**Effect of withholding phonetics cues to English-Spanish code-switching**

Jiawei Shao

**Introduction**

Code-switching (CS) is the linguistic phenomenon when more than one language is used in one utterance. In comparison to a monolingual discourse, it is reported in a line of studies that more complex processes are involved the production, recognition and comprehension of CS utterances: in Grainger and Beauvillain’s study (1987) performance costs was in their lexical decision task when bilingual participants are involved in switching languages when recognizing lexical items; Soares & Grosjean (1984) reported the bilingual speakers who can perform similar to monolinguals in monolingual context still showed a slower lexicon access in the bilingual speech; Olson’s study (2017) by using eye-tracking paradigm extended the line of production-oriented switch costs researches to auditory comprehension.

In Grainger and Beauvillain’s study (1987) they also suggested that the switch cost they found could be reduced or eliminated when the bilinguals were exposed to language-specific orthographic cues. In line with their study, Thomas & Allport (2000) also, reported switching cost for visual word recognition, however, they found that the orthographic cues weren’t effective for bilinguals to use to cope with switch cost: they claimed that the result in Grainer and Beauvillain (1987) might be caused by a missing control condition. Whether and how bilinguals make use of semantic, acoustic, morphologic and/or other cues to manage the potentially difficult processing task of recognizing code-switched utterances? Another line of recent researches made some great progress trying to answer this question.

Fricke, Kroll and Dussias (2016) report subtle shifts in voice onset time (VOT) before English-to-Spanish code-switches, while other studies reported opposite results suggesting that there’s no difference between phonetic productions in monolingual versus code-switching utterances (Grosjean & Miller, 1994). Furthermore, In Piccinini & Garellek’s study (2014), it is reported that the listeners may be able to rely on the anticipatory phonetic cues to mitigate the increased processing cost: it is reported in their study that, in either Spanish-English or English Spanish direction, there were subtle shift prior to code-switches, and bilingual listeners have access to these acoustic cues to help themselves anticipating and comprehending CS utterances. Phonetic cues can mitigate switch cost by playing as an anticipatory role in CS utterances. On the other hand, Shen et al. (2020) pointed out that, code-switching pronunciation (phonetic cues to upcoming code-switches) could potentially be another barrier for the listeners when recognizing and comprehending CS utterances due to preservative coarticulation of matrix language phonetics into the code-switch, and vice versa from the switched item into the matrix language, which could be detrimental to recognition.

By looking into how do pronunciation of matrix language and of switch items interact could provide us with a possible explanation to the former addressed conflict between the possible role of code-switching pronunciation in CS production and comprehension. Shen et al. (2020) summed up three possible mechanisms:

…*blending mechanism* by which code-switching pronunciation might represent a blend of the phonetic features of both languages (Grosjean, 2012; Olson, 2013) … *preparation mechanism* by which code-switching pronunciation might reflect articulatory gestures that are preparatory to the production of a specific code-switched target… A third possibility is that code-switching pronunciation might reflect *global cognitive costs* of code-switching.

As formerly mentioned, many previous studies have reported segmental properties going through alternation when produced in CS context in both matrix language and switched items (eg, VOT in Fricke, Kroll and Dussias, 2016), and there were reported suprasegmental features involved in similar phenomenon (intonation in Piccinini & Garellek’s study, 2014). Also, Olsen (2012) reported that insertional code-switched tokens are produced with a degree of hyper-articulation, evidenced by an increase in pitch height and duration. Furthermore, Olsen in his 2015 work also suggested that the suprasegmental realizations of code-switched tokens correspond to a degree of contextually driven predictability. Shen et al.(2020) showed that in English-Chinese code switching withholding acoustic cues can cause slower recognition of switched item, reflecting the possible predictability lies in tonal cues in matrix language as the latter language shows lexical tone while the other does not, and they concluded that bilingual listeners have access to phonetic cues in the matrix language, furthermore they discussed the implication of tonal aspect of the phonetic cues to code-switching based on their acoustic analysis.

The present study will be in line with Shen et al. (2020) study, examining the role of phonetic cues in auditory recognition of English-Spanish CS utterances. As Spanish and English are both intonational language, the tonal variation between these two languages is less compared with English-Chinese. As formerly mentioned, Piccinini & Garellek’s study (2014) reported intonational variation in English-Spanish CS and other studies reported VOT changes. Are listeners of English-Spanish CS utterance able to, either consciously or unconsciously, make use of the cues to cope with switch cost? We designed two experiments, a concept monitor experiment and an eye-tracking experiment, to answer these questions by comparing listeners reaction time to monolingual sentences, CS sentences and CS sentences without possible acoustic cues. Out hypothesis is that, in naturally produced English-Spanish utterances, there is anticipatory phonetic cues in the matrix language, which the listeners are able to (consciously or unconsciously) recognize and make use of so that the delay in recognizing switched items caused by switch cost will be shorter. And when the phonetic cues were eliminated, there will be an increase in reaction time.

