Advisor: Matthew Goldrick
Second Reader: Ann Bradlow

Will knowing more English reduce the perceptual illusions in Mandarin Chinese?

Qile Chen, Matthew Goldrick

November 18, 2021

Abstract: Word-medial consonant clusters (e.g. /ʃp/ in 'dishpan') are illicit in Mandarin Chinese. When Mandarin listeners perceive this sound sequence, it has been claimed that their speech perception processes will 'repair' it by adding an illusory medial vowel /i/. In an AX discrimination task (where participants decide whether two stimuli are the same or different), we ask whether the knowledge of a language with a less restrictive phonotactic system, in this study, English, will reduce the perceptual illusions resulting from the knowledge of a language with a more restrictive phonotactic system, Mandarin. Critical trials have a short vowel [i] or [y] and a long vowel with the same quality (/i/ or /y/). We predict that, for [y] trials, the perceptual repair process replaces [y] with /i/, leading to fewer incorrect 'same' responses; for [i] trials, [i] is replaced with /i/, leading to more incorrect 'same' responses. Critically, for [i] trials, the stronger perceptual repair process for Mandarin-dominant bilinguals will lead to more incorrect 'same' responses relative to English-dominant bilinguals. However, the results did not follow our predictions. There was no perceptual repair effect measured by the proportion of incorrect 'same' response of [i] trials over that of the control group, the [y] trials, and the degree of dominance in English was neither found to affect the results. We conclude by discussing directions for future research.

Introduction

Along with many other methods, the phenomenon of perceptual vowel illusions—when a vowel is inserted into a syllable during the process of speech perception—has been instrumental in understanding speech processing mechanisms. Previous research revealed some sources of perceptual vowel illusions such as violations of language universals that argue for broad phonological restrictions cross-linguistically (Berent et al., 2007; Berent, Lennertz, Jun, Moreno, & Smolensky, 2008; Berent, Lennertz, Smolensky, & Vaknin-Nusbaum, 2009) and confounding acoustic cues. With respect to the latter, some tokens in Dupoux, Parlato, Frota, Hirose, & Peperkamp (2011) and its shadowing experiment in Wilson, Davidson, & Martin (2014) indicated the influence of co-articulatory cues, which did not explain the effect of vowel epenthesis per se. Not only can the nature of perceptual vowel illusions be comprehended through the lenses above, it can also be seen in the consonant clusters that are illegal to certain language phonology, vindicated on monolingual as well as bilingual listeners.

Listeners utilize language-specific repair strategies when they perceive illegal phoneme sequences according to the rules that govern those sequences, namely, phonotactics. Previous research studies have revealed the illusory perception of a prothetic or epenthetic vowel as one repair strategy to consonant clusters that violate phonotactic constraints in a particular language (Dupoux, Kakehi, Hirose, Pallier, & Mehler, 1999; Berent et al., 2007; Durvasula & Kahng, 2016; Kabak & Idsardi, 2007). Taking Japanese as an example, Dupoux et al. found that when native listeners hear a non-word such as [ebna], their perceptual systems were more likely to resolve it as /ebɯna/ than French listeners due to the unacceptability of consonant cluster [bn] in Japanese language phonotactics. Similar to Japanese, Mandarin phonotactics does not permit complex consonant clusters. Compared to English monolinguals, Mandarin listeners perceived an illusory /i/ between [tɕ^{h}] and [m] (Durvasula, Huang, Uehara, Luo, & Lin, 2018[1]). These two studies are both based on the results of vowel identification and ABX discrimination task, in which Mandarin speakers made more errors than English speakers when asked to pick whether [atɕ^{h}ima] or [atɕ^{h}ma] was more similar to [atɕ^{h}ima].

Furthermore, another aspect of the literature on the quality of the illusory vowel has added the condition of phonological alternations to epenthetic vowel perception; in other words, language-specific phonological context will influence listeners' expectation of the quality of the illusory vowel. Presented with the same stimuli, listeners of different native languages heard different illusory vowels, and the choice of the illusory vowel is determined by the minimal vowel in the inventory (Dupoux et al., 2011). Based on the patterns of Japanese loanwords, Monahan, Takahashi, Nakao, & Idsardi (2009) and Mattingley, Hume, & Currie (2015) concluded that there can be more than one illusory vowel in a language, such as Japanese to have /ɯ/ as illusory vowel in non-palatal and non-coronal contexts but /i/ in palatal context and /o/ before coronal consonant. Durvasula and Kahng (2015) looked at the same issue. Tested on Korean speakers, they found that Korean listeners perceive /ɨ/ as the illusory vowel in alveolar contexts, but /i/ next to a palatal fricative, and both [ɨ] and [i] next to palatal stops. In addition, according to the results from ABX task and vowel identification task in Durvasula et al. (2018), Mandarin speakers confused [atɕ^{h}ima~atɕ^{h}ma] and [at^{h}əma~at^{h}ma] more than English speakers, suggesting that they perceived an illusory /i/ in the illegal alveopalatal consonant contexts, but /ə/ after illegal alveolar stop coda contexts.

So far only the cases of monolinguals have been introduced, but evidence also revealed that L2-dominant late bilinguals can apply repair strategies in L2, suggesting that language phonotactics could be learned in a second language. For example, Parlato-Oliveira (2010) found in the results of a

---

[1] Other modifications in response to the word-medial consonant cluster violation in Mandarin have not been thoroughly investigated. We acknowledge that vowel insertion/epenthesis is not the only method to deal with disfavored consonant clusters, and there are other perceptual modifications (e.g. vowel deletion process: [dripi] → /ripi/ in Lin, 2003; Davidson and Shawn, 2012), but in this paper we mainly focus on epenthetic vowel insertion.

vowel identification task that Japanese-Brazilian Portuguese bilinguals' choice of repair strategy depended on their age of acquisition and the task. Unlike the first-generation immigrants who were Japanese-dominant and chose to repair the illicit consonant clusters with /ɯ/, the second-generation immigrants who were Portuguese-dominant chose /i/ as the epenthetic vowel. Moreover, bilinguals exhibit gradient influence from the phonotactic systems of the two languages they know. Since word-initial /s/-consonant clusters are illegal in Spanish but permissible in English, Carlson, Goldrick, Blasingame, & Fink (2016) tested Spanish monolinguals and Spanish-English bilinguals' perception of the illusory vowel [e] preceding such clusters in vowel identification task. The results showed that the repair effect was weaker in bilinguals than in monolinguals; furthermore, it was weaker in English-dominant bilinguals than in Spanish-dominant bilinguals. They concluded on this issue that more knowledge of English will reduce the perceptual repair risen from ambiguous and illegal #VsC sequence in Spanish.

These studies indicate that perceptual repair effects are subject to the composition of language experience, thus generalization from a group of bilinguals to another requires considerations of heterogeneity. This insight aligns with the usage-based view of language knowledge that draws evidence from psycholinguistics and functional/cognitive linguistics (Bybee and Hopper 2001; Hall, Cheng, & Carlson, 2006). It was also put forward that the more constant and frequent a particular language pattern appears in one's interaction with others, the more likely the pattern will be stored, and as a result the speaker's language knowledge is shaped. One group difference of the patterns falls on how bilinguals use their languages; more specifically, how frequently bilinguals switch and mix their languages everyday vary across groups, and in some literature it is correlated with performance on lexical control (Basnight-Brown & Altarriba, 2007; Heredia & Altarriba, 2001; Prior & Gollan, 2011; Verreyt, Woumans, Vandelanotte, Szmalec, & Duyck, 2016). For instance, Mandarin-English bilinguals and Spanish-English bilinguals differ in their self-reported frequency of language switching, and the former code-switch less in daily language use than the latter in the U.S. (Prior & Gollan, 2011). Given all the indications for the importance of heterogeneity, we want to conduct a study on Mandarin-English bilinguals to confirm the generalizability of the results of Carlson et al. (2016).

We will then introduce the phonological patterns in Mandarin and English that explain the rationale behind the perceptual repair strategy in the current experiment. Alveo-palatal consonants [tɕ, tɕ^h, ɕ] in Mandarin can only precede the high front vowels [i] and its rounded counterpart [y] (Cheng, 1966). At the same time, the consonant-vowel phonotactic structure in Mandarin prohibits the occurrence of a consonant cluster such as [atɕ^{h}ma], resulting in listeners' biased perceptions of /i/ at the word-medial position (Durvasula et al., 2018).

