Journal of Phonetics 52 (2015) 121–137

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Research Article

Voice onset time in Spanish–English spontaneous code-switching

Page Piccinini *, Amalia Arvaniti [1]

University of California, San Diego, Department of Linguistics, 9500 Gilman Drive, La Jolla, CA 92093-0108, USA

ARTICLE INFO

Article history:
Received 23 January 2014
Received in revised form
7 July 2015
Accepted 14 July 2015
Available online 6 August 2015

Keywords:
Code-switching
Bilingualism
Cross-language activation
Phonological organization
Voice onset time

ABSTRACT

Research on the phonetics of code-switching has focused on voice onset time (VOT) and has yielded mixed results regarding cross-language interaction, possibly due to differences in data used (scripted vs. spontaneous speech) and populations examined (L1 vs. L2 dominant, early vs. late bilinguals). Here VOT was measured in a corpus of spontaneous code-switching speech elicited from a homogeneous group of early bilinguals in conversation with and without distraction (completion of jigsaw puzzles). The distraction meant to increase cognitive load, a manipulation that could affect phonetic realization. Both English and Spanish VOT were shorter at code-switching points than in comparable monolingual utterances. English VOT lengthened overall under increased cognitive load (but remained shorter in code-switching as compared to the monolingual context). These results support previous findings of VOT shortening in code-switching for both English and Spanish, and confirm that the effect applies in the natural speech of early bilinguals.

1. Introduction

Code-switching is a practice common among bilinguals whereby speakers use both languages in a single utterance (Gumperz, 1977; Bullock & Toribio, 2009). Code-switching is particularly prevalent among fluent early or simultaneous bilinguals, defined as those who learned both languages before the age of six and continue to use them both in everyday life (McLaughlin, 1978; Poplack, 1980; Padilla & Lindholm, 1984; Flege, 1991; Flege, Munro, & MacKay, 1995; Hamers & Blanc, 2000; Genesee, Paradis & Crago, 2004; Gildersleeve-Neumann & Wright, 2010; Lee & Iverson, 2012). In bilingual research, code-switching has been well studied in regards to grammatical structure (Pfaff, 1979; Poplack, 1980, 1987; Woolford, 1983; Belazi, Rubin, & Toribio, 1994; Myers-Scotton, 2008). Relatively fewer studies, however, have examined the phonetics of code-switching (but see Grosjean & Miller, 1994; Bullock, Toribio, González, & Dalola, 2006; Antoniou, Best, Tyler, & Kroos, 2011; López, 2012; Olson, 2013; Balukas & Koops, 2014; Olson, to appear). Yet code-switching provides an interesting context in which to examine bilingual speech production, as it offers a window into bilingual processing in a natural context.

Past studies on the phonetics of code-switching have produced mixed results. Some found no difference between phonetic productions in monolingual vs. code-switching utterances (Grosjean & Miller, 1994; López, 2012). In others, differences were found for one of the languages. Antoniou et al. (2011) used Greek-English bilinguals whose L1 was Greek and found that these speakers' English VOT became shorter when produced in a code-switching context as compared to a monolingual context; in contrast, they did not find a similar effect of English on Greek VOT. Similar results are reported by Balukas and Koops (2014): the VOT of their Spanish–English bilingual speakers' English (their L2) was shorter when produced closer to a code-switch but there was no effect for Spanish. In yet a third set of studies, effects were found for both languages but of different types, depending on the population. Specifically, Bullock et al. (2006) found that the English VOT of L1 Spanish speakers in their study was shorter in code-switching, but the VOT of their Spanish was not affected; their L1 English speakers, on the other hand, showed a shortening of *both* English and Spanish VOT in code-switching contexts. The results of Olson (2013) indicate that effects may depend on language dominance: in his study the VOT of the speakers' dominant language (English or Spanish) shifted towards the non-dominant language under code-switching, while the non-dominant language was not affected; e.g., English VOT shortened in the speech of English-dominant

* Corresponding author. Tel.: +1 415 990 7616.
E-mail addresses: ppiccinini@ucsd.edu (P. Piccinini), a.arvaniti@kent.ac.uk (A. Arvaniti).
[1] *Present address*: University of Kent, Department of English Language & Linguistics, Canterbury, Kent CT2 7NF, UK.

0095-4470/$ - see front matter © 2015 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2015.07.004

speakers, but was unaffected by code-switching in the speech of Spanish-dominant speakers. These results are partially confirmed by Olson (to appear): with respect to English dominant speakers, this study replicated the effect found in Olson (2013): their English VOT shortened, but Spanish VOT was unaffected. On the other hand, Olson (to appear) found a bidirectional effect for Spanish-dominant speakers: their English VOT shortened and their Spanish VOT lengthened in code-switching.

There are several possible reasons for this lack of agreement among studies. First, not all studies tested the same populations: Grosjean and Miller (1994), Bullock et al. (2006), Olson (2013), and Olson (to appear) tested late bilinguals, while Antoniou et al. (2011) and Balukas and Koops (2014) tested early bilinguals. In Antoniou et al. (2011) and Olson (2013) speakers were L2 dominant, while in other studies they were L1 dominant (e.g. Bullock et al., 2006; Olson, to appear), or dominance was unclear (e.g. Balukas & Koops, 2014). In some studies the participant population was relatively uniform in age and other social characteristics (e.g. Grosjean & Miller, 1994; Bullock et al., 2006), but in others participants varied significantly in age (e.g. Balukas & Koops, 2014). This is important given that studies on VOT, such as Nagy and Kochetov (2013), have documented inter-generational changes in bilingual immigrant populations similar to the population in Balukas and Koops (2014). Olson (2013) tested both English and Spanish dominant bilinguals but it is arguable whether the participants in his study qualified as early bilinguals, having all learned their L2 after the age of 12. Furthermore, factors such as age and order of acquisition, language dominance, and language mode have all been found to affect bilingual production and processing indicating that these differences across studies may be responsible for the differences in the results (for age and order of acquisition, see, inter alia, Flege et al., 1995; Birdsong, 2001; Hakuta, Bialystok, & Wiley, 2003; for language dominance, see Cutler, Mehler, Norris, & Segui, 1992; Mok, 2011; Olson, 2013; Piccinini & Arvaniti, 2014; for language mode, see Dijkstra & Van Hell, 2003; Marian & Spivey, 2003; Soares & Grosjean, 1984).

An additional reason for the discrepancies could well be that studies used widely different tasks. Several studies relied on scripted materials of varying degrees of naturalness. Grosjean and Miller (1994) instructed participants to read names in French and English sentences pronouncing the names in English or French respectively. Bullock et al. (2006) and Olson (to appear), on the other hand, used more realistic sentences that switched from one language to the other in various ways, while Antoniou et al. (2011) elicited nonce monosyllables (e.g. [pa]) in a typical phonetic frame indicating the language switch by a change in alphabet (Greek vs. Latin). Other researchers tested spontaneous code-switching (e.g. Khattab, 2009; Balukas & Koops, 2014), and yet others relied on the production of isolated words (Olson, 2013). These different experimental paradigms are likely to have consequences for the realization of phonetic categories. Scripted code-switching allows for ample preplanning; e.g. the participants of Antoniou et al. (2011) were familiarized with the switched materials before recordings began. Code-switching in spontaneous speech, meanwhile, is likely taking place with less preplanning, potentially resulting in a greater affect of one language on the other in phonetic productions. Such fundamental differences between tasks could well have affected the phonetics of VOT (see Khattab (2002) and Olson (2013) and references therein). The use of different tasks in combination with different populations of bilinguals is likely to have further compounded discrepancies among studies.

The current study addresses these concerns by examining the effects of code-switching on VOT (1) in spontaneous speech, and (2) with a homogeneous group of early Spanish–English bilinguals who are now English (L2) dominant. By using spontaneous speech we can determine whether previously reported effects of code-switching on phonetic parameters are task artifacts, or whether they are real and observable in ecologically valid studies that take into account the social, spontaneous and interactive nature of code-switching. By focusing on a homogeneous group of early bilinguals, we can further test if effects are present in this specific bilingual population. The combination of these two elements allows us also to shed light on the reasons for the discrepancies in the results of previous studies. Finally, here effects are examined in two conditions, (a) natural code-switching and (b) code-switching with increased cognitive load. This additional parameter allows us to observe possible effects of increased cognitive load on changes resulting from code-switching itself.

VOT was selected both because it has been used in many previous studies, as noted, and thus its study would facilitate comparisons with previous literature, but also because the phonetics of VOT in Spanish and English are well understood. It is well established that Spanish has significantly shorter VOT than English, especially word-initially (Lisker & Abramson, 1964). This distinction has been documented in the speech of Spanish–English bilinguals, demonstrating that they are capable of maintaining distinct VOT distributions for each language (Flege & Eefting, 1987). Nevertheless, recent studies have found that while bilinguals are often able to produce VOT durations appropriate for each of their languages, they still do not perform exactly like monolinguals. The differences are manifested either as longer or shorter durations for a specific VOT category, or as more variable productions of a specific category (see Khattab, 2002, for Arabic-English; Kehoe, Lleó, & Rakow, 2004, for German-Spanish; Sundara, Polka, & Baum, 2006, and Lev-Ari & Peperkamp, 2013, for French-English; Lee & Iverson, 2012, for Korean-English).

Our predictions for the current study are based on previous work according to which bilinguals operate in a continuum with many intermediate stages between a fully monolingual mode in one language and a fully monolingual mode in the other (Grosjean, 2001). In this continuum bilinguals do not fully deactivate either language (Green, 1998), while different contexts can induce different degrees of activation of each language. For example, bilinguals are slower at naming pictures in one of their languages when a distractor from the other language is present, as compared to when the distractor is from the same language as the picture to name (Ehri & Ryan, 1980; Bijeljac-Babic, Biardeau, & Grainger, 1997; Costa, Miozzo, & Caramazza, 1999; Perea, Duñabeitia, & Carreiras, 2008). Increasing cognitive load (e.g. in the form of auditory feedback) can also result in more heavily accented speech, as bilinguals have difficulty suppressing the inactive language (Howell & Dworzynski, 2001). Similar effects are reported when bilinguals are tested in different language modes. For example, Simonet (2014) found that the Catalan vowels of highly proficient Catalan–Spanish bilinguals were affected by whether Spanish words were included in a task; when Spanish words were present, the Catalan vowels /o, ɔ/ moved closer to Spanish /o/. Even when comparing within language, bilingual productions are affected by the inactive

Table 1
Language profiles of the speakers including average (mode) age of acquisition, age of full fluency in understanding and speaking, and current exposure to English and Spanish.

