**The Effects of withholding acoustic cues to English-Spanish codeswitching in Wh-questions**

Jiawei Shao

Department of Spanish and Portuguese, Rutgers University

April 10, 2024

**Abstract**

Code-switching is the linguistic phenomenon in which more than one language is used in one utterance. A line of studies showed that there can be an additional cognitive cost for bilinguals in a code-switching context (Grainger & Beauvillain, 1987; Soares & Grosjean, 1984; Olson, 2017) and Shen et al. (2020) found that in English-Chinese code-switching, tonal cues can help mitigate switch cost. The present study aligns with Shen et al. (2020), examining the role of acoustic cues in auditory recognition of English-Spanish CS utterances and expands on the conclusions made in Shen et al. (2020) by gathering evidence from a different type of language pair. Due to the limited number of participants and the nature of the small size of effect of our topic, the evidence we found was not able to expand the former study’s conclusion, though it does raise new questions about the role of acoustic cues during bilingual speech perception, particularly with regard to the naturalness of (re)synthesized code-switching speech and the roles of different acoustic properties.

**Introduction**

Bilingual communication is a common practice in almost all the countries around the world and constitutes an important area in linguistic studies (Schmidt, 2014). Bilingual communications differ from monolingual ones in that bilinguals can alter their ways of communication according to the counterpart: when bilinguals are with monolinguals, they may avoid using their other language(s) while when they communicate with other bilinguals who share their language(s), they may change completely to the other language or bring elements of the other language(s) into the one already in use (Grosjean, 2013). This difference in using language(s) has been discussed by Grosjean (2013) as bilingual and monolingual language modes. Especially for bilinguals language mode is the state of activation of their languages (say language A and language B) and language processing mechanism at a given point of time, or a specific occasion of communication: the level of activation for language A and B along the language-mode continuum, as illustrated in his work (Figure 1). As a bilingual speaker moves along the language-mode continuum in different linguistic contexts, the amount of use of the other language (guest language), the amount and type of mixed language use, the ease of processing the two languages and the frequency of the base-language change will undergo changes accordingly. One common practice when a bilingual individual communicates in bilingual mode is switching from one language to another, and often in one utterance.

**Figure 1**

*A visual representation of the language-mode continuum*

Note. Two positions on the continuum are illustrated for a person with two languages (A and B): toward the monolingual end (on the left) and toward the bilingual end (on the right). The level of activation of a language (square) is depicted by the degree of darkness (black represents an active language and white an inactive or deactivated one). From “Bilingual and Monolingual Languages Modes,” by François Grosjean, 2013, *The Encyclopedia of Applied Linguistics,* Copyright 2015 by Blackwell Publishing Ltd, 489-493.

Diagram, schematic

Description automatically generated

**Code-switching Communications**

Code-switching (CS) is the linguistic phenomenon during which more than one language is used in one utterance. As a result of language contact, in US Latino communities the alternating use of Spanish and English in the same conversational event is quite common (Toribio, 2002). Toribio’s case studies (2002), combining linguistic form and function in specific social contexts, has shown that despite some low prestige associated with code alternation, the practice of code-switching is commonly used and even some use the alternation of English and Spanish in their speech to signal social identities. Below are some of the naturally produced code-switching sentences reported in Toribio’s study (2002):

1. *La madre de caperucita le da* a jar of honey.

Little Red Riding Hood’s mom gave her a jar of honey.

1. They both finish talking *y el lobo camina en otra dirección opuesta a la de caperucita.*

They both finish talking and the wolf walks in the opposite direction of Little Red Riding Hood.

These two examples were collected from the language user in a code-switching narrative writing task. A code-switching narrative retelling task was also conducted in their study. It was shown that the language user had a strong sensitivity to code-switching well-formedness; in other words, as noted by Shana Poplack (1980) among others, a prerequisite for successful code-switching is competence in the component languages. The code-switchers have access to linguistic knowledge in both languages at the same time, hence the cognitive process should be different from monolingual or unilingual communication.

In comparison to a monolingual (unilingual) discourse, it is reported in a line of studies that more complex processes are involved in recognition, production, and comprehension in a code-mixing or code-switching context.

In Grainger and Beauvillain’s (1987) study compared the effects of mixed- and pure-language lists on lexical decision times with English-French bilinguals through two experiments, and it is reported in one experiment that pure-language (unilingual) presentation received faster reaction time than mixed-language presentation; together with the second experiment they found that the effect of language alternation depends language-specific orthographic information: if the word is not language specific, it takes longer for participants to recognize.

Soares & Grosjean (1984) conducted a study comparing Portuguese-English bilingual speakers’ monolingual language mode versus bilingual mode through an on-line task with also English monolinguals as controlled group: the phoneme-triggered lexical decision task (Blank, 1980). Their results showed that the bilinguals’ response times to code-switched word targets in the bilingual mode were significantly slower compared to monolinguals, even though in monolingual tasks the two groups were identical. In addition, in detecting pseudowords, the bilinguals’ response times were longer than those in the monolingual group in both monolingual (unilingual) speech and bilingual speech. They therefore hypothesized that bilinguals search both lexicons when confronted with nonwords, even when in a totally monolingual mode.