**Methodology**

In order to our hypothesis that listeners have access to anticipatory phonetic cues in matrix language that can aid them to cope with switch costs when processing code-switches, we designed a concept monitoring experiment and an eye tracking experiment. The preparation of stimulus is involved in a procedure called splicing.

***Splicing***

The current study adopts splicing manipulation to test the predictions. When the speaker records auditory stimulus sentences, he/she will repeat the sentences in CS twice and unilingual English sentences once. Both the repeated CS sentences and unilingual English sentences are used to create spliced materials.

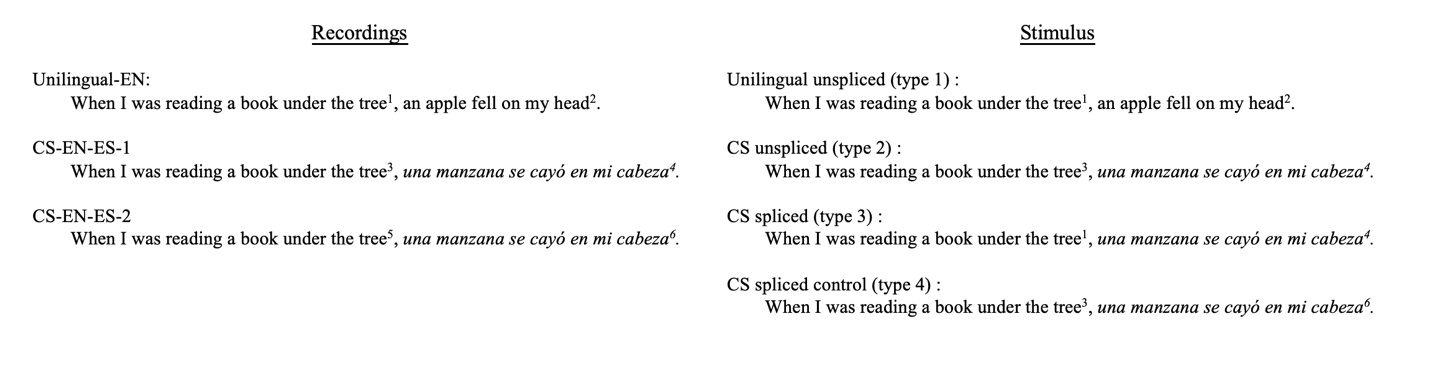
When preparing auditory materials, we spliced for both experiments: we spliced Spanish code-switched items from English-Spanish code-switched sentences (e.g., When I was reading a book under the tree, *una manzana se cayó en mi cabeza.)* into English sentences that were originally unilingual (e.g., When I was reading a book under the tree, an apple fell on my head.) to withhold any anticipatory phonetic cues to the code-switch. By doing so, bias will be created that the listeners may not be excepting the utterance that started in English will suddenly switch to Spanish, as all the potential acoustic cue were eliminated from the matrix language. We then will compare, respectively in experiment 1 and experiment 2, the reaction time to press the button indicating the recognition of the visual stimuli in the recording, and fixation time on the target item. We prepared four conditions for analysis: unilingual unspliced, CS unspliced, CS spliced and CS spliced control (figure 1).

Figure 1 - How we designed to making spliced and unspliced auditory stimulus using recordings.

***Experiment 1: concept monitoring***

The goal of this experiments is to test whether eliminating the anticipatory phonetic in the matrix language (English) can cause more reaction time for the listeners to recognize Spanish switched items. This goal is designed to be met by comparing reaction times to spliced and unsplised stimuli in a concept monitory experiment. In such experiment, the participant should be seated and they’ll be presented with pictured object. When they hear the object named in the auditory stimuli, they should press a button to show they’ve recognized the object.

