CAMBRIDGE
UNIVERSITY PRESS

ORIGINAL ARTICLE

How bidialectalism affects non-native speech acquisition: Evidence from Shanghai and Mandarin Chinese

Xiaoluan Liu[1] and Paola Escudero[2]

[1]Department of English, School of Foreign languages, East China Normal University, Zhongshan, China and
[2]The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Penrith, Australia
Corresponding author: Xiaoluan Liu; Email: LXL0803@outlook.com

Abstract
The current study examines how bidialectalism influences non-native speech production. We compared monodialectal Mandarin Chinese with bidialectal Shanghai-Mandarin Chinese speakers in terms of their ability to produce easy and difficult American English vowels. The results showed a general advantage for the bidialectal group compared with the monodialectal group in the production of the vowel formants and duration of the easy English vowels [i] and [u]. However, for the English vowels [ɪ] and [ʊ] known to be difficult for Chinese learners of English, both groups experienced the same challenges in terms of accurately producing the formants of the target vowels. Nevertheless, the bidialectal Shanghai-Mandarin speakers were still better than the monodialectal Mandarin speakers in the durational aspect of the two difficult English vowels. The results are explained by the Second Language Linguistic Perception (L2LP) model and suggest that the bidialectal advantage in non-native speech acquisition is subject to the modulation of cross-linguistic difficulty of the target speech sounds.

Introduction

A growing body of research has shown that bilinguals tend to have advantages in learning additional languages compared to monolinguals (Abu-Rabia & Sanitsky, 2010; Hirosh & Degani, 2018), with regard to both language-general proficiency (Swain et al., 1990) and language-specific skills (Klein, 1995). Nevertheless, in the domain of non-native phonological/phonetic acquisition, studies on the influence of bilingualism have rendered mixed results (Antoniou et al., 2015; Elvin et al., 2018; Escudero et al.. 2016), which therefore requires further exploration. The current study presents a unique contribution to this line of research by examining the influence of bilingualism on non-native phonetic learning through the lens of

bidialectalism (i.e., speaking a dialect besides a standard language) in the context of non-native speech production. This is because compared with the majority of research on bilingualism in foreign language learning, relatively little attention has been paid to the influence of bidialectalism on non-native phonetic learning, especially in terms of how bidialectalism interacts with cross-language similarities/difficulties of the phonetic targets in non-native speech production. In the present study, we focused on the comparison between bidialectal Shanghai-Mandarin Chinese speakers and monodialectal Mandarin Chinese speakers when producing American English vowels that were judged to be either easy or difficult for Chinese learners of English, depending on the cross-linguistic relationships between English and Chinese vowels.

Bilingualism effects

Bilingualism refers to one's ability to understand, speak, and frequently use two languages (Luk & Bialystok, 2013). So far, there is no consensus as to the influence of bilingualism on non-native phonetic learning, probably because compared with lexical and grammatical aspects of language learning, phonetic learning in a foreign language is more complicated due to the complexities and difficulties in learning non-native speech contrasts (Colantoni et al., 2015). Some studies have shown that bilinguals have advantages over monolinguals in learning non-native speech sounds. For example, Cohen et al. (1967) found that bilinguals generally were more accurate than monolinguals in producing non-native phoneme sequences. Similarly, Enomoto (1994) found that bilinguals outperformed monolinguals in perceptually differentiating between Japanese phonemic contrasts. Recent studies such as Singh et al. (2016) also found a bilingual advantage in integrating lexical tones into novel word learning.

Nevertheless, some studies have failed to identify a consistent advantage for bilinguals in distinguishing non-native speech contrasts. For example, Werker (1986) found no significant difference between bilinguals (L1 English with different L2 backgrounds) and English monolinguals in their ability to differentiate between Hindi retroflex and dental contrasts, as well as velar and uvular contrasts. Similarly, Patihis et al. (2015) found that Spanish–English bilingual individuals were no better than English monolinguals and worse than Armenian–English bilinguals in discrimination of L3 Korean stop consonants. Escudero et al. (2013) also found that bilingualism (L1 Spanish–L2 English) did not help the learning of L3 Dutch vowels. In addition, Kopečková (2016) found that the bilingual advantage (L1 German–L2 English) in L3 learning of Spanish rhotic sounds is not broad-based; rather, it is subject to the difficulty and learnability of the non-native phonetic features. This finding echoes the results in Antoniou et al. (2015) where the bilingual advantage did not apply to learning all non-native speech contrasts: the advantage was more obvious when the target foreign sound contrasts were easy (e.g., retroflex); for difficult contrasts (e.g., lenition), the bilingual advantage was not sufficient, because other factors such as phonetic similarity between languages also played a significant role. Similarly, Escudero et al. (2016) found an overall advantage for Singaporean English–Mandarin bilinguals when learning CVC words that formed non-minimal pairs but no specific advantages for vowel minimal pairs compared with Australian