Studies have found consistently a small reaction time delay as proof of switch cost in production and such cost can be modulated by both individual differences and contextual factors (Meuter & Allport, 1999; Costa & Santesteban, 2004; Olson, 2016). Amengual (2018) studied the acoustic realization of voiced lateral approximants in the Spanish and English of heritage Spanish speakers and L2 Spanish learners. One of implications of the results of their study is that different immigrant generations and late L2 Spanish learners produce a less target-like lateral when in bilingual mode, in that the non-dominant language shifts towards the dominant language. Olson (2017) used an eye-tracking paradigm and extended the line of production-oriented switch costs research to auditory comprehension.

**Cues to Switch**

As discussed so far, since there is a performance cost for language users when involved in code-switching communications, we can naturally ask the following question: is there information in the discourse/context to help bilinguals to activate the bilingual mode and cope with the higher performance cost and cognitive cost? Grainger and Beauvillain (1987) reported language-specific orthographic cues in language mixing context that could help bilinguals to recognize the alternation in languages. Thomas and Allport (2000) looked further into the orthographic cues by using non-words in their experiment; they found that orthographic cues alone are not sufficient to mitigate switching cost and they argue that visual word recognition is not language selective.

Beyond written language, 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, Piccinini & Garellek (2014) found that 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 anticipate and comprehend CS utterances. Phonetic cues can mitigate switch cost by playing 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.

Looking into how the pronunciation of the matrix language and that of the switch items interact could provide us with a possible explanation of conflict between the 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.

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 (2012) 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 the 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. 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. In Shen et al. (2020) the language pair under study consisted of a tonal language and a non-tonal language: Mandarin Chinese that has true tonal system (McCawley, 1978) including lexical tones while English doesn’t have a true tonal system. They found evidence in the linguistic property (tone) that plays a much more important role in one language than the other. The present study aims to extend their finding to another language pair that are both non-tonal languages: English – Spanish.

**Intonational Changes in English and Spanish Bilinguals**

In monolingual speech, wh-questions in English and Spanish show rather similar syntactic structure:

1. ¿Dónde compraste estos libros?

*Where buyPAST-2S these books*

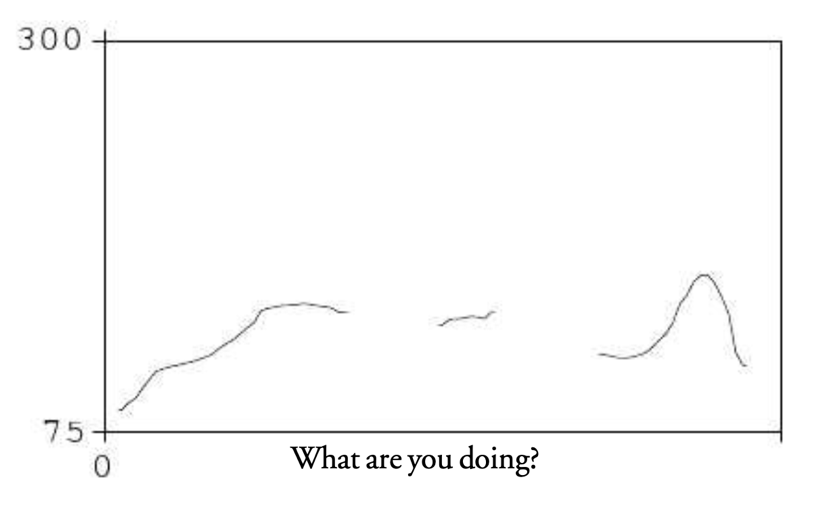
1. Where did you buy these books?

Even in sentences with similar syntactic structure, we can see a difference in pitch contours when comparing Figure 2 and Figure 3. Figure 2 showed pitch contours of wh-questions in American English (Minnesota), reported in Farías’ (2013) study. Figure 3 showed Mexican Spanish (Mexico City) wh-questions’ pitch contour. Even though the general tendencies are similar (rising-plateau-rising-fallying), we can see in that the final rising-falling peak is more drastic in Figure 2 (English).

**Figure 2**

*Example of the pitch contour of a wh-question produced by an English speaker reading a sentence in English*

Note. English wh-question produced by an English speaker. From “A Comparative Analysis of Intonation Between Spanish and English Speakers in Tag Questions, Wh-Questions, Inverted Questions, and Repetition Questions,” by Maria Gabriela Valenzuela Farías, 2013, *Revista Brasileira de Linguística Aplicada.*



**Figure 3**

*Example of the pitch contour of a wh-question produced by a Spanish speaker*

Note. Waveform, spectrogram and F0 trace for the information‐seeking wh‐ question ¿Y tú de qué pueblo vienes*? ‘And you, whereabouts are you from?’* produced with a L+H\* HL% nuclear configuration. From “Mexican Spanish Intonation” by Carme de‐la‐Mota, Pedro Martín Butragueño and Pilar Prieto, 2010, *Transcription of intonation of the Spanish language,* Copyright 2012 byLINCOM publisher.

