, 54.8, range, 41-65). This suggests that, for these participants, representations of English words have greater lexical robustness (e.g., Costa, Santesteban, & Ivanova, 2006; Schwieter & Sunderman, 2008) than representations of Spanish words.

Materials

Appendix B provides target item details. Sixteen non-cognate words in both English and Spanish[2] were selected along with eight cognates, yielding a total of 48 target lexemes. All words began with /p, b, t/ or /d/, allowing us to examine how VOT was impacted by language context, and were followed by /i/ or /e/ in Spanish and /i, ɪ, e/ and /ɛ/ in English, allowing us to examine language context impacted vowel contrasts. While English distinguishes lax (/ɪ, ɛ/ as in "bit" and "bet", respectively) from tense vowels (/i, e/, as in "beet" and "bait"), Spanish only uses tense vowels (e.g., "piso" "beso"; Bradlow, 1995). Sound combinations were as equally distributed as possible, although given the constraints on the lexicons of Spanish and English, it was not possible to have an equal distribution. After controlling for the phonetic environment of the initial stops and picturability of the items, we were unable to match cognates and noncognates for frequency and length (within each language, cognates were more frequent and longer than noncognates; these differences were not matched across languages[3]). Given this lack of control, we omitted this factor from our analyses. Thirty-two filler lexemes were selected, including 8 non-cognates and 4 cognates in each language, as well as the translation equivalents of the non-cognate targets.

A colored picture that depicts each word was taken from the Bank of Standardized Stimuli (Brodeur, Guérard, & Bouras, 2014) or Google Images (images.google.com). Target pictures were normed by Spanish speakers from Mexico and native English speakers from the U.S. (see Supplemental Materials for more details).

Procedure

Participants completed a picture naming task. Target labels were introduced in two familiarization blocks. First, each picture in the experiment was shown in random order with its English and Spanish label for four seconds. Over each label there was the flag that corresponded with the cued language. Following Kleinman and Gollan (2018), an American flag was used to cue English, and a Mexican flag was used to cue Spanish. Flags were used instead of color to minimize the difficulty of learning the association between cue and language. In the second familiarization block, participants named each picture in both English and Spanish, and were cued by flags on which language to name the picture. After naming a picture, participants were always given orthographic feedback showing the picture's target label. This was done to help ensure that participants remembered the target label for each of the pictures.

After the familiarization blocks, participants completed the three experimental blocks. Participants were asked to name each picture as quickly as possible while using the labels from the familiarization task. Each trial consisted of the presentation of a fixation cross (350ms), a blank screen (150ms), the flag cue by itself (250ms), the flag cue with the target picture (maximum of 3000ms), and an intertrial blank screen interval (850ms). The picture disappeared and the experiment moved on to the next trial once a response was detected. The experiment was implemented in Max/MSP (Cycling'74, 2016), and responses were recorded using Audacity (Audacity Team, 2018).

Participants completed two single language blocks (one English only and one Spanish only) and a mixed language block. The ordering of single language blocks was counterbalanced across participants. The mixed language block was always the last block participants completed. Breaks were offered between blocks and during the last and longest block. In single language blocks, the entire set of pictures was repeated four times. There were 24 target words (total of 96 critical trials) and 28 filler words (total of 112 filler trials). In the mixed language block, pictures were named eight times, half in stay and half in switch trials, distributed throughout the block. There were 48 target words (total of 384 critical trials) and 56 filler (total of 448 filler trials). This yielded a total of 576 critical tokens per participant. There were two fixed lists for all three blocks that were pseudo randomized such that all trial types were evenly distributed throughout the block. A filler trial always happened before a target trial, and there were never two target trials in a row. Pictures were not repeated on adjacent trials.

Phonetic analysis

A machine learning algorithm (Goldrick et al., 2021; Shrem, Goldrick, & Keshet, 2019) was used to detect the onset and offset of VOT. Reaction time was set to the difference between picture onset and the onset of VOT. Vowels were segmented from the

[1]While no participant is dominant in Spanish, a small minority (four participants) had a 1 point difference between English and Spanish. To verify that this small group did not bias the results, the voiced stop model was re-run without these 4 participants. Critically, we observed the same dominance effects, making us confident that the results reflect English dominance.

[2]Two of the Spanish target words (i.e., dije and betún) were specific to Mexican Spanish. This is because the phonological constraint on the structure of the first syllable were prioritized and the first author, who chose the stimuli, is a speaker of Mexican Spanish. With the wide range of varieties of Spanish spoken by the participants, this could have caused production difficulties. To minimize this issue, speakers were pretrained on the target labels (see Procedure section). In addition, many exploratory analyses were conducted on the errors produced; these failed to show reliable effects of dialect-specific words and/or dialect of Spanish spoken.

[3]English non-cognate targets were an average of four phonemes long and had a frequency of 39; cognate were six phonemes long and had a frequency of 84. Spanish noncognates were an average of six phonemes long and had a frequency of 34; cognate targets were an average of seven phonemes long and had a frequency of 74.

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

speech using the Montreal Forced Aligner (McAuliffe, Socolof, Mihuc, Wagner, & Sonderegger, 2017). Praat (Boersma & Weenink, 2018) was used to analyze the phonetic properties that signal acoustic vowel contrasts: the first (F1) and second (F2) vowel formants (resonant frequencies of the vocal tract; Peterson & Barney, 1952). Following standard analysis methods (e.g., Mack, 1989), we focus on measurements of formants at vowel midpoint. (See Supplementary Materials for additional analyses of other phonetic properties of vowels; no interactions with language context were found in these analyses). F1 is used to distinguish vowel HEIGHT (e.g., the contrast between the vowels in beet vs. bat), while F2 is used to distinguish vowel backness (e.g., beet vs boot; Peterson & Barney, 1952). The contrasts between the target vowels will be explained in more detail in the "Results" section.

Exclusion criteria

Errors were identified by the experimenter while the study was conducted, and then hand-checked using the recorded speech. A total of 554 production errors were identified (5%, $N = 10,368$). All of these errors were excluded from analyses of RT and acoustic measures.