In the current study, we follow Carlson et al. (2016) and use an AX discrimination task to test for perceptual illusions. This task will ask participants to compare the two stimuli and decide whether they are the same or not. Ideally, use of this task will lead participants to rely on pre-lexical perceptual processing of the acoustic cues more than metalinguistic knowledge. Critical trials have a short vowel [i] or [y] and a long vowel with the same quality (/i/ or /y/). We predict that, for [y] trials, the perceptual repair process replaces [y] with /i/, leading to fewer incorrect 'same' responses. For [i] trials, [i] is replaced with /i/, leading to more incorrect 'same' responses. Critically, for [i] trials, the stronger perceptual repair process for Mandarin-dominant bilinguals will lead to more incorrect 'same' responses relative to English-dominant bilinguals.

Note that our stimuli include [y] and [ɕ], which are not native to English inventory. We do not believe this to be a problem since the bilinguals should have substantial experience with both of them when they use Mandarin. However, if it was ever controversial that these phonemes were non-native to late bilinguals of Mandarin, we proposed that the phonemes would be assimilated to listeners' L1 (English) phoneme categories, according to Perceptual Assimilation Theory (PAM) (Best, 1994; Best & Tyler, 2007). That is to say, if /u/ and /y/ in Mandarin are perceived as good exemplars of the same category [u] in English, it is very likely that native speakers of English will not tell /u/ and /y/ apart due to assimilation, supported by the evidence of English listeners perceiving French /y/ (Levy, 2009). Similarly, the alveo-palatal consonants in Mandarin are close enough to the palato-alveolar consonants [tʃ, ʃ] in English, yielding the perceptions of /ʃi/ and /ʃu/ instead of [ɕi] and [ɕy]. This complexity of English perception of non-native phonemes should not affect the main results because neither /i/ nor /u/ are illusory vowels in English.

To summarize, we expected that among the Mandarin-dominant bilinguals, the error rate for the /i/ condition would be higher than for the /y/ condition in our AX task, on account of /i/ being the illusory vowel after alveo-palatal consonants in Mandarin. Furthermore, we supposed the performance difference between /i/ and /y/ conditions to be on the decrease as participants' knowledge of English increases, for English is phonotactically less restrictive than Mandarin.

Methods

Participants

We decided to recruit 66 participants after conducting Monte Carlo Power analysis (Gelman & Hill, 2006). We took the effect size and variance estimates based on the regression for AX 'same-vowel' trials from Carlson et al. (2016). We used this model to generate a set of simulated data, specifying the number of simulated participants. A logistic regression model predicting "same" responses from language dominance, medial vowel, and their interaction was fit to the data with the random effects of intercepts by subjects and a random slope for vowel by subjects. And then we used a likelihood ratio test to evaluate whether there was significant effect of the interaction between dominance and medial vowel. We stimulated the analyses at least 1000 times for varying numbers of participants. The estimated power exceeded 0.8 at 66 participants (assuming half were Mandarin-dominant and half were English-dominant). Sixty-six Mandarin-English bilingual participants in the U.S. (Age range = 18-30, mean = 22.85) were recruited through word of mouth. Due to tax purposes and other travel regulations during COVID-19, we did not use Monolingual Mandarin speakers as a control group. The bilinguals recruited here are divided into Mandarin- and English-dominant groups depending on their language proficiency.

Participants performed MINT Sprint (Garcia & Gollan, 2021), a time-pressured picture naming task that was expanded from Multilingual Naming Test (Gollan, Weissberger, Runnqvist, Montoya, & Cera, 2012) to measure their language proficiency objectively. The same 80 pictures were presented in both Mandarin and English with the order of the language counterbalanced. Among the bilinguals, 40 were Mandarin-dominant (Mean age = 23.53, SD = 2.47) who reported using English for mean years of 15.14 (SD = 4.57), and 26 were English-dominant (Mean age = 21.81, SD = 2.79) who reported using Mandarin for mean years of 19.5 (SD = 5.72).

Besides one participant who did not fill out their age of acquisition for listening, in the rest of the bilinguals, 48 reported learning Mandarin first (Mean AoA for English = 9.63, SD = 6.10), and 12 of them became dominant in English (MINT Sprint English score was higher than Mandarin score). Four bilinguals started listening to English first, and another 13 bilinguals reported the same age of acquisition for Mandarin and English (Mean AoA for Mandarin = 3.06, SD = 6.20). Out of these 17 bilinguals, 14 were English-dominant.

Participants' linguistic background was assessed using LHQ3.0 (Li, Zhang, Yu, & Zhao, 2020) which included self-reported language proficiency and patterns of language usage in each language they know. For age of acquisition (AOA), since some answers conflicted with the answers in year of use (YoU), we standardized AoA by subtracting the YoU from age. The numbers in self-ratings of Mandarin and English abilities were transformed from a 7-point scale ranging from "very poor" to "excellent." The daily engagement of each language was calculated by summing up the total hours participants spent on six common activities such as watching TV. A summary of raw MINT scores and the descriptive data are presented in Table 1.

		Mandarin-dominant bilinguals (Mean-SD)	English-dominant bilinguals (Mean-SD)
AOA (Listening)	Mandarin	0.98 (3.23)	2.31 (4.96)
	English	9.54 (6.67)	3.88 (5.10)
Years of use	Mandarin	22.55 (3.99)	19.50 (5.72)
	English	15.14 (4.57)	18.64 (4.60)
Self-ratings (Mandarin)	Listening	6.63 (0.74)	6.12 (0.99)
	Speaking	6.48 (0.85)	5.54 (1.24)
	Reading	6.53 (0.78)	4.96 (1.46)
	Writing	6.08 (1.12)	4.52 (1.87)
Self-ratings (English)	Listening	5.59 (1.23)	6.15 (1.12)
	Speaking	4.92 (1.29)	5.88 (1.40)
	Reading	5.36 (1.16)	5.50 (1.82)
	Writing	4.90 (1.37)	5.00 (2.21)
Daily engagement (hrs)	Mandarin	5.63 (4.16)	1.65 (1.72)
	English	10.85 (10.06)	10.38 (3.36)
Raw MINT scores (out of 80)	Mandarin	63.63 (3.81)	45.38 (11.20)
	English	45.5 (8.13)	68.31 (5.04)

Table 1. *Participants' mean age of use (AoA), years of use (YoU), self-ratings (on a scale of 1-7), daily engagement (hours), and MINT scores with standard deviation.*

Materials

Modeled after the AX task in Carlson et al. (2016), the materials consisted of 240 trials. A female native speaker of Mandarin Chinese produced non-words of the format *aɕVCo*. [ɕ] is a palatal-fricative that can be followed by [i] or [y] (*V*) in Mandarin. *C* consisted of 10 different consonants that can precede [o] and can follow [i]/[y] in Mandarin. Combining the two vowels and 10 consonants generated 20 items: [aɕipo], [aɕito], [aɕiko], [aɕibo], [aɕido], [aɕigo], [aɕifo], [aɕilo], [aɕimo], [aɕino], [aɕypo], [aɕyto], [aɕyko], [aɕybo], [aɕydo], [aɕygo], [aɕyfo], [aɕylo], [aɕymo], [aɕyno]. Regarding the lexical tones in Mandarin, the speaker produced the first and second syllables with first/level tone, and the third syllable with neutral tone (āɕV̄Co).

Following the design of Carlson et al. (2016), the stimuli varied in the length of the medial vowels in the critical trials (Same Vowel and Different Vowels). Each trial in the AX task had two non-words with the same consonant after the medial vowels (see Table 2). The medial vowel lengths in critical trials were always different. On 'Same Vowel' trials, the medial vowels had the same quality; while on 'Different Vowel' trials, they were different. The medial vowel lengths in filler trials were always the same. There were two types of filler trials with the medial vowels being either identical or non-identical.

Trial types	Medial vowel	Medial vowel length	Trial count (Total: 240)
Critical trials	Same vowel	Short - Long	20
		Long - Short	20
	Different vowels	Short - Long	20
		Long - Short	20
Fillers	Non-identical	Short - Short	20
		Long - Long	20
	Identical	Short - Short	60
		Long - Long	60

Table 2. *Trial types and counts in the AX discrimination task.*

In order to create the stimuli, the full-length medial vowel was first excised from the onset to the end of the medial vowel's periodicity (e.g. from [aɕipo] to [aɕpo]). The stimuli with short medial vowels (shown with a diacritic ĭ) were then generated by adding 4 periods of the excised vowel (e.g. [aɕĭpo]), which is approximately 15.4 ms. The stimuli with long medial vowels were generated by adding 16 periods, which is approximately 61.6 ms.