		English	Spanish
Age of acquisition	Average	1	1
	Range	0–14[a]	0–2
Age fluent	Average	5	4
	Range	1–16	1–12
Current exposure	Average	80%	20%
	Range	49–90%	10–50%

[a] All speakers were exposed to English before age 6 even if they did not fully "acquire" it by that age.

language; e.g. Amengual (2012) and Goldrick, Runnqvist, and Costa (2014) found that bilinguals move phonetic categories towards the inactive language when producing words that have a cognate in that language. Code-switching speech can serve as another context to study this effect of degree of cross-language activation and in particular to test how such activation is manifested in running natural speech.

With this background in mind, we predicted that the VOT of English code-switching tokens would be shorter (more Spanish-like) than that of English monolingual tokens; in contrast, the VOT of Spanish code-switching tokens would be longer (more English-like) than that of Spanish monolingual tokens. In addition, we predicted that in the condition with increased cognitive load, the VOT of English voiceless stops would be shorter than that of English stops in the condition without such increase. We predicted that the reverse would be obtained for Spanish: increased cognitive load would lead to longer VOT for Spanish voiceless stops compared to the condition without the cognitive load increase. Finally, we predicted a cumulative effect, such that the VOT values to show the greatest effect from the other language would be those of code-switching tokens in the context with increased cognitive load (i.e. the shortest English VOTs and longest Spanish VOTs would be found in code-switching tokens produced in the condition with increased cognitive load).

2. Materials and methods

Early Spanish–English bilinguals of Mexican-American heritage were recorded in dyads while conversing on topics particular to Mexican and Mexican-American culture. They did so both with and without a distractor (the completion of jigsaw puzzles during conversation), on the assumption that the presence of the distractor would increase cognitive load.

2.1. Materials

Prompts based on Mexican-American culture were chosen with the help of a Mexican-American Spanish–English bilingual undergraduate researcher; the prompts were selected to be culturally appropriate and thus elicit as natural a conversation as possible. Three conversational prompts and accompanying pictures were used to elicit spontaneous speech: *Quinceañera*, a girl's 15th birthday party that marks an important milestone in Mexican-American culture; *Chavo del 8*, a popular Mexican TV show also shown in the United States on Spanish TV channels; and *Día de los Muertos* or Day of the Dead, an important holiday in Mexican and Mexican-American culture to honor and celebrate the dead. Prompts asked speakers to talk about their thoughts on the topic and posed specific questions about their experiences. Written versions of the prompts were provided on a piece of paper in both English and Spanish; Spanish was at the top of the page with the English translation below it. The Spanish text was presented first as a way to help speakers to get into a bilingual mode and thus facilitate code-switching during the conversation, as up until that point speakers had been using English. On a second piece of paper speakers were provided with a picture related to the prompt (for *Día de los Muertos*, speakers were given two pictures). If the speakers were two females, *Quinceañera* and *Chavo del 8* were used; if the speakers were two males or one female and one male *Chavo del 8* and *Día de los Muertos* were used, as males would have less to talk about for the *Quinceañera* prompt. As a result of this set-up no dyad had both the *Quinceañera* and *Día de los Muertos* prompts; nevertheless, prompts were evenly used across tasks (except that *Chavo del 8* was used four times in the task without distraction and three times in the task with distraction).

In one of the tasks, jigsaw puzzles were used as a form of distraction. There were four puzzles in total, each consisting of 12 2 in. × 2 in. pieces. Each puzzle was of a different animal one would find at the zoo; all puzzles were designed for children ages three and up. The puzzles were deliberately selected to be easy, as the aim was to provide a mild distraction, not stifle conversation due to the demands of the puzzles.

2.2. Speakers

Fourteen Spanish–English bilinguals of Mexican-American heritage participated in the experiment; one female speaker turned out to be a speaker of Puerto Rican Spanish, so her data were discarded. Speakers participated in dyads. Four dyads were female–

Table 2
Speakers' average (mode) self-reported proficiency in speaking, understanding, and reading in English and Spanish. Ratings were on a Likert scale (0 – none to 10 – perfect).

		English	Spanish
Speaking	Average	8	8
	Range	5–10	6–10
Understanding	Average	8	8
	Range	8–10	6–10
Reading	Average	8	8
	Range	7–10	4–9

female and three female–male (total 11 females, 3 males). The speakers were all first generation Mexican Americans; their average age was 20.2 years and ranged from 18 to 24 years. All speakers were UCSD undergraduates who were given course credit in exchange for participation. They all self-identified as fluent speakers of both languages, and said they were exposed to both languages before the age of six and continued to use them both in everyday life.

In order to corroborate these statements, before participating in the experiment speakers filled out the Language Experience and Proficiency Questionnaire (LEAP-Q; Marian, Blumenfeld, & Kaushanskaya, 2007) in English, answering questions about their language background and their use of English and Spanish. Speakers were allowed to mark both English and Spanish as their first language. Five speakers marked English as their first language and eight marked it as their second; 11 speakers marked Spanish as their first and two as their second language (in other words, three speakers marked both English and Spanish as first language). None marked anything but English or Spanish as their first or second language. Nine out of the thirteen speakers whose data were included in the study reported English as their dominant language, with the other four choosing Spanish. All speakers marked Hispanic for ethnicity, except one who declined to answer. None reported any speaking or hearing disorders. Average ages of acquisition and current exposure are reported in Table 1. Data by each speaker individually is presented in Table A1. Average self-reported proficiency measures are reported in Table 2, individual responses in Table A2.

2.3. Procedure

The study included two tasks: (1) directed conversation (henceforth referred to as the *Conversation Task*), and (2) directed conversation with distraction (henceforth the Conversation with Puzzle Task, or *Puzzle Task* for short); the distraction was the requirement that speakers complete individually the four jigsaw puzzles mentioned above while holding a conversation. All speakers were greeted in English and completed the language questionnaire in English before participating in the study. Speakers were given one of the three conversational prompts with accompanying picture(s) and told to read the prompt and discuss it using the pictures (different prompts were used for the two tasks of each dyad of speakers). Prompts and task order were counterbalanced between dyads of speakers, to the extent that gender-related requirements permitted (see Section 2.1.). In the Puzzle Task speakers were instructed to independently complete each of the four puzzles while talking about the prompt. Speakers were given no restrictions regarding turn-taking or about which language to use, and were not directed or interrupted by the experimenter until the end of the task. All conversations took place in the sound booth of the UCSD Phonetics Lab. The experimenter was not present for the conversations. For the Conversation Task, after 15 min passed, the experimenter went into the sound booth and told the speakers to end their conversation (the average duration of conversations from the point when speakers started discussing the prompt to the point when the experimenter returned to the sound booth was 14.89 min [standard deviation 0.13 min]). For the Puzzle Task, conversation ended when both speakers had completed all four puzzles, at which point they alerted the experimenter (the average duration of conversations from the point when speakers started discussing the prompt to their leaving the sound booth to get the experimenter was 9.23 min [standard deviation 1.86 min]).

All speakers knew their partner before the experiment. This was deliberate, as a pilot study showed that speakers would not code-switch unless they were already familiar with their partner. Conversations were recorded using Praat (Boersma & Weenink, 2001) and an A-to-D converter (at a sampling rate of 48 Hz with a quantization rate of 16-bit). The recordings were in stereo using two Earthworks SR77 microphones.

2.4. Annotation and measurements

All conversations were transcribed in standard orthography using the annotation facility in Praat. Four transcription tiers were used, two per speaker, with one for English orthographic transcription and the other for Spanish orthographic transcription (two speakers × two languages). This was done to keep productions by each speaker and language separate. Utterance boundaries were annotated in Praat by the first author, who is a native speaker of English and an L2 Spanish speaker. If the first author was unsure about the language an utterance belonged to, it was marked on both tiers and checked with a second transcriber, a native speaker of Spanish. Once all utterance boundaries were annotated, all English utterances were transcribed by the first author. Spanish utterances were transcribed by either the first author or a native Spanish speaker; any Spanish transcription by the first author was double checked by a native Spanish speaker.

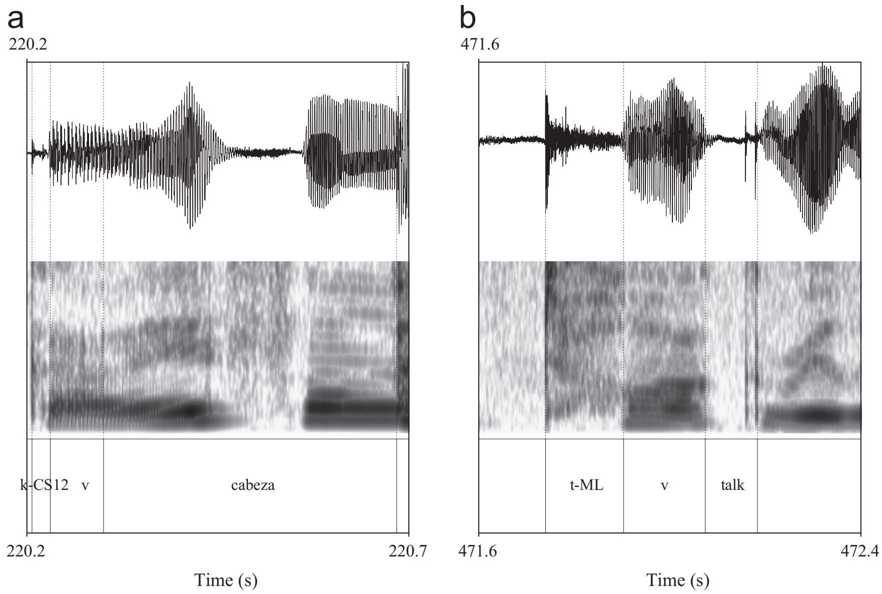

Fig. 1. Examples of English and Spanish VOT segmentations. In (a) the Spanish word “cabeza” (“head”) in a code-switching utterance; in (b) the English word “talk” in a monolingual utterance.

Table 3
Examples of utterance coding. Example words are bolded.

