Didn't hear that coming: Effects of withholding phonetic cues to code-switching

Alice Shen, Susanne Gahl and Keith Johnson

University of California Berkeley

Research Article

Abstract

Code-switching has been found to incur a processing cost in auditory comprehension. However, listeners may have access to anticipatory phonetic cues to code-switches (Piccinini & Garellek, 2014; Fricke et al., 2016), thus mitigating switch cost. We investigated effects of withholding anticipatory phonetic cues on code-switched word recognition by splicing English-to-Mandarin code-switches into unilingual English sentences. In a concept monitoring experiment, Mandarin–English bilinguals took longer to recognize code-switches, suggesting a switch cost. In an eye tracking experiment, the average proportion of all participants' looks to pictures corresponding to sentence-medial code-switches decreased when cues were withheld. Acoustic analysis of stimuli revealed tone-specific pitch contours before English-to-Mandarin code-switches, consistent with previous work on tonal coarticulation. We conclude that withholding anticipatory phonetic cues can negatively affect code-switched recognition: therefore, bilingual listeners use phonetic cues in processing code-switches under normal conditions. We discuss the implications of tonal coarticulation for mechanisms underlying phonetic cues to code-switching.

Introduction

Bilinguals frequently switch between languages mid-utterance. Many psycholinguistic studies on code-switching have reported a 'switch cost', i.e., an increased processing difficulty, in production (Meuter & Allport, 1999; Thomas & Allport, 2000; Costa & Santesteban, 2004; Gollan & Ferreira, 2009, although see Kleinman & Gollan, 2016), recognition (Soares & Grosjean, 1984), and comprehension (Olson, 2017). How then do bilingual listeners manage the potentially difficult processing task of recognizing a code-switched word? A recent line of research points to subtle details of pronunciation as a possible key to this question.

For instance, Fricke, Kroll and Dussias (2016) report subtle shifts in voice onset time (VOT) before an English-to-Spanish code-switch, while Piccinini and Garellek (2014) report subtle shifts in intonation prior to code-switches in either direction. They further found that bilingual listeners use shifts in VOT and intonation as cues to anticipate code-switches. Phonetic cues to upcoming code-switches (henceforth 'code-switching pronunciation') may thus mitigate switch cost.

On the other hand, code-switching pronunciation could potentially make the comprehension process more difficult: perseverative coarticulation of matrix language phonetics into the code-switch – or indeed of the switch language back into the matrix language – might be detrimental to recognition.

There are at least three possible mechanisms by which code-switching pronunciation might arise. One is a 'blending' mechanism by which code-switching pronunciation might represent a blend of the phonetic features of both languages (Grosjean, 2012; Olson, 2013): the matrix language may come to sound more like the switch language, or vice versa. For example, Piccinini and Garellek (2014) observed that stressed syllable pitch patterns in Spanish/English code-switched contexts were intermediate between those observed in unilingual contexts in either language. If such 'intermediate' prosodic contours are characteristic of utterances containing code-switches, they could serve as cues to an upcoming code-switch.

Another possibility is a 'preparation' mechanism by which code-switching pronunciation might reflect articulatory gestures that are preparatory to the production of a specific code-switched target.

These two explanations are mutually compatible, but entail slightly different empirical predictions. Under blending, code-switching pronunciation would be independent of specific upcoming code-switched targets. Under preparation, by contrast, the acoustic consequences of speakers preparing code-switched targets would depend on the articulatory gestures needed to prepare a specific target. Of course, code-switched utterances might very well be characterized both by general code-switching pronunciation patterns, such as the prosodic contours found in Piccinini and Garellek's (2014) study, as well as by context-specific pronunciations arising in preparation for a specific code-switching target.

Sentence frame 1:   We saw the [cat] in the tree.

Sentence frame 2:   We saw the [cat] in the tree.   We saw the [猫] in the tree.
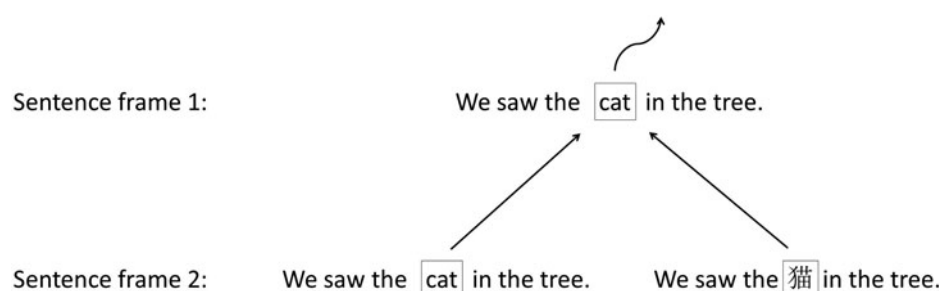
Fig. 1. Splicing auditory stimuli. The speaker recorded two sentence frames per experimental item: unilingual English sentences were recorded twice, and code-switched sentences were additionally recorded as unilingual English sentences. Target words were then cut from the unilingual or code-switched sentence frame and spliced into the fully English sentence frame.

A third possibility is that code-switching pronunciation might reflect global cognitive costs of code-switching: if code-switching incurs a processing cost for the speaker, that increased processing load might cause an overall slowed speaking rate, for example. Under this scenario, the existence and degree of 'code-switching pronunciation' would depend on the degree of processing load (see e.g., Gollan, Kleinman & Wierenga, 2014, for evidence showing that code-switching does not necessarily or consistently entail a processing cost in production). We will not pursue this possibility further here, except to note that it is in principle compatible with both the blending and preparation scenarios: code-switching pronunciation may be a variable phenomenon modulated by processing demands of a specific code-switching context.

Phonetic consequences of code-switching may also differ across language pairs. The literature on phonetic reflexes of code-switching has so far been limited to English–Spanish, English–French, and English–Greek code-switching. One goal of the current study is to widen the evidence-base on the possible role in comprehension of phonetic reflexes of code-switching, by examining English-Mandarin code-switches.

We hypothesized that the comprehension of code-switched targets would differ depending on whether code-switched targets were spliced into utterances that were originally unilingual vs. utterances that originally contained code-switches. If that is the case, it would strongly suggest that there must be phonetic differences between unilingual vs. code-switched utterances, and that listeners use these differences as cues to upcoming code-switches. In other words, if it is true that bilingual speakers produce phonetic cues and listeners use them in comprehension, then manipulating the acoustic signal to remove those cues should impede recognition of the code-switch: if phonetic preparation acts as a ramp to ease the gradual transition to another language or to highlight the phonetic contrast between the languages, then removing the phonetic 'ramp' should make code-switches phonetically abrupt and difficult to anticipate.

While the current study was primarily designed to target the possible role of phonetic reflexes of code-switches on the comprehension process, we also analyzed the pitch contours of our stimuli, as a step towards pinpointing what acoustic events might be responsible for effects of the splicing manipulation on comprehension and to explore whether phonetic cues to code-switching were target-specific.

We focus on pitch contours because Mandarin has lexical tone while English does not. It is conceivable therefore that pitch patterns in English contexts preceding switches into Mandarin might reflect tonal properties of the Mandarin target. For example, pitch might dip in anticipation of a low tone, such that there is assimilatory anticipatory coarticulation with pitch ramping to meet the low onset of that low tone.

Tonal coarticulation has been observed in unilingual Mandarin speech (Xu, 1997), which we describe in detail in the Acoustic Analysis section later on. English-to-Mandarin code-switching pronunciation might result in patterns resembling patterns of unilingual Mandarin tonal coarticulation. Alternatively, English-to-Mandarin code-switching pronunciation might differ from unilingual Mandarin tonal coarticulation: English does not have lexical tone, so pitch contours can in principle vary more freely in English than in Mandarin.

Tone-specific patterns are expected under the 'preparation' explanation for code-switching pronunciation, but not under the 'blending' explanation. Exploring these patterns can aid us in understanding the potential role of anticipatory coarticulation in code-switching pronunciation.