Procedure

Participants were first asked to play an audio file to ensure that their Internet and browser supported the experiment, plus that their headphones/speakers were functioning properly. They then reviewed a consent form in English, but instructions during the AX discrimination task were only in Mandarin Chinese for the sake of maximizing the activation of Mandarin. Due to the varying levels of literacy amongst Mandarin-English bilinguals, we recorded the instructions so people with less knowledge of the written form (such as heritage speakers) could understand as well.

At the beginning of each AX trial, participants encountered a fixation cross at the center of the screen for 500 ms, then the Chinese characters of "please listen" *(*请听*)* was presented together with the first auditory stimulus (the *A*). After a silent inter-stimulus interval (ISI) of 250 ms, the second stimulus (the *X*) played. The Chinese characters stayed on the screen until the end of *X*. For each trial, participants responded by pressing 'F' if they thought A and X stimuli were exactly the same, or 'J' if different.

With regard to the implementation of ISI, our intention was to encourage direct acoustic-level comparisons between the two sound tokens in a trial. After hearing stimulus A, the short ISI potentially limited listeners' phonetic categorizations, rendering the interpretation of the responses possible to reflect low-level perceptual consequences of Mandarin phonotactic under the influence of

English knowledge. Therefore, we followed the 250 ms ISI in some previous work (Carlson et al., 2016; Davidson, 2011; Pisoni, 1973).

All bilinguals were tested through online platforms (Firebase, Qualtrics, and LHQ3.0) with their own equipment in the U.S.

Results

Critical trials - Same Vowel

The logistic mixed effects regression was used to analyze the incorrect "same" responses in Same Vowel condition in the critical trials. The correct response was "different" because the two sounds in an AX trial differed only on the length of the medial vowel, being either long or short. In order to compare to the findings in Carlson et al. (2016), we used orthogonal contrasts to code participants' language groups (either Mandarin-dominant or English-dominant; referred to as Model Group). We also used the centered ratios of MINT scores (English over Mandarin) as a continuous independent variable (Model Ratio). In both models, the factor "medial vowel" was also contrast-coded based on the quality of the medial vowel in the AX trial being [i] or [y], which interacted with the measure of dominance. By-participant varying intercept, by-participant varying medial vowel slope, as well as a correlation of these terms were included as random effects in both of the models described above.

There was no perceptual repair effect. This effect would be reflected by greater errors on /i/ relative to the baseline condition /y/. However, Figure 1 suggested either the opposite (error rate for /y/ was higher than /i/) or no significant difference between the two conditions. For the main effect of medial vowel, which indexes the perceptual repair effect, the model failed to reject the null hypothesis (Model Ratio: $\beta = -0.25$, SE = 0.15, $\chi^2(1) = 2.72$, p = 0.0994; Model Group: -0.25, SE = 0.15, $\chi^2(1) = 2.83$, p = 0.0925). There was no significant main effect of the dominance measures (Model Ratio: $\beta = -0.13$, SE = 0.22, $\chi^2(1) = 0.38$, p = 0.5391; Model Group: $\beta = 0.07$, SE = 0.24, $\chi^2(1) = 0.09$, p = 0.7606), nor were the interactions significant (Model Group: $\beta = 0.07$, SE = 0.28, $\chi^2(1) = 0.06$, p = 0.8114; Model Ratio: $\beta = -0.14$, SE = 0.26, $\chi^2(1) = 0.29$, p = 0.5873), suggesting that perceptual repair effect was not reduced as the knowledge of English increased. Compared to Carlson et al. (2016), the mean error rates on AX discrimination were higher in the current study (Carlson et al. baseline condition [a] < 0.7 error vs. current study [y] > 0.8 error) .

We also examined logged RTs for correct responses in the Same Vowel condition using linear mixed effects regression with dominance group or centered dominance ratio, and medial vowel being the independent variables, including the interaction between the dominance measure and medial vowel (see Fig. 2). The same random effect structure as in the previous models was used. Outliers (any reaction time two standard deviations above or below each participant's mean within each condition) were removed, but we found no significant main effects when the dominance measure was centered ratio (Centered ratio: β = -0.01, SE = 0.04, t = -0.24, p = 0.8488; Medial vowel: β = 0.02, SE = 0.03, t = 0.58, p = 0.6654) nor when it was dominance group (Main effect of dominance group: β = 0.03, SE = 0.05, t = 0.65, p = 0.6331; Main effect of medial vowel: β = -0.05, SE = 0.09, t = -0.57, p = 0.6702). The results also indicated that the RTs were not significantly different across medial vowel conditions within each dominance group (Model Ratio: β = -0.004, SE = 0.05, t = -0.07, p = 0.9555; Model Group: β = 0.04, SE = 0.05, t = 0.78, p = 0.5783).

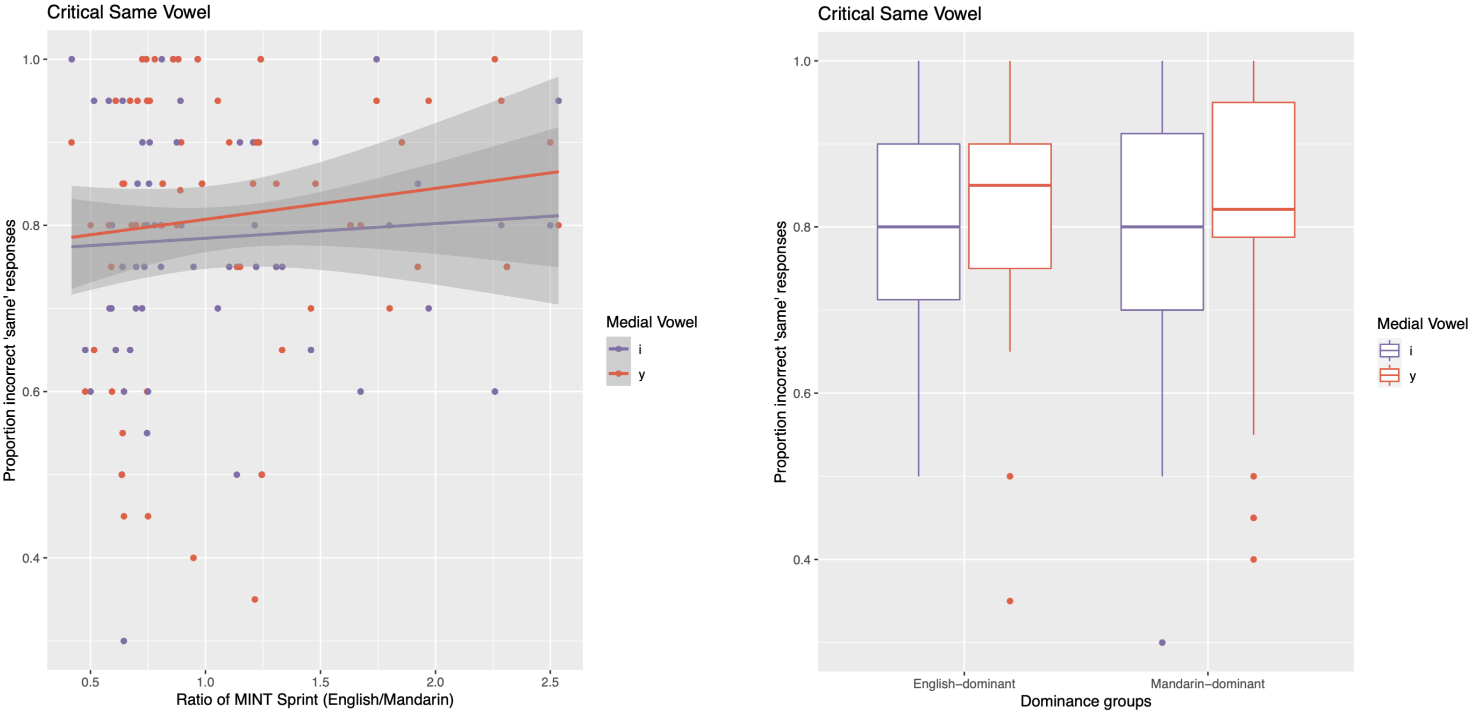

Figure 1. *Left: proportion of critical AX pairs incorrectly judged to be the "same", by score ratio of MINT Sprint (a picture naming task). Right: proportion of critical AX pairs incorrectly judged to be the "same", by dominance group. Stimuli in Same Vowel were acoustically identical except for the duration of the medial vowels. The grey area shows 95% confidence intervals.*