Trials were excluded if the automatically measured VOT was likely an error (voiceless targets: VOTs ≤ 5 msec or > 120 msec; 8.6%, $N = 4,967$; voiced targets: VOTs ≤ -200 msec, > 50 msec, $|\text{VOT}| < 5$ msec; 9.4%, $N = 4,886$). For vowels, any tokens for which either F1 or F2 were 3 standard deviations away from the participant's mean within each language were excluded (5.1%, $N = 9,843$). Finally, for reaction time (RT) analyses trials with RTs < 250 msec or > 3000 msec were excluded (total RTs excluded: 10.8%, $N = 9,853$).

Results

Data for study can be downloaded from the Open Science Foundation at: https://osf.io/a52yg/

Reaction time (RT)

R (R Core Team, 2013) was used to run a linear mixed-effects model that examined log-transformed RT depending on contrast-coded language context (single as -0.5 versus mix trials (stay, switch) as 0.25 and stay as -0.5 versus switch as 0.5) and contrast-coded language (English as -0.5 versus Spanish as 0.5) along with the interaction of these factors. This same fixed effects structure was used across all analyses reported below. Log-transformed RT was used because the residuals of the raw reaction time model were less well approximated by a normal distribution. For this model and every other model reported in the paper, random effects were fitted using an iterative procedure. We attempted to fit the maximal random effects structure, eliminating correlations and terms in order of complexity until the model converged and did not have a singular fit. To guard against overfitting, this structure was critically examined following the procedure described by Bates, Kliegl, Vasishth, and Baayen (2015). For the RT analysis, there were two sets of correlated random effects factors: (1) by participant, with a random intercept and random slopes for single vs mixed condition; (2) by item, with a random intercept, and a random slope for condition. Significance of fixed effects was assessed by Satterthwaite-corrected t-tests (*lmerTest v3.1*; Kuznetsova, Brockhoff, & Christensen, 2017).

As shown in Figure 1, RTs replicated typical results reported in the literature. Bilinguals showed a mixing cost, with shorter

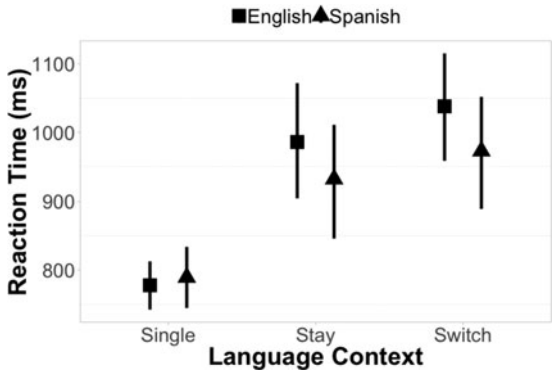


Figure 1. By participant mean raw reaction time by English and Spanish (wings show bootstrapped 95% confidence intervals

reaction times in single versus mixed language contexts ($\beta = 0.27$, *SE* $\beta = 0.03$, $t = 8.67$, $p < 0.001$). While there was no main effect of language ($\beta = -0.037$, *SE* $\beta = 0.02$, $t = -1.5$, $p > 0.05$), the mixing cost was not uniform across target language (significant language by context interaction; $\beta = -0.098$, *SE* $\beta = 0.02$, $t = -4.03$, $p < 0.001$). Follow up regressions within each language showed the mixing cost was larger in English ($\beta = 0.32$, *SE* $\beta = 0.04$, $t = 9.3$, $p < 0.001$) than Spanish ($\beta = 0.22$, *SE* $\beta = 0.04$, $t = 6.1$, $p < 0.001$). Another set of follow up regressions within condition indicated that there is non-significant advantage of English over Spanish in single language contexts ($\beta = 0.01$, *SE* $\beta = 0.03$, $t = 0.43$, $p > 0.05$), which shows a trend towards a reversal in mixed language contexts ($\beta = -0.06$, *SE* $\beta = 0.03$, $t = -1.97$, $p > 0.05$).

Participants also showed a robust switching cost, producing words more quickly in stay vs. switch contexts ($\beta = 0.05$, *SE* $\beta = 0.008$, $t = 6.47$, $p < 0.001$). There was no significant interaction of switching and language ($\beta = 0.009$, *SE* $\beta = 0.02$, $t = 0.59$, $p > 0.05$).

Error rates

In order to confirm that these reaction time effects did not reflect a speed-accuracy tradeoff, a logistic mixed-effects model was run to examine the predictability probability that a trial is correct (1) vs incorrect (0). Our fitting process yielded two sets of correlated random effects factors: (1) by participants with a random intercept and single versus mix and language as random slopes and (2) by word with a random intercept and single versus mix as a random slope. For all subsequent logistic regressions, significance of fixed effects was assessed by the likelihood ratio test (Barr, Levy, Scheepers, & Tily, 2013).

Bilinguals showed a mixing cost since they produced more correct productions in single language contexts (mean: 96%) than in mixed language contexts (mean: 95%; $\beta = -0.87$, *SE* $\beta = 0.37$, $\chi^2(1) = 4.97$, $p < 0.05$). They also showed a switching cost, with more correct productions in stay (mean: 95%) vs. switch contexts (mean: 94%; $\beta = -0.42$, *SE* $\beta = 0.11$, $\chi^2(1) = 13.68$, $p < 0.01$). These results show that reaction time effects did not reflect a speed-accuracy tradeoff. There were no other main effects (χ^2s $(1) < 1.29$, ps > 0.05), nor any interactions (χ^2s $(1) < 0.54$, ps > 0.05).