English monolinguals. These mixed findings collectively indicate that the impact of bilingualism on phonetic learning in a foreign language may be influenced by the acoustic properties of the non-native speech sounds in relation to the learner's native language. Additionally, it suggests that certain speech sounds may pose universal challenges in learning, irrespective of one's linguistic background (i.e., bilingual or not) (Antoniou et al., 2015). Hence, these studies suggest a need to further investigate how the relations between L1 and L2 acoustic properties modulate the effect of bilingualism on learning non-native sounds.

Bidialectalism effects

Bidialectals are those who can fluently speak a standard language and a regional dialect. The existing literature mainly focuses on the relations between bidialectalism and executive functions, which so far have presented mixed results. Some studies have reported a potential advantage for bidialectals. For example, Antoniou et al. (2016) found that bidialectals (Cypriot and Standard Modern Greek) were similar to bilinguals and outperformed monolinguals in working memory and inhibitory control tasks. Some studies also suggest that bidialectalism may specifically impact certain aspects of executive functions. For example, Blom et al. (2017) found that Limburgish-Dutch bidialectal children were significantly different from monolingual Dutch children in a selective attention task but not in a flanker task. Similarly, Oschwald et al. (2018) only found a positive relation between bidialectalism and working memory but failed to find such an association in other measures of executive functions. Furthermore, the frequency of language usage may also play a role, for example, as found in Poarch et al. (2019), bidialectal language usage patterns can influence the relations between bidialectalism and executive functions, that is, those who used the nonstandard dialect more frequently had better executive control skills than monolinguals. On the other hand, some studies have failed to discover cognitive advantages for bidialectals. For instance, in Ross and Melinger (2017), no significant differences were found between bidialectal and monolingual children in inhibitory control and shifting tasks. In studies where bidialectal participants were older adults, results have shown that bidialectals were similar to monolinguals in executive control tasks (Kirk et al., 2014; Scaltritti et al., 2017).

In the domain of speech acquisition, the impact of bidialectalism on non-native phonetic learning still calls for more research efforts. The existing studies are mainly focused on speech perception, with the results suggesting that dialectal differences can significantly affect one's accuracy in perception of non-native vowels. For example, Escudero and Williams (2012) compared Peruvian Spanish (PS) and Iberian Spanish (IS) learners regarding non-native Dutch vowel discrimination. They found that IS learners were better than PS learners at differentiating between the Dutch vowel contrasts. The results suggest that acoustic characteristics of vowels of one's native language or dialect have a direct impact on L2 vowel perception. Similarly, Escudero et al. (2012) found that non-native speech perception was significantly influenced by regional/dialectal differences in the listener's L1. Specifically, they compared native speakers of North Holland Dutch with those of Flemish Dutch in terms of their perception of English vowel contrast (/ɛ/ vs. /æ/).

The results showed that the dialectal differences in vowel production by two groups of speakers led to different vowel categorization responses. Some studies have also demonstrated the impact of a possible activation switch between different modes of languages/dialects on learning an additional language. For instance, Williams and Escudero (2014a) compared Northern and Southern British listeners in their perceptual categorization of non-native Dutch vowels. Interestingly, they found that the Northern listeners' categorization of Dutch vowels was influenced by their knowledge about the acoustic patterns of the standard Southern British vowels, possibly due to the activation of the Southern British English mode of speech perception during the laboratory testing session.