To test the hypothesis that anticipatory phonetic cues aid in processing code-switches, we conducted a concept monitoring experiment and an eye tracking experiment.

For both experiments, we spliced Mandarin code-switched target words from English-Mandarin code-switched sentences (e.g., *I saw a màozi*) into English sentences that were originally unilingual (e.g., *I saw a hat*) to withhold any anticipatory phonetic cues to the code-switch. The resulting spliced stimulus should bias the listener toward expecting the utterance to continue in English, as code-switch cues are absent. We compared listeners' reaction times and proportions of looks to English and Mandarin targets spliced into English utterances that originally did vs. did not contain Mandarin targets, as illustrated in Figure 1. This resulted in four conditions: code-switched spliced, code-switched unspliced, unilingual spliced, and unilingual unspliced.

Our prediction was that listeners will take longer to recognize code-switched target words, especially when spliced into unilingual utterances, since there would be no code-switching pronunciation to cue listeners to the upcoming code-switch.

Experiment 1: concept monitoring

This experiment tests whether listeners are slower to recognize Mandarin target words in English sentences if anticipatory phonetic cues to the code-switch are absent from the acoustic signal. This is tested by comparing reaction times to spliced and unspliced stimuli, in a concept monitoring experiment where participants see a pictured object and press a button when they hear the object named in an auditorily presented sentence. Spliced code-switched stimuli consist of a Mandarin target word spliced into an originally unilingual English utterance, so that the pronunciation in the portion of the utterance leading up to the target word will incorrectly bias the listener toward expecting an English target word. Unspliced code-switched stimuli consist of an originally code-switched English sentence with a Mandarin target word, so that the code-switching pronunciation leading up to the

Table 1. Speaker self-rated proficiency.

	English	Mandarin
Speaking	5	6
Understanding	6	6
Reading	5	6
Writing	5	4

Table 2. Mean participant age of acquisition and self-rated proficiency. Standard deviations are in parentheses.

	English	Mandarin
Age of acquisition (years)	5.4 (2.7)	1.2 (0.6)
Speaking	5.12 (0.89)	5.6 (0.73)
Understanding	5.36 (0.76)	5.7 (0.6)
Reading	5.3 (0.82)	5.5 (1.1)
Writing	5 (1)	4.98 (1.55)

code-switched target word might aid in recognition of the code-switch. The prediction is that listeners will be slower to recognize the target when the phonetic information available is incongruent with the code-switch, so reaction times to spliced code-switched stimuli will be slower than to unspliced code-switched stimuli.

Method

Speaker

A 21-year-old female Mandarin–English bilingual produced all of the auditory stimuli. She self-reported balanced usage of both languages in home and school environments, having acquired Mandarin from birth and English around age four. The speaker completed a written language background questionnaire asking for speaking, listening, reading, and writing proficiency self-ratings in both languages. She rated herself as proficient in English and Mandarin on a scale of 0–6, with 0 being low and 6 being high, as shown in Table 1. The speaker read over the list of stimuli before recording, to check for grammaticality, and to ensure familiarity with the sentences to avoid hesitations during recording. The speaker was also administered the Bilingual Language Profile (Birdsong, Gertken & Amengual, 2012), on which she scored −23 on a scale from −218 (very Mandarin-dominant) to 218 (very English-dominant), suggesting that she is a relatively balanced bilingual, though slightly more dominant in Mandarin. In addition, she reported having a positive attitude toward code-switching, frequently code-switching with friends, and occasionally with family.

Participant screening

Participants were screened for proficiency prior to the experiments with two tasks. First, they were administered the same written language background questionnaire as was given to the speaker. They then completed a familiarization task, to check vocabulary size and to ensure association of the appropriate Mandarin and English names with the pictured objects. Participants were presented all visual stimuli one by one on a computer screen, along with printed English and Mandarin names for the pictured objects. The positions of the English and Mandarin names (left or right underneath the picture) were randomized. The task was self-paced, and participants were given an index card to note down any English and Mandarin words they were unfamiliar with, or if the words were not ones that they would typically use to name the pictured object. If the participant was not proficient enough according to the questionnaire (i.e., scoring below 3 on the 1 (low) – 4 (high) understanding and speaking proficiency scales on the language background questionnaire) or their vocabulary was too limited based on the familiarization task, they were disqualified from participating. A substantial vocabulary in both English and Mandarin, as well as familiarity with specific names of pictured objects, was

desirable, as the study relied on participants' being able to associate pictures with their spoken names in both languages. Therefore, any participants who marked more than ten words (of a total of 224) as unfamiliar or not their primary choice for describing the picture in either language (e.g., due to dialectal differences) was disqualified from participating. The entire screening process for each participant lasted approximately twenty minutes.

Participant language background

A total of 42 Mandarin–English bilinguals (35 female, 7 male) with no reported speech or hearing defects qualified for participation in this study. All participants but one completed both this experiment and Experiment 2. The participants' linguistic backgrounds and, consequently, their language dominance, varied. Thirty-five participants were L1 Mandarin speakers, one participant was an L1 English speaker, while six participants were simultaneous bilinguals. Twenty-three participants reported also speaking other languages, and four participants reported both Mandarin and other Chinese languages as their L1s: Wu (Shanghainese), Yue (Cantonese), and Southern Min. The average age was 20.4 years ($SD = 2.2$). While most participants were 18–24, one male participant was 31 years of age. The average age of arrival to the U.S. was 15 years ($SD = 7$), although two participants first lived in Canada starting at ages four and eight, before moving to the U.S. at ages 12 and 18, respectively. Additionally, several participants grew up in Singapore, where English is an official language and most of the population code-switches frequently. Most participants moved from China to the U.S. for college, while two each moved from Malaysia and Singapore, and one each from Taiwan and Hong Kong. Four participants were born and raised in the U.S. All participants reported occasionally or regularly code-switching with friends or family. Three participants were left-handed.

We quantified participants' language dominance using the Bilingual Language Profile (Birdsong et al., 2012), a questionnaire that assesses language dominance. The participants' scores ranged from -159 to 96, averaging -31 ($SD = 59$), meaning that most participants leaned Mandarin-dominant. Twenty-seven participants had negative scores, suggesting Mandarin dominance, while the other fifteen had positive scores, suggesting English dominance. Following a reviewer's suggestion, these dominance scores are included as part of a separate model in the Results section, to ascertain whether this diversity affected findings.

Table 2 provides participants' average age of acquisition of English and Mandarin, as well as their self-rated proficiency in each language on a scale of 0–6, where 0 means "not well at all" and 6 means "very well." Participants rated themselves as being almost equally proficient in speaking, understanding, reading, and writing both languages.

Visual stimuli

Visual stimuli consisted of 80 pictures from the Rossion and Pourtois (2004) colored line drawing database, or other public domain colored line drawings that visually resembled the Rossion and Pourtois (2004) set. All pictures depicted common objects, and were modified to the same dimensions. Of the 80 pictures, 64 were target experimental items, in that the pictured objects were mentioned in the corresponding auditory stimulus sentence. The other 16 pictures functioned as part of catch trials, where the pictured object was not mentioned in the corresponding auditory stimulus.

Auditory stimuli

Auditory stimuli consisted of 144 spoken English sentences: (a) 64 sentences that mentioned the paired visual stimulus, (b) the spliced versions of those 64 sentences that mentioned the paired visual stimulus, and (c) 16 sentences that functioned as catch trials, thereby not mentioning the paired visual stimulus.