Diagram

Description automatically generated

According to Spaai and Hermes (1993), pitch variations are essential components not only to distinguish the speaker's intention, but also to identify non-linguistic tasks such as emotions, social status, and personalities. Studies suggested that non-native speakers of English showed a hard time acquiring the intonational patterns in English: Farías (2013) explained that since Spanish is a language with a narrow variation in intonation, ESL learners may tend to transfer their pitch into English, having as a result a "flat" sound (Celce-Murcia et al., 1996). In addition, Bowen (1956) suggested that Spanish speakers reading utterances in English will negatively transfer the intonation patterns of their L1. For example, a Spanish emphatic sentence, such as "he does eat pasta" (él sí come pasta) to an English speaker is perceived as annoying. This annoyance is attributed to negative intonation transfers. The same effect occurs if an English speaker speaks the same sentences, transferring English intonation into Spanish utterances.

Farías (2013) tested the differences and similarities in intonation when producing tag questions, wh-questions, inverted questions, and repetition questions among native English speakers and ESL Spanish speakers. Her results showed that 100% of the participants who were native English speakers ended the questions with a falling contour, while more than half (66%) of the L1 Spanish speakers can produce such English sentences with the same falling contour. Spanish wh-questions produced by the participants had the tendency to end with rising intonation, as opposed to the falling contour given by the Spanish speakers to English sentences. She found another difference in producing wh-question in English and Spanish for the L1Spanish speakers in their pitch: the mean pitch is 180Hz when they are producing English target items and 143Hz in Spanish.

In Piccinini and Garellek’s work (2015), they specifically investigated listeners’ ability to anticipate code-switches in speech-in-noise and the role of prosodic cues in code-switching speech. They also ran an F0 analysis of the stimuli created by a 22-year-old early Spanish-English bilingual female who speaks both Mexican and Peninsular Spanish. Figure 4 demonstrates that her speech displays an interesting patter, where it seems that the code-switching utterances shows a more similar F0 contour to the embedded language rather than the matrix language, which is in line with the findings from Shen et al. (2020) where matrix language is going through some tonal changes and thus language users can use such information to anticipate code-switch and mitigate the switch cost.

**Figure 4**

*Normalized F0 contour for full sentence by context. Error bars indicate one standard error*

Note. From “Prosodic Cues to Monolingual versus Code-switching Sentences in English and Spanish” by Page Piccininio and Marc Garellek, 2014, *Proceedings of the 7th Speech Prosody Conference.*

Chart

Description automatically generated

**Present Study**

The present study seeks to expand the conclusions made in Shen et al. (2020) by gathering evidence from a different type of language pair: English-Spanish (intonational languages). More specifically, we are looking into late learners of Spanish who are native speakers of American English and their manner in recognizing and comprehending English-Spanish code-switching utterances.

The first objective of this study is to investigate if native speakers of English who are late learners of Spanish can make use of acoustic cues to cope with a higher cost on cognitive ability when a switch of language happens in a sentence that starts in English and ends in Spanish. More specifically, we are taking a further look at both the segmental level of the acoustics and the suprasegmental level using the resynthesized audios: we used the splicing method to manipulate the recording, forming code-switch sentences that are not naturally produced. The splicing method will be explained in later sections.

In accordance to our first objective, we proposed our Research Question 1: *Are English speaking learners of Spanish able to make use of acoustic cues to cope with switch cost in code-swtiching sentences? More specifically, when perceiving Wh-questions start in English and end in Spanish?* Previous studies have reported segmental properties going through alternation when produced in a CS context in both the matrix language and switched items (e.g., VOT in Fricke, Kroll and Dussias, 2016), and there were reported suprasegmental features involved in similar phenomenon (intonation in Piccinini & Garellek, 2014); Shen et al. (2020) showed that in English-Chinese code switching withholding acoustic cues can cause slower recognition of switched items. In line with Shen et al. (2020), we hypothesize that English speaking learners of Spanish can make use of acoustic cues to cope with switch cost in English-Spanish code switching Wh-questions.

The second objective of this study is to investigate how native speakers of English who are learners of Spanish make use of the segmental and suprasegmental levels of acoustics to mitigate the difficulty in a language mixing context. More specifically, we are testing if the acoustic cues are acting as anticipatory cues to upcoming switching.

In accordance to our second objective, we proposed our Research Question 2: *Do acoustic cues (segmental and suprasegmental) help mitigate switch cost by providing anticipatory information?* In line with previous studies, we hypothesize that our participants will make use of both types of acoustic cues (segmental and suprasegmental) to predict upcoming code switching, which will suggest the change of language mode, in other words, anticipatory activation of the other language.