The present study

The above literature review on phonetic learning and bilingualism/bidialectalism is mainly centered around the research question of whether knowing a second language/dialect would benefit phonetic learning of a third language, that is, does knowing one more language/dialect lead to an advantage in learning speech sounds of a new language? The mixed results of previous studies as reviewed above suggest that the answer should take into account the cross-linguistic influences between native and non-native sounds. Specifically, the acoustic characteristics and learning difficulty of the non-native speech sounds in relation to the native sound system could play a significant role in determining how bilingualism/bidialectalism influences non-native speech learning (Antoniou et al., 2015; Elvin et al., 2018; Escudero et al. 2013, 2016; Kopečková, 2016). For adult learners, the acquisition of sounds in a new language is usually influenced by the learner's experience with speech sounds in previously acquired languages.

Indeed, well-established models of L2 perception/production such as the Second Language Linguistic Perception (L2LP) model (Escudero, 2005, 2009; Escudero & Yazawa, in press; van Leussen & Escudero, 2015) state that the acquisition of non-native speech sounds is related to the influence of L1. In the context of the present study, the L2LP model is suitable because it applies to both monolingual and bilingual/bidialectal learners (Escudero et al., 2016; Escudero & Yazawa, in press). Furthermore, L2LP strives to comprehensively model the whole developmental trajectory in non-native speech learning, spanning from novice to advanced learners (for more details see Escudero 2005, 2009; Escudero & Yazawa, in press; van Leussen & Escudero, 2015), and is thus suitable for the present study where participants were learners with prior exposure to the target non-native speech sounds. Particularly, the L2LP model can provide explanations as to why, despite years of dedicated efforts, the ultimate mastery of L2 production and perception may not be fully attained, due to the activation of L1 for sequential bilinguals (i.e., L2 learners) whose onset of L2 learning is after early childhood (Escudero & Yazawa, in press).

Since the L2LP model has the word "perception" in it, one may wonder whether it is appropriate to use this model to explain non-native speech production. Admittedly, the L2LP model was initially developed for speech perception, but it has been extended to explain speech production and lexical development (e.g., Elvin et al., 2016; 2020; Elvin & Escudero, 2019; Escudero et al., 2022; Escudero & Yazawa,

in press; van Leussen & Escudero, 2015; Yazawa et al., 2020; Yazawa et al., 2023). Crucially, other models of L2 speech do not make explicit predictions about the possible shifts between different language or dialect modes for bilingual or bidialectal speakers because they assume a single phonetic space for an L2 learner's two languages (see Colantoni et al., 2015 for a thorough comparison between L2 speech models). In contrast, the L2LP model explicitly predicts that bilinguals and bidialectals have two separate systems readily accessible, including separate perception and production grammars (Escudero, 2005, 2009; Escudero & Yazawa, in press; van Leussen & Escudero, 2015; Yazawa et al., 2020, 2023). Therefore, when bilinguals and bidialectals learn an additional language (L3, L4, etc.), the speech sounds of the additional language could be mapped to either their first or second language or dialect (Escudero et al., 2013; Williams & Escudero, 2014a). This further suggests that bilinguals and bidialectals may switch between different modes when learning an additional language because their separate systems could be activated selectively (Escudero, 2005, 2009; Escudero & Yazawa, in press; Williams & Escudero, 2014a). Whether this holds true for bidialectals' non-native speech production remains to be explored.

In addition, one may wonder why bidialectalism is worth examination. This is because firstly, compared with bilingualism, the effect of bidialectalism is largely under-recognized and undervalued, leaving much room for future research (Antoniou et al., 2016; Oschwald et al., 2018; Poarch et al., 2019). Secondly, bidialectals could be different from bilinguals because of the "ubiquitous usage of both dialects in their environment compared to bilinguals who may display a more compartmentalized language usage pattern" (Poarch et al., 2019: 613). This may reveal how frequency of language use affects the learning of a subsequent language in bilinguals versus bidialectals, the understanding of which is currently unclear (cf. Antoniou et al., 2016; Oschwald et al., 2018). Research of this kind could also make people appreciate the effect of bidialectalism, which could further contribute to people's understanding of their own identity as well as how dialects could relate to learning a foreign language (Antoniou et al., 2016).