Voiceless stops

A linear mixed-effects model examined log-transformed voice onset time (VOT) with the same set of fixed effects predictors as above (residuals of the raw VOT model were less well approximated by a normal distribution). Our fitting procedure yielded

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

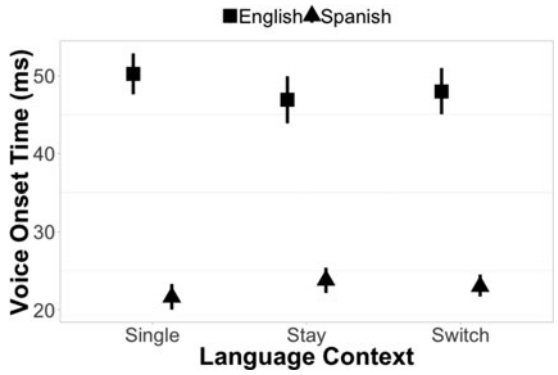


Figure 2. By participant mean voiceless stop VOT (ms) by English and Spanish (wings show bootstrapped 95% confidence intervals)

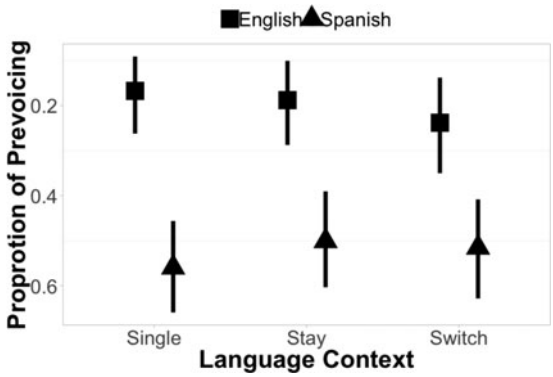


Figure 3. By participant proportion of producing pre-voiced VOT by English and Spanish (wings show bootstrapped 95% confidence interval).

two sets of correlated random effects factors: (1) by participants with a random intercept and single versus mix and language as random intercepts and (2) by word with no slopes.

As shown in Figure 2, bilinguals successfully switched between their two languages when producing voiceless stops, with English target words produced with longer VOT than Spanish target words ($\beta = -0.79$, *SE* $\beta = 0.09$, $t = -8.6$, $p < 0.001$). Although there was no main effect of single versus mixed language context ($\beta = -0.003$, *SE* $\beta = 0.03$, $t = -0.11$, $p > 0.05$), there was a significant interaction between single versus mixed language context and language ($\beta = 0.2178$, *SE* $\beta = 0.03$, $t = 6.1$, $p < 0.001$), suggesting that there was a mixing cost with regards to the production of voiceless stops' VOT[4]. Follow up regressions conducted on English and Spanish subsets of the data indicated that this interaction reflected the opposing effects of stay versus mixed contexts in each language, as seen in Figure 2. In English, VOTs shortened when mixing (i.e., became more Spanish-like; $\beta = -0.009$, *SE* $\beta = 0.03$, $t = -3.22$, $p < 0.01$). In contrast, Spanish VOTs were lengthened when mixing (i.e., became more English-like; $\beta = 0.09$, *SE* $\beta = 0.04$, $t = 2.2$ $p < 0.01$). Note that these mixing costs were symmetric (i.e., there was no significant difference in the mixing effect across languages), whereas there were larger mixing costs for English than for Spanish in RT. Another set of follow up regressions conducted on single and mixed language contexts subsets of the data indicated that while bilinguals successfully switched between their two languages in both single and mixed contexts, the contrast between languages was larger in single ($\beta = -0.88$, *SE* $\beta = 0.09$, $t = -9.37$, $p < 0.001$) versus mixed language contexts ($\beta = -0.06$, *SE* $\beta = 0.03$, $t = -1.99$, $p > 0.05$).

While there was a main effect of switch versus stay contexts ($\beta = 0.03$, *SE* $\beta = 0.01$, $t = -2.21$, $p < 0.05$), there was no significant switch cost; the interaction between stay versus switch language contexts and language was not significant ($\beta = -0.048$, *SE* $\beta = 0.026$, $t = -1.88$, $p > 0.05$).

Voiced stops

A logistic mixed-effects model examined the odds of producing pre-voiced VOT. There were two sets of uncorrelated random effects factors: (1) by participants with a random intercept and single versus mix and language as random slopes and (2) by word with a random intercept and single versus mix as a random slope.

Similar to voiceless stops, bilinguals successfully switched between their two languages when producing voiced stops. As seen in Figure 3, English target words were produced with more English-appropriate short-lag VOT than Spanish ($\beta = 2.1$, *SE* $\beta = 0.34$, χ^2 $(1) = 25.96$, $p < 0.001$). Although there was no main effect of single versus mixed language context ($\beta = 0.06$, *SE* $\beta = 0.18$, χ^2 $(1) = 0.12$, $p > 0.05$), there was a significant interaction between single versus mixed contexts and language ($\beta = -0.83$, *SE* $\beta = 0.27$, χ^2 $(1) = 8.29$, $p < 0.01$) suggesting a mixing cost for voiced stops. Follow up regressions conducted on English and Spanish subsets of the data indicated that bilinguals were more likely to produce pre-voiced VOT for English target words in mixed contexts than in single contexts ($\beta = 0.52$, *SE* $\beta = 0.25$, χ^2 $(1) = 3.78$, $p < 0.05$). In contrast, bilinguals showed no significant difference in pre-voicing for Spanish target words in single versus mixed contexts ($\beta = -0.36$, *SE* $\beta = 0.27$, χ^2 $(1) = 1.67$, $p > 0.05$). Additional regressions conducted on single and mixed context subsets of the data indicated that bilinguals successfully switched between their two languages in single and mixed contexts, with a larger distinction in single ($\beta = 2.55$, *SE* $\beta = 0.45$, χ^2 $(1) = 25.5$, $p < 0.001$) than in mixed ($\beta = 1.87$, *SE* $\beta = 0.3$, χ^2 $(1) = 23.47$, $p < 0.001$) contexts. Overall, these results suggest a mixing cost for English but not Spanish productions, parallel to the dominance reversal found in RT mixing costs.

There was no main effect of switching ($\beta = 0.19$, *SE* $\beta = 0.09$, χ^2 $(1) = 3.36$, $p > 0.05$), or a significant interaction between stay versus switch language contexts and language ($\beta = -0.12$, *SE* $\beta = 0.19$, χ^2 $(1) = 0.38$, $p > 0.05$).