		Language	
		English	Spanish
Context	ML	I I saw it but just like a really long **[t]ime** ago.	***[p]orqure*** *no* ***[t]e*** *gusta si ves* ***[p]areces [k]ue*** *eres niña* ***[k]ue*** *le gusta eso* GLOSS: *because you don't like if you see that you look like the girl who likes that*
	CS	Kinda like you know how they **[p]ut** on *esos aretes*… GLOSS: Kinda like you know how they put on *those earrings*…	***[p]orqu****é* *este* why is he like carrying a flowers did he like the girl? GLOSS: *why this* why is he like carrying a flowers did he like the girl?
		…*no sé mucho* like a lot of **[k]ountries**… GLOSS: *I don't know* like a lot of countries…	…sweet sixteen *nomás es de* ***[k]ue*** *gracias y ahora party* GLOSS: …sweet sixteen *isn't about giving thanks just party*
		…*pero* like about his **[k]aracter** and stuff *no me acuerdo mucho*. GLOSS: *but* like about his character and stuff *I don't remember much*	…he would like crack a joke or whatever like *alguien le* ***[p]egaba*** *en la* ***[k]abeza*** or something… GLOSS: …he would like crack a joke or whatever like *someone hit him on the head* or something…

VOT was measured for word initial voiceless stops /p/, /t/, and /k/ in English and Spanish by simultaneous inspection of the waveform and spectrogram provided by Praat. Only word initial voiceless stops followed by a vowel were selected, since clustering and a stop's position in a word can affect VOT duration or lead to lenition, such as /t/-flapping in American English (Ladefoged, 2003). Measurements were taken from the onset of the burst to the onset of the following vowel. VOT was measured from the point when a vertical striation in the spectrogram and amplitude spike in the waveform were evident to the point when the waveform became consistently periodic and the spectrogram showed clear formant structure. Tokens of /t/ were not included when /t/ was intervocalic due to the preceding word ending in a vowel, and the /t/ was judged to be a flap based on short closure duration and lack of a clear burst and VOT.

In addition to VOT, the vowel following the stop and the remainder of the word were segmented and the durations of VOT, the following vowel and the entire word were calculated. These durations were used to provide two measures of speaking rate. First, the total duration of the word (VOT plus remainder of the word after VOT) was divided by the number of phonemes in the word; this provided an “average” phoneme duration for phonemes within a given word. Second, following Balukas and Koops (2014), the duration of the vowel following each stop was used as an additional measure of speaking rate. All segmental annotation was done on four additional tiers, one per language and speaker. All measurements were done by the first author but unclear cases were discussed with the second author. Two examples are provided in Figs. 1(a) and (b).

All VOT measurements were coded with information that was considered likely to affect VOT duration. Thus, tokens were coded for: (1) specific word; (2) presence of stress on the first syllable of the word; (3) word type (content or function); and (4) the quality of the vowel following the stop. Studies have found that these factors can affect VOT duration in English (Klatt, 1975; Neiman, Klich, & Shuey, 1983; Higgins, Netsell, & Schulte, 1998; Cho & Ladefoged, 1999; Whiteside, Henry, & Dobbin, 2004; Yao, 2009). The same applies to Spanish though to a lesser extent (Magloire & Green, 1999; Schmidt & Flege, 1996).

In addition, tokens were coded according to whether the stop occurred in a monolingual utterance, either English or Spanish, or a code-switching utterance. Due to a lack of clearly defined norms for code-switching, for the purposes of this study a “code-switching utterance” was operationally defined as an utterance that included both languages, had a pause of less than 300 ms between languages at switch points, and had no false starts. Coding was based on whether the coder (the first author) considered the code-switch and VOT token to be contained within the same utterance. Although this can be seen as subjective, given that 70% of tokens were within three words or less of a switch point (see Section 2.5), we can reasonably infer that code-switching tokens did occur within the scope of a code-switching utterance. In all coding “code-switch” words were considered to be the result of true code-switching, and not words of one of the two languages borrowed into the other with phonology and phonetics matching the frame language (see Pfaff, 1979, for discussion of code-switching vs. borrowing). Three code-switching contexts were defined: (1) pre-switch contexts (e.g. for English words “English–Spanish”), (2) post-switch contexts (e.g. for English words “Spanish–English”), and (3) dual-switch contexts (e.g. for English words “Spanish–English–Spanish”). Table 3 provides examples of monolingual utterances in both languages and examples of all three code-switching types by language.

2.5. *Description of the corpus*

The collected corpus included 159.62 min of conversation, 104.25 min from the Conversation Task and 55.37 min from the Puzzle Task (for one dyad of speakers the recording of the Puzzle Task was lost due to experimenter error).

This corpus was analyzed for a variety of characteristics. Our primary concerns were as follows. Did task affect code-switching behavior and, more generally, the speakers' speaking patterns? This was a possibility, since the distraction used could have rendered conversation fragmented and desultory. Did speakers regularly code-switch and if so, were they influenced by each other, since they operated in dyads and speakers have been found to influence each other's speech patterns in paired tasks (Pardo, 2006; Kim, Horton, & Bradlow, 2011)? Finally, did the amount of code-switching differ by gender and prompt (since prompts were gendered, to some extent)? Below we present some general information about the corpus and then focus on the part of the corpus investigated here, word-initial voiceless stop tokens.

First, the corpus was analyzed for amount of code-switching. This analysis included the speaker whose data were discarded because she was a speaker of Puerto Rican Spanish; this was done to get a full picture of the conversations as a whole. Speakers' utterances were marked off such that any pause of 300 ms resulted in a new utterance. For example, if a speaker was speaking for a stretch of 1000 ms, but had a 400 ms pause 200 ms into the turn, the whole stretch would be marked as two utterances, one from 0 to 200 ms and one from 600 to 1000 ms. If there was a change in language, even if languages were separated by less than 300 ms, each language was coded as its own utterance. Utterances were then coded for code-switching. This coding followed the same conventions as the coding of VOT tokens (see Section 2.4). If an utterance in one language was followed or preceded by the other language with a pause of less than 300 ms between utterances, it was coded as a code-switching utterance; all other utterances were coded as monolingual. Code-switching utterances were coded as pre-switch (an utterance preceding a switch into the other language), post-switch (an utterance following a switch from the other language), or dual-switch (an utterance both preceding and following a switch).[2]

Two variables were examined: (1) the raw number and (2) the total duration of monolingual and code-switching utterances in the corpus. The corpus contained a total of 5309 utterances; 3839 of these (72%) were monolingual (2415 English, 1424 Spanish), and the other 1470 (28%) were code-switching utterances (775 English, 695 Spanish). Within the code-switching part of the corpus, 596 utterances (40.5% of all code-switching utterances) were pre-switch, with 347 being English utterances and 249 Spanish utterances; 596 (40.5%) were post-switch, with 329 being English utterances and 267 Spanish utterances; finally 278 utterances (19% of switches) were dual-switch, with 99 being English utterances and 179 Spanish utterances. In regards to duration, 75% of the conversation was produced in a monolingual utterance and 25% in a code-switching utterance. This general pattern also applied when the data were separated by task. In the Conversation Task, 73% of utterances were monolingual and 27% code-switching; in the Puzzle Task the percentages were 72% and 28% respectively. A paired *t*-test by speaker for percentage of code-switching utterances found no significant effect of task [$t(11) = -0.20$, *n.s.*]. When the data was analyzed in regards to duration, in the Conversation Task 77% of speech was in monolingual utterances and 23% in code-switching utterances; in the Puzzle task, the percentages were 72% and 28% respectively. A paired *t*-test by speaker for percentage of duration of code-switching speech again found no significant effect of task [$t(11) = -0.75$, *n.s.*]. See Table B1 for more details.

To examine how dyads of speakers compared with respect to amount of code-switching, we computed the difference of the percentage of code-switching within each dyad, both in regards to number of utterances and duration of speech, and took its absolute value. A 0% difference would mean that each speaker in the dyad code-switched exactly the same amount; a large difference would mean that one speaker in the dyad code-switched much more than the other. For example, if one speaker in a dyad code-switched 20% of the time and the other speaker 21% of the time the difference between speakers would be 1%. If speakers in a dyad converged towards each other in their amount of code-switching, then the standard deviation of this difference across dyads should be lower than the standard deviation for all speakers together. For example, if for Dyad 1, the percentage of utterances that were code-switching was 20% for Speaker #1 and 21% for Speaker #2 (difference of 1%), while for Dyad 2, the percentages were 50% and 52% respectively (difference of 2%), the standard deviation for all four speakers together would be 18%, but the standard deviation of the difference between speakers in a given dyad would only be 1%. For number of code-switching utterances the standard deviation of the difference between speakers in a given dyad was 9%, and the standard deviation of all speakers was 10%;

[2] This coding scheme does create a tautology such that any time there is a pre-switch utterance there is also a post-switch utterance. As a result there are the same number of pre- and post-switch utterances. However, due to the presence of occasional intervening dual-switch utterances, the breakdown by language is not symmetrical, as can be seen by the numbers of English and Spanish pre- and post-switch utterances reported.

Table 4
Total number of monolingual and code-switching tokens across and within tasks by language.

	English		**Spanish**		Grand Total
	Monolingual	Code-switching	Monolingual	Code-switching	
Conversation Task	373	91	746	191	1401
Puzzle Task	147	30	318	126	621
Total	520	121	1064	317	2022

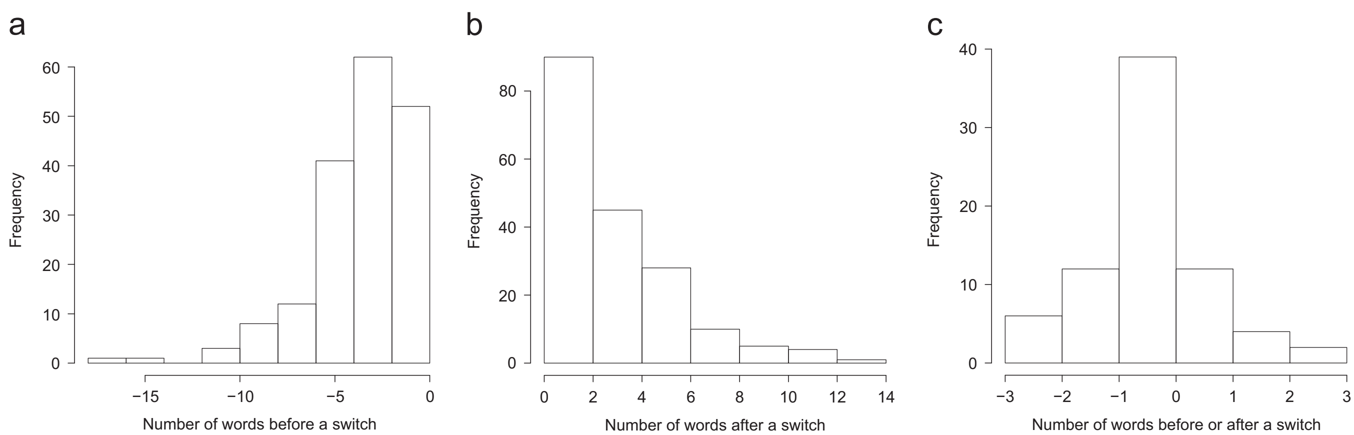

Fig. 2. Histograms of number of words before or after a code-switch for (a) pre-switch, (b) post-switch, and (c) dual-switch tokens.

for duration it was 10% for both the difference between speakers in a given dyad and for all speakers. This suggests that speakers within a dyad were no more similar in their amount of code-switching to their partner than to the rest of the speakers.