***Participants***

A total of 40 Spanish heritage speakers (20 male, 20 female) were recruited for the present study. All of them are righthanded and showed no speech or hearing defect during the experiment. All participants but one completed both this experiment and Experiment 2. When recruiting, all of them self-reported as intermediate-advanced level Spanish speaker who have at least the level of B2 under DELE scheme. Each of the participants finished an adapted language history questionnaire based on LHQ3 (Li, Zhang, Yu, Zhao, 2020), an adapted DELE exam testing their reading comprehension, grammar, listening comprehension and speaking skills, and an adapted BSWQ (Rodriguez-Fornells et al., 2012) after the two experiments. The language proficiency test result shows that all the participants are as fluent as their claim when being recruited in comprehension and speaking tests, while 2 of them had a beginner-intermediate level of proficiency for grammar test. The participants all live in the same community where the most commonly used variate of Spanish is Mexican Spanish, they are all born in Mexico and moved to the US with their entire family before age of 8, and have been living in the US for at least 15 years. The average age was 20.4 years (SD = 2.2). All participants reported regularly code-switching with friends or family.

The participants’ language dominance is evaluated by Bilingual Language Profile (Birdsong et al., 2012) integrated in the LHQ3. The qualitative questions shows that all the participants tend to speak in Spanish within their families and English is more dominant in working and studying environments. The Bilingual Language Profile result shows that, they are all English dominant. BSWQ result shows that they are frequent code switchers that equally switch in either of the directions. The family and community they live in also is reported to be an environment where CS regularly happens.

***Visual stimuli***

Altogether 128 distinct pictures are included in the visual stimuli. They are from the Rossion and Pourtois (2004) colored line drawing database were, or from other public domain colored line drawings that visually resembled the Rossion and Pourtois (2004) set. All pictures depicted common objects or easily recognizable figures, and were modified to the same dimensions. All the 128 pictures are paired with an auditory stimulus, 64 were target experimental items, in that the pictured objects were mentioned in the corresponding auditory stimulus sentence, meanwhile the other 64 pictures are paired with filler sentences.

***Auditory stimuli***

A 35-year-old male Spanish–English bilingual volunteered to make the recording for all of the auditory stimuli. He self-reported being born in a Spanish-English bilingual family as his father is Mexican and his mother is American. He started acquiring both languages in an early age and has been conducting CS for a long time within his household as well as in his work in their community. The speaker also takes the three tests as the participants: LHQ, DELE and BSWQ, besides, his speaking, listening, reading, and writing proficiency of English were also proven with researchers. The result of tests should suggest that the speaker is a balanced bilingual with high proficiency in both languages, and also a balanced frequent switcher in both English - Spanish and Spanish - English directions. The speaker reads over stimulus and fillers before recording, to check for grammaticality, and to ensure familiarity with the sentences to avoid hesitations during recording, also he should check if the CS sentences were acceptable.

Auditory stimuli consisted of four passages telling the two different stories: for each story there will be two versions of recordings, one in unilingual English and the other in CS. Half of the formerly mention 64 stimuli will be in one story and the other half will be in the other story, in both unilingual version and CS version that pairs with visual stimulus; 64 sentences that functioned as fillers that either don’t involve CS (solely in English or Spanish) or in the opposite direction (Spanish - English) will also be divided into half and half for each story. The object of the experiment won’t be clarified to the speaker until he finishes all the recording.

Target sentences were constructed so that each mentioned a picturable noun that can be illustrated and easy to be recognized. All the target sentences are narratives with no strong emotional cues formed by two parts, the first half is an adverbial phrase indicating time, location, mood, etc., and the second half will start with a single countable noun introduced by either a definite article, indefinite article. Such similar syntactic is able to control for intonational patterns. Spliced versions of these 128 sentences were also constructed, as described in the Splicing section.

***Procedure***

Data collection took place in a sound-attenuated booth. Before the experiment starts, instructions would be given in the form of recording in both English and Spanish, informing the participants what they were supposed to do in the experiment. They were told to be pay attention the whole time and press the button once they hear the picture on the screen is mentioned. In the instruction section the participants are also listening to a self-presentation of the speaker as part of the same community, the self-presentation will be in English with some English-Spanish switched phrases. Also, the auditory stimuli will start with “I had this interesting dream last night. It was so real but some details were just so strange.”, so that the context can permit phrases with a wider semantic possibility, which will force the listeners to rely less on the prediction based on the contextual information.