High vowels (/i,ɪ/)

Separate linear mixed-effects models were constructed to examine the F1 and F2 of high vowels at 50 percent vowel duration as dependent variables[5]. The fixed effects structure of previous models was extended to include vowel type (English /i/ versus Spanish /i/ and English /ɪ/ versus Spanish /i/; each factor was treatment-coded with Spanish /i/ as the reference-level). For the F1 model,

[4]Casillas (2021) discusses how mixtures of categorically different productions can yield apparently gradient effects (e.g., the mean of productions from a mixture of two discretely categories, short- vs. long-lag VOT, can yield a value intermediate between the means of the two categories). Inspection of the distributions of English and Spanish positive-lag VOTs reveals no evidence of bimodality; there is a graded shift in the center of the VOT distributions.

[5]The mid and high vowel analyses differ from the pre-registered analysis, which was not sufficient for assessing our research questions. Since it did not compare English and Spanish vowel categories in the same model, it did not provide information about changes in the contrast between the two languages of bilinguals.

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

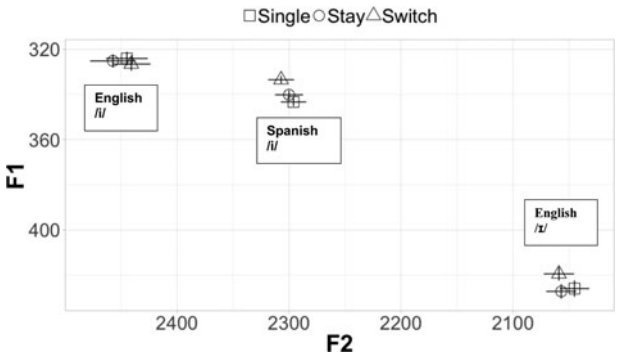


Figure 4. Mean formant values for high English and Spanish vowels at 50 percent vowel duration (vertical wings show standard error for F1 and horizontal wings show standard error for F2).

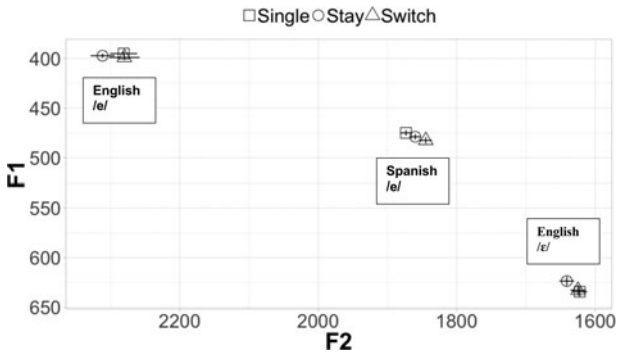


Figure 5. Mean formant values for mid English and Spanish vowels at 50 percent vowel duration (vertical wings show standard error for F1 and horizontal wings show standard error for F2).

our fitting procedure yielded two random effects: (1) by participants with a random intercept and (2) by word with a random intercept. For the F2 at 50 percent duration model, there were two random effects: (1) by participants with a random intercept and (2) by word with a random intercept.

For ease of discussion, we summarize the two analyses simultaneously, interpreting variation in phonetic properties in terms of position in the traditional vowel space (a graphical visualization of vowel contrasts with F1 as the Y axis and F2 as the X axis). We describe F1 as indicating whether a vowel is "raised" versus "lowered" (i.e., smaller versus larger F1 values) and F2 as indicating a change in whether the vowel is more "back" versus "front" (i.e., smaller versus larger F2 values). As a reminder, in single language contexts, English /i/ is expected to be slightly higher and fronter than Spanish /i/, and English /ɪ/ should be lower and slightly more back than both English /i/ and Spanish /i/.

As seen in Figure 4, in single language blocks, bilinguals produced English /i/ more front ($\beta = 146.98$, *SE* $\beta = 54.41$, $t = 2.7$, $p < 0.05$) and higher than Spanish /i/ ($\beta = -28.16$, *SE* $\beta = 12.54$, t= -2.25, $p < 0.05$). In mixing contexts, there was significant raising ($\beta = -6.85$, *SE* $\beta = 2.17$, t= -3.16, $p < 0.01$), decreasing the contrast between Spanish /i/ and English /i/ ($\beta = 7.91$, *SE* $\beta = 3.98$, $t = 1.99$, $p < 0.05$). While there was also significant fronting of Spanish /i/ when mixing ($\beta = 22.06$, *SE* $\beta = 7.81$, $t = 2.83$, $p < 0.01$), there was no significant change in the contrast with English /i/ ($\beta = -20.32$, *SE* $\beta = 14.33$, t= -1.41, $p > 0.05$). Switching

Table 1: Results for F1 (height) linear mixed effects model for high vowels

Fixed effects	β	*SE* β	t	p
single vs mixed context	−6.85	2.17	−3.16	<0.01**
stay vs switch context	−2.2	1.98	−1.11	0.27
Spanish /i/ vs English /i/	−28.16	12.54	−2.25	<0.05*
Spanish /i/ vs English /ɪ/	47.54	2.58	18.4	<0.001***
single vs mixed context X Spanish /i/ vs English /i/	7.91	3.98	1.99	<0.05*
single vs mixed context X Spanish /i/ vs English /ɪ/	4.11	3.55	1.16	0.25
stay vs switch context X Spanish /i/ vs English /i/	5.37	3.65	1.5	0.14
stay vs switch context X Spanish /i/ vs English /ɪ/	6.16	3.23	1.91	0.057

Table 2: *Results for F2 (front/back) linear mixed effects model for high vowels*

Fixed effects	β	*SE* β	t	p
single vs mixed context	22.06	7.81	2.83	<0.05*
stay vs switch context	−2.19	7.1340	−0.31	0.76
Spanish /i/ vs English /i/	146.98	54.41	2.7	<0.05*
Spanish /i/ vs English /ɪ/	−286.35	47.03	−6.08	<0.001***
single vs mixed context X Spanish /i/ vs English /i/	−20.32	14.33	−1.42	0.16
single vs mixed context X Spanish /i/ vs English /ɪ/	−4.19	12.8	−0.33	0.74
stay vs switch context X Spanish /i/ vs English /i/	0.41	13.15	0.03	0.98
stay vs switch context X Spanish /i/ vs English /ɪ/	4.03	11.64	0.35	0.73