Breaking down the data by gender, we found that females and males produced roughly the same proportion of code-switching utterances within their speech (females 26% of utterances, males 27% of utterances). Males and females were also similar in the proportion of speech duration that included a code-switching utterance (females 24%, males 22%). Examining the amount of code-switching by prompt, we found the *Quinceañera* and *Chavo del Ocho* prompts had similar amounts of code-switching both in regards to number of utterances (29% of utterances for *Quinceañera* and 30% of utterances for *Chavo del Ocho*) and in regards to durations (26% of speech for both). The *Día de los Muertos* prompt resulted in somewhat less code-switching (21% of utterances and 18% of the duration of conversations).

The corpus was additionally analyzed for pauses and disfluencies in order to ensure that the Puzzle Task—for which conversations were shorter, as noted—had not resulted in desultory or disfluent conversation among long pauses during which speakers worked on the puzzles. Any period of silence of more than 300 ms was considered a pause. Both pauses within one speaker's turn and those between speaker turns were included (again, this analysis included the speaker whose data were discarded because she was not a speaker of Mexican Spanish; this was done so that between-turn pauses between her and her partner could be calculated). Pause durations were expressed as a percentage of the total duration of the conversation they were part of, since conversations differed in duration. A paired t-test showed no significant difference in the percentage of time filled by pauses of any type in the Puzzle Task compared to the Conversation Task [$t(5) = -1.21$, *n.s.*]. Differences across tasks were not found when pauses were separated into within- and between-turn pauses [$t(5) = -2.16$, *n.s.* and $t(5) = -0.28$, *n.s.*, respectively]. Finally, no speaker-specific differences were found for the duration of within-speaker pauses between the two tasks [$t(11) = -1.04$, *n.s.*]. See Table B2 for more details.

As noted, disfluencies were also examined. A disfluency was defined as any instance where a speaker stopped producing a word before completing it. In the Conversation Task, 211 disfluencies were observed overall (for 14 speakers, including the speaker not analyzed in the final data set); this is approximately 15 disfluencies per speaker or 2 per minute. In the Puzzle Task, 104 disfluencies were observed overall (for 12 speakers, as one dyad did not have a recording for the Conversation Task); this is approximately 9 disfluencies per speaker or 2 per minute. A paired t-test for disfluencies per minute for each speaker found no difference between the two tasks [$t(11) = 0.17$, *n.s.*]. See Table B3 for more details.

The main variable in this study was word-initial voiceless stops. The corpus included a total of 2022 instances of word-initial voiceless stops that were measured for analysis. The total number of stops measured by language, task, and context (monolingual or code-switching) is reported in Table 4. Of all word-initial voiceless stops measured, 78% were monolingual and 22% code-switching. As can be seen, there were fewer tokens in the Puzzle Task than the Conversation Task. This is likely due in large part to the shorter duration of the conversations in the Puzzle Task, since the percentage of tokens that are code-switches is similar across the two tasks (20% in the Conversation Task, 25% in the Puzzle Task). We also note that these numbers do not reflect the number of times code-switching occurred, only the number of times a word beginning with a voiceless stop occurred in each context. Thus the fact that there are fewer code-switching *tokens* in the Puzzle Task does not mean there was less code-switching *in general* in that task.

Table 5
Average (mean) duration of tokens for each language, context, task; data are presented both pooled and by place of articulation. Standard deviations are presented in parentheses.

		English		Spanish	
		Monolingual	Code-switching	Monolingual	Code-switching
Conversation	p	56 (22)	51 (23)	26 (18)	22 (13)
	t	62 (23)	56 (18)	29 (15)	24 (11)
	k	59 (21)	55 (26)	35 (15)	33 (13)
	Pooled means	59 (22)	54 (23)	32 (16)	29 (13)
Puzzle	p	61 (24)	45 (16)	24 (11)	23 (09)
	t	68 (24)	58 (17)	26 (11)	27 (14)
	k	64 (26)	62 (23)	33 (14)	29 (12)
	Pooled means	65 (25)	57 (19)	30 (13)	27 (12)

This is confirmed by the analysis of code-switching in the entire corpus discussed earlier in this section; this analysis shows very comparable percentages of code-switching to those that pertain to word-initial voiceless stops in particular.

Fig. 2(a)–(c) gives the distribution of the code-switching voiceless stop tokens in terms of their distance from the switching point. Negative numbers indicate pre-switch tokens. For dual-switch tokens one of two codings was used: if a token was equally distant from both switch points (e.g. if it was two words after a switch and two words before a switch), it was coded as zero; otherwise the shortest distance from the nearest switch point was used (e.g. if a token was two words after a switch and three words before the next switch it was coded as “2” not “−3”). The pre-switch token farthest from a code-switching point was 18 words before the switch, and the post-switch token farthest was 14 words after the switch. The median for pre-switch tokens was three words before the switch, for post-switch tokens three words after the switch, and for dual-switch tokens directly between two languages switches (i.e. zero words). Although sometimes tokens were relatively far from the switch this did not apply to most tokens, as can be seen in the histograms; 70% of code-switch tokens were within three words before or after the switch (−3 to 3). This compares well with scripted data, such as that of Bullock et al. (2006), in which the scripted code-switches were two or three words before or after a switch. To the extent that comparison is possible, the data are also comparable with the spontaneous code-switching corpus of Balukas and Koops (2014), who measured distance from a code-switching point only in terms of duration.

Additional analysis of the distribution of code-switching stop tokens by gender indicated that the female speakers produced a higher percentage of code-switching tokens than males (23% vs. 18% of their total number of word-initial voiceless stops respectively). Tokens were somewhat differently distributed for males and females: for females, pre-switches were 38% of the total, post-switches were 46%, and dual-switches were 16%; for males the percentages were 51%, 27%, and 22% respectively. Analysis of the tokens by prompt indicated that the *Quinceañera* and *Chavo del 8* prompts elicited more code-switching tokens (26% and 23% of voiceless stop tokens respectively) than the *Día de los Muertos* prompt (13%). As the overall code-switching patterns indicate, these differences were likely due to chance, the number of word-initial stop tokens in the sample, rather than to overall frequency of code-switching.

Based on these results, we can reasonably assume that the code-switching data that pertain to the study of VOT were representative of the code-switching in the conversations. Of all utterances in the corpus, 28% were code-switching utterances, and of all VOT tokens coded 22% were code-switching, thus supporting the idea that the VOT data is representative of the corpus as a whole. Differences by task were similar between the full corpus and the VOT subset of data. Regarding a possible effect of dyad, the results show that a given speaker's amount of code-switching was no more similar to that of their partner than to any other speaker in the corpus, therefore there is no strong evidence of convergence between speakers of each dyad. Task and particularly the distraction did not seem to have negatively influenced conversations. Speakers were not pausing significantly more in the Puzzle Task, and thus they were most likely working on the puzzle while talking. This conclusion is supported by the fact that that the speakers discussed the puzzles during the recordings and even commented on the difficulty of talking while working on the puzzles. At the same time, it is clear that working on the puzzles did not affect their speech to the point of stopping them from conversing or affecting their speech, as it did not result in either a greater number of disfluencies or in longer pauses. This in turn suggests that the manipulation was successful: working on the puzzles provided a light distraction without stopping conversation or leading to disfluent speech. Our conclusion is also indirectly supported by results on speaking rate, which also did not show a difference between the Puzzle and Conversation Tasks (for details, see Section 3.1.).

3. Results

3.1. Monolingual vs. code-switching VOT

Average VOT duration by task, language, context and place of articulation are presented in Table 5. To test for significant effects, linear mixed effects models (henceforth LMEMs) were run in R (R Development Core Team, 2013) using the lme4 package (Bates, Maechler, Bolker, & Walker, 2014). LMEMs were chosen because our dataset was not balanced across conditions (an inevitable consequence of using natural speech), and LMEMs are robust to the effects of unbalanced datasets.

The dependent variable was VOT in log-transformed ms. The log transform was conducted to make the distribution more normal and decrease the degree of positive skew (Keene, 1995; Limpert, Stahel, & Abbt, 2001).[3] The fixed effects were language (English or Spanish), context (monolingual or code-switch), task (Conversation or Puzzle), and place of articulation (bilabial, alveolar, velar). In order to take possible speaking rate effects on VOT duration into account, the (log-transformed) average phoneme duration in the words from which VOT was extracted was included in the model as covariate (see also Section 2.4.).[4] Language, context, and task were included as interactions; the three types of code-switching were pooled to increase power. No other variables coded for were found to significantly improve the model and thus some of these (e.g. "word type") were removed from the final model. All categorical variables were coded using contrast coding; as such, place of articulation was included as two fixed effects, bilabial vs. lingual (alveolar and velar), and alveolar vs. velar. A random slope for speaker by language, task, and context was included, allowing us to factor out individual differences. A random slope for dyad of speakers by language, task, and context was also included, allowing us to factor out any effects due to individuals being paired with other specific individuals. This was the maximal, uncorrelated random-effects structure that converged. There were no interactions in the random effects structure; only random slopes for each main effect included. Significance of fixed effects, which was set at $p<0.05$, was assessed using model comparison.