One of the two stories will be presented to one participant. An experimenter was present to answer questions, as well as to clarify that: a) it’s not a logic test, the story may not make sense as it’s a monologue of an interesting dream the speaker had the other day, and b) participants, when hearing the word for the image presented in the screen, in either language, should press the button. Auditory stimuli were presented through headphones. During the experiments, participants saw a picture in the center of the computer screen, and as the story being told, listeners should press a button as soon as they heard the object mentioned in the sentence. The software was programmed as when a sentence contains a noun that paired with the picture and in 3000 ms the participant still doesn’t have any reaction, it will be skipped automatically

This experiment (concept monitoring) was counter-balanced with the next experiment (eye tracking); as the stimulus are presented in the context of a story, the order of the stimulus won’t be changed, but when and where a target sentences will show in a certain condition is randomized, while making sure that for each participant they’ll hear 8 sentences of each condition. Figure 2 shows the procedure of experiment 1.

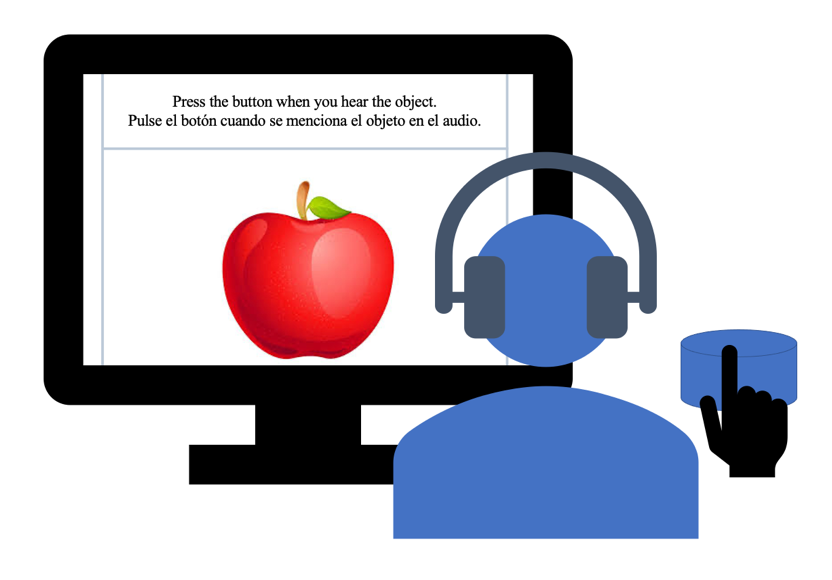


Figure 2 Setting of experiment 1

After completion of both experiments, participants can take a break and before getting the monetary compensation they should finish the three short tests: LHQ3, DELE and BSWQ. The participants are allowed to stay longer in the lab with the researchers to talk about the experiment and to give feedback on the experiment. The entire study lasted around 45 minutes, and participants were compensated $15 for the completion of all of the three components.

Reaction times will be measured from the onset of the target region (the article of the noun phrase) until the participants’ response was given by pressing the button. Data was trimmed to remove trials with reaction times that were under 200 ms(Shen et al., 2020) or longer than the trial duration. Additionally, trials with target words that participants noted as unfamiliar during the familiarization task were excluded.

***Experiment 2: Eye tracking***

Experiment 1 should unveil the switch cost in English-Spanish CS utterances as the reaction time for type 1 sentences will be shorter than type 2 sentences. In Shen et al’s study (2020) their concept monitoring experiment but failed to show an effect of the absence of anticipatory phonetic cues on concept monitoring times. In addition to this offline task, they also conducted an online task, eye tracking, to get some insight into the time course of recognition and whether and when phonetic cues are incorporated.

In experiment 2, in addition to splicing method, we will also use the visual world eye tracking to investigate, with a more precise tool, whether eliminating potentially anticipatory phonetic cues can cause delay in code-switched recognition. The participant will be listening to the other story they didn’t listened to in experiment 1. While listening to the story, they’ll have to choose one from four pictures that was mentioned in the story. In this way, the time course of lexical access is elucidated by the participant’s fixations to pictures during perception of that continuous speech.

***Participants***

All the participants who have finished experiment 1 should also finish experiments 2 after taking a 5-10 minutes of break. All the participants have normal or adjusted to normal visions. No color blind or any other defective vision is reported in the experiment.

***Visual stimuli***

For each trial there will be one picture paired with the target region and three other pictures that won’t be mentioned in either of the stories on the screen. All the visual style and category of the object/figure will be the same as formerly explained.

***Auditory stimuli***

The creation and standard of the auditory stimuli is the same as explained in the experiment 1, as in experiment 2 the audio recording in use for each participant is the one that hasn’t been used in experiment 1.

***Procedure***

A computer screen is mounted with a comfortable distance from the participants, meanwhile the space between the screen and the participant is enough for a desk mounted eye tracker (EyeLink2000), which was then calibrated with a nine-point calibration. Sampling frequency of the gaze location was 60 Hz.