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

Table 3: *Results for F1 linear mixed effects model for mid vowels*

Fixed effects	β	*SE* β	t	p
single vs mixed context	9.37	4.57	2.05	<0.05*
stay vs switch context	−2.16	2.59	−0.84	0.4
Spanish /e/ vs English /e/	−79.79	16.92	−4.72	<0.001***
Spanish /e/ vs English /ɛ/	150.79	15.12	9.97	<0.001***
single vs mixed context X Spanish /e/ vs English /e/	−7.29	5.33	−1.37	0.17
single vs mixed context X Spanish /e/ vs English /ɛ/	−18.14	4.75	−3.82	<0.001***
stay vs switch context X Spanish /e/ vs English /e/	−0.15	4.85	−0.03	0.97
stay vs switch context X Spanish /e/ vs English /ɛ/	−0.38	4.33	−0.09	0.93

Table 4: *Results for F2 linear mixed effects model for mid vowels*

Fixed effects	β	*SE* β	t	p
single vs mixed context	−16.49	7.53	−2.19	<0.05*
stay vs switch context	12.02	6.76	1.78	0.08
Spanish /e/ vs English /e/	435.2	43.41	10.03	<0.001***
Spanish /e/ vs English /ɛ/	−233.33	38.78	−6.02	<0.001***
single vs mixed context X Spanish /e/ vs English /e/	25.06	13.9	1.8	0.07
single vs mixed context X Spanish /e/ vs English /ɛ/	31.93	12.39	2.58	<0.05*
stay vs switch context X Spanish /e/ vs English /e/	−2.44	12.66	−0.19	0.85
stay vs switch context X Spanish /e/ vs English /ɛ/	−7.83	11.3	−0.69	0.49

did not significantly impact the height contrast (see Table 1) or induce fronting (see Table 2).

English /ɪ/ was realized backer ($\beta = -286.35$, *SE* $\beta = 47.03$, $t = -6.09$, $p < 0.001$) and lower ($\beta = 47.54$, *SE* $\beta = 2.58$, $t = 18.4$, $p < 0.001$) than Spanish /i/. Neither mixing nor switching significantly impacted the height contrast (see Table 1) or induce fronting (see Table 2).

Mid vowels (/e,ɛ/)

The analysis of mid vowels followed that of high vowels, contrasting vowel type with Spanish /e/ as the reference level (English /e/ versus Spanish /e/ and English /ɛ/ versus Spanish /e/). For the F1 model, there were two sets of correlated random effects factors: (1) by participants with a random intercept and single versus mix as a random slope and (2) by word with a random intercept. For the F2 model, there were two sets of uncorrelated random effects factors: (1) by participants with a random intercept and (2) by word with a random intercept.

As a reminder, in single language contexts, English /e/ is expected to be higher and slightly fronter than Spanish /e/, and English /ɛ/ should be lower and slightly more back than both English /e/ and Spanish /e/.

As seen in Figure 5, in single language blocks, bilinguals made a significant distinction between Spanish /e/ and English /e/, with English /e/ produced more raised ($\beta = -79.79$, *SE* $\beta = 16.92$, $t = -4.72$, $p < 0.001$) and fronted ($\beta = 435.2$, *SE* $\beta = 43.41$, $t = 10.03$, $p < 0.001$). They also distinguished Spanish /e/ and English /ɛ/, producing the English vowel lower ($\beta = 150.79$, *SE* $\beta = 15.12$, $t = 9.97$, $p < 0.001$) and backer ($\beta = -233.33$, *SE* $\beta = 38.78$, $t = -6.02$, $p < 0.001$). There was significant lowering for Spanish /e/ ($\beta = 9.37$, *SE* $\beta = 4.57$, $t = 2.05$, $p < 0.05$) and raising for English /ɛ/ in mixing contexts, reducing the contrast between the vowels ($\beta = -18.14$, *SE* $\beta = 4.75$, $t = -3.82$, $p < 0.001$). There was significant backing of Spanish /e/ when mixing ($\beta = -16.49$, *SE* $\beta = 7.53$, $t = -2.19$, $p < 0.05$), significantly decreasing the contrast between the two vowels on this dimension as well ($\beta = 31.93$, *SE* $\beta = 13.9$, $t = 2.58$, $p < 0.05$). However, the reduction of contrast between Spanish /e/ and English /ɛ/ did not result in a larger contrast between Spanish /e/ and English /e/ in height ($\beta = -7.2$, *SE* $\beta = 5.33$, $t = -1.37$, $p > 0.05$) and backness ($\beta = 25.06$, *SE* $\beta = 13.9$, $t = 1.8$, $p > 0.05$) in mixing contexts. There were no significant effects of switching on raising (see Table 3) or backness (see Table 4).

Relationship between reaction time and phonetic measures

A key advantage of our study, relative to previous work, is that we can directly assess whether difficulties in retrieval were related to difficulties in phonetic processing by examining the by-trial relationships between the two measures. We did this via a series of follow up regression analyses. The final model for each phonetic measure was extended by including RT (centered) and its interactions with other predictors. Model tables are included in Supplementary Materials.

Across the majority of measures, the results showed that the phonetic contrast between English and Spanish was reduced on trials with longer RTs (i.e., significant interactions of RT and language; $ts > -8.09$, $\chi^2(1) > 4.38$, $ps < .05$). This suggests that, controlling for the effect of context, difficulty in retrieval disrupts phonetic processing.

RT modulated a mixing effect for height (i.e., F1) in mid vowels and a switching effect for voiced stops. For mid vowel

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

height, there were two 3-way interactions: one of RT, single vs. mix, and Spanish /e/ and English /e/ ($\beta = -35.33$, *SE* $\beta = 18$, $t = -1.96$, $p < 0.05$) and another of RT, single vs. mix, and Spanish /e/ and English /ɛ/ ($\beta = -65.16$, *SE* $\beta = 17.45$, $t = -3.74$, $p < 0.001$). These interactions reflect a decrease in contrast between Spanish /e/ and English /ɛ/ in mixed contexts (single context mean: 163 hertz; mixed context mean: 141 hertz), with the effect being magnified in trials where RTs are longer (longer RT trials single context mean: 171 hertz; longer RT trials mix context mean: 143 hertz; shorter RT trials single context mean: 156 hertz; shorter RT trials mixed context mean: 140 hertz). The lowering of Spanish /e/ impacts the contrast between it and English /e/, increasing the contrast between Spanish /e/ and English /e/ in mixing contexts (single context mean: -70.9 hertz; mixed context mean: −81.9 hertz), with the effect being magnified in trials where RTs are longer (longer RT trials single context mean: −60.1 hertz; longer RT trials mix context mean: −77.3 hertz; shorter RT trials single context mean: −81.8 hertz; shorter RT trials mixed context mean: -86.6 hertz).