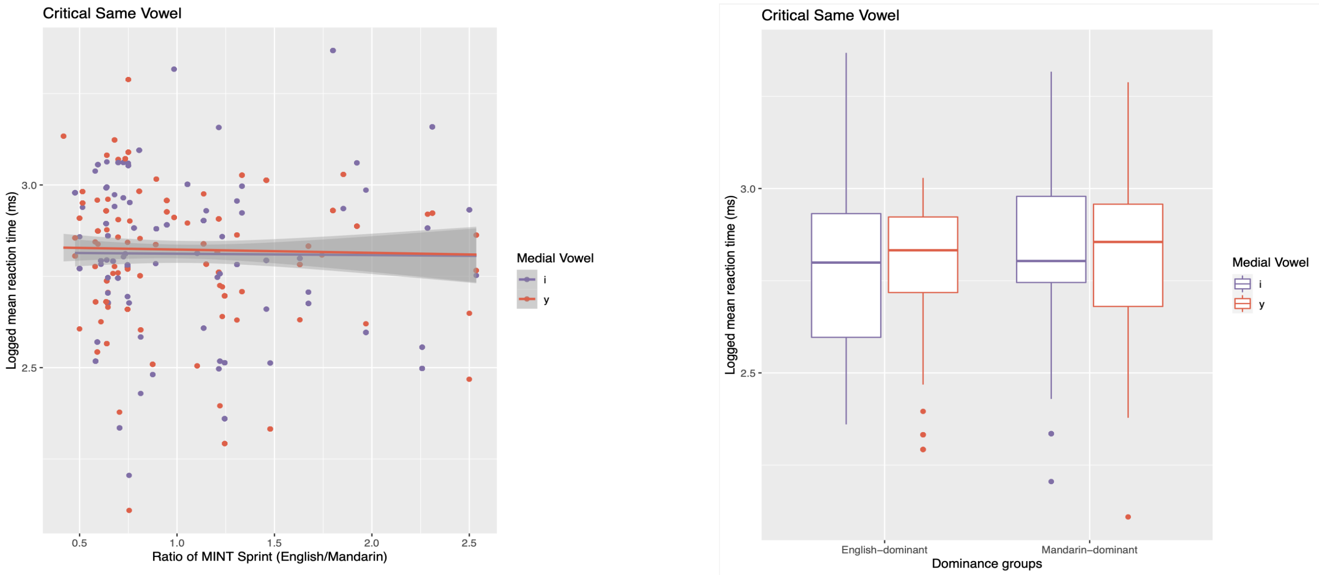

Figure 2. *Logged mean RTs by score ratio of MINT Sprint (left), and by two dominance groups (right) in Same Vowel condition.*

Critical trials - Different Vowels

Similar models were applied to the analyses of the incorrect "same" responses in Different Vowels. The correct response was "different" because both the quality ([i] or [y]) as well as the length (long or short) of the medial vowel varied in an AX trial. Together with the factors of language group (for Model Group) and centered MINT score ratio (for Model Ratio), we contrast-coded the factor "longer vowel" to indicate whether the longer vowel in an AX trial was [i] or [y] ([aɕy̆po] vs. [aɕipo] or [aɕĭpo] vs. [aɕypo]). By-participant varying intercept, by-participant varying longer vowel slope, as well as a correlation of these terms were included as random effects.

The discrimination performance was not significantly improved when the longer vowel was [y]. Even though the regression line for [y] in Figure 3 was below the one for [i], the main effect of longer vowel was not significant in either model (Model Ratio: $\beta = 0.22$, SE = 0.12, $\chi^2(1) = 3.31$, p = 0.0690; Model Group: $\beta = 0.21$, SE = 0.12, $\chi^2(1) = 3.01$, p = 0.0827). Based on the results of the main effect of dominance measures (Model Ratio: $\beta = -0.18$, SE = 0.23, $\chi^2(1) = 0.61$, p = 0.4354; Model Group: $\beta = 0.46$, SE = 0.25, $\chi^2(1) = 3.28$, p = 0.0701) and interactions (Model Ratio: $\beta = -0.08$, SE = 0.21, $\chi^2(1) = 0.14$, p = 0.7054; Model Group: $\beta = 0.06$, SE = 0.23, $\chi^2(1) = 0.08$, p = 0.781), dominance was not found to modulate the effects of the longer vowel in Different Vowels trials. The mean error rates in this study are much higher (> 0.6 for either vowel) than those in Carlson et al. (2016) (< 0.5 for [e] and < 0.25 for baseline condition [a])

Examinations on Logged RTs for correct responses in the Different Vowels condition used linear mixed effects regression with dominance group and longer vowel being the independent variables, including the interaction between the dominance measure and medial vowel (see Fig. 4). Outliers were removed according to the same criterion as in Same Vowel trials, and the same random effect was used. We did not find any significant difference of RT for dominance group (β = 0.02, SE = 0.04, t = 0.47, p = 0.7203), centered ratio (β = -0.01, SE = 0.04, t = -0.27, p = 0.8321), nor longer vowel (Model Ratio: β = 0.02, SE = 0.02, t = 0.67, p = 0.6242; Model Group: β = -0.04, SE = 0.08, t = -0.52, p = 0.6947). The increase of participants' knowledge of English did not change the effect of longer vowel (Model Ratio: β = 0.002, SE = 0.04, t = 0.05, p = 0.9682; Model Group: 0.04, SE = 0.05, t = 0.76, p = 0.5863).