There was a significant effect of language, such that Spanish tokens had shorter VOTs than English tokens [$\beta=-0.31$, $SE=0.02$; $\chi^2(1)=24.93$, $p<0.001$]. There was also a significant effect of context, with code-switching tokens having shorter VOTs than monolingual tokens [$\beta=-0.04$, $SE=0.01$; $\chi^2(1)=6.97$, $p<0.01$]. The main effects of language and context are presented in Fig. 3. Three other fixed effects were also significant. First, as expected, bilabial voiceless stops had shorter VOTs than lingual voiceless stops [$\beta=0.13$, $SE=0.01$; $\chi^2(1)=116.97$, $p<0.001$], and alveolar voiceless stops had shorter VOT than velar voiceless stops [$\beta=-0.06$, $SE=0.01$; $\chi^2(1)=43.50$, $p<0.001$]. Average phoneme duration was also significant, with tokens showing a positive correlation between VOT and average phoneme duration [$\beta=0.24$, $SE=0.02$; $\chi^2(1)=164.28$, $p<0.001$]. A simple linear regression with VOT duration and average phoneme duration, both log-transformed, had a significant positive slope, indicating that VOT increased at slower speaking rates [$r=0.31$, $p<0.001$].

In addition to main effects, the model showed a significant interaction of language and task [$\beta=-0.06$, $SE=0.02$; $\chi^2(1)=5.76$, $p<0.05$]; see Fig. 4. Follow up regressions run separately by language found that VOTs were longer in the Puzzle Task for English [$r=0.09$, $p<0.05$], but the effect was not present for Spanish [$r=0.03$, *n.s*]. No other interactions were significant. Most notably the language by context interaction was not significant, suggesting that the effect of context was the same for both English and Spanish tokens (see Fig. 3). This result was followed by post-hoc analyses conducted on English and Spanish separately, since a context effect but no context by language interaction was unexpected, and we wanted to be sure the context effect did apply to both English and Spanish. The analyses confirmed that the context effect was significant for both English and Spanish at $p<0.05$. The three-way interaction of language, context, and task was also not significant. For individual speaker data separated by language, context, and place of articulation see Table C1.

3.2. Distance from code-switch point

In order to explore the effect of distance from the switch on VOT durations, an LMEM was run with only the code-switching data. The model was the same as the one with all data, except that context was no longer included as a fixed effect (as only code-switching tokens were examined) and distance from the code-switch (−18 to 14 words)[5] was included as a fixed effect and as an interaction with language and task. The results indicate that distance from the code-switch point did not affect VOT. However, there was a significant interaction of distance from the code-switch and language [$\beta=-0.01$, $SE=0.004$; $\chi^2(1)=3.89$, $p<0.05$]. Follow up simple linear regressions, however, did not find a significant effect of distance from the code-switch for either English or Spanish. Fig. 5(a)–(c) shows the transition over time for both languages separated by place of articulation. Visual inspection of the figures suggests that for English tokens, the greater the distance from a post code-switch point, the higher (more English-like) VOT durations become. Spanish tokens however appear to have a steady duration regardless of distance from code-switching. While this was not confirmed with the follow-up regressions, the fact that the interaction of language and distance from the code-switch was significant suggests that with greater power the regressions could become significant for English tokens.

3.3. Individual differences

In addition to pooled data, we also examined the extent to which the pooled results applied for individual speakers. To this effect, ANOVAs were run on the data of individual speakers (though we note that ANOVA does not give the most representative picture of individual speaker results since it assumes a normal distribution and a balanced data set; the spontaneous nature of the corpus meant that these conditions were not met for all speakers). The models had the same independent variables as the full LMEM. Only results found to be significant in the full model with all speakers that were of particular interest to this study are reported: language,

[3] A Shapiro–Wilk normality test found that the raw durations were not normally distributed [$S–W=0.90$, $p<0.0001$]. The log-transform data were not normally distributed either, but were significantly less skewed than the raw data [$S–W=0.99$, $p<0.001$].

[4] The same analysis was run with duration of following vowel log-transformed as the covariate to account for speaking rate, following Balukas and Koops (2014). It produced the same results, but the model reported here with average segment duration as the covariate produced a better fit for the data.

[5] Two further analyses were conducted. The first used absolute time as a measure of distance from the code-switching point rather than number of words; this analysis produced a model with a worse fit than that using number of words. The other involved only tokens within 10 words of a switch so as to avoid possible artifacts from including words too far from the switch point to be affected by it; this trimming did not affect the results.

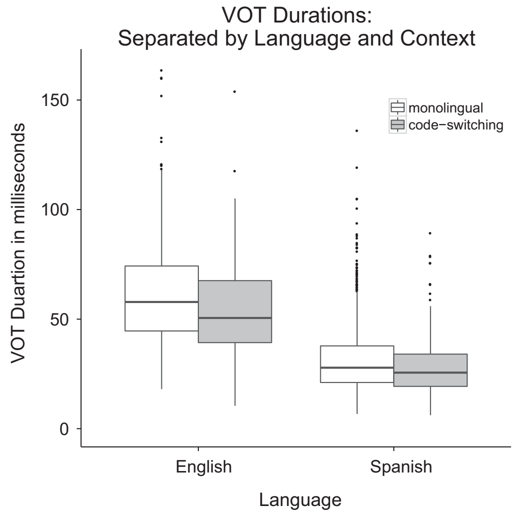

Fig. 3. Boxplots of tokens separated by language and context. Thick bands represent the second quartile (median) VOT durations, top and bottom bands of the box represent first and third quartiles of the data, whiskers represent the lowest and highest data points still within 1.5 interquartile range (IQR) of the lower and higher quartile; data points outside of the whiskers can be considered outliers.

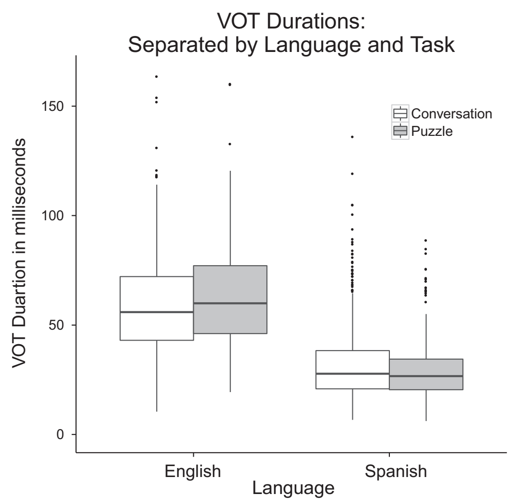

Fig. 4. Boxplots of tokens separated by language and task. Thick bands represent the second quartile (median) VOT durations, top and bottom bands of the box represent first and third quartiles of the data, whiskers represent the lowest and highest data points still within 1.5 interquartile range (IQR) of the lower and higher quartile; data points outside of the whiskers can be considered outliers.

context, and the language by task interaction. Since not all speakers had data points for every data cell (e.g. Spanish code-switching in the Puzzle Task), only speakers who had a full paradigm of data points were analyzed. This reduced the number of speakers analyzed with individual ANOVAs to six.

The effect of language was significant [$p<0.05$] for all speakers, with Spanish tokens having shorter VOTs than English tokens. The effect of context was significant [$p<0.05$] for three of the six speakers. The interaction of language and task was significant for four out of six speakers. The lack of these two effects for some speakers may be plausibly attributed to data scarcity. This tentative explanation is supported by the fact that the four speakers who had the significant interaction of language and task were also in the top five speakers in the study for most number of data points. See Table 6 for summary of individual models. While these results are not as robust as one would hope, it is important to remember that the original LMEM does account for individual differences by including speaker as a random effect, and thus present a more accurate picture of overall effects of code-switching and increased cognitive load on VOT.

4. Discussion

The present study examined spontaneous code-switching as produced by a homogeneous group of largely L2 dominant bilinguals. The results showed that although speakers did maintain distinct VOT categories in English and Spanish, their code-switching tokens were different from tokens produced in monolingual utterances. This is prima facie evidence that code-switching does affect the phonetic production of bilinguals even when they are in what can be seen as a generally bilingual mode – i.e. a mode

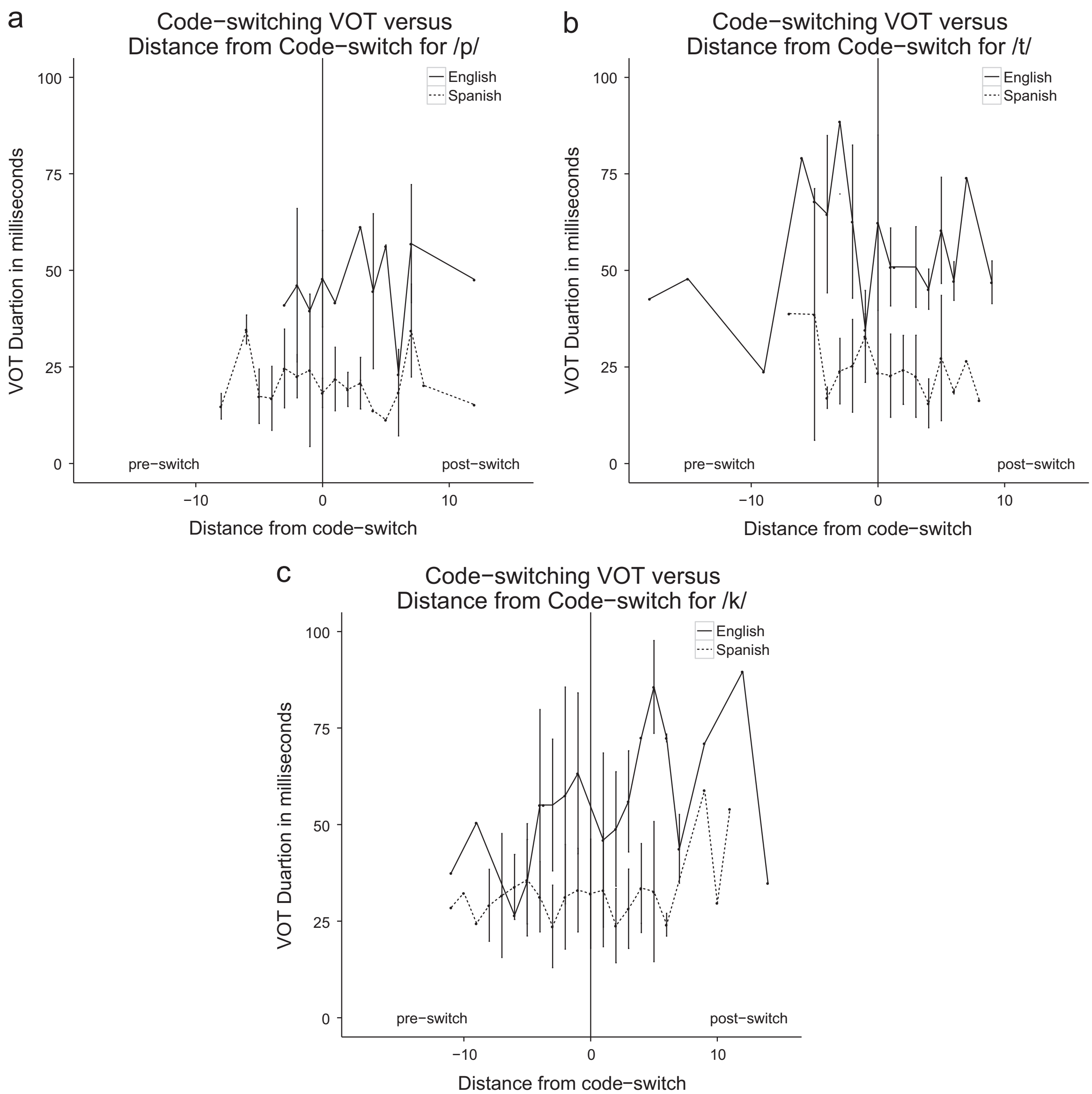

Fig. 5. Mean VOT by place of articulation for English and Spanish by distance from code-switch with standard deviations for (a) /p/, (b) /t/, and (c) /k/. Negative numbers correspond to pre-switch and positive numbers post-switch; zero is dual-switch.

in which both languages are in use – as in the present study. Thus the results confirm that, at least among early bilinguals, code-switching effects reported in earlier studies are not an experimental artifact but apply in spontaneous speech as well.