For voiced stops, there was a 3-way interaction of RT, switch vs. stay, and language ($\beta = 1.44$, *SE* $\beta = 0.56$, χ^2 (1) = 6.52, $p < 0.01$). This reflects a floor effect for longer RTs. For trials with quicker-than-average RTs, the mean difference between proportion of prevoicing for Spanish vs. English was larger for stay versus switch trials (33.8% versus 27.2%), reflecting a switch cost. However, for trials with slower-than-average RTs, the reduction in the contrast between languages was already so reduced that switching had no additional effect (stay trial mean: 27.1%; switch mean: 27.7%).

However, across the remaining measures, RT did not significantly modulate mixing or switching costs (i.e., the RT by context by language interactions were not significant; $ts > -1.65$, χ^2 (1) > 0.61, $ps > .05$). This suggests that the mixing effect observed in the analyses in prior sections reflects an independent impact on phonetic processing, over and above disruptions to lexical processing.

Discussion

Several theories of speech production integrate the processes of retrieving words from memory with processes specifying the phonetic detail of words. These theories predict that language context (language mixing and switching) should simultaneously influence reaction times and the phonetic properties of speech. However, previous work has not documented these effects simultaneously, within the same participants and trials. To test this claim, we simultaneously measured, on the same trials, how mixing and switching impacted Spanish–English bilinguals' reaction times (RTs) as well as phonetic measures of consonants (VOT) and vowels (F1/F2 at midpoint). Our key findings are summarized in Table 5. We found robust mixing and switching effects for RTs, but only found consistent mixing effects on our phonetic variables – a divergence that is not predicted by current accounts. However, over and above these condition-level effects, analysis of the trial-by-trial relationship between RT and phonetic measures showed that overall retrieval difficulty leads to a reduction in the phonetic contrast between languages – consistent with some degree of integration between lexical retrieval and phonetic processing. This mixed pattern of results suggests that lexical retrieval and phonetic processing are neither tightly linked (as claimed by current proposals) nor completely independent.

Given that we replicated 'standard' effects in lexical retrieval, we are confident that the divergences between RT and phonetic effects are not due to unusual properties of the stimuli or task. We found mixing and switching costs (e.g., Gollan & Ferreira, 2009; Hernandez & Kohnert, 1999; Kleinman & Gollan, 2018; Prior & Gollan, 2013) as well as reversed dominance effects (e.g., Declerck et al., 2020; Gollan & Ferreira, 2009; Kleinman & Gollan, 2018). In contrast, the phonetic measures only showed consistent mixing effects. Furthermore, as summarized in Table 5, the directionality of the effect varied. The failure to find significant switching effects in phonetic measures is inconsistent with some previous work using cued switching (e.g., Goldrick et al., 2014); however, other studies examining reading aloud of sentences with code switches have failed to find significant effects (e.g., Grosjean & Miller, 1994; Šimáčková & Podlipský, 2018). As noted in the introduction, the difference across previous phonetic studies may simply reflect their small sample size. In contrast, our study is high-powered for a phonetic study, with the sample size determined by a power analysis based

Table 5: *Summary of study results*

Measure	Mixing cost	Switching cost	Reduced distinction between languages for trials with longer RT
Lexical retrieval (RT)	✓ Dominant > Nondominant	✓ Dominant ≈ Nondominant	*n/a*
Voiceless VOT	✓ Dominant ≈ Nondominant	×	✓
Voiced VOT	✓ Dominant > Nondominant	×	✓
High Vowels	✓ Nondominant > Dominant	×	✓ Spanish /i/ vs. English /i/, height Spanish /i/ vs. English /i/, backness Spanish /i/ vs. English /ɪ/, backness × Spanish /i/ vs. English /ɪ/, height
Mid Vowels	✓ Height: Dominant > Nondominant; Backness: Nondominant > Dominant	×	✓ Spanish /e/ vs. English /e/, height × Spanish /e/ vs. English /e/, backness Spanish /e/ vs. English /ɛ/, height Spanish /e/ vs. English /ɛ/, backness

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

on previous work showing significant switching effects (Goldrick et al., 2014).

The divergence of language context effects in lexical retrieval and phonetic production is not predicted by theories claiming that processes retrieving words from memory are strongly integrated with processes specifying the phonetic detail of words. At the same time, the finding that trial-level difficulty modulates the phonetic contrast between languages suggests that there must be some degree of integration between these processes. To account for the full set of data, theories must be revised to allow for weaker coupling of these two aspects of speech production. In the context of cascading activation theories, one such proposal claims that interactions between lexical access and phonetic processing continue after the initiation of the response (Fink, Oppenheim, & Goldrick, 2018; Goldrick, McClain, Cibelli, Adi, Gustafson, Moers, & Keshet, 2019). According to this account, participants are able to resolve processing conflicts in lexical access while phonetic processing is ongoing. Relatively small disruptions to lexical access (e.g., the increased time required for switch versus stay trials in our experiment) may be fully resolved before phonetic processing is complete, reducing their impact on phonetic measures. In contrast, larger disruptions (e.g., the increased time required for stay versus single trials) will be more difficult to resolve before articulation begins, yielding interactive effects. Dynamically varying interaction may also be a promising area of development for exemplar models (e.g., Clopper & Pierrehumbert, 2008; see Fink & Goldrick, 2015, for review and discussion). Longer delays in word retrieval (e.g., mixing effects) might allow a greater number of exemplars from the non-target language to become activated, further reducing the phonetic contrast between languages (relative to contexts with less delays, as induced by switching). Accounting for the full pattern of results observed here will require greater elaboration and specification of these proposals.