The target experimental items included the 64 spoken English sentences with either English target words (32 unilingual sentences) or Mandarin target words (32 code-switched sentences), recorded by the speaker in random order. Sentences were constructed so that each mentioned a picturable noun. Picturable nouns occurred sentence-medially in half of the sentences and sentence-finally in the other half. This gave a total of 16 English sentences with medial nouns, 16 English sentences with final nouns, 16 code-switched sentences with medial nouns, and 16 code-switched sentences with final nouns. Sentences were designed with similar syntactic structures to control for intonational patterns: either 1) a main clause beginning with a subject pronoun, followed by a transitive verb and direct object, ending with a prepositional phrase, or 2) a subject pronoun, main verb, and embedded clause. In the former case, medial targets occupied the direct object position, while final targets were located in the prepositional phrase. In the latter case, final targets were located in the embedded clause. Target words were introduced by either a definite article, indefinite article, or possessive pronoun. Spliced versions of these 64 sentences were also constructed, as described in the Splicing section.

Additionally, 16 sentences were not target trials but functioned as catch trials instead, in that none of the picturable nouns heard in the auditory stimuli matched the pictured objects on the screen. For instance, participants might hear "I saw a raccoon behind the plant," while being presented a picture of a zebra. The inclusion of these catch trials was to ensure that target loci were not predictable from the similar syntactic structures of the stimulus sentences. The catch trials were split evenly among the four kinds of stimuli, regarding position of the picturable noun and whether there was a code-switch. The intention was to prevent participants possibly using syntactic or contextual predictability to respond whenever they expected to hear a noun, e.g., pushing a button when they heard the determiner preceding the target noun.

These sentences can be found in Appendix S1.

Splicing

This study utilizes a splicing manipulation in both experiments to test the prediction that listeners will have relatively more difficulty recognizing a code-switch (manifesting as slower reaction time) if anticipatory phonetic cues to the code-switch are withheld. The speaker recorded multiple repetitions of each auditory stimulus sentence, including English-only versions of code-switched sentences to use as frames in the splicing condition.

To eliminate any phonetic information provided in the sentence leading up to the target word that could cue the language of the target word, stimuli were cross-spliced, so that a Mandarin target originally recorded in a code-switched sentence was spliced into what was originally a unilingual English sentence. To control for any effects of the splicing manipulation itself, English sentence stimuli were recorded twice, and English targets were identity-spliced into a separate repetition of the same English sentence. This procedure is illustrated in Figure 1.

Since the spliced and unspliced versions of each sentence were identical content-wise and would sound identical aside from the splicing effect, two lists were created in each experiment to avoid participants hearing the same sentence both spliced and unspliced. In each list, half of the items were spliced. The concept monitoring experiment had 64 distinct target sentences, so that each list had 32 spliced items (along with the 16 catch trial sentences). Participants were randomly assigned to one of two lists at the start of each experiment, with an equal number of participants assigned to each list.

Procedure

Data collection took place in a sound-attenuated booth in the PhonLab in the Department of Linguistics at the University of California Berkeley. Prior to the experiment, participants were presented with printed English instructions on a computer screen, informing them that they would hear a sentence while an image is displayed on the screen. Instructions stated that participants would hear both English and Mandarin throughout the experiment, and asked that they press a button if they heard the pictured object mentioned in the sentence. An experimenter was present to answer questions, as well as to clarify that: a) the pictured object would sometimes not be mentioned (i.e., in catch trials), and in that case, not to press a button, and b) participants were to press a button if the pictured object were named at all, in either language. Auditory stimuli were presented through headphones. During each trial, participants saw a picture in the center of the computer screen, and heard a spoken sentence that mentioned the pictured object. The task was to press a button as soon as they heard the object mentioned in the sentence. Presentation of trials was randomized, and a 1000 ms delay occurred between trials. Each trial lasted 3000 ms. The experiment lasted approximately fifteen minutes.

This experiment (concept monitoring) was counter-balanced with the next experiment (eye tracking); participants were randomly assigned the order in which to complete the two experiments. After completion of both experiments, participants were administered the Bilingual Language Profile (Birdsong et al., 2012) as well as a questionnaire asking about their code-switching attitudes and behaviors. The entire study lasted around 45 minutes, and participants were compensated $5 for the completion of each of the three components.

Reaction times were measured as the latency between the onset of the target word and the subject's keypress response. Catch trials were first excluded from analysis, so that there were a total of 2688 target trials (64 unique stimuli x 42 participants). Data was then trimmed to remove trials with reaction times that were under 200 ms or longer than the trial duration. This resulted in the loss of 47 observations. Additionally, trials with target words that participants noted as unfamiliar during the familiarization task were excluded. Finally, each participant's mean was calculated, and any reaction times that were more than two standard deviations from that participant's mean were excluded from analysis. Only two observations were removed as outliers in this

Table 3. Mean raw reaction times (ms), as a function of switch, splice, and position. Standard deviations are in parentheses.

	No code-switch		Code-switch	
	Medial	Final	Medial	Final
Unspliced	1004 (549)	819 (410)	1097 (616)	798 (378)
Spliced	1033 (631)	826 (473)	1126 (611)	890 (474)
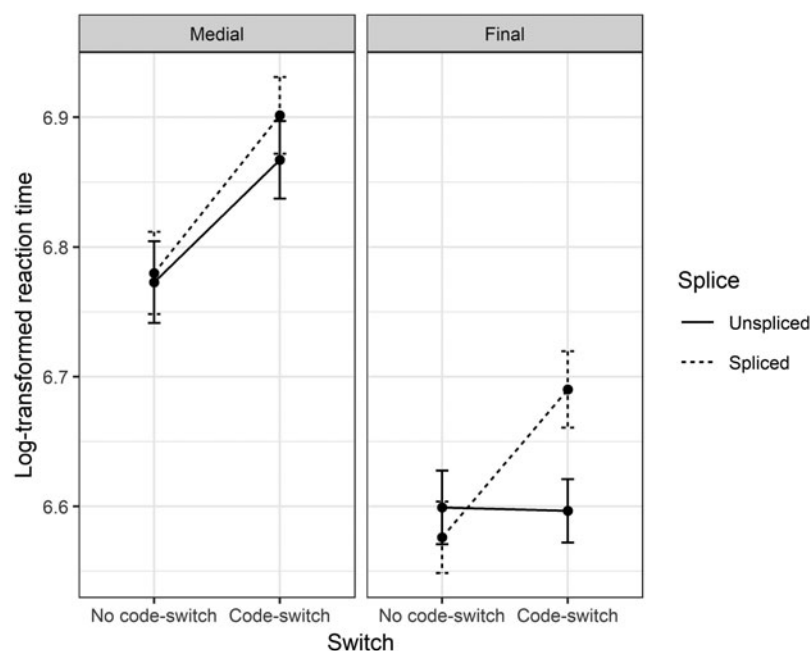
Fig. 2. Log-transformed mean reaction times, by position, switch, and splicing conditions. Vertical lines represent standard errors.

manner. After trimming, 2506 observations remained for analysis, so that approximately 7% of the target data was excluded.

Data analysis
The log-transformed data was modeled with a linear mixed effects regression, shown in Table S1. The model considers an interaction between whether a target word is a code-switch or not (Switch), spliced or unspliced (Splice), and sentence-medial or sentence-final (Position), and includes random slopes for Splice-by-item and Switch-by-subject (Baayen, Davidson & Bates, 2008). As a follow-up analysis, we fitted an alternate model including an interaction of participants' BLP scores (Dominance) with Switch, Splice, and Position.

Results

Concept monitoring
Table 3 shows average reaction time (in milliseconds) as a function of Switch, Splice, and Position. Generally, reaction times to code-switched targets were slower than to English targets (with the exception of final, unspliced targets), and reaction times to spliced targets were slower than to unspliced targets. However, the most noticeable difference is between reaction times to sentence-medial and sentence-final targets.

Since the data distribution was right-skewed, reaction times were log-transformed.

The linear mixed effects regression model summarized in Table S1 and plotted in Figure 2 suggests that there was a significant effect of Position, and a tendency for Switch to affect reaction

time. The target being code-switched is associated with longer reaction times ($\beta = .091$, $t = 1.912$, $p = .059$). Reaction times to sentence-medial words were significantly longer than those to sentence-final words ($\beta = .217$, $t = 4.705$, $p < .001$). However, Splice was not a significant effect ($\beta = .047$, $t = 1.382$, $p = .172$). Additionally, the interaction between Switch and Splice is not significant, suggesting that reaction times for code-switched trials are not predicted to differ significantly depending on whether they were spliced or unspliced.