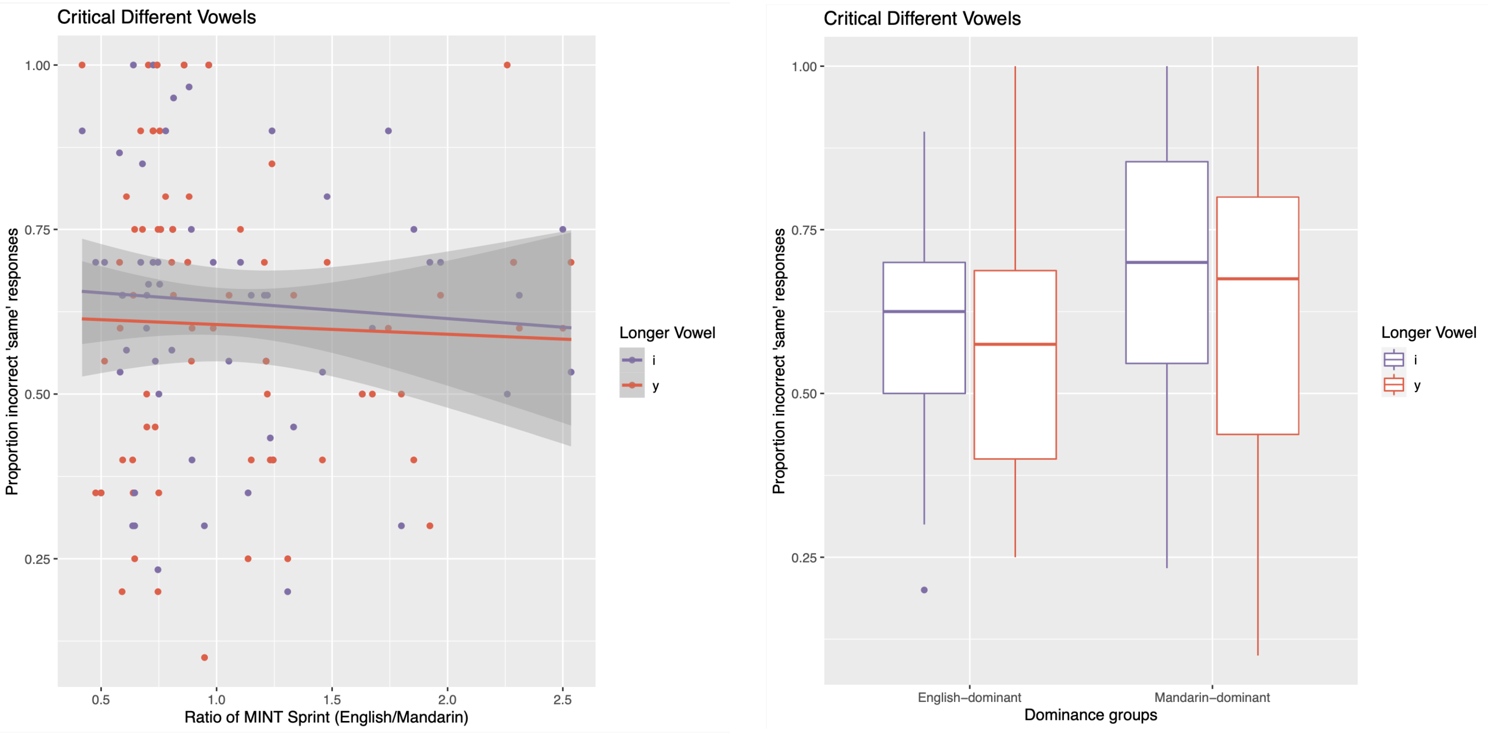

Figure 3. *Left: proportion of critical AX pairs incorrectly judged to be the "same", by score ratio of MINT Sprint (a picture naming task). Right: proportion of critical AX pairs incorrectly judged to be the "same", by dominance group. Stimuli in Different Vowels were different both on the duration and the quality of the medial vowels. The grey area shows 95% confidence intervals.*

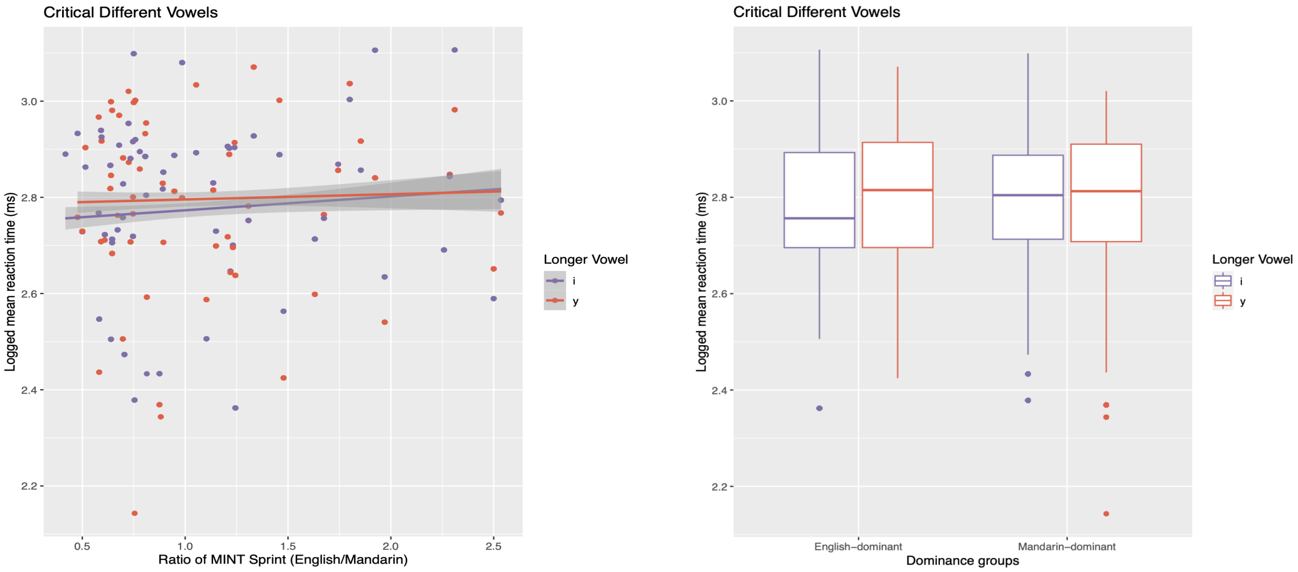

Figure 4. *Logged mean RTs by score ratio of MINT Sprint (left), and by two dominance groups (right) in Different Vowels condition.*

Signal detection analyses

In order to know whether the filler-identical trials had any effect on the main results, we examined the filler-identical trials along with the critical-same trials and conducted signal detection analyses. The question we were probing was whether participants actually differentiated the medial vowel length difference, which could be tested by comparing the responses between the "critical-same" and "filler-identical" trials. In addition, this analysis allowed us to examine whether the perceptual bias to respond "same" for each vowel in the two trial types was different across English-dominant versus Mandarin-dominant bilinguals. Contrast-coded factors included participants' response of "same" or "different," medial vowel being either [i] or [y], and the trial type being either "critical-same" or "filler-identical." We centered both the dominance score (MINT Sprint score) in English (Model E) and Mandarin (Model M). For random effects, we applied by-participant varying intercept, correlated with by-participant varying medial vowel slope and trial type slope.

Overall, the intercepts of both models above were biased toward responding "same" over "different" (Model E: $\beta = 2.77$, SE = 0.13; Model M: $\beta = -0.27$, SE = 0.13). For each medial vowel, the significant main effect of medial vowel revealed that participants were biased to answer "same" more under [y] condition than [i] condition (Model E: $\beta = -0.27$, SE = 0.13, $\chi^2(1) = 4.70$, p = 0.0324; Model M: $\beta = -0.27$, SE = 0.13, $\chi^2(1) = 4.47$, p = 0.0345). Participants with different dominance scores of the same language were not significantly different in terms of their bias to answer "same" over "different" (Model E: $\beta = 0.01$, SE = 0.01, $\chi^2(1) = 0.80$, p = 0.3713; Model M: $\beta = 0.002$, SE = 0.01, $\chi^2(1) = 0.03$, p

= 0.8673), nor was their bias of answering “same” more for [y], which is derived from the results of the interaction between medial vowel and dominance (Model E: $\beta = -0.01$, SE = 0.01, $\chi^2(1) = 1.62$, $p = 0.2024$; Model M: $\beta = 0.01$, SE = 0.01, $\chi^2(1) = 0.83$, $p = 0.3617$).

The significant main effect of trial type showed that participants did discriminate the medial vowel length difference in the Same Vowel condition of critical trials (Model E: $\beta = 2.36$, SE = 0.18, $\chi^2(1) = 92.12$, $p < 2.22e-16$; Model M: $\beta = 2.36$, SE = 0.18, $\chi^2(1) = 92.76$, $p < 2.22e-16$). This is equivalent to *d'*, the measure of listeners' sensitivity in signal detection theory. This discrimination was not significantly different across the two medial vowel conditions (Model E: $\beta = 0.05$, SE = 0.18, $\chi^2(1) = 0.06$, $p = 0.8144$; Model M: $\beta = 0.07$, SE = 0.18, $\chi^2(1) = 0.13$, $p = 0.7203$), which indicates there is no effect of perceptual bias. Nor was the discrimination different across different levels of dominance in English vs. Mandarin (Model E: $\beta = 0.01$, SE = 0.01, $\chi^2(1) = 0.09$, $p = 0.7638$; Model M: $\beta = 0.01$, SE = 0.01, $\chi^2(1) = 0.71$, $p = 0.3985$). Eventually, the three-way interaction showed no significant effect in either model (Model E: $\beta = -0.02$, SE = 0.01, $\chi^2(1) = 1.91$, $p = 0.1671$; Model M: $\beta = 0.01$, SE = 0.02, $\chi^2(1) = 0.66$, $p = 0.417$), suggesting that there was not a different perceptual bias effect across participants with different dominances in English and Mandarin. In a word, the main findings on the critical trials above were not shifted by the results in signal detection analyses.