The above review suggests that it is still not clear how bilingualism and bidialectalism could influence individuals acquiring sound categories in a new language. Particularly, the impact of bidialectalism on non-native speech production remains largely unexplored. It is important to address this issue, as a substantial portion (approximately 50% to 70%) of the global population possesses proficiency in multiple languages or dialects (Grosjean, 2021). This percentage further increases in regions where multiple dialects are prevalent. However, there is a prevailing issue in English as a Second Language research that inaccurately portrays English learners as monolingual speakers, thus failing to represent the reality of English learners worldwide (Leivada et al., 2023). Therefore, the present study is among the few that confronts this bias through a comparative study between Chinese speakers in Beijing where the majority are monodialectal in Mandarin Chinese and Chinese speakers in Shanghai where individuals use two dialects (Mandarin and Shanghai Chinese) in their daily life. Specifically, this study examines the production of American English vowels by monodialectal Mandarin Chinese speakers compared with bidialectal Shanghai-Mandarin Chinese speakers.

The choice of American English aligns with the current English teaching environment in China, where American English is the dominant target L2 variety.

Mandarin Chinese is the official standard language of China, while Shanghai Chinese is mainly spoken in the city of Shanghai. Shanghai Chinese belongs to the Wu family of Chinese dialects. As noted in Chao (1967), Chinese dialects are "primarily different in phonology, secondarily in lexicon and least in grammatical structure" (pp. 92–93). In terms of phonology, one of the most prominent distinctions between Shanghai and Mandarin Chinese is that Shanghai Chinese has a larger vowel inventory containing 15 monophthongs (6 monophthongs also found in Mandarin: /i, y, a, u, ɤ, ə/ and 9 only found in Shanghai Chinese: /ɛ, ø, o, ɔ, ɪ, ʏ, ɐ, ʊ, ɑ/) and 8 diphthongs (3 diphthongs also found in Mandarin Chinese /ia, ie, ua/ and 5 only found in Shanghai Chinese: /iɔ, iɤ, ue, uø, yø/) (Chen, 2008; Chen & Gussenhoven, 2015; Yu et al., 2004), while Mandarin Chinese has a smaller vowel inventory containing 6 monophthongs (/i, y, a, u, ɤ, ə/) and 11 diphthongs (/ai, au, ou, uo, ei, ye, ie, ia, ua, uə, iu/) (Lee & Zee, 2003). In addition, only Shanghai Chinese contains short vowels such as [ɪ] and [ʊ], which sound similar to (but are not exactly the same as) the English [ɪ] and [ʊ] vowels, as detailed in the next paragraph. These "short" vowels only occur in closed syllables that end with a glottal stop coda in Shanghai Chinese (Chen, 2008), while Mandarin does not have "short" vowels because the oral stop coda has been lost historically, resulting in vowel length variation being more relevant for Shanghai Chinese than for Mandarin. Therefore, the contrast between the vowel inventory of Mandarin Chinese and Shanghai Chinese makes an ideal test case for our study.

As reviewed, how bilingualism and bidialectalism influence non-native speech learning could be related to the learning difficulty of the non-native speech sounds. Therefore, in the present study, the target American English vowels were classified into two categories of difficulty for Chinese speakers: easy and difficult. Based on previous research (Chen et al., 2001; Jia et al., 2006), the easy American English vowels chosen for the present study were [i] and [u] because: a) they are found in Chinese (Shanghai and Mandarin) and English, and b) Chinese speakers produce these two English vowels with high accuracy (Jia et al., 2006). The difficult American English vowels chosen for the present study were [ɪ] and [ʊ] because: (a) they are unfamiliar to Mandarin speakers (Chen et al., 2001), and (b) Mandarin speakers produce these vowels differently from native American English speakers (Chen et al., 2001; Jia et al., 2006). For Shanghai Chinese speakers, the two vowels could also be difficult because according to the L2LP model, bidialectals could map the incoming non-native speech targets to either their first or second language/dialect (Escudero et al., 2013; Williams & Escudero, 2014a). This means that bidialectals such as Shanghai-Mandarin Chinese could map the American English [ɪ] and [ʊ] to either the [ɪ] and [ʊ] in Shanghai Chinese or the [i] and [u] in Mandarin Chinese, which as a result could interfere with the effective establishment of the target non-native sounds.