We designed a two-experiment setting that will provide us with more evidence of whether listeners are able to use acoustic cues to mitigate the higher cognitive cost in a code-switching context in an anticipatory manner or not. In addition, we also designed a 4-version stimuli that would provide us with preliminary understanding in how these two levels of acoustics work together in a code-switching context.

**Methodology**

**Participants**

A total of 29 participants took part in the study. All participants were assigned to Group 1 or Group 2. After reviewing their language background and their experiment completion status, we discarded participants who reported having knowledge of a 3rd language, or didn’t complete the experiment, we ended up having 22 participants (7 males and 15 females) with valid data for our analysis.

Participants were between 18 and 26 years old, mostly righthanded (one lefthanded), educated (they had at least completed high school education). No participant reported or demonstrated having hearing or visual impairment that would have stoped them from completing all the tasks in the experiment. Participants were compensated with course credit or monetary reward.

Participants’ linguistic background information was collected using a language background questionnaire we compiled, covering language proficiency, age of acquisition of the second language, language use, language exposure, bilingual profile, and language proficiency in Spanish, etc. We adapted questions from Language History Questionnaire (LHQ) (Li et al., 2020), Bilingual Language Profile (BLP) (Gertken, Birdsong & Amengual, 2014) and Bilingual Switch Questionnaire (BSWQ) (Rodriguez-Fornells, Krämer, Lorenzo-Seva, Festman & Münte, 2012). At the end of the questionnaire, we also included a language proficiency test (Lextale-Esp, Izura & Brysbaert, 2014).

All the participants were English native speakers living in the United States who reported using English as their daily dominant language (percentage of the time using English: mean = 97.95, SD = 3.87), and, at the time of their participation, reported using and/or learning Spanish as their second language. In addition, the learners had an average proficiency score (Lextale) of 8.64 (SD = 13.33) points. The Participants were randomly divided in to two groups and finished one of the two versions of the tasks.

**Materials and Procedure**

Prior to the recruitment of participants, written materials were created first bearing in mind that they would be used in both experiments. Then visual and auditory materials were collected, created and modified in accordance to the written materials.

***Visual Stimuli***

There were in total 72 different images used in both experiments. Images were downloaded from open-source websites. Then every single image was adjusted to the resolution rate of 72\*72 with a white background. In PsychoPy (Pierce et al., 2019) images appeared in pairs in Size (0.5, 0.5) and Position (-0.5, 0) and (0.5, 0) respectively (Anchor center).

***Auditory Stimuli and Splicing***

The stimuli used in this study are Wh-question created by us. The target sentences are designed to be intrasentential code-switching Wh-questions that have English as matrix language and only the last lexical item in the sentence is switched to Spanish (e.g., Where is my *perro*?). To avoid conflict, all switched items are preceded by a possessive adjective in English. All the switched items in target sentences are designed to be easy to visualize in an illustration and are not culturally embedded in either Hispanic or English language. Additional sentences including Spanish unilingual, English unilingual, Spanish-English code-switching were also created for splicing and filler use. A 30-year-old female speaker of Mexican (Mexico City) Spanish recorded all the stimuli through Zoom (2016), recoded using the built-in function of Zoom (2016) and Voice Memos App on iPhone. The sentences that the speaker was supposed to record were presented one by one on her screen and she could repeat them as many times as needed to create a most naturalistic sound. All the recordings were manipulated using Praat (Boersma & Weenink, 2022).

Splicing and F0 manipulations are applied to the creation of auditory stimuli. Two conditions of the target sentences require such manipulation:

1. Spliced English-Spanish CS F0 not controlled, non-naturally produced code-switching (CS) sentences created by us without manipulation on intonation (F0);
2. Spliced English-Spanish CS F0 controlled non-naturally produced code-switching (CS) sentences created by us with manipulated intonation (F0).

To create the spliced English-Spanish CS F0 not controlled condition, we used two recordings: one English unilingual recording, and one English-Spanish code-switch (CS) recording. We replaced the lexical item in the same position in English unilingual sentence as the switched item in CS sentence with the word cut from CS sentence. In this way, we obtained one English-Spanish CS F0 not controlled sentence which in the matrix language maintained the acoustics from a unilingual context and the switched item from a CS context (Figure 5).

**Figure 5**

*Splicing procedure visualized 1*

Note. Splicing auditory stimuli. The double underline and wavy underline respectively represent the F0 contour of each sentence. In this case, the resynthesized sentence’s F0 isn’t manipulated, maintaining the original pattern from the recorded sentence of each part.

Diagram

Description automatically generated

To create the spliced English-Spanish CS F0 controlled condition, we repeated the same process as shown in Figure 5, and we added another step of F0 manipulation. More specifically, we altered the F0 contour using the pitch contour cloning function in Praat and changing the F0 of the whole matrix sentence to the one of English-Spanish CS, naturally produced (Figure 6). In this way we obtain one English-Spanish CS F0 controlled sentence which in the matrix language maintained the segmental properties from a unilingual context and the switched item from a CS context, while maintaining the suprasegmental property entirely from the naturally produced CS sentence.