A limitation of this work, and previous experimental phonetic studies, is that most interactions have been examined with a small number of phonetic parameters. It's possible that intrinsic properties of these phonetic measures could be skewing results. Most phonetic studies use VOT, specifically focusing on the contrast between short-lag and long-lag positive VOTs (as in the contrast between Spanish versus English voiceless stops in the current study). The greater range of variation in long-lag VOTs (Kessinger & Blumstein, 1997) may provide greater power for observing effects. This confound between language and phonetic properties complicates the interpretation of results. For example, effects may have been observed for English stops not because English was the dominant language, but because English stops have longer VOTs than Spanish stops. To address this confound, future work should examine other phonetic measures and language pairings that do not include English. Similarly, there is a possibility that the higher variability in vowel acoustics as compared to VOT is affecting our ability to measure the effects for vowels. Investigating a greater range of speech sounds within and across languages will help clarify whether there are systematic differences in the relationship between phonetic processing and control mechanisms.

Conclusion

Bilingual lexical access and phonetic processing have typically been studied separately. As theoretical approaches work to bridge this divide, it is critical that we extend empirical research to encompass methods used in both domains. Our direct examination of the link between lexical access and phonetic processes revealed consistent, robust effects of mixing and switching in reaction time, but only consistent mixing effects in phonetic measures. The divergent results suggest there are constraints on the degree of interaction between lexical access and phonetic processes. These initial efforts show that richer datasets can provide important constraints on theories of these processes.

Acknowledgments. Thanks to Rosemary Dong for her assistance in data collection and analysis, and to Yosi Shrem and Joseph Keshet for assistance with analysis of voice onset times. Supported in part by NSF grant 2219843.

Competing interests. The author(s) declare none

Supplementary Material. For supplementary material accompanying this paper, visit https://doi.org/10.1017/S1366728922000682

Power Analysis: a description of how the power analysis was done and a table showing results.

Stimulus Norming: a description of how stimulus norming was done, as well as the results of the norming.

Filler Items: a table listing the filler items used in the study.

Results of follow up RT and phonetic measure models: a list of 6 model output tables for the RT and phonetic measure models. The models are: RT and voiceless VOT; RT and voiced VOT; RT and F1 for high vowels; RT and F2 for high vowels; RT and F1 for mid vowels; and RT and F2 for mid vowels.

Degree of diphthongization and monophthongization of /e/: a description of an analysis conducted to see if there was an effect of language context on the degree of diphthongization of /e/ in English and the degree of monophthongization of /e/ in Spanish. A table and figure of the results are included.

References

Amengual M (2018) Asymmetrical interlingual influence in the production of Spanish and English laterals as a result of competing activation in bilingual language processing. *Journal of Phonetics* **69**, 12–28.

Amengual M (2021) The acoustic realization of language-specific phonological categories despite dynamic cross-linguistic influence in bilingual and trilingual speech. *Journal of the Acoustical Society of America* **149**, 1271–1284.

Audacity Team (2018) Audacity [Software]. Version 2.3.1. Available at: https://www.audacitycityteam.org

Barr DJ, Levy R, Scheepers C and Tily HJ (2013) Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language* **68**, 255–278.

Bates B, Kliegl R, Vasishth S and Baayen H (2015) *Parsimonious Mixed Models*. arXiv preprint available at https://arxiv.org/pdf/1506.04967.pdf

Bobb SC and Wodniecka Z (2013) Language switching in picture naming: What asymmetric switch costs (do not) tell us about inhibition in bilingual speech planning. *Journal of Cognitive Psychology* **25**, 568–585.

Boersma P and Weenink D (2018) *Praat: Doing Phonetics by Computer [Computer Program]. Version 6.0.43*. Available at: http://www.praat.org/

Bradlow AR (1995) A comparative acoustic study of English and Spanish vowels. *The Journal of the Acoustical Society of America* **97**, 1916–1924.

Branzi FM, Martin CD, Abutalebi J and Costa A (2014) The after-effects of bilingual language production. *Neuropsychologia* **52**, 102–116.

Brodeur MB, Guérard K and Bouras M (2014) Bank of standardized stimuli (BOSS) phase II: 930 new normative photos. *PLoS One* **9**, e106953.

Casillas JV (2021) Interlingual interactions elicit performance mismatches not "compromise" categories in early bilinguals: Evidence from meta-analysis and coronal stops. *Languages* **6**, 9.

Christoffels IK, Firk C and Schiller NO (2007) Bilingual language control: An event-related brain potential study. *Brain Research* **1147**, 192–208.

Clopper CG and Pierrehumbert JB (2008) Effects of semantic predictability and regional dialect on vowel space reduction. *The Journal of the Acoustical Society of America* **124**, 1682–1688.

Costa A, Santesteban M and Ivanova I (2006) How do highly proficient bilinguals control their lexicalization process? Inhibitory and language-specific selection mechanisms are both functional. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **32**, 1057–1074.

Cycling'74. (2016) *Max/MSP [Software]. Version 7.3.6.* Available at: https://cycling74.com/downloads/older

Declerck M (2020) What about proactive language control? *Psychonomic bulletin & review* **27**, 24–35.

Declerck M, Kleinman D and Gollan TH (2020) Which bilinguals reverse language dominance and why? *Cognition* **204**, 104384.

Fink A and Goldrick M (2015) The influence of word retrieval and planning on phonetic variation: Implications for exemplar models. *Linguistics Vanguard* **1**, 215–225.

Fink A, Oppenheim G and Goldrick M (2018) Interactions between lexical access and articulation. *Language, Cognition, and Neuroscience* **33**, 12–24. PMC5793891.

Goldrick M, McClain R, Cibelli E, Adi Y, Gustafson E, Moers C and Keshet J (2019) The influence of lexical selection disruptions on articulation. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **46**, 1107–1141.

Goldrick M, Runnqvist E and Costa A (2014) Language switching makes pronunciation less nativelike. *Psychological science* **25**, 1031–1036.

Goldrick M, Shrem Y, Kilbourn-Ceron O, Baus C and Keshet J (2021) Using automated acoustic analysis to explore the link between planning and articulation in second-language speech production. *Language, Cognition, and Neuroscience* **36**, 824–839.