In sum, the present study is concerned with how bidialectalism interacts with cross-language difficulties of the phonetic targets in non-native speech acquisition. We aimed at answering the following research questions: (1) Do Shanghai-Mandarin Chinese bidialectal speakers differ from monodialectal Mandarin Chinese speakers in their production of easy and difficult American English vowels? If so, in

which acoustic dimension do the two groups differ, vowel formants or duration? (2) How does the vowel system of the participants' Chinese dialects influence their production of the non-native English vowels? The two groups of Chinese participants were asked to produce the target American English vowels [i], [ɪ], [u], [ʊ], and their native Mandarin Chinese [i] and [u] vowels; additionally, the bidialectal Shanghai-Mandarin speakers were also asked to produce their native Shanghai Chinese [ɪ] and [ʊ] vowels.

Based on previous research where the bilingual advantage was more evident in learning easy non-native speech sounds (Antoniou et al., 2015), we hypothesize that bidialectal Shanghai-Mandarin Chinese speakers could outperform monodialectal Mandarin Chinese speakers in accurately producing the easy English vowels [i] and [u], which will be reflected in smaller formant and durational differences from American English speakers' production. The bidialectal advantage of Shanghai-Mandarin Chinese speakers in non-native English vowel production may become less apparent for the difficult English vowels [ɪ] and [ʊ] in certain acoustic aspects due to the influence of their native languages. Specifically, given that Shanghai Chinese contains short vowels whereas Mandarin does not, we speculate that the bidialectal Shanghai Chinese production of the difficult English vowels [ɪ] and [ʊ] may approach American speakers' production more closely in duration than Mandarin Chinese speakers would do. Nevertheless, both groups (Shanghai and Mandarin Chinese) could be similarly deviant from American speakers in terms of the formants of the two difficult English vowels due to the influence of their native Chinese. Following the L2LP model's proposal, the bidialectal speakers may switch between the two languages/dialects when learning an additional language, resulting in their mapping of the non-native English vowels to either Shanghai or Mandarin Chinese depending on the specific language mode they are in. This suggests that their production of the difficult English vowels could be closer to Mandarin vowels in formants and duration if they are in their Mandarin Chinese mode, or closer to Shanghai vowels if they are in their Shanghai Chinese mode.

Methods

Participants

Forty adult native Chinese speakers (20 females and 20 males, aged between 19 and 26 years) without hearing or speech impairments participated in the present study. Twenty of them (10 females and 10 males) were monodialectal speakers of Mandarin Chinese; the other 20 of them (10 females and 10 males) were bidialectal speakers of Shanghai and Mandarin Chinese, that is, they were proficient in both Shanghai dialect and Mandarin Chinese and used the two language varieties on a daily basis. Specifically, the participants in the monodialectal group grew up in Beijing where only Mandarin Chinese is used in daily life. They came to Shanghai for higher education but could not understand or speak the Shanghai dialect at the time of the experiment nor could they speak any other Chinese dialects. The participants in the bidialectal group grew up in Shanghai, with daily exposure to and frequent usage of both Shanghai and Mandarin Chinese. Participants completed a language background survey where they rated their language proficiency (i.e., daily

usage of and lifetime exposure to the target language) on a scale of 1–5 (1 = not familiar; 2 = familiar; 3 = fair; 4 = proficient; 5 = very proficient). The monodialectal group's average Mandarin proficiency was 4.9, while for the bidialectal group, the average proficiency in Mandarin and Shanghai Chinese was 4.85 and 4.9, respectively. Proficiency in Mandarin and Shanghai Chinese was comparable for the bidialectal group (i.e., the differences were nonsignificant [$F(1,19) = 0.192$, $p = 0.67$]), and Mandarin proficiency was comparable between the monodialectal and bidialectal groups [$F(1,19) = 0.192$, $p = 0.67$].

All Chinese participants had studied English as a foreign language in China for an average of 14 years, with no history of living in an English-speaking country for more than one month. They all reported speaking American English only. In the same language background survey (as mentioned in the previous paragraph), they were also asked to indicate how often they used English and Chinese (Mandarin for the monodialectal group; Mandarin and Shanghai Chinese for the bidialectal group) in their daily communication on a scale of 1–5 (1 = not at all; 2 = only occasionally; 3 = sometimes; 4 = frequently; 5 = very frequently). For English, the average score was 2.03, while for Chinese the average score was 4.73, and the difference was significant [$F (1, 39) = 466.08$, $p < 0.001$, $\eta_p^2 = 0.92$]. Therefore, Chinese was mainly used for their daily life and English was only used occasionally. Participants also indicated that when they spoke English, it was with their Chinese peers and teachers rather than with English native speakers. Six adult native speakers of General American English (three females and three males, mean age = 35) were recruited in the U.S. to produce the American English stimuli. They did not understand or speak any form of Chinese (Mandarin or Shanghai or other Chinese dialects) at all. The acoustic characteristics of the target English vowels produced by the six American speakers for the present study (detailed in Fig. 1) were consistent with previous studies on American English vowels (Figure 3 of Hillenbrand et al., 1995).