**Figure 6**

*Splicing procedure visualized 2*

Note. Splicing auditory stimuli. The double underline and wavy underline respectively represent the F0 contour of each sentence. In this case, the resynthesized sentence’s F0 is manipulated: the double underlined part of the sentence goes through a resynthesized intonational change that clones the F0 contour of the matrix sentence from naturally produced CS, thus maintaining the segmental properties from English unilingual sentence yet having suprasegments from naturally produced CS sentence.

Diagram

Description automatically generated

Each list contains 20 target sentences and 35 fillers sentences. Four conditions of target sentences are included: natrually produced English unilingual (na.en), naturally produced English-Spanish CS (na.es), spliced English-Spanish CS without F0 manipulation (f0u.es), spliced English-Spanish CS with F0 manipulation (f0c.es).

***Procedure***

Participants finished 2 experiments individually in one session in the following order: consent procedure, Experiment 1 (concept monitoring), optional break, Experiment 2 (switch prediction), optional break, language background questionnaire. optional feedback. All the participants chose to finish all the steps in their L1.

**Experiment 1: Concept Monitoring.** Experiment 1 is a visual world paradigm task. In this task, participants were asked to sit down in front of a computer screen, paired with a keyboard, a mouse, and noise canceling headphones. During the experiment, participants were instructed to make a corresponding choice between two pictures as soon as they hear the audio playing in the headphone mention one of the two illustrated items using the keyboard. After the instructions, participants completed the practice trials and, when they felt confident to start the experiment, they could press a key to start. Each trial started with a 250-ms blank screen with a cross in the center. The visual stimulus appeared 1,250 ms before the onset of the auditory stimuli. Each trial ended when participants made a choice by pressing a key, and if no key was pressed, the program moved on to the next trial 4,000 ms after the onset of the auditory stimuli. The choice and the reaction time were collected. No break was given until the end of the Experiment. Group 1 got List 1 and Group 2 got List 2 in this experiment.

**Experiment 2: Switch Prediction.** Experiment 2 was a prediction task which aimed to overtly test participants’ ability to predict a language switch. The setting was the same as in Experiment 1. Participants saw two flags on screen, one of the US representing English and the other of Spain representing Spanish[[1]](#footnote-1). The participants were instructed to choose a language as soon as they could hear the language using the keyboard. Once they heard a switch in language, they had to switch their choice by pressing the other key. After the instructions, participants completed the practice trials and, when they felt confident to start the experiment, they could press a key to start. Each trial started with a blank screen with a cross in the center during 250-ms. The visual stimulus appeared 1,250 ms before the onset of the auditory stimuli. Each trial ended after 4,000 ms of the onset of the auditory stimuli. The choice and the reaction time were collected. No break was given until the end of this Experiment. Group 1 got List 2 and Group 2 got List 1 in this experiment.

After the two experiments, the participant could take another optional break and continue with the Language Background Questionnaire where the linguistic background and Spanish proficiency was registered.

**Results**

**Experiment 1**

The data collected during Experiment 1 were tidied in following steps: we left out participants who were not part of the target population according to their responses in the language background questionnaire, then we excluded incomplete data sets, we also calculated the rate of making the correct choice during this experiment and only included participants who scored above 75% in correction rate. We then calculated reaction time (rt) subtracting the duration of matrix sentence from the recorded reaction when participants pressed the key. We submitted the data to a linear mixed-effect model analysis to determine whether the participants reacted at differently to different conditions of stimulus. We trimmed down the data points per participant where rt is 2 standard deviations longer than their mean rt or more. Rection time data were log transformed to ensure normal distribution.

The reaction time data and analyzed using a linear mixed-effects model using R (R Core Team, 2022). Estimated log converted reaction time was the criterion with language proficiency (standardized Lextale score) as predictor, conditions (na.es, na.en, f0c.es, f0u.es) as covariate, and a by-participant random intercept. The condition group factor was dummy coded with condition group na.es (naturally produced code-switching condition) set as the reference group. Main effects and the language proficiency by condition group interaction were assessed using nested model comparisons. Experiment-wise alpha was set at 0.05.

No main effect of including proficiency was found (X2 = 1.233, p = 0.2669), nor was there a condition group by proficiency interaction (Χ2 = 0.925, p = 0.819). There was a main effect of condition group (Χ2= 67.326, p < 0.001). The model containing only the condition group factor provided the best fit of the data. Overall, reaction time decreased as a function of proficiency (estimate = -0.01733, t value = -0.947, p = 0.356), yet the effect was negligible. Reaction time changed among conditions. Specifically, holding proficiency constant and na.es as the reference group (estimate = 0.508, t value = 23.333, p < 0.001), na.en condition showed a decrease reaction time (estimate = -0.102, t value = - 5.459, p < 0.001) which indicates participants reacted at faster speed when hearing naturally produced unilingual English utterance. Meanwhile es.f0c condition showed an increase of reaction time with a smaller size of effect (estimate = 0.078, t value = 3.097, p < 0.01), which suggested that participants reacted slower in this condition. For the simple spliced condition (es.f0u), the reaction time increased by a small amount (estimate = 0.019, t value = 0.811, p = 0.418), but the change was not statistically significant.