Consonants

In addition to the confirmatory analyses and signal detection analyses, we also explored some other factors that potentially have influenced the results. As introduced in the methods section, ten different consonants followed the medial vowel in the stimuli, thus we examine whether the distribution of incorrect “same” responses were different across the consonants (see Fig. 5). The figure showed us that only the consonants [p], [t], and [k] fell into our prediction that participants would have a higher error rate for [i] condition than for [y] condition. After filtering out the rest of the consonant trials, we ran the analyses on Same Vowel condition in the critical trials. The insignificant main effect of medial vowel in the two models failed to reject the null hypothesis (Model Ratio: $\beta = 0.01$, SE = 0.20, $\chi^2(1) = 0$, $p = 1$; Model Group: $\beta = 0.01$, SE = 0.20, $\chi^2(1) = 0$, $p = 1$). Performance difference between the two medial vowels was not reduced as the knowledge of English increased as shown in both the main effect of dominance (Model Ratio: $\beta = 0.20$, SE = 0.22, $\chi^2(1) = 0.82$, $p = 0.3659$; Model Group: $\beta = 0.02$, SE = 0.24, $\chi^2(1) = 0.01$, $p = 0.926$) and the interactions (Model Ratio: $\beta = -0.14$, SE = 0.26, $\chi^2(1) = 0.18$, $p = 0.673$; Model Group: $\beta = 0.07$, SE = 0.28, $\chi^2(1) = 0.01$, $p = 0.9255$).

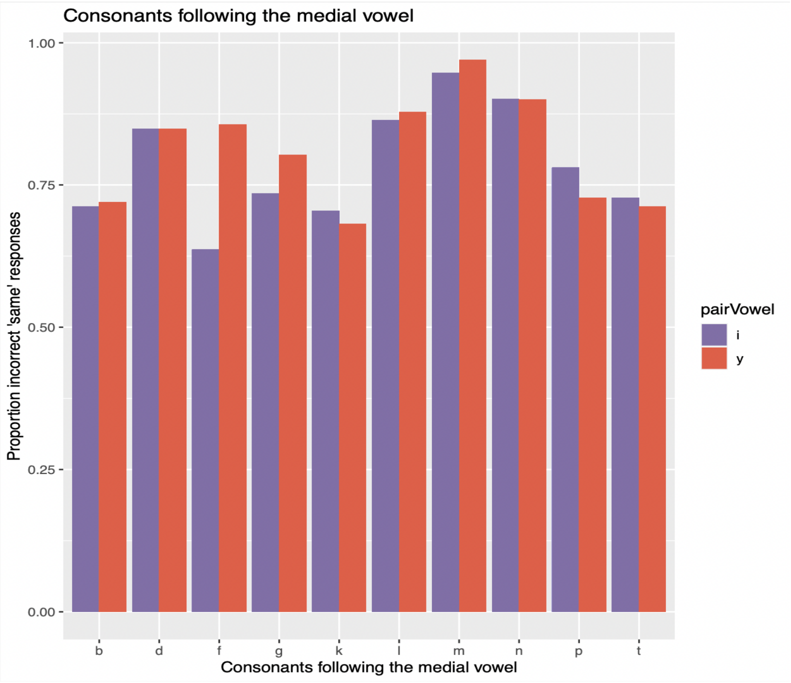

Figure 5. *The proportion of incorrect "same" responses by different consonants following the medial vowel.*

Daily engagement

Apart from the consonants, we investigated measurements other than picture naming task scores that could possibly represent the participants' knowledge of English and Mandarin. Hours of daily engagement with the target language might better reflect the amount of language exposure at the time when they participated (shown in Fig. 6). The hour of engagement was calculated by summing up the self-reported hours the participants spent on six common daily activities: watching TV, listening to radio, reading, writing, social media, and Internet.

Following the models we built in the critical trial (Same Vowel), we separated the participants into engagement groups depending on which language they had more hours of engagement with and contrast-coded the variables. The ratio of daily engagement (English/Mandarin) as a continuous variable was also used, but answers that were larger than 20 hours as well as ratios that were more than 100 were excluded. We found neither significant main effects nor interactions, again failing to support a perceptual repair effect, and this statement did not change as the relative daily engagement of English increased in both Model Ratio (Medial vowel: $\beta = -0.26$, SE = 0.33, $\chi^2(1) = 0.56$, p = 0.4542; Interaction: $\beta = 0.05$, SE = 0.11, $\chi^2(1) = 0.19$, p = 0.6665) and Model Group (Medial vowel: $\beta = -0.72$, SE = 0.43, $\chi^2(1) = 2.41$, p = 0.1207; Interaction: $\beta = -1.22$, SE = 0.83, $\chi^2(1) = 1.83$, p = 0.176).

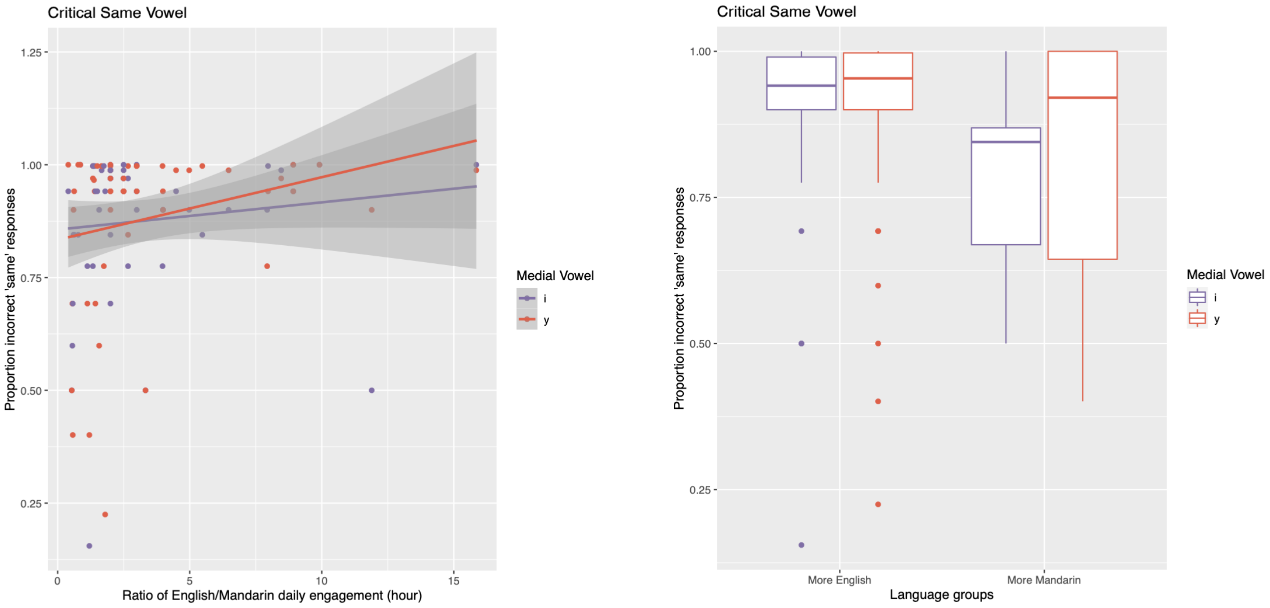

Figure 6. *Left: proportion of incorrect "same" responses by the ratio of daily engagement (English/Mandarin; only ratio less than 100). Right: proportion of incorrect "same" responses by language engagement groups.*

Age of acquisition (AoA)

In order to see if the perceptual repair effect was present within a particular range of early AoA, we filtered the age of listening acquisition for English in the samples we had according to different cut-offs, and we found from the figures that the means for [y] condition were above [i] condition across all the cut-offs (see Fig. 7). This finding was consistent with the main result in critical trials (Same Vowel), and it suggested that the absence of the perceptual repair effect was most likely not due to the variance of AoA.