The keyboard used in experiment 1 was removed and for this experiment a mouse is connected to the computer so that the participants can use it to choose the picture.

The rest of the settings are the same as in experiment 1. Figure 3 shows the setting of experiment 2.



Figure 3 Setting of experiment 2

The experiment starts with instructions in both English and Spanish. During the experiment, the participants should, like in experiment 1, listen to a story and choose the mentioned object/figure on the screen from four colored line drawings corresponding to four picturable nouns (target, and three distractors). One picture was centered in each of the four quadrants of the screen. The positions of the four types of pictured objects in the visual world display were randomized across the four fixed quadrant positions for each trial, so that the same type of picture (e.g., target) was not always presented in the same quadrant. When the participant makes a decision by clicking, no matter the result is correct or not, next screen of four picture will show. Same as in Experiment 1, if the participant don’t click until the end of the target phrase, the screen will automatically switch to the next screen

**Result prediction and discussion**

***Result of experiment 1***

We predict that the RT of type 1 and type 2 show significant difference, as the in comparison with RT-type-1, the type 2 are naturally produced CS sentences that can cause switch cost for the listeners when comprehending. RT-type-2 and RT-type-4 won’t show significant difference as the type 4 sentences were created using splicing without eliminating the phonetic cues. RT-type-2 and RT-type-4 will show a marginally significant differences, the RT-type-4 will be longer than RT-type-2 as the phonetic cue is absent.

***Result of experiment 2***

We predict that the fixation of type 1 and type 2 show significant difference, as the in comparison with type 1, then switch cost in type 2 can result in longer recognition time for listeners, thus reporting a shorter fixation time. Type 2 and type 4 won’t show significant difference in fixation time in line with the result in experiment 1. For type 2 and type 3 there will be a more significant difference in comparison with the same type pair in experiment 1 as the online method can provide with more precise data.

The results of the experiments proof the existence of switch cost in auditory comprehension even for the regular English-Spanish code switchers who are proficient in both languages. The results also may suggest the existence of acoustic cues in matrix language in English-Spanish CS utterances. Also, Bilinguals are faced with higher processing cost recognizing CS utterances and processing them when the phonetic cues were eliminated from the matrix language. Further researches are required to exam which of the segmental and/or suprasegmental properties of the matrix language is most salient in helping listeners to cope with switching costs.

**References**

Birdsong, D, Gertken, LM and Amengual, M (2012) Bilingual language profile: An easy-to-use instrument to assess bilingualism. COERLL, University of Texas at Austin. Retrieved from https://sites.la.utexas.edu/bilingual/

Fricke, M., Kroll, J. F., & Dussias, P. E. (2016). Phonetic variation in bilingual speech: A lens for studying the production–comprehension link. *Journal of memory and language*, *89*, 110-137.

Grainger, J., & Beauvillain, C. (1987). Language blocking and lexical access in bilinguals. *The Quarterly Journal of Experimental Psychology Section A*, *39*(2), 295-319.

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

Li, P., Zhang, F., Yu, A., & Zhao, X. (2020). Language History Questionnaire (LHQ3): An enhanced tool for assessing multilingual experience. *Bilingualism: Language and Cognition*, 23(5), 938-944.

Olson, D. J. (2012). The phonetics of insertional code-switching: Suprasegmental analysis and a case for hyper-articulation. *Linguistic Approaches to Bilingualism, 2*(4), 439-457.

Olson, D. J. (2016). The impact of code-switching, language context, and language dominance on suprasegmental phonetics: Evidence for the role of predictability. *International Journal of Bilingualism, 20*(4), 453–472.

Olson, D. J. (2017). Bilingual language switching costs in auditory comprehension. *Language, Cognition and Neuroscience*, *32*(4), 494-513.

Piccinini, P. E., & Garellek, M. (2014). Prosodic cues to monolingual versus code-switching sentences in English and Spanish. In *Proceedings of the 7th Speech Prosody Conference* (pp. 885-889).

Rodriguez-Fornells, A., Krmer, U. M., Lorenzo-Seva, U., Festman, J., & Mnte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology*, 2, 115.

Shen A, Gahl S, Johnson K (2020). Didn’t hear that coming: Effects of withholding phonetic cues to code-switching. *Bilingualism: Language and Cognition* 23, 1020–1031.

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.

Thomas, M. S., & Allport, A. (2000). Language switching costs in bilingual visual word recognition. *Journal of memory and language*, *43*(1), 44-66.