Gollan T and Ferreira VS (2009) Should I stay or should I switch? A cost-benefit analysis of voluntary language switching in young and aging bilinguals. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **35**, 640–665.

Gollan T and Goldrick M (2018) A switch is not a switch: Syntactically-driven bilingual language control. *Journal of Experimental Psychology: Learning, Memory, and Cognition* **44**, 143–156.

Gollan T, Weissberger GH, Runnqvist E, Montoya RI and Cera CM (2012) Self-ratings of spoken language dominance: a multi- lingual naming test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism Language and Cognition* **15**, 594–615.

Grosjean F and Miller JL (1994) Going in and out of languages: An example of bilingual flexibility. *Psychological science* **5**, 201–206.

Gustafson E and Goldrick M (2018) The role of linguistic experience in the processing of probabilistic information in production. *Language, Cognition, and Neuroscience* **33**, 211–226.

Hernandez AE and Kohnert KJ (1999) Aging and language switching in bilinguals. *Aging, Neuropsychology, and Cognition* **6**, 69–83.

Jacobs A, Fricke M and Kroll JF (2016) Cross-language activation begins during speech planning and extends into second language speech. *Language learning* **66**, 324–353.

Kessinger RH and Blumstein SE (1997) Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics* **25**, 143–168.

Kleinman D and Gollan TH (2018) Inhibition accumulates over time at multiple processing levels in bilingual language control. *Cognition* **178**, 115–132.

Kuznetsova A, Brockhoff PB and Christensen RHB (2017) lmerTest Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**, 1–26.

Levelt WJM (1989) *Speaking: From intention to articulation.* Boston, MA: MIT Press.

Lisker L and Abramson AS (1964) A cross-language study of voicing in initial stops: Acoustical measurements. *Word* **20**, 384–422.

Mack M (1989) Consonant and vowel perception and production: Early English-French bilinguals and English monolinguals. *Perception & Psychophysics* **46**, 187–200.

McAuliffe M, Socolof M, Mihuc S, Wagner M and Sonderegger M (2017) Montreal Forced Aligner: trainable text-speech alignment using Kaldi. In *Proceedings of the 18th Conference of the International Speech Communication Association.*

Melinger A, Branigan HP and Pickering MJ (2014) Parallel processing in language production. *Language, Cognition and Neuroscience* **29**, 663–683.

Mertzen D, Lago S and Vasishth S (2021) The benefits of preregistration for hypothesis-driven bilingualism research. *Bilingualism: Language and Cognition* **24**, 807–812.

Meuter RFI and Allport A (1999) Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of Memory and Language* **40**, 25–40.

Olson DJ (2013) Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics* **41**, 407–420.

Peterson GE and Barney HL (1952) Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* **24**, 175–184.

Pierrehumbert J (2002) Word-specific phonetics. In C Gussenhoven and N Warner (Eds.), *Laboratory phonology VII* Berlin: Mouton, pp. 101–140.

Prior A and Gollan TH (2013) Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin-English bilinguals. *Journal of the International Neuropsychological Society* **17**, 682–691.

R Core Team (2013) R: a language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing. Available at: http://www.R-project.org/

Schwieter JW and Sunderman G (2008) Language switching in bilingual speech production: In search of the language-specific selection mechanism. *The Mental Lexicon* **3**, 214–238.

Shrem Y, Goldrick M and Keshet Y (2019) Dr. VOT : Measuring positive and negative voice onset time in the wild. In T. Hain & B. Schuller (Eds.) *Proceedings of Interspeech 2019.*

Šimáčková S and Podlipský VJ (2018) Patterns of short-term phonetic interference in bilingual speech. *Languages* **34**.

Tsui RKY, Tong X and Chan CSK (2019) Impact of language dominance on phonetic transfer in Cantonese–English bilingual language switching. *Applied Psycholinguistics* **40**, 29–58.

Weissberger GH, Wierenga CE, Bondi MW and Gollan TH (2012) Partially overlapping mechanisms of language and task control in young and older bilinguals. *Psychology and Aging* **27**, 959–974.

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press

Appendix A: Participant language background

Participant	Age	AOA[a] English	MINT English[b]	MINT Spanish	Spanish dialect
1	20	0	68	47	Mexican
2	18	6	66	61	Mexican
3	21	0	68	55	Ecuadorian
4	20	8	58	51	Mexican
5	25	9	61	60	Peruvian
6	27	4	64	56	Mexican
7	22	4	60	59	Colombian
8	18	8	62	61	Mexican
9	21	0	68	61	Peninsular
10	26	3	68	53	Mexican
11	21	6	64	59	Colombian
12	23	0	67	41	Mexican
13	28	4	66	65	Paraguayan
14	19	3	62	58	Chilean
15	19	0	65	44	Mexican
16	18	0	68	42	Colombian
17	21	3	67	51	Colombian
18	21	0	67	63	Costa Rican

[a]AoA = Age of Acquistion
[b]MINT = Multilingual Naming Test (Gollan et al., 2012). Maximum possible score: 68

Appendix B: Target words

English Only	Spanish Only	Cognate: English	Cognate: Spanish
beaver	pico (*beak*)	bikini	bikini
beach	piso (*floor*)	baby	bebé
peacock	bigote (*mustache*)	dictionary	diccionario
teapot	biblioteca (*library*)	desert	desierto
bib	tiburón (*shark*)	picnic	picnic
pig	tijeras (*scissors*)	telephone	teléfono
dishes	dibujo (*drawing*)	teepee	tipi
ticket	dije (*pendant*)	petals	pétalos
bacon	perro (*dog*)		
pastry	pez (*fish*)		
daisy	betún (*frosting*)		
table	berenjena (*eggplant*)		
bed	teclado (*keyboard*)		
pepper	techo (*roof*)		
desk	dedo (*finger*)		
test	desayuno (*breakfast*)		

Note: English translation for Spanish Only words are in parentheses. English translation for Cognate: Spanish words are in corresponding Cognate: English row.

https://doi.org/10.1017/S1366728922000682 Published online by Cambridge University Press