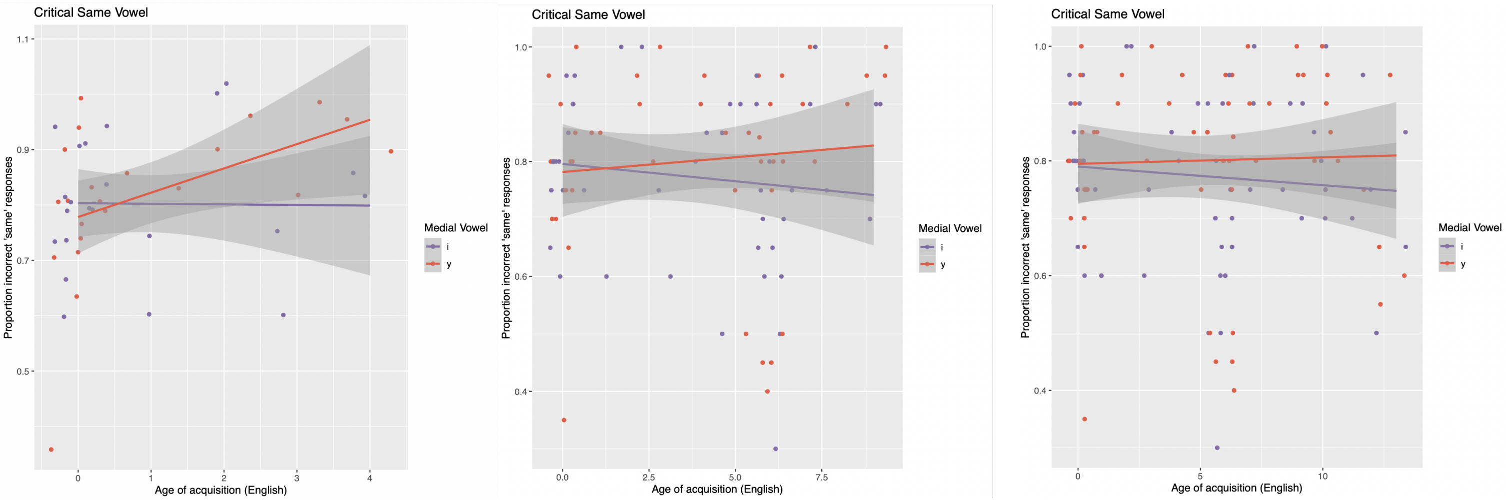

Figure 7. *Proportion of incorrect "same" responses by different cut-offs for age of acquisition (AoA) in English listening. Left: AoA less than 5 years. Middle: AoA less than 10 years. Right: AoA less than 15 years.*

Mixing frequency

Finally, we looked into the language mixing frequency measured in a seven-point scale, since mixing frequency has shown to be correlated with inhibitory control and phonological processing ability, which could influence perceptual illusions. Point 1 represents "never" and point 7 represents "always" as the frequency of mixing English and Mandarin (see Fig. 8). The logistic mixed-effect model used contrast-coded medial vowel and the centered mixing frequency between English and Mandarin as independent variables, then the random effect with by-participant varying intercept, medial slope, and a correlation of these two was added to the fixed effect model. The results showed no significant main effects nor interaction, again failing to support a perceptual illusion effect (Mixing frequency: $\beta = -0.08$, SE = 0.11, $\chi^2(1) = 0.60$, p = 0.4389; Medial vowel: $\beta = -0.21$, SE = 0.21, $\chi^2(1) = 3.40$, p = 0.065; Interaction: $\beta = 0.10$, SE = 0.08, $\chi^2(1) = 1.36$, p = 0.2428).

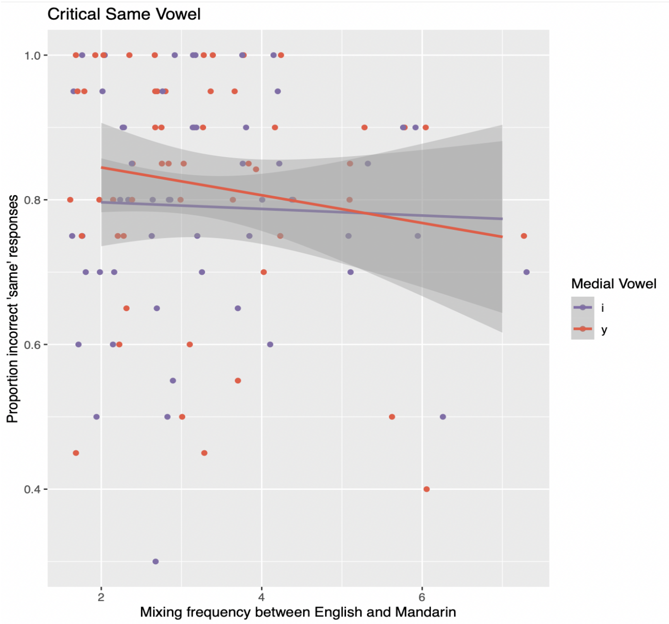

Figure 8. *Proportion of incorrect "same" responses by mixing frequency between English and Mandarin (1 = never, 7 = always).*

General Discussion

The current research inquired into the relationship between phonotactic constraints and perceptual illusions influenced by bilingual language experience. Certain groups of bilinguals perceive illusory vowels as a language-specific repair strategy in response to the violation of phonotactic constraints. For example, the Mandarin language disallows consonant clusters while the English language allows their occurrence. In theory, Mandarin speakers are likely to hear an epenthetic vowel /i/ in alveo-palatal consonant clusters. Critically, in bilingual speakers the knowledge of English should mitigate this perceptual illusion (Carlson et al., 2016). Mandarin-English bilinguals with more knowledge of English are proposed to be less affected by the perceptual repair effect in Mandarin phonotactics. We used an AX task to test this prediction. However, the work failed to yield the predicted results using a non-illusory vowel /y/ as the baseline to measure the perceptual repair effect of the illusory vowel /i/.

Although the findings in the current experiment did not support our hypothesis that the knowledge of English will modulate Mandarin-English bilinguals' illusory perception when listening to phonotactically illicit sound sequences in Mandarin, there are some possible reasons that point to the directions for future research studies. First, the difficulty of the task might have reached the ceiling of the participants' ability. As it came out as a surprise that there was no significant difference between the two vowel conditions ([i] and [y]) in critical trials, we also observed that the mean error rates of the current study were higher than Carlson et al. (2016) (see Fig. 1&2). Changes could be made to make the current task easier, such as increasing the vowel length difference between the short and long stimuli in an AX trial, or making the medial vowel absent or present. This would be consistent with the approach in Durvasula et al. (2018). They supported the claim that [i] is the illusory vowel after alveo-palatal consonants in Mandarin from the results of ABX task and vowel identification task in Durvasula et al. (2018). Critically, in the stimuli they compared [i] with no vowel instead of short [i] with long [i] (and short [y] with long [y]).

Another possibility is that the perceptual repair effect on [i] (compared to [y]) might occur at different stages of perceptual processing. For example, vowel identification task (e.g. Exp. 1 in Carlson et al.) or syllable counting (e.g. de Jong & Park, 2012) are relatively offline tasks that involve metalinguistic knowledge, occuring in the later stages of perceptual processing. In the contrary, the discrimination task with short inter-stimulus interval used here occurs in the earlier stages. Future confirmation of whether participants perceptually repair /i/ more than /y/ in later stages of processing can therefore be conducted using paradigms other than discrimination tasks.

The hypothetical illusory vowel in the current study is modulated by phonological context. In other words, [i] is the illusory vowel in Mandarin after alveo-palatal consonant, but the result of vowel identification in Durvasula et al. (2018) found that Mandarin listeners heard [ə] after a stop consonant in word-medial position. Therefore, even though there was no direct comparison between [i] and [ə] in AX discrimination task, we do not eliminate the possibility that using [ə] could result in different results from the current ones.

Still another possibility is that the results in Carlson et al. (2016) might be unreplicable due to small sample size in the original AX task. The power analysis we conducted was based on their experimental design, but with doubled the number of participants (32 in Carlson et al. and 66 in current study) we did not find significant effect of perceptual repair, leading to this feasible explanation. For future replications of Carlson et al. on Spanish-English bilinguals, increasing the sample size would test this explanation, especially if the results turned out to be different from the original study.

Furthermore, the reason that we did not see a perceptual repair effect could perhaps be the stronger perceptual repair effect for word-initial position compared to word-medial position. A recent study Leung et al. (2021) found that in an AX task among the Spanish- and Mandarin-speaking English learners, while Spanish speakers performed worse with prothesis compared to epenthesis at word initial position, they did better with prothesis at word medial position. For Mandarin speakers, responses did not reflect such a distinction. Based on their results, the word-initial place of perceptual repair (#sC) in Carlson et al. (2016) could allow Spanish-English bilinguals to show stronger effects relative to Mandarin-English bilinguals at word-medial position. To test this claim, we could replicate the current study with stimuli with epenthetic vowel at word initial position (ɕV̄Co) to verify if word position also influence epenthesis in Mandarin.