We also fit a model that included participants' language dominance scores from the Bilingual Language Profile, since our participants' linguistic backgrounds varied. This model had an interaction between Switch, Splice, Position, and Dominance. While this model performed worse than the one in Table S1 as evaluated by both models' Akaike Information Criteria and log likelihoods, Switch was significant in this model ($\beta = .013$, $t = 2.62$, $p = .01$), as was the interaction between Switch and Dominance ($\beta = .0001$, $t = 2.52$, $p = .01$), and Switch, Position, and Dominance ($\beta = .0001$, $t = -2.98$, $p = .003$). This suggests that participants with a more positive BLP score (i.e., more English-dominant participants) had slower reaction times to code-switches, especially sentence-medial code-switches.

Due to participants' different backgrounds in country of origin, age of arrival to the U.S., and age of acquisition of English, we performed further analyses with the original model to determine whether excluding the ten participants who were not born and raised in China (i.e., from Singapore, Malaysia, Taiwan, Hong Kong, or the U.S.) affected the results. This was not the case; the pattern of the results was unchanged. Exclusion of the five simultaneous bilinguals,

who were also not born and raised in China, but rather in the U.S., Singapore, and Hong Kong, also did not affect results.

Discussion

The results of this experiment are consistent with the switch cost findings in previous studies: Listeners were slower to recognize code-switched words compared to words in a unilingual utterance. However, the absence of anticipatory phonetic cues did not have an apparent effect on the recognition of the code-switch, contrary to our initial hypothesis.

Assuming the intended anticipatory phonetic cues are present in the speech signal, this result suggests that perhaps Mandarin–English bilingual listeners did not detect or use such cues. However, while the reaction time measure used in this experiment revealed that Mandarin–English bilinguals are slower overall to recognize code-switches, it is possible that phonetic cues did affect the recognition process prior to and at the beginning of the code-switch, but that these effects had already dissipated before the button-press in the concept monitoring task.

The position of the target word had an interesting influence on recognition of code-switches. Though target word position was originally varied to prevent participants from predicting its location in the sentence by using syntactic cues like determiners and possessive pronouns, listeners took longer to recognize sentence-medial targets compared to sentence-final targets, regardless of whether the target was a code-switch or not. This difference could potentially be attributed to the reduction of uncertainty as the sentence progresses. After participants experience several trials, it might become clear that targets only occur medially, finally, or not at all, especially because sentences are controlled for syntactic structure. If participants are strategically expecting targets by syntactic position, rather than monitoring for the concept, then sentence-final targets might be easier. For example, if the participant has already heard the main clause but not the target, then the target is either sentence-final or will not occur.

Alternatively, listeners' use of phonetic information in sentence processing could be affected by the amount of time they have to incorporate such information; all sentence stimuli were similar lengths so that trials with sentence-final targets are preceded by a longer utterance than trials with sentence-medial targets. Future work can manipulate sentence length, word position, and number of catch trials to investigate the difference between medial and final targets.

The model including Dominance suggests that dominant language is a factor in code-switched recognition. English-dominant bilinguals were slower to respond to code-switched, i.e., Mandarin, targets. One interpretation is that switching out of one's dominant language and into the non-dominant language is difficult. Perhaps bilinguals can recognize code-switches more easily if the switch occurs in their dominant language. This pattern is reminiscent of the Inhibitory Control Model (Green, 1998), and what Olson (2017) found in comprehension, though with a different effect of dominant language: instead of switching back into the dominant language being more costly due to the fact dominant language requires stronger inhibition, our bilinguals took longer to switch into their non-dominant language.

Experiment 2: eye tracking

Experiment 1 showed that Mandarin–English bilinguals are slower to recognize code-switched words, but failed to show an effect of the absence of anticipatory phonetic cues on concept monitoring times. While an offline task like the concept monitoring experiment can reveal whether code-switched recognition incurs a switch cost, it may not give insight into the time course of recognition and whether and when phonetic cues are incorporated.

Online tasks such as eye tracking are advantageous for understanding the time course of lexical activation during spoken language comprehension (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). The visual world paradigm in eye tracking is a particularly good method for studying spoken word recognition (Allopenna, Magnuson & Tanenhaus, 1998; Altmann & Kamide, 2009; Huettig, Rommers & Meyer, 2011).

The visual world paradigm involves a visual display of pictures, with a simultaneous auditory stimulus naming one of the pictures. The pictures represent the target word and various lexical competitors, with participants' eye movements revealing when certain lexical items are activated during spoken word recognition. The auditory stimulus can be manipulated to test the role of different phonetic details in the process of recognizing a spoken word.

Experiment 2 uses the visual world eye tracking paradigm and splicing to investigate whether withholding anticipatory phonetic cues affects code-switched recognition. The visual world involves a display of four pictures, each corresponding to a different type of lexical candidate, and a simultaneous auditory stimulus so that the time course of lexical access is elucidated by the participant's fixations to pictures during perception of that continuous speech. The goal of this experiment is to probe which lexical candidates are considered during the processing of a code-switch, and whether bilingual listeners use phonetic information to constrain recognition to candidates in the expected language.

We predict that recognition of a code-switch will be hindered by a lack of phonetic cues to that switch. Therefore, in the spliced code-switched condition, we predict that listeners will fixate less on the target as compared to the unspliced code-switched condition, because the phonetic context will lack switch cues and bias them away from Mandarin. Listeners might therefore look at an English competitor early on, expecting a target in the same language as the sentence frame. In the unspliced code-switched condition, listeners will fixate more on the target, since available phonetic cues will bias them toward expecting a Mandarin code-switch. Listeners might also look toward the Mandarin competitor more than in any other condition, since only the unspliced code-switched condition involves phonetic cues signaling an upcoming Mandarin word.

Method

Speaker and Participants

The speaker who recorded the auditory stimuli for Experiment 1 also recorded the auditory stimuli for this experiment.

Of the 42 participants who completed Experiment 1, data from one participant was excluded in Experiment 2 due to their corrective lens interfering with the eye tracker's calibration process.

Visual stimuli

36 picturable nouns (18 Mandarin nouns, 18 English nouns) that have both picturable Mandarin and English noun cohort competitors were selected, for 18 sets of three picturable nouns. To each set, a distractor that was not a cohort competitor, i.e., did not share an onset, was added. This resulted in 36 sets of four

picturable nouns. Colored line drawings in the Rossion and Pourtois ([2004](#)) database or available in public domain were selected for the picturable nouns.

Auditory stimuli

A sentence was constructed for each set of four picturable nouns, resulting in 36 total sentences. The target noun was located sentence-medially in 18 sentences, and sentence-finally in the other 18 sentences. The portions of these sentences preceding the target were constructed so that any of the four picturable nouns in the set were semantically congruous with the verb. For example, a code-switched trial might have visual stimuli where the Mandarin target *màozi* [maʊ⁵¹tsɨ] corresponds to a picture of a hat, the cohort competitors in English and Mandarin, *mouse* and *máojīn* [maʊ⁵¹tɕɪn ⁵⁵], respectively correspond to pictures of a mouse and a towel, and the distractor corresponds to a picture of flower (*huā* [xwa⁵⁵] in Mandarin). The corresponding auditory stimulus is the sentence *We saw the màozi in a tree* where any of the four picturable nouns in the set are semantically congruous as direct objects of the verb *saw*. Figure S1 shows example sets of visual world stimuli with a corresponding auditory sentence (where the target is sentence-medial) for both the code-switch and no code-switch conditions.