The overall difference between Spanish and English VOT replicates previous findings that early bilinguals are able to maintain distinct phonetic categories for both of their languages. At the same time, the data support the finding that bilingual productions are not always identical to those of monolinguals and may present more variation. This is illustrated in Table 7 which reports the values for monolingual tokens from the Conversation Task and compares them to VOT durations from early Spanish–English bilinguals (Flege & Eefting, 1987), early-Spanish mixed-English (some early, some late) bilinguals (Balukas & Koops, 2014), and monolingual English and Spanish speakers (Lisker & Abramson, 1964). These comparisons are less than ideal as they involve different populations and types of data: Flege and Eefting (1987) and Lisker and Abramson (1964) used scripted speech which shows less VOT variability than the spontaneous speech used here (Yao, 2009); Balukas and Koops (2014) had a much more variable population in terms of age than the present study (see Nagy & Kochetov, 2013 on this point). Nevertheless, as Table 7 indicates, the English values from the present study were closer on average to Flege and Eefting (1987) than to Lisker and Abramson (1964) or Balukas and Koops (2014); the same partly applies to Spanish as well. In addition, the present speakers had longer Spanish VOTs than the speakers in the other studies, possibly because they are L1-Spanish but English-dominant. A comparable effect was found by Balukas and Koops (2014) for English VOT: their speakers, for whom dominance was not clear, had rather short VOTs in English compared to monolinguals (cf. Sundara et al., 2006). As Figs. 3 and 4 also indicate, the Spanish VOT values in the present study were more variable than the values for English; this is rather unusual, given the generally low variability of short-lag VOT (which is also reflected in the standard deviations in Table 7; cf. Kessinger & Blumstein, 1997; Arvaniti, 1999).

Table 6
Results of ANOVAs for individual speakers for three main significant effects from full model, language, context, and the language by task interaction (*$p<0.05$, **$p<0.01$, ***$p<0.001$).

	Num.	Language		Context		Language × Task	
		F	*Significance*	*F*	*Significance*	*F*	*Significance*
Female 2	182	163.21	***	5.72	*	1.87	n.s.
Female 3	190	31.91	***	1.23	n.s.	4.81	*
Female 7	318	89.72	***	21.78	***	4.83	*
Female 9	163	116.72	***	6.09	**	5.65	*
Female 10	115	139.00	***	1.30	n.s.	1.29	n.s.
Male 2	369	322.73	***	0.14	n.s.	4.06	*

Table 7
Mean VOT durations and standard deviations (in parentheses) for monolingual utterances in the Conversation Task compared to Flege and Eefting (1987), Balukas and Koops (2014), and Lisker and Abramson (1964). Δ is the average absolute difference in VOT duration between the present data and each of the other studies; averages pooled over place of articulation but separate for English and Spanish.

	English				Spanish			
	Present study	F&E (1987)	B&K (2014)	L&A (1964)	Present Study	F&E (1987)	B&K (2014)	L&A (1964)
p	56 (22)	57 (14)	30 (18)	58	26 (18)	17 (7)	20 (11)	4
t	62 (23)	74 (16)	35 (59)	70	29 (15)	19 (5)	20 (12)	9
k	59 (21)	75 (17)	45 (20)	80	35 (15)	31 (6)	26 (12)	29
Δ		6	22	10		8	8	16

The results also showed that VOT was affected by the study's manipulations, code-switching and cognitive load increase (Conversation vs. Puzzle Task). We discuss each effect in turn.

First, in both tasks, Conversation and Puzzle, code-switching resulted in shorter VOTs in both English and Spanish. This result did not match our hypotheses that shortening would be observed only in English, with Spanish showing VOT lengthening instead. The result for English can be interpreted as the outcome of interference from Spanish, as per our hypotheses and previous studies, such as Bullock et al. (2006) and Olson (2013), which suggest that L1 influences L2 even when L2 is dominant. It is clear, however, that the same explanation cannot apply to Spanish. For Spanish, our results most closely match those reported by Bullock et al. (2006), who found that their L1 English speakers (who were also L1-dominant) had shorter VOT in both their English code-switch tokens and their Spanish pre-switch tokens. Bullock et al. (2006) argued that these comparable results across the two languages are due to different causes: they interpret English VOT shortening as evidence of English being influenced by Spanish, but attribute the shortening of Spanish VOT to hyperarticulation. The reason offered for the latter explanation is that these L1 English speakers had learned Spanish through formal training and were Spanish language teachers, so they would be likely to accentuate phonetic characteristics of Spanish when under pressure to differentiate the two languages, such as during code-switching. Something comparable may have applied to our participants as well: while L1 speakers of Spanish, they were also English dominant and less secure about their Spanish than their English, so they may have hyperarticulated features like VOT in order to signal their competence in Spanish. Using a range of phonetic values, some of which are closer to one or the other language, is part of a bilingual's repertoire even when in monolingual mode, as Khattab (2002) has demonstrated. As Olson (to appear) argues, however, there are likely limits to this type of variability, so that speakers remain intelligible and within the norms of their community. This could apply here as well: our speakers had relatively long VOTs in Spanish (see Table 7), so they would be less likely to lengthen VOT further. Though clearly further research is necessary with diverse bilingual populations and a variety of tasks and phonetic parameters, the present data serve to illustrate the more general point that emerges from recent research, namely that changes during code-switching may not have a unique cause, be it cognitive or sociolinguistic, even if the end result is the same (e.g. VOT shortening as in the present data).

With respect to the cognitive load manipulation, our study showed effects only for English. We had expected that the effect of code-switching would be enhanced in the Puzzle Task due to the added cognitive load, leading to English code-switching tokens having *even shorter VOT* in this task. Instead, under these conditions, English VOT lengthened relative to the values in the Conversation Task. As demonstrated in Section 3.1, this result could be attributed at least in part to a decrease in speaking rate. However, this decrease cannot be the sole cause, as the interaction of language and task was significant despite including speaking rate in the model. Though the reason for the lengthening is not clear, we note that the difference between code-switching and monolingual tokens (with the former having shorter VOT than the latter) was not only present but enhanced in the Puzzle Task (cf. Table 5). This is a significant finding if one considers that in the present study all conversations took place in a dual language mode and thus not only the code-switching but also the monolingual VOT tokens were produced with increased cross-language activation. For Spanish, on the other hand, no effect was found for the cognitive load manipulation possibly for similar reasons to those mentioned above: the speakers already had VOTs that were rather long for Spanish, making further lengthening undesirable (cf. Olson, to appear).

One general outcome of the study has been the greater effects observed on English VOT relative to Spanish: the cognitive load manipulation did not affect Spanish, while the observed shortening due to code-switching was very small in Spanish compared to

English (across tasks, the difference between code-switching and monolingual tokens for English was 6 ms and for Spanish 3 ms). A more general question then is why that would be the case. A possible explanation is that the difference reflects the fact that short-lag VOT, due to its small range, is not easily amenable to durational changes (cf. Olson, to appear). This is documented by studies with both monolinguals (Kessinger & Blumstein, 1997; Arvaniti, 1999) and bilinguals (Schmidt & Flege, 1996; Magloire & Green, 1999; Olson, to appear) in a variety of languages. This characteristic of short-lag VOT may well be the reason why effects on Spanish VOT have been inconsistent across studies (cf. Bullock et al., 2006; Olson, 2013; Balukas & Koops, 2014). Whether this explanation holds can only be determined by testing bilinguals speaking languages with similar VOT categories but different distributions, such as English (Lisker & Abramson, 1964) and Navajo (Cho & Ladefoged, 1999), both of which have long-lag VOT but with Navajo values being substantially longer.

Finally, we note that the present study examined a group of early bilinguals who are now dominant in their L2 and form a homogeneous population in that they were raised in the same area and were of similar age. The fact that our results have broad similarities with but are not identical to those of previous studies – such as Bullock et al. (2006), Antoniou et al. (2011), Olson (2013), Balukas and Koops (2014), Olson (to appear) inter alia – indicates that both age of acquisition and current language dominance play a part in code-switching speech production: both languages in the present study were affected by code-switching, with phonetic productions shortening in code-switching contexts for both the dominant L2 and non-dominant L1. Future work should thus consider not only age and order of acquisition but current dominance as well to ensure homogeneity in groups of bilinguals studied. Doing so will allow us to better compare results across studies, and understand how different factors affect bilingual speech production. Similarly, using spontaneous or semi-controlled conversational data, as is done here, will take into account the cooperative and interactive nature of code-switching; this is vital in obtaining an ecologically valid picture of this phenomenon. The present study, though small in size, can serve as an example of this practice.