In addition to all the reasons above, it is possible that intergroup differences could potentially influence the results. Due to the different dominance measures used in Carlson et al. (2016), parallel comparisons could not be made; however, we do not believe this to be the case. For example, if a higher proportion of Mandarin-dominant bilinguals switched dominance compared to the Spanish-dominant bilinguals, we would predict that the Mandarin-dominant bilinguals were better at inhibiting English when processing Mandarin, thus less influenced by English. However, we observed the opposite in our data. While 28 out of thirty-two participants in Carlson et al. learned Spanish first (Mean AoA for English = 7.46, SD = 4.37), 13 of which became English-dominant, 48 participants in the current study learned Mandarin first (Mean AoA for English = 9.63, SD = 6.10), and one fourth of them became English-dominant. These differences would suggest that there should have been more of the perceptual repair effect in Mandarin-dominant bilinguals. The absence of repair effect could not be explained as the result of the switch of dominance.

There may be other differences in experience that could lead to the different patterns of results., Carlson and McAllister (2019) reported evidence on the phenomenon that Andalusian Spanish speakers reduced the intensity of the vowel in #VsC word production, and in some words they completely deleted it. This finding supported phonetic reduction as a hypothesized contribution to the perceptual repair effect, since the reduced vowel encountered in speech would need to be restored during perception. Spanish-English bilinguals in Carlson et al. (2016) may therefore have learned to repair /e/ in perception due to phonetic reduction. In contrast, there is no evidence that Mandarin-English bilinguals would drop [i] or [y] in these word-medial clusters.

Conclusion

For bilinguals who have two contrasting phonotactic systems, the AX discrimination results reported here failed to reject the null hypothesis about perceptual repair effect on word-medial illicit consonant clusters. Alterations of our paradigm in future studies may help to answer the question of the degree of interaction between two phonotactically contrasting languages in a bilingual mind.

References

Basnight-Brown, D. M., & Altarriba, J. (2007). Differences in semantic and translation priming across languages: The role of language direction and language dominance. *Memory & cognition*, *35*(5), 953-965.

Berent, I., Steriade, D., Lennertz, T., & Vaknin, V. (2007). What we know about what we have never heard: Evidence from perceptual illusions. *Cognition*, *104*(3), 591-630.

Berent, I., Lennertz, T., Jun, J., Moreno, M. A., & Smolensky, P. (2008). Language universals in human brains. *Proceedings of the National Academy of Sciences*, *105*(14), 5321-5325.

Berent, I., Lennertz, T., Smolensky, P., & Vaknin-Nusbaum, V. (2009). Listeners' knowledge of phonological universals: Evidence from nasal clusters. *Phonology*, *26*(1), 75-108.

Best, C. T. (1994). The emergence of native-language phonological influences in infants: A perceptual assimilation model. *The development of speech perception: The transition from speech sounds to spoken words*, *167*(224), 233-277.

Best, C. T., & Tyler, M. D. (2007). Nonnative and second-language speech. *Language experience in second language speech learning: In honor of James Emil Flege*, *17*, 13.

Bybee, J. L., & Hopper, P. J. (Eds.). (2001). *Frequency and the emergence of linguistic structure* (Vol. 45). John Benjamins Publishing.

Carlson, M. T., Goldrick, M., Blasingame, M., & Fink, A. (2016). Navigating conflicting phonotactic constraints in bilingual speech perception. *Bilingualism: Language and Cognition*, *19*(5), 939-954.

Carlson, M. T., & McAllister, A. (2019). I've heard that one before: Phonetic reduction in speech production as a possible contributing factor in perceptual illusory vowel effects. *Studies in Hispanic and Lusophone Linguistics*, *12*(2), 281-308.

Cheng, R. L. (1966). Mandarin phonological structure. *Journal of Linguistics*, *2*(2), 135-158.

de Jong, K., & Park, H. (2012). Vowel epenthesis and segment identity in Korean learners of English. *Studies in Second Language Acquisition*, *34*(1), 127-155.

Davidson, L. (2011). Phonetic and phonological factors in the second language production of phonemes and phonotactics. *Language and Linguistics Compass*, *5*(3), 126-139.

Davidson, L., & Shaw, J. A. (2012). Sources of illusion in consonant cluster perception. *Journal of phonetics*, *40*(2), 234-248.

Dupoux, E., Kakehi, K., Hirose, Y., Pallier, C., & Mehler, J. (1999). Epenthetic vowels in Japanese: A perceptual illusion?. *Journal of experimental psychology: human perception and performance*, *25*(6), 1568.

Dupoux, E., Parlato, E., Frota, S., Hirose, Y., & Peperkamp, S. (2011). Where do illusory vowels come from?. *Journal of memory and language*, *64*(3), 199-210.

Durvasula, K., & Kahng, J. (2016). The role of phrasal phonology in speech perception: What perceptual epenthesis shows us. *Journal of phonetics*, *54*, 15-34.

Durvasula, K., Huang, H. H., Uehara, S., Luo, Q., & Lin, Y. H. (2018). Phonology modulates the illusory vowels in perceptual illusions: Evidence from Mandarin and English. *Laboratory Phonology Journal of the Association for Laboratory Phonology*.

Garcia, D. L., & Gollan, T. H. (2021). The MINT Sprint: Exploring a Fast Administration Procedure with an Expanded Multilingual Naming Test. *Journal of the International Neuropsychological Society*, 1-17.

Gelman, A., & Hill, J. (2006). *Data analysis using regression and multilevel/hierarchical models*. Cambridge university press.

Gollan, T. H., Weissberger, G. H., Runnqvist, E., Montoya, R. I., & Cera, C. M. (2012). Self-ratings of spoken language dominance: A Multilingual Naming Test (MINT) and preliminary norms for young and aging Spanish–English bilinguals. *Bilingualism: language and cognition*, *15*(3), 594-615.

Hall, J. K., Cheng, A., & Carlson, M. T. (2006). Reconceptual- izing multicompetence as a theory of language knowledge. *Applied Linguistics, 27*(*2*), 220–240.

Heredia, R. R., & Altarriba, J. (2001). Bilingual language mixing: Why do bilinguals code-switch?. *Current Directions in Psychological Science*, *10*(5), 164-168.

Kabak, B., & Idsardi, W. J. (2007). Perceptual distortions in the adaptation of English consonant clusters: Syllable structure or consonantal contact constraints?. *Language and speech*, *50*(1), 23-52.

Leung, A. H. C., Young-Scholten, M., Almurashi, W., Ghadanfari, S., Nash, C., & Outhwaite, O. (2021).

(Mis) perception of consonant clusters and short vowels in English as a foreign language. *International Review of Applied Linguistics in Language Teaching*.

Levy, E. S. (2009). On the assimilation-discrimination relationship in American English adults' French vowel learning. *The Journal of the Acoustical Society of America*, *126*(5), 2670-2682.

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

Lin, Yuh-Huey. 2003. Interphonology variability: Sociolinguistic factors affecting L2 simplification strategies. Applied Linguistics 24(4). 439–464.

Mattingley, W., Hume, E., & Currie Hall, K. 2015. The influence of preceding consonant on perceptual epenthesis in japanese. In: T. S. C. for ICPhS 2015 (ed.), *Proceedings of the 18th International Congress of Phonetic Sciences, 0888.1–5.*

Monahan, P. J., Takahashi, E., Nakao, C., & Idsardi, W. J. (2009). Not all epenthetic contexts are equal: Differential effects in Japanese illusory vowel perception. *Japanese/Korean Linguistics*, *17*, 391-405.

Parlato-Oliveira, E., Christophe, A., Hirose, Y., & Dupoux, E. (2010). Plasticity of illusory vowel perception in Brazilian-Japanese bilinguals. *The Journal of the Acoustical Society of America*, *127*(6), 3738-3748.

Pisoni, D. B. (1973). Auditory and phonetic memory codes in the discrimination of consonants and vowels. *Perception & psychophysics*, *13*(2), 253-260.

Prior, A., & Gollan, T. H. (2011). Good language-switchers are good task-switchers: Evidence from Spanish–English and Mandarin–English bilinguals. *Journal of the International Neuropsychological Society*, *17*(4), 682-691.

Verreyt, N., Woumans, E. V. Y., Vandelanotte, D., Szmalec, A., & Duyck, W. (2016). The influence of language-switching experience on the bilingual executive control advantage. *Bilingualism: Language and Cognition*, *19*(1), 181-190.

Wilson, C., Davidson, L., & Martin, S. (2014). Effects of acoustic–phonetic detail on cross-language speech production. *Journal of Memory and Language*, *77*, 1-24.