**Figure 7**

*Experiment 1 plot*

Note: This plot shows data collected in experiment 1, Y-axis is Log transformed reaction time and X-axis consists of all four conditions: the naturally produced code-switching condition(na.es) as reference, naturally produced English unilingual condition(na.en), spliced code-switching with manipulated intonation(f0c.es) and spliced code-switching without manipulated intonation(f0u.es).

A graph of a graph showing the condition of a function

Description automatically generated

**Experiment 2**

The data collected during Experiment 2 were trimmed following similar steps as in Experiment 1: we left out participants whose data were not included in Experiment 1 analysis, then we excluded incomplete data sets. We did not include any incorrect trials: when no key was pressed, when the wrong key was pressed or when more key(s) than necessary were pressed. In Experiment 2 we collected two reaction times in the target items, reaction time 1 (rt1) when the participants pressed the first key choosing the first language they recognized in the utterance, and reaction time 2 (rt2) when they pressed the second key when they hear a language switch in the utterance. We also calculated the duration of each matrix sentence (durmat). We calculated switching reaction time (rtcs = rt2 – rt1) and matrix reaction time (rtmat = rt2 - durmat). The data was submitted to an analysis to test if participants were anticipating a language switch using the acoustic cues. We trimmed down the data points per participant where rt1 or rt2 is 2 standard deviations longer than their mean rt1 or rt2 or more. The two different kinds of rection time data were then log transformed to ensure normal distribution.

The switching reaction time (rtcs = rt2 – rt1) and matrix reaction time (rtmat = rt2 - durmat) data were both analyzed using a linear mixed-effects model using R (R Core Team, 2022). Estimated log converted matrix or switching reaction time was the criterion with language proficiency (standardized Lextale score) as predictor, conditions (na.es, f0c.es, f0u.es) as covariate, and a by-participant random intercept. The condition group factor was dummy coded with condition group na.es (naturally produced code-switching condition) set as the reference group. Main effects and the language proficiency by condition group interaction were assessed using nested model comparisons. Experiment-wise alpha was set at 0.05.

We firstly sent matrix reaction time data to analysis. There was no main effect of proficiency predictor (Χ2 = 2.233, p = 0.135), no main effect of including condition covariate (X2 = 3.728, p = 0.155) nor a condition group by proficiency interaction (Χ2 = 0.658, p = 0.720). By analyzing the model with both factors and their interaction, we found the following results. Holding proficiency constant, participant showed a slightly longer rtmat in es.f0c condition (estimate = 0.017, t value = 0.776, p = 0.438) yet they showed a slightly shorter rtmat in es.f0u condition (estimate = -0.042, t value = - 1.626, p = 0.105). Matrix reaction time decreased slowly as a function of proficiency (estimate = -0.062, t value = -1.621, p = 0.121) but this tendency was modulated by condition group: es.f0c condition showed an additional increase in rtmat (estimate = 0.011, t value = 0.605, p = 0.545) while es.f0u condition showed an additional decrease in rtmat (estimate = -0.018, t value = -0.462, p = 0.645). However, none of these increases nor decreases has a salient size of effect, indicating that the changes were minor and had no statistical significance.

We then sent switching reaction time to similar analysis. There was a main effect of condition group although the size of effect was small (Χ2 = 7.265, p = 0.026), no main effect of including proficiency (X2 = 0.255, p = 0.614) nor condition group by proficiency interaction (Χ2 = 1.806, p = 0.405). The model containing proficiency as a continuous predictor and condition as a categorical covariate provided the best fit of the data. Overall, reaction time increased as a function of proficiency (estimate = 0.031, t value = 0.537, p = 0.598) but this change was minor. In es.f0c condition we observed a decreased reaction time (estimate = -0.139, t value = -2.352, p = 0.019) which indicates participants reacted at faster speed in this condition, even though the size of effect was small. In es.f0u condition, we also observed a decreased reaction time (estimate = -0.089, t value = -1.481, p = 0.140), but it has no statistical significance.

**Figure 8**

*Experiment 2 plot matrix reaction time*

Note: This plot shows data collected in experiment 2, Y-axis is Log transformed matrix reaction time and X-axis consists of three conditions: the naturally produced code-switching condition(na.es) as reference, spliced code-switching with manipulated intonation(f0c.es) and spliced code-switching without manipulated intonation(f0u.es).

A graph of a function of conditions

Description automatically generated

**Figure 9**

*Experiment 2 plot switching reaction time*

Note: This plot shows data collected in experiment 2, Y-axis is Log transformed *switching* reaction time and X-axis consists of three conditions: the naturally produced code-switching condition(na.es) as reference, spliced code-switching with manipulated intonation(f0c.es) and spliced code-switching without manipulated intonation(f0u.es).