Stimuli were spliced as in Experiment 1, so that a spliced version of each sentence was created. There was thus a total of 72 auditory stimuli: the spliced and unspliced versions of nine unilingual stimuli with sentence-medial targets, nine unilingual stimuli with sentence-final targets, nine bilingual stimuli with sentence-medial targets, and nine bilingual stimuli with sentence-final targets. Participants only heard one version of each sentence, depending on their experimental list, given that spliced and unspliced versions of a sentence were identical aside from the phonetic manipulation. The sets of picturable nouns and their corresponding sentences can be found respectively in Appendices S2 and S3.

Procedure

Participants were seated a comfortable distance from the computer screen and an eye tracker (The Eye Tribe), which was then calibrated with a nine-point calibration. Sampling frequency of the gaze location was 60 Hz. Participants wore headphones for presentation of auditory stimuli. Text instructions displayed on the computer screen prior to the experiment informed participants that they would see images while hearing English and Mandarin throughout the experiment.

During each trial, participants saw a visual world display of four colored line drawings corresponding to four picturable nouns (target, English cohort competitor, Mandarin cohort competitor, and distractor). One picture was centered in each of the four quadrants of the screen. Then after a delay averaging 250 ms, participants heard a spoken sentence. Their task was to press a button as soon as they heard any pictured object in the display be named in the sentence. Each trial lasted 4000 ms. 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.

The presentation of trials was randomized, and a 1000 ms delay occurred between trials with a central fixation cross. The eye tracking task lasted approximately ten minutes.

Data analysis

We excluded trials with visual displays that included any pictures that corresponded to nouns that participants marked as unfamiliar during screening.

Looks to the quadrant of each type of picture in the visual world display (Mandarin or English target, English cohort competitor, Mandarin cohort competitor, distractor) were counted as fixations to that picture. To calculate the average proportion of fixations for a condition, the number of fixations toward a type of picture were summed across all trials in that condition and all participants, and then divided by the total number of trials in that condition.

The following analyses focus on the time window corresponding to increasing activation: from target word onset to 1200 ms, which is when target fixations plateaued. Following Mirman ([2014](#)), growth curve analysis with orthogonal polynomials was used to model the time course of fixations to the pictures corresponding to the target word and competitors.

Growth curve analysis is well-suited for analysis of eye tracking data, in that time is treated as a continuous variable. The addition of orthogonal polynomials allows modeling the shape of the time course of fixations. Upon visual inspection of the time course data, cubic orthogonal polynomials were chosen as the best approximation of the shape of the curve for proportion of looks over time. The random effects structure for each model included by-participant random slopes for Switch (Baayen et al., [2008](#)).

To assess the best-fitting models for the data, a baseline model was used as a starting point. Variables were added gradually to produce several models varying in complexity, and ANOVA was used to compare the baseline model and these models. Log likelihood and Akaike information criterion (AIC) were then used to assess the best-fitting models for the data. Alpha levels of 0.05 were used to evaluate the significance of each predictor. The interactions between linear, quadratic, and cubic orthogonal polynomials with all fixed and random variables were included.

Results

Looks to target

The model for looks to the target included the fixed effects of Position (whether the target occurred sentence-medially or -finally), Switch (whether the target was a code-switch), and Splice (whether the target was spliced). It treated sentence-final, no code-switch, and unspliced as the reference points, and statistical significance was calculated using the normal approximation. The model is shown in Table S2, and plotted in [Figure 3](#).

All three fixed effects were significant in this model. The main effect of Switch was significant; there were fewer looks to code-switched than not switched targets ($\beta = -.0769$, $t = 3.83$, $p < .001$). For example, participants would look toward the image of the tiger less often if they heard "She saw a picture of the *lǎohǔ*" than if they heard "She saw a picture of the *tiger*." Position was also significant; there were fewer looks to sentence-medial than sentence-final targets ($\beta = -.1849$, $t = -10.74$, $p < .001$). Finally, Splice was significant: there were fewer looks to spliced than unspliced targets ($\beta = -.0833$, $t = -5.23$, $p < .001$).

The only interaction with an orthogonal polynomial to be significant was that between the quadratic term and Position. Although there were initially fewer looks to a sentence-medial target, the rate of looks to that target increased faster compared to a sentence-final target ($\beta = .1706$, $t = 2.88$, $p = .004$).
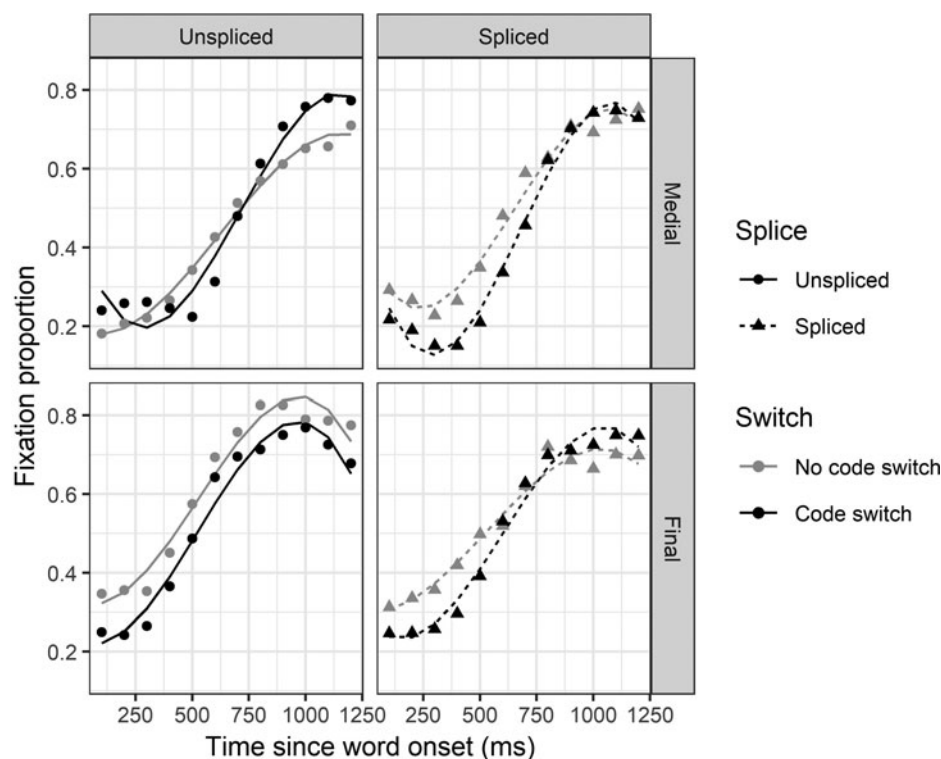
The interaction between Position and Switch was significant. Participants looked at a medial target more when it was code-switched than when it was not code-switched ($\beta = .1023$, $t = 3.44$, $p < .001$). The significant interaction between Position and Splice indicates more looks to a medial target when it was spliced than when unspliced ($\beta = .1395$, $t = 5.52$, $p < .001$). Finally, the three-way interaction was significant; there were fewer looks to a sentence-medial code-switched target when it was spliced than when unspliced ($\beta = −.1328$, $t = -3.03$, $p = .002$). For instance, presented with a code-switched utterance "We saw the *màozi* in a tree," listeners looked toward the target image of the hat less if it was spliced into that frame.

Since participants differed in country of origin, age of arrival to the U.S., and age of acquisition of English, further analyses were performed to check whether participants could appropriately be analyzed as a single group. Excluding the nine participants who were not born and raised in China (i.e., from Singapore, Malaysia, Taiwan, Hong Kong, or the U.S.) from this analysis did not affect results. Exclusion of the five simultaneous bilinguals, a group which overlapped heavily with those born and raised outside of China, also did not affect results. Examination of individual results did not reveal any pattern among simultaneous bilinguals.