5. Conclusions

The present study examined VOT in spontaneous code-switching speech elicited from a homogeneous group of English–Spanish early bilinguals. The results show that speakers had shorter VOT in code-switching contexts in both English and Spanish, though the effects were more pronounced for English (a difference that could indicate that general phonetic factors – in this instance the greater resistance of short-lag VOT to durational variability – may also play a part in determining phonetic parameters in bilingual speech). The shortening of English long-lag VOT found in code-switching contexts could suggest an effect of L1 on L2, while the shortening of Spanish short-lag VOT (instead of the expected lengthening) could possibly be due to sociolinguistic factors coupled with the relative large duration of Spanish VOT in the present data. Taken together, the results from English and Spanish indicate that the effects of code-switching may have distinct origins even when they produce the same outcome. Effects of cognitive load were also found, but only for English, with VOT being produced with longer durations when cognitive load was increased. The present results, which have broad similarities with those of previous studies, overall confirm that code-switching does affect the productions of early bilinguals in spontaneous speech and is not the result of artificial experimental tasks.

Acknowledgments

Thank you to Rosa Quezada for help with stimuli development. Thank you to Criccely Grijalva, Mireya Pinell-Cruz, and Rosalia Schultze-Kraft for help with the Spanish transcriptions. Finally, thank you to the members of the UCSD Phonetics Lab, the Phon Company, and three anonymous reviewers for advice and comments.

Appendix A

See Tables A1 and A2.

Table A1
Language profiles of speakers individually for age of acquisition, age of full fluency in understanding and speaking, and current exposure to both English and Spanish.

	Age of acquisition (yr)		Age when fluent (yr)		Current exposure	
	English	Spanish	English	Spanish	English (%)	Spanish (%)
Female 1	3	1	7	1	70	30
Female 2	1	1	1	1	80	19
Female 3	1	1	5	5	90	10
Female 4	9	0	15	2	50	50
Female 5	5	1	10	4	65	45
Female 6	4	2	6	4	80	20
Female 7	4	1	5	4	60	40
Female 8	4	1	6	6	80	20
Female 9	0	0	4	3	55	45
Female 10	1	1	4	4	60	40
Male 1	2	0	4	4	80	20
Male 2	14	0	16	4	49	50
Male 3	1	2	5	12	70	30

Table A2
Speakers' individually self-reported proficiency in speaking, understanding, and reading in English and Spanish. Ratings were on a Likert scale (0 – none to 10 – perfect).

	Speaking		Understanding		Reading	
	English	Spanish	English	Spanish	English	Spanish
Female 1	8	8	8	7	8	6
Female 2	8	8	8	8	8	4
Female 3	8	8	8	8	8	8
Female 4	5	8	8	8	7	7
Female 5	7	8	8	8	7	8
Female 6	9	8	9	9	9	8
Female 7	9	8	8	8	8	8
Female 8	7	7	8	9	8	8
Female 9	8	9	9	10	9	9
Female 10	8	8	8	8	9	8
Male 1	10	8	10	9	10	8
Male 2	8	10	8	9	8	8
Male 3	10	6	10	6	10	5

Appendix B

See Tables B1–B3.

Table B1
Percentage of utterances labeled code-switch and total percentage of speech produced in code-switching utterances by speaker.

	Conversation task		Puzzle task	
	Percentage utterances (%)	Percentage speech (%)	Percentage utterances (%)	Percentage speech (%)
Female 1	31.40	31.86	n.a.	n.a.
Female 2	31.99	36.04	49.01	55.61
Female 3	13.81	11.86	14.53	23.99
Female 4	19.85	17.12	22.22	26.83
Female 5	9.78	8.04	9.35	5.91
Female 5's partner	34.15	30.74	30.31	26.86
Female 6	9.01	6.14	37.23	44.38
Female 7	28.20	19.10	31.91	26.41
Female 8	19.71	17.76	17.36	11.96
Female 9	38.93	36.61	36.11	29.31
Female 10	32.68	26.91	29.63	22.74
Male 1	38.03	27.73	n.a.	n.a.
Male 2	31.53	26.71	20.00	19.93
Male 3	19.52	17.63	0.00	0.00

Table B2
Percentage of conversation filled with pauses, both within and between turns, by task, speaker, and conversation.

	Between speakers		Within speaker		Pooled	
	Conversation (%)	Puzzle (%)	Conversation (%)	Puzzle (%)	Conversation (%)	Puzzle (%)
F1 & M1	11.41	n.a	12.37	n.a	23.78	n.a
F2 & F3	15.84	8.29	18.17	16.13	34.01	24.42
F4 & M2	11.07	18.13	21.68	22.29	32.75	40.42
F5 & F5's partner	2.34	2.50	14.66	19.57	17.00	22.07
F6 & M3	16.19	13.66	20.40	26.71	36.59	40.37
F7 & F8	6.66	15.58	16.11	18.18	22.77	33.76

Table B3
Average number of disfluencies per minute by speaker and task.

	Conversation	Puzzle
Female 1	1.44	n.a.
Female 2	0.74	1.46
Female 3	1.21	1.46
Female 4	0.07	0.10
Female 5	1.00	0.51
Female 5's partner	1.14	1.52
Female 6	0.87	1.28
Female 7	1.46	0.98
Female 8	0.40	1.09

Table B3 (*continued*)

	Conversation	Puzzle
Female 9	1.28	0.48
Female 10	1.95	0.96
Male 1	1.23	n.a.
Male 2	1.00	0.89
Male 3	0.40	0.48

Appendix C

See Table C1.

Table C1
Mean VOT values by speaker, by language, by task, and by context.

		Conversation				Puzzle			
		English		Spanish		English		Spanish	
		ML	CS	ML	CS	ML	CS	ML	CS
Female 1	p	47	57	25	18	n.a.	n.a.	n.a.	n.a.
	t	55	48	29	16	n.a.	n.a.	n.a.	n.a.
	k	72	94	45	49	n.a.	n.a.	n.a.	n.a.
	Pooled means	64	80	37	32	n.a.	n.a.	n.a.	n.a.
Female 2	p	54	n.a.	22	20	44	44	22	21
	t	97	n.a.	24	21	66	57	22	27
	k	74	66	38	28	65	43	42	29
	Pooled means	81	66	29	25	59	53	31	27
Female 3	p	n.a.	46	37	13	72	n.a.	30	32
	t	40	50	36	36	65	n.a.	33	30
	k	76	33	35	31	92	90	38	46
	Pooled means	52	41	36	30	80	90	35	38
Female 4	p	72	n.a.	23	11	n.a.	n.a.	23	18
	t	44	n.a.	24	22	31	n.a.	17	24
	k	n.a.	n.a.	30	30	n.a.	n.a.	29	25
	Pooled means	58	n.a.	28	24	31	n.a.	26	20
Female 5	p	51	30	16	22	80	n.a.	20	6
	t	56	43	18	n.a.	88	n.a.	23	n.a.
	k	62	n.a.	33	33	51	n.a.	31	n.a.
	Pooled means	59	34	23	31	64	n.a.	26	6
Female 6	p	58	n.a.	15	12	73	n.a.	20	24
	t	69	n.a.	24	25	74	n.a.	n.a.	18
	k	52	39	28	22	61	n.a.	32	27
	Pooled means	58	39	24	20	68	n.a.	29	24
Female 7	p	59	32	40	24	50	n.a.	32	25
	t	66	54	46	29	32	66	36	50
	k	57	42	46	35	66	n.a.	32	32
	Pooled means	60	47	45	31	55	66	33	32
Female 8	p	85	n.a.	33	n.a.	65	n.a.	24	22
	t	68	50	26	n.a.	41	n.a.	24	15
	k	54	55	39	37	102	n.a.	35	32
	Pooled means	72	52	32	37	73	n.a.	29	28
Female 9	p	44	45	46	31	54	23	33	16
	t	55	35	24	23	64	46	n.a.	13
	k	51	60	33	33	58	40	24	23
	Pooled means	51	46	32	31	59	37	28	22
Female 10	p	36	118	n.a.	12	107	68	12	23
	t	51	63	22	16	78	65	n.a.	47
	k	58	57	32	30	65	71	24	22
	Pooled means	53	63	29	27	70	69	20	25
Male 1	p	53	58	n.a.	29	n.a.	n.a.	n.a.	n.a.
	t	63	69	52	21	n.a.	n.a.	n.a.	n.a.
	k	70	41	40	41	n.a.	n.a.	n.a.	n.a.
	Pooled means	67	58	41	36	n.a.	n.a.	n.a.	n.a.
Male 2	p	46	46	18	18	63	n.a.	16	n.a.
	t	72	70	18	25	71	n.a.	17	19
	k	55	39	30	28	69	78	31	28
	Pooled means	58	48	24	25	68	78	26	26

Table C1 (*continued*)

		Conversation				Puzzle			
		English		Spanish		English		Spanish	
		ML	CS	ML	CS	ML	CS	ML	CS
Male 3	p	52	n.a.	34	n.a.	50	n.a.	n.a.	n.a.
	t	59	n.a.	35	n.a.	75	n.a.	33	n.a.
	k	55	n.a.	39	53	61	n.a.	65	n.a.
	Pooled means	56	n.a.	37	53	64	n.a.	49	n.a.

References

Amengual, M. (2012). Interlingual influence in bilingual speech: Cognate status effect in a continuum of bilingualism. *Bilingualism: Language and Cognition*, *15*(3), 517–530.

Antoniou, M., Best, C. T., Tyler, M. D., & Kroos, C. (2011). Inter-language interference in VOT production by L2-dominant bilinguals: Asymmetries in phonetic code-switching. *Journal of Phonetics*, *39*(4), 558–570.

Arvaniti, A. (1999). Effects of speaking rate on the timing of single and geminate sonorants. In J. J. Ohala (Ed.), *Proceedings of the XIVth International Congress of Phonetic Sciences* (pp. 599-602). Berkeley: University of California.

Balukas, C., & Koops, C. (2014). Spanish–English bilingual voice onset time in spontaneous code-switching. *International Journal of Bilingualism*(January).

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2014). *_lme4: Linear mixed-effects models using Eigen and S4_. R package version 1.1-7*, ⟨http://CRAN.R-project.org/package=lme4⟩.

Bijeljac-Babic, R., Biardeau, A., & Grainger, J. (1997). Masked orthographic priming in bilingual word recognition. *Memory & Cognition*, *25*(4), 447–457.

Birdsong, D. (2001). On the evidence for maturational constraints in second-language acquisition. *Journal of Memory and Language*, *44*(2), 235–249.

Boersma, P., & Weenink, D. (2001). Praat, a system for doing phonetics by computer. *Glot International*, *5*(9), 341–345.

Belazi, H. M., Rubin, E. J., & Toribio, A. J. (1994). Code switching and X-Bar Theory: The functional head constraint. *Linguistic Inquiry*, *25*(2), 221–237.