**~~A graph of a function of conditions

Description automatically generated~~**

**Discussion**

Evidence was found to support the theory of switch cost: in Experiment 1 main effect was found when including condition factor, while no main effect was found when taking proficiency or proficiency-condition interaction. More specifically, a salient difference of reaction time was reported when comparing na.en group and na.es group, suggesting that indeed participants reacted at a slower pace when hearing an English-Spanish code-switching utterance than in English unilingual than. This result is in line with what was reported in Olson’s (2017) study.

We got mixed results when trying to answer our research questions of whether the acoustic cues would help mitigate switch cost, and whether participants were predicting based on the acoustic cues.

The reported data and analysis for Experiment 1 led us to a result that’s partially different from what we predicted. The estimate is below zero when we in other words, as predicted, participants’ rt lowers as they show more advancement in Spanish language; however, the size of effect was negligible. A similar unexpected result reported in the f0u.es condition: as predicted, participants reacted at a slower pace in f0u.es condition as the splicing procedure eliminated most of the possible acoustic cues to switch. This finding is consistent with finding of English-Mandarin (Shen et, al., 2020), yet in their study the spliced condition showed significantly longer rt, while in our study, the difference was not statistically significant. This expected result might come from the fact that our study didn’t manage to recruit a large number of participants (N = 22) when in Shen’s study they recruited in total 42 participants. Shen’s study recruited Mandarin Chinese speakers in the University of California Berkeley with less restrictions on participants’ multilingual profile, less restrictions on participants’ language proficiency as well. In addition, the language pair (English – Spanish) we are investigating in the current study shows less difference than their language pair (English – Mandarin Chinese). As Figure 4 shows, the intonational contour has a very small difference in English and Spanish sentences that has a similar syntactic structure. This could lead to that the rt change in our study tend to be smaller, and in order to test the difference, we need even more participants than Shen’s study. Also due to the design of the materials, we are analyzing between subject factors, which results in less power (Charness et, al., 2012) thus demands for an even larger pool of participants. The small size of effect of the finding in Experiment 2 can also suggest that we are in need for more participants.

The data and analysis for the f0c.es condition in Experiment 1 revealed unexpected result as well, in addition to the same possible cause, it brings up new challenges. In this condition we resynthesized recordings to adjust their intonation resulting in a more similar F0 contour in matrix sentence to naturally produced CS sentences. However, we found a positive small effect for this condition: participants were reacting slower in this condition. No participants have reported that any of the recordings sounded unnatural or weird, which indicates that the participants weren’t explicitly aware of the manipulation that the recordings received, however, during this fast-paced online experiment, acoustic information could still be perceived and affect participants behavior. In our case, even carefully handled, the resynthesis can introduce features and traits that can sound unnatural to listeners, and without being able to explicitly perceive these traits, L2 learners can still be affected: they did not hear it, but they sensed it. Our intonational resynthesis was conducted solely without any segmental changes and it can affect the perception of naturalness of the speech: the perception of naturalness of synthetic speech lies on both segmental quality and intonation, and these two levels are mutually dependent (Vaino et, al., 2002). The tendencies shown in Experiment 2 rtmat data can support this explanation: participants tend to react faster than na.es condition (mean = 1.57) in f0c.es condition (mean = 1.49), while slower in f0u.es (mean = 1.63). These two conditions went through the same splicing process, resulting in the same level of unnaturalness in both segmental and suprasegmental levels; f0c.es then got one more step of resynthesis, the f0 manipulation, resulting in an unbalanced change in these two levels, which could cause the opposite effect to participants’ behavior.

**Conclusion**

The present study intended to exam whether English native speakers who are actively using and/or learning Spanish could perceive the acoustic cues in the matrix sentence of English – Spanish code-switching utterances and predict code-switching based on the acoustic cues, especially the intonational cues. The two-experiment design of the study failed to find any salient evidence to support our hypothesis based on the previous studies that our participant would be able to perceive the acoustic cues and predict an upcoming code-switch, although more evidence was found to support the existence of switch cost. We discussed possible explanations of such an outcome. Firstly, our study tries to observe on a relatively small size of effect, which requires a larger pool of participants, yet our study only managed to recruit a relatively small number of qualified participants who completed the study. Also, the between subject design of the study could avoid confound yet again demands even a larger number of participants due to its relatively smaller power.

In addition, we discussed on a third possible reason of such outcome that the naturalness of resynthesized auditory stimulus used in the experiment could have an impact on the results. This presents new challenges for future studies that aim to examine code-switching phonetics and phonology using resynthesis methodology: as shown in former studies on synthetic speech, naturalness is a multidimensional and attempts were made to separate contributions of different properties (Nusbaum et, al., 1997), the segmental properties were found to be dependent on the suprasegmental properties and the other way round as well (Vaino et, al., 2002). This opens new questions for future studies to investigate into the naturalness of (re)synthesized code-switching utterances, as well as the roles of different acoustic properties in code-switching speech.