To account for differences in participants' backgrounds as in Experiment 1, we added participants' dominance scores from the BLP questionnaire as a continuous covariate in the model, such that we had an interaction between the orthogonal polynomials, Position, Switch, Splice, and Dominance. This model did not perform as well as the one shown in Table S2, as evaluated by their Akaike information criteria (AIC) and Bayesian information criteria (BIC). Significance did not change for any of the original effects, but there was an additional significant interaction between Switch and Dominance: if the target was a final, unspliced code-switch, participants with more positive dominance scores (English-dominant) looked at the target less compared to if it were unswitched ($\beta = −.001$, $t = −2.95$, $p = .003$). There was also a significant interaction between the linear orthogonal polynomial, Switch, and Dominance, such that the rate of fixations to final, unspliced code-switched targets increased faster for more English-dominant bilinguals than for more Mandarin-dominant bilinguals ($\beta = .002$, $t = 2.26$, $p = .02$).

Looks to the Mandarin competitor

Looks to the Mandarin competitor were modeled with cubic orthogonal polynomials, fixed effects of Switch and Splice (baseline: no code-switch, unspliced), and by-participant random slopes. The model can be found in Table S3, and plotted in Figure S2.

There was a main effect of Switch, showing that there were more looks to the Mandarin competitor in code-switched than in English unspliced trials ($\beta = .0641$, $t = 5.44$, $p < .001$). Interactions between Switch and both the linear and quadratic terms were significant. The decay in looks to the Mandarin competitor was steeper in code-switched than in English unspliced trials ($\beta = −.1035$, $t = −2.76$, $p = .006$; $\beta = −.1114$, $t = −3.48$, $p < .001$). There was no main effect of Splice, although the interaction between the cubic term and Splice is significant ($\beta = −.0697$, $t = −2.5$, $p = .01$). Therefore, the shape of the function capturing fixations to the Mandarin competitor differed in spliced versus unspliced trials without switches, although there was no difference between those conditions in proportion of fixations. Finally, the interaction between Switch and Splice was significant ($\beta = −.0232$, $t = −1.98$, $p = .048$). There were fewer looks to the Mandarin competitor in code-switched trials when the target was spliced, compared to when the target was unspliced. For instance, if *màozi* was spliced into "We saw the *màozi* in a tree," listeners looked toward Mandarin competitor *máojīn* less, than if *màozi* were not spliced.

Looks to the English competitor

Looks to the English competitor were modeled in the same way as looks to the Mandarin competitor were, with cubic orthogonal polynomials, Switch, Splice, and by-participant random slopes. This model can be found in Table S4, and plotted in Figure S3, with model fits as lines and empirical data as points.

The near-significant effect of Splice suggests an increase in looks to the English competitor when the target word was spliced and unswitched ($\beta = .0354$, $t = 1.95$, $p = .0512$), but all of the main effects failed to reach significance.

Discussion

We found that withholding anticipatory phonetic cues affects code-switched recognition. These results suggest that removing anticipatory phonetic cues to a Mandarin code-switch in an English utterance can affect the processing of that code-switch. However, this effect is mediated by the position of the code-switch. Specifically, our study indicated fewer looks toward a spliced code-switched target in the sentence-medial condition, than an unspliced code-switched target in the same condition.

This is consistent with what studies of Spanish–English bilingual listeners have found: that bilinguals can use phonetic cues (intonation and VOT) to anticipate an upcoming code-switch (Piccinini & Garellek, 2014; Fricke et al., 2016). In conjunction with these previous findings, our results suggest that while the presence of anticipatory switch cues is facilitatory, the absence of such cues can hinder code-switched recognition.

When anticipatory phonetic cues were withheld, there were fewer looks to a medial code-switch (compared to when cues were present), and also fewer looks to the Mandarin competitor during the switch condition. This suggests that cues point the listener to the switch language. Moreover, without anticipatory cues, listeners were more likely to look toward an English competitor overall; with a lack of cues to another language, listeners expect English, the matrix language.

Our findings hinge on the position of the target word, as there is no interaction between Switch and Splice for sentence-final code-switched targets. Word position appears to play an important role in recognition, whether code-switched or not, since all sentence-final targets received more looks than sentence-medial targets. As mentioned in the discussion of Experiment 1, this effect could be due to context, processing time, and various other factors.

Moreover, it appears that participants' different language backgrounds, including age of arrival to the U.S., age of acquisition of English, and country of origin (which might affect the variety of Mandarin spoken), did not affect results. The alternate model that includes dominance does suggest that while bilinguals who were more English-dominant looked at the target less when it was a code-switch, the role of anticipatory cues in recognition was not affected by language dominance. Perhaps it is more difficult to switch into the less dominant language, but sensitivity to phonetic cues is unaffected by dominance.

Acoustic analysis

Background

This acoustic analysis considers the potential mechanisms for code-switching pronunciation: 'blending' of the phonetic features of both languages and 'preparation' of articulatory gestures for

Table 4. Number of Mandarin words with tones 1–4 in each experiment.

	Experiment 1	Experiment 2	Total
Tone 1 (high level)	9	6	15
Tone 2 (rising)	8	4	12
Tone 3 (falling-rising)	6	3	9
Tone 4 (falling)	9	5	14

production of a specific code-switched target. Experiment 2 suggested that Mandarin–English bilingual listeners are sensitive to some phonetic nuance in the acoustic signal leading up to sentence-medial Mandarin code-switches. The set of anticipatory phonetic cues for Mandarin code-switches could consist of a bundle of suprasegmental and segmental features. Given the difference in lexical tone between Mandarin and English, and Piccinini and Garellek's (2014) study showing intonation functioning as a cue, we will focus on the fundamental frequency of vocal fold vibration (f0), which is the primary acoustic correlate of perceived pitch.

We analyzed the pitch contours of all unspliced stimuli in both experiments, by using Praat (Boersma & Weenink, 2001) to extract f0 measurements in 10 millisecond intervals from each sentence produced by the speaker. The main comparison is between the unilingual English unspliced stimuli and the code-switched unspliced stimuli: if the pitch preceding English target words differs from the pitch preceding Mandarin code-switched target words, then pitch might be responsible for the differences in perception found in Experiment 2. If realized via 'blending', unilingual and code-switched pitch will generally differ. However, if realized via 'preparation', there will be target-specific differences between the pitch preceding English words and the pitch preceding Mandarin words of each tone, such as the mostly dissimilatory tone-specific anticipatory coarticulation found in unilingual Mandarin speech (Xu, 1997). If code-switched tonal coarticulation patterns with unilingual tonal coarticulation, then we will find dissimilatory anticipatory effects before all tones except the falling-rising tone 3. Alternatively, tonal coarticulation patterns in code-switched utterances might differ, since the English portion of the utterances is unconstrained by lexical tone, whether assuming different contours or similar but more extreme contours.

The 50 Mandarin target words were not balanced with respect to tone; Table 4 shows the number of Mandarin words that have initial syllables of each tone, by experiment.

We analyzed the unspliced stimuli from each experiment separately, since only Experiment 2 showed a splicing effect for sentence-medial code-switches. The experiments used separate sets of stimuli and neither used stimuli balanced by Mandarin tone, which might have resulted in differences in f0 by experiment.