Bullock, B. E., & Toribio, A. J. (Eds.). (2009). *Cambridge handbook of linguistic code-switching*. Cambridge: Cambridge University Press.

Bullock, B. E., Toribio, A. J., González, V., & Dalola, A. (2006). Language dominance and performance outcomes in bilingual pronunciation. In M. G. O'Brien, C. Shea, & J. Archibald (Eds.), *Proceedings of the 8th Generative Approaches to Second Language Acquisition Conference (GASLA 2006)* (pp. 9–16). Somerville: Cascadilla Proceedings Project.

Cho, T., & Ladefoged, P. (1999). Variations and universals in VOT: Evidence from 18 languages. *Journal of Phonetics*, *27*(2), 207–229.

Costa, A., Miozzo, M., & Caramazza, A. (1999). Lexical selection in bilinguals: Do words in the bilingual's two lexicons compete for selection?. *Journal of Memory and Language*, *41*(3), 365–397.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1992). The monolingual nature of speech by bilinguals. *Cognitive Psychology*, *24*(3), 381–410.

Dijkstra, T., & Van Hell, J. G. (2003). Testing the language mode hypothesis using trilinguals. *International Journal of Bilingual Education and Bilingualism*, *6*(1), 2–16.

Ehri, L. C., & Ryan, E. B. (1980). Performance of bilinguals in a picture-word interference task. *Journal of Psycholinguistic Research*, *9*(3), 285–302.

Flege, J. E. (1991). Age of learning affects the authenticity of voice onset time (VOT) in stop consonants produced in a second language. *Journal of the Acoustical Society of America*, *89*(1), 395–411.

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

Flege, J. E., Munro, M. J., & MacKay, I. R. A. (1995). Effects of age of second-language learning on the production of English consonants. *Speech Communication*, *16*(1), 1–26.

Genesee, F., Paradis, J., & Crago, M. B. (2004). *Dual language development and disorders: A handbook on bilingualism and second language learning*. Baltimore, MD: Brookes.

Gildersleeve-Neumann, C. E., & Wright, K. L. (2010). English speech acquisition in 3- to 5-year-old children learning Russian and English. *Language, Speech, and Hearing Services in Schools*, *41*(4), 429–444.

Goldrick, M., Runnqvist, E., & Costa, A. (2014). Language switching makes pronunciation less nativelike. *Psychological Science*, *25*(4), 1031–1036.

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

Grosjean F., The bilingual's language modes, In: J. Nicol One Mind, Two Languages: Bilingual Language Processing (Ed), 2001, Blackwell; Oxford, 1–22, (Also in Li Wei (Ed.). The bilingual reader (2nd ed.). London: Routledge, 2007).

Grosjean, F., & Miller, J. L. (1994). Going in and out of languages: An example of bilingual flexibility. *Psychological Science*, *5*(4), 201–206.

Gumperz, J. J. (1977). The sociolinguistic significance of conversational code-switching. *RELC Journal*, *8*(2), 1–34.

Hakuta, K., Bialystok, E., & Wiley, E. (2003). Critical evidence: A test of the Critical-Period Hypothesis for second-language acquisition. *Psychological Science*, *14*(1), 31–38.

Hamers, J. F., & Blanc, M. H. A. (2000). *Bilinguality and bilingualism*. Cambridge: Cambridge University Press.

Higgins, M. B., Netsell, R., & Schulte, L. (1998). Vowel related differences in laryngeal articulatory and phonatory function. *Journal of Speech, Language, and Hearing Research*, *41*(4), 712–724.

Howell, P., & Dworzynski, K. (2001). Strength of German accent under altered auditory feedback. *Perception & Psychophysics*, *63*(3), 501–513.

Keene, O. N. (1995). The log transformation is special. *Statistics in Medicine*, *14*(8), 811–819.

Kehoe, M. M., Lleó, C., & Rakow, M. (2004). Voice onset time in bilingual German-Spanish children. *Bilingualism: Language and Cognition*, *7*(1), 71–88.

Kessinger, R. H., & Blumstein, S. E. (1997). Effects of speaking rate on voice-onset time in Thai, French, and English. *Journal of Phonetics*, *25*(2), 143–168.

Khattab, G. (2002). VOT production in English and Arabic bilingual and monolingual children. In D. Parkinson, & E. Benmamoun (Eds.), *Perspectives on Arabic Linguistics XIII-XIV. Papers from the Thirteenth and Fourteenth Annual Symposia on Arabic Linguistics* (pp. 1–38). Amsterdam, Philadelphia: John Benjamins Publishing Co.

Khattab, G. (2009). Phonetic accommodation in children's code-switching. In B. E. Bullock, & A. J. Toribio (Eds.), *Cambridge Handbook of Linguistic Code-Switching* (pp. 142–159). Cambridge: Cambridge University Press.

Kim, M., Horton, W. S., & Bradlow, A. R. (2011). Phonetic convergence in spontaneous conversations as a function of interlocutor language distance. *Laboratory Phonology*, *2*(1), 125–156.

Klatt, D. H. (1975). Voice onset time, frication, and aspiration in word-initial consonants clusters. *Journal of Speech and Hearing Research*, *18*(4), 686–706.

Ladefoged, P. (2003). *Phonetic data analysis: An introduction to instrumental phonetic fieldwork*. Oxford: Blackwell.

Lee, S. A. S., & Iverson, G. K. (2012). Stop consonant productions of Korean–English bilingual children. *Bilingualism: Language and Cognition*, *15*(2), 275–287.

Lev-Ari, S., & Peperkamp, S. (2013). Low inhibitory skill leads to non-native perception and production in bilinguals' native language. *Journal of Phonetics*, *41*(5), 320–331.

Limpert, E., Stahel, W. A., & Abbt, M. (2001). Log-normal distributions across the sciences: Keys and clues. *BioScience*, *51*(5), 341–352.

Lisker, L., & Abramson, A. S. (1964). A dual language study of voicing in initial stops: Acoustical measurements. *Word*, *20*(3), 384–422.

López, V. G. (2012). Spanish and English word-initial voiceless stop production in code-switched vs. monolingual structures. *Second Language Research*, *28*(2), 243–263.

Magloire, J., & Green, K. P. (1999). A dual language comparison of speaking rate effects on the production of voice onset time in English and Spanish. *Phonetica*, *56*(3-4), 158–185.

Marian, V., Blumenfeld, H. K., & Kaushanskaya, M. (2007). The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals. *Journal of Speech, Language, and Hearing Research*, *50*(4), 940–967.

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

McLaughlin, B. (1978). *Second-language acquisition in childhood*. Mahway: Lawrence Erlbaum.

Mok, P. P. K. (2011). The acquisition of speech rhythm by three-year-old bilingual and monolingual children: Cantonese and English. *Bilingualism: Language and Cognition*, *14*(4), 458–472.

Myers-Scotton, C. (2008). Language contact: Why outsider system morphemes resist transfer. *Journal of Language Contact – TEMA*, *2*(1), 21–41.

Nagy, N., & Kochetov, A. (2013). VOT across the generations: A dual linguistic study of contact-induced change. In P. Siemund, I. Gogolin, M. Schulz, & J. Davydova (Eds.), *Multilingualism and Language Contact in Urban Areas: Acquisition – Development – Teaching – Communication* (pp. 19–38). Amsterdam: John Benjamins.

Neiman, G. S., Klich, R. J., & Shuey, E. M. (1983). Voice onset time in young and 70-year-old women. *Journal of Speech and Hearing Research*, *26*(1), 118–123.

Olson, D. J. (2013). Bilingual language switching and selection at the phonetic level: Asymmetrical transfer in VOT production. *Journal of Phonetics*, *41*(6), 407–420.

Olson, D. (to appear). The role of code-switching and language context in bilingual phonetic transfer. *Journal of the International Phonetic Association*.

Padilla, A. M., & Lindholm, K. (1984). Child bilingualism: The same old issues revisited. In J. Martinez, & R. Mendoza (Eds.), *Chicano Psychology* (pp. 369–408). Orlando: Academic Press.
Pardo, J. S. (2006). On phonetic convergence during conversational interaction. *Journal of Acoustic Society of America*, *119*(4), 2382–2393.
Perea, M., Duñabeitia, J. A., & Carreiras, M. (2008). Masked associative/semantic priming effects across languages with highly proficient bilinguals. *Journal of Memory and Language*, *58*(4), 916–930.
Pfaff, C. W. (1979). Constraints on language mixing: Intrasentential code-switching and borrowing in Spanish/English. *Language*, *55*(2), 291–318.
Piccinini, P., & Arvaniti, A. (2014). Accessing cross language categories in learning a third language. In W. Orman, & M. J. Valleau (Eds.), *Proceedings of the 38th Boston University Conference on Language Development* (pp. 342–354). Boston: Cascadilla Press.
Poplack, S. (1980). Sometimes I'll start a sentence in Spanish Y TERMINO EN ESPAÑOL: Toward a typology of code-switching. *Linguistics*, *18*(7-8), 581–618.
Poplack, S. (1987). Contrasting patterns of code-switching in two communities. In E. Wande, J. Anward, B. Nordberg, L. Steensland, & M. Thelander (Eds.), *Aspects of Multilingualism* (pp. 51–77). Borgstroms, Motala: Uppsala.
R Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing. Vienna, Austria. ⟨http://www.R-project.org/⟩.
Schmidt, A. M., & Flege, J. E. (1996). Speaking rate effects on stops produced by Spanish and English monolinguals and Spanish/English bilinguals. *Phonetica*, *53*(3), 162–179.
Simonet, M. (2014). Phonetic consequences of dynamic dual linguistic interference in proficient bilinguals. *Journal of Phonetics*, *43*(1), 26–37.
Soares, C., & Grosjean, F. (1984). Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory & Cognition*, *12*(4), 380–386.
Sundara, M., Polka, L., & Baum, S. (2006). Production of coronal stops by simultaneous bilingual adults. *Bilingualism: Language and Cognition*, *9*(1), 97–114.
Whiteside, S. P., Henry, L., & Dobbin, R. (2004). Sex differences in voice onset time: A developmental study of phonetic context effects in British English. *Journal of the Acoustical Society of America*, *116*(2), 1179–1183.
Woolford, E. (1983). Bilingual code-switching and syntactic theory. *Linguistic Inquiry*, *14*(3), 520–536.
Yao, Y. (2009). Understanding VOT variation in spontaneous speech. In *Proceedings of the 18th international congress of linguists (CIL XVIII)*. Seoul: Korea University.