**References**

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

Beardsmore, H.B. (1986). *Bilingualism: Basic principles* (Vol.1), England: Multilingual Matters.

Blank, M. A. (1980). Measuring lexical access during sentence processing. *Perception & Psychophysics, 28*(1), 1-8.

Birdsong, D., Gertken, L.M., & Amengual, M. *Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism.* COERLL, University of Texas at Austin. Web. 20 Jan. 2012. <https://sites.la.utexas.edu/bilingual/>.

Boersma, P & Weenink, D. (2022). Praat: doing phonetics by computer [Computer program]. Version 6.3.03, retrieved 17 December 2022 from http://www.praat.org/

Bowen, J. D. (1956). A comparison of the intonation patterns of English and Spanish. *Hispania*, *39*(1), 30-35.

Celce-Murcia, M., Brinton, D. M., & Goodwin, J. M. (1996). *Teaching pronunciation: A reference for teachers of English to speakers of other languages*. Cambridge University Press.

Charness, G., Gneezy, U., & Kuhn, M. A. (2012). Experimental methods: Between-subject and within-subject design. *Journal of economic behavior & organization*, *81*(1), 1-8.

Costa, A., & Santesteban, M. (2004). Lexical access in bilingual speech production: Evidence from language switching in highly proficient bilinguals and L2 learners. *Journal of memory and Language*, *50*(4), 491-511.

Farías, M. G. V. (2013). A comparative analysis of intonation between Spanish and English speakers in tag questions, wh-questions, inverted questions, and repetition questions. *Revista Brasileira de Linguística Aplicada*, *13*, 1061-1083.

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.

Gertken, L. M., Amengual, M., & Birdsong, D. (2014). Assessing language dominance with the bilingual language profile. *Measuring L2 proficiency: Perspectives from SLA*, *208*, 225.

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.

Grosjean, F. (2013). Bilingual and monolingual language modes. *The encyclopedia of applied linguistics*, 489-493.

Izura, C., Cuetos, F., & Brysbaert, M. (2014). Lextale-Esp: A test to rapidly and efficiently assess the Spanish vocabulary size. *Psicológica*, *35*(1), 49-66.

Li, P., Zhang, F. A. N., Tsai, E., & Puls, B. (2014). Language history questionnaire (LHQ 2.0): A new dynamic web-based research tool. *Bilingualism: Language and Cognition*, *17*(3), 673-680.

McCawley, J. D. (1978). What is a tone language?. In *Tone* (pp. 113-131). Academic Press.

Meuter, R. F., & Allport, A. (1999). Bilingual language switching in naming: Asymmetrical costs of language selection. *Journal of memory and language*, *40*(1), 25-40.

Nusbaum, H.C., Francis, A.L. & Henly, A.S. Measuring the naturalness of synthetic speech. *Int J Speech Technol* **2**, 7–19 (1997). https://doi.org/10.1007/BF02215800

Olson, D. (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 gradient effect of context on language switching and lexical access in bilingual production. *Applied Psycholinguistics*, *37*(3), 725-756.

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

Peirce, J., Gray, J. R., Simpson, S., MacAskill, M., Höchenberger, R., Sogo, H., ... & Lindeløv, J. K. (2019). PsychoPy2: Experiments in behavior made easy. *Behavior research methods*, *51*, 195-203.

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).

Poplack, S. (1980). Sometimes i’ll start a sentence in spanish y termino en espanol: toward a typology of code-switching1.

R Core Team (2022). “ R: A language and environment for statistical computing,” R Foundation for Statistical Computing, Vienna, Austria, https://www.R-project.org/ (Last viewed 3/28/2023).

Rodriguez-Fornells, A., Krämer, U. M., Lorenzo-Seva, U., Festman, J., & Münte, T. F. (2012). Self-assessment of individual differences in language switching. *Frontiers in Psychology, 2*, 388.

Schmidt, A. (2014). *Between the languages: Code-switching in bilingual communication*. Anchor Academic Publishing (aap\_verlag).

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*(5), 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.

Spaai, G. W., & Hermes, D. J. (1993). A visual display for the teaching of intonation. *Calico Journal*, 19-30.

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

Toribio, A. J. (2002). Spanish-English code-switching among US Latinos.

Vainio, M., Jarvikivi, J., Werner, S., Volk, N., & Valikangas, J. (2002, September). Effect of prosodic naturalness on segmental acceptability in synthetic speech. In *Proceedings of 2002 IEEE Workshop on Speech Synthesis, 2002.* (pp. 143-146). IEEE.

Zoom Video Communications Inc. (2016). Security guide. ZoomVideo Communications Inc. Retrieved from https://d24cgw3uvb9a9h.cloudfront.net/static/81625/doc/Zoom-Security-White-Paper.pdf8

1. The choice of flags is made regarding the country of residence of the participants and other philological considerations in order to avoid any possible extra cognitive cost during this online-task based experiment. [↑](#footnote-ref-1)