Acoustic analysis of Experiment 1 stimuli

Experiment 1 found no splicing effect on reaction times to code-switched stimuli, so reaction times are plotted by target word tone to examine whether there might have been tone-specific differences in pronunciation (Figure S4). Sentence-medial and -final reaction times are averaged together since they showed the same general pattern, although there were
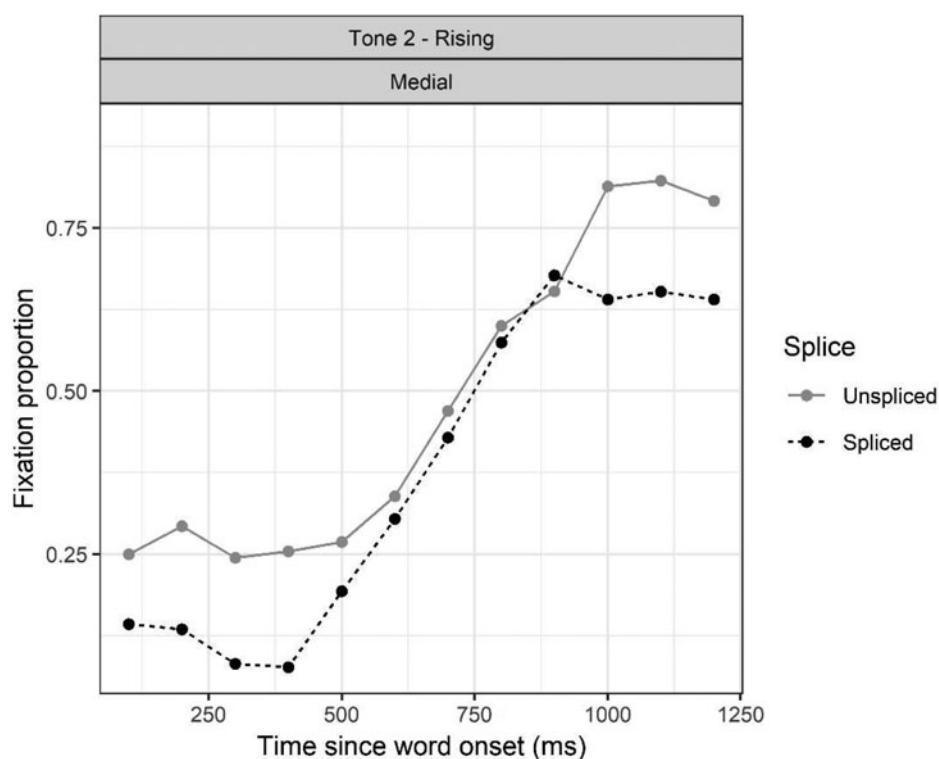
Fig. 4. Looks to sentence-medial code-switched (Mandarin) tone 2 targets.

no tone 3 sentence-medial words. Figure S4 indicates that reaction times to code-switches with tones 1 and 3 are visibly shorter when unspliced, while code-switches with tones 2 and 4 seem similar regardless of splicing, suggesting possible differences in cuing for tones 1 and 3 vs. 2 and 4.

We plotted F0 measurements for 500ms before and after the target onset for unspliced stimuli from Experiment 1. Figure S5 shows evidence for a preparation mechanism, with tone-specific patterns such that the entire pitch contours before tones 2 and 4 are relatively high compared with before English targets, while the pitch contours before tones 1 and 3 are overall lower than the pitch contour before English targets. If listeners categorize pitch contours as high versus low, then English contours might be categorized with tone 2 and 4 contours, which would mean that code-switches of tones 1 and 3 might be cued to a greater degree than code-switches of tones 2 and 4. This is one possible explanation for the tone-specific reaction times in Figure S4 and the lack of a general splicing effect in the experiment.

Acoustic analysis of Experiment 2 stimuli

Experiment 2 found an effect of splicing, but only on sentence-medial code-switched targets, suggesting that phonetic cues can affect recognition depending on word position. We plotted looks to target by both position and tone of the target word in Figure S6, to determine whether cuing might vary by position and tone as well.

Figure 4 plots an instance of listeners looking more toward unspliced than spliced sentence-medial code-switched targets: those with tone 2. When comparing sentence-medial code-switched targets of all tones, this pattern is only present for tone 2 and 4 targets, as seen in Figure S6. However, listeners look more toward sentence-final targets of nearly all tones when they are unspliced.

The f0 of these stimuli are plotted by tone, with sentence-medial targets in Figure 5 and sentence-final targets in Figure S7. There were no sentence-medial tone 3 words. Both figures provide evidence for a preparation mechanism of code-switching pronunciation of f0, with tone-specific pitch patterns for stimuli in Experiment 2.

Figure 5 demonstrates the expected dissimilatory anticipatory tonal coarticulation patterns for sentence-medial Experiment 2 stimuli, but with no obvious difference between the pitch contours preceding tones 2 and 4 and those preceding tone 1, despite the former receiving more looks when unspliced. Pitch therefore might not be the primary phonetic cue; segmental cues might be playing a larger role, especially since some competitors differ slightly in onset, e.g., voiceless retroflex fricative in Mandarin vs. voiceless post-alveolar fricative in English.

Figure S7 shows either reduced or absent tonal coarticulation into sentence-final code-switched targets. The pitch contour before final English words was higher than before any of the Mandarin tones, across the entire 500ms duration before target onset. The f0 range for pitch contours is also below 220 Hz, in contrast to pitch contours before sentence-medial words, which occur above 220 Hz. This lower f0 range could have acted as a cue to the listener that the sentence was nearing its end: if the participant has heard the entire main clause with no mention of any pictured objects, then process of elimination and lower f0 should lead them to expect a sentence-final target. Additionally, all of the Mandarin tones on the code-switched targets had a rising contour, including tone 4, which should be a falling tone. In contrast, the f0 of final English targets slopes downwards. If listeners categorize contours as rising or falling, then this difference of f0 on the target words themselves might have contributed to the Experiment 2 result that showed no splicing effect on sentence-final code-switched stimuli.
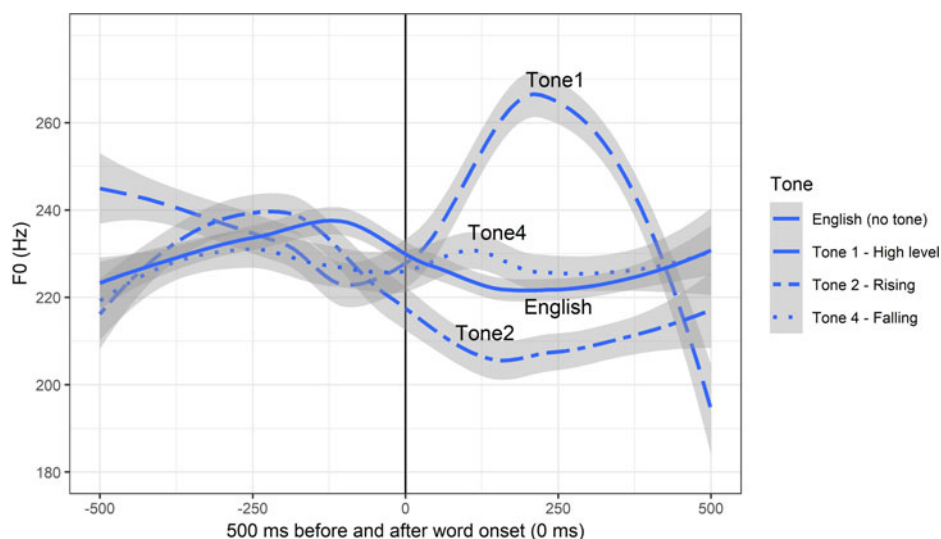
Fig. 5. Pitch contours for sentence-medial unspliced stimuli in Experiment 2, by tone. Contours for code-switched stimuli are labeled by tone, and unilingual stimuli are labeled as English.

Summary

This exploratory acoustic analysis suggests that code-switching pronunciation of f0 depends on the tone of the Mandarin code-switch and whether the code-switch is sentence-medial or -final. This points to a preparation explanation of code-switching pronunciation, as realization of anticipatory f0 contours depends on aspects of the target, rather than characteristics of the language. Nevertheless, a blending mechanism is not completely ruled out as we do not have unilingual Mandarin productions to compare with the unilingual English productions.

The role of code-switching pronunciation of f0 as an anticipatory phonetic cue to code-switching is difficult to determine through this analysis. While the eye tracking results suggest that sentence-medial code-switches are more easily processed when phonetic context is retained, the f0 contours before these target code-switched stimuli do not seem to exhibit any obvious cue-like features. Regardless, listeners are evidently sensitive to something in the acoustic signal, so this is work for future studies investigating the phonetic features of Mandarin–English code-switching pronunciation, with stimuli balanced across each Mandarin tone.

General discussion

Using concept monitoring and eye tracking experiments, we found evidence for a switch cost in Mandarin–English code-switches. Crucially, we also found that phonetic cues mitigated switch cost in sentence-medial switches.

The finding of a switch cost in auditory word recognition is consistent with previous studies: Soares and Grosjean (1984) found that English-Portuguese bilinguals took longer to recognize code-switched words in carrier sentences in a lexical decision task. Similarly, Olson (2017) found longer fixation times in Spanish–English code-switched vs. unilingual sentences in one experimental condition (the 'monolingual language mode', i.e., stimuli with few switches). We found, in our concept monitoring task, that Mandarin–English bilinguals took longer to respond to Mandarin code-switched words. The similarity of our results with the monolingual language mode in Olson (2017), combined with the fact that our code-switched targets were syntactically integrated in the English matrix sentences, invites the question of whether they were phonologically integrated into the English

context, i.e., whether they were more similar to borrowings or to nonce loans (Cacoullos & Aaron, 2003; Poplack & Dion, 2012). Although there are no unilingual Mandarin utterances from our speaker for comparison, the acoustic analysis suggests that the Mandarin tone contours were intact. Hence, our stimuli could not be considered borrowings.

The reaction times to the (off-line) concept monitoring task did not reveal any benefit of retaining the phonetic context leading up to code-switches, but eye-tracking did: there were more looks to sentence-medial code-switches when the phonetic context was present. These results suggest that the phonetic context can facilitate the recognition process as it unfolds, but that it does so to different degrees in different contexts (medial vs. final).

Our acoustic analyses may shed some light on the mechanisms underlying the pronunciation of code-switched utterances. Recall from the Introduction that acoustic properties of code-switched targets may arise as a result of 'blending' of two phonetic systems on the one hand and of 'preparation' for the target pronunciation on the other. A key difference between these two possibilities is that the 'preparation' scenario, but not the 'blending' scenario, leads one to expect the pronunciation of the region preceding the code-switched target to depend on the specific target. Our acoustic analyses revealed that the phonetic context preceding code-switches depended on the target word's lexical tone: we observed dissimilatory anticipatory tonal coarticulation before tones 1, 2, and 4, and assimilatory coarticulation before tone 3. We regard these observations to be tentative, because of the small numbers of items. If the pattern we observed can be confirmed with a larger set of items of each tone, it would suggest the presence of target-specific code-switching pronunciation. Such target-specific tonal coarticulation would be expected under the 'preparation' account of phonetic cues to code-switching, rather than the 'blending' account.

It is impossible to say, on the basis of the available data, which aspects of tonal coarticulation – f0 contours, range, and/or extrema, or other suprasegmental and segmental features – might act as phonetic cues. We regard this as a question for future research.

The position-dependent (medial vs. final) pattern we observed has theoretical and methodological implications. Recognition of code-switches, as with words in unilingual utterances, might be affected by lexical content, structural position, contextual information, etc., all of which might affect switch cost.

Sentence-final code-switched targets might additionally be cued by the gradual lowering of f0 present throughout all of our English sentences signaling the end of the stimulus and a last possible location for the target to occur. Information structure and task demand might therefore result in no cost to recognizing a code-switch at the end of a sentence, since the position of a final target word becomes more and more predictable as the utterance progresses.

A limitation of this study is a possible confound between splicing and the effect of withholding phonetic cues. Our splicing manipulation was confined to identity-splicing unilingual English sentences and cross-splicing Mandarin code-switches into English sentences that were originally unilingual. A complete splicing design would also include cross-splicing English words into code-switched sentences and identity-splicing Mandarin code-switches into code-switched sentences. As the study stands, any code-switched stimuli that are spliced also lack anticipatory phonetic cues to a code-switch. The inclusion of spliced code-switched stimuli that retain phonetic cues (i.e., identity-spliced code-switched sentences) would allow a more certain assessment of whether it was purely withholding phonetic cues that resulted in slower bilingual recognition of the code-switch. The tone-specific patterns we observed argue against a uniform effect of splicing. Nevertheless, a complete splicing paradigm would be desirable, for teasing apart effects of splicing and tonal coarticulation for each target tone and sentence position.

Conclusion

Our study adds to previous literature on phonetic reflexes and cuing of code-switches, in showing that removing the preceding phonetic context through splicing can make recognition of a code-switch more difficult in certain contexts. Thus, the present study contributes to a rapidly growing body of literature investigating whether code-switches must necessarily incur a processing cost for listeners. It is becoming apparent that phonetic cues play a role in code-switched recognition (Piccinini & Garellek, 2014; Fricke et al., 2016), though this sensitivity depends on the phonetic features of the languages being switched. Previous studies have indicated the facilitatory role of phonetic cues on recognition of a code-switch, and we have further shown that Mandarin–English bilingual listeners can be negatively affected by the removal of phonetic cues, as listeners expect to continue hearing the matrix language unless they are given phonetic reasons to expect otherwise. Mandarin–English bilinguals are sensitive to code-switching pronunciation, and pitch is possibly one of many interacting anticipatory phonetic cues to a code-switch.

References

Allopenna, PD, Magnuson, JS and Tanenhaus, MK (1998) Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language* **38**, 419–439.

Altmann, GT and Kamide, Y (2009) Discourse-mediation of the mapping between language and the visual world: Eye movements and mental representation. *Cognition* **111**, 55–71.

Baayen, RH, Davidson, DJ and Bates, DM (2008) Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language* **59**, 390–412.

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/

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

Cacoullos, RT and Aaron, JE (2003) Bare English-origin nouns in Spanish: Rates, constraints, and discourse functions. *Language Variation and Change* **15**, 289–328.

Cooper, RM (1974) The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology* **6**, 84–107.

Costa, A and 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**, 491–511.

Fricke, M, Kroll, JF and Dussias, PE (2016) Phonetic variation in bilingual speech: A lens for studying the production-comprehension link. *Journal of Memory and Language* **89**, 110–137.

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

Gollan, TH, Kleinman, D and Wierenga, CE (2014) What's easier: Doing what you want, or being told what to do? Cued versus voluntary language and task switching. *Journal of Experimental Psychology: General* **143**, 2167–2195.

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

Grosjean, F (2012) An attempt to isolate, and then differentiate, transfer and interference. *International Journal of Bilingualism* **16**, 11–21.

Huettig, F, Rommers, J and Meyer, AS (2011) Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica* **137**, 151–171.

Kleinman, D and Gollan, TH (2016) Speaking two languages for the price of one: Bypassing language control mechanisms via accessibility-driven switches. *Psychological science* **27**, 700–714.

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

Mirman, D (2014) *Growth curve analysis and visualization using R*. Boca Raton, FL: CRC Press.

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

Olson, DJ (2017) Bilingual language switching costs in auditory comprehension. *Language, Cognition, and Neuroscience* **32**, 494–513.

Piccinini, PE and Garellek, M (2014) Prosodic cues to monolingual versus code-switching sentences in English and Spanish. In Campbell N, Gibbon D and Hirst D (eds), *Proceedings of the 7th International Conference on Speech Prosody*. Dublin, Ireland: Trinity College Dublin, pp. 885–889.

Poplack, S and Dion, N (2012) Myths and facts about loanword development, *Language Variation and Change* **24**, 279–315.

Rossion, B and Pourtois, G (2004) Revisiting Snodgrass and Vanderwart's object pictorial set: The role of surface detail in basic-level object recognition. *Perception* **33**, 217–236.

Soares, C and Grosjean, F (1984) Bilinguals in a monolingual and a bilingual speech mode: The effect on lexical access. *Memory and Cognition* **12**, 380–386.

Tanenhaus, MK, Spivey-Knowlton, MJ, Eberhard, KM and Sedivy, JC (1995) Integration of visual and linguistic information in spoken language comprehension. *Science* **268**, 1632–1634.

Thomas, MS and Allport, A (2000) Language switching costs in bilingual visual word recognition. *Journal of Memory and Language* **43**, 44–66.

Xu, Y (1997) Contextual tonal variations in Mandarin. *Journal of Phonetics* **25**, 61–83.