Stimuli

The American English stimuli included two target English words (*deed, goose*) containing the easy vowels ([i], [u]) and two English words (*did, good*) containing the difficult vowels ([ɪ], [ʊ]). The Chinese stimuli included two Chinese words (/di/ <*brother* 弟>, /gu/ <*old* 故>) containing two vowels ([i], [u]) in both Mandarin and Shanghai Chinese and two Chinese words (/tɪʔ/ <*drop* 滴>, /kʊʔ/ <*country* 国>) containing Shanghai Chinese vowels ([ɪ], [ʊ]). Filler items that were not analyzed were *bird, bait, brown, dice, joy, gold, door,* and *fate* for English and *ren* <*people*人>, *hua* <*flower* 花>, *xing* <*star* 星>, *lan* <*blue* 蓝>, *niao* <*bird* 鸟>, *xian* <*fresh* 鲜>, *jiu* <*wine* 酒>, and *nuan* <*warm* 暖> for Chinese.

Procedure

The Chinese participants were asked to produce the English and Chinese speech stimuli (presented randomly on a screen) three times each; the American participants were asked to produce the English stimuli only, three times each. Their
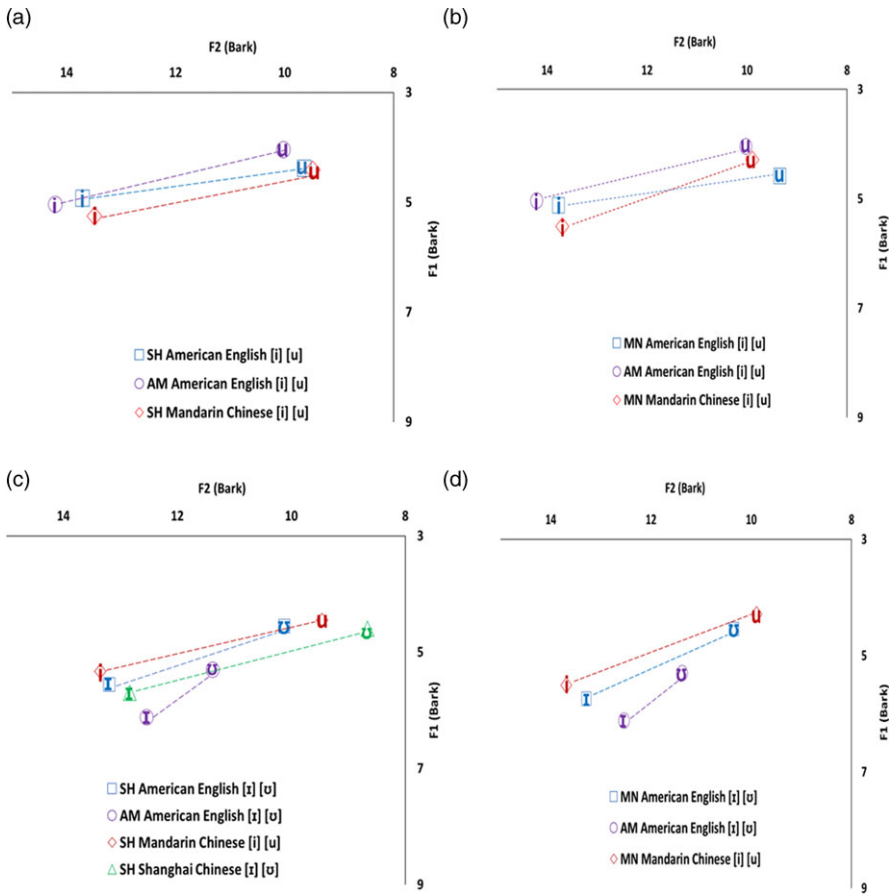
(a)

F2 (Bark)

14   12   10   8

SH American English [i] [u]
AM American English [i] [u]
SH Mandarin Chinese [i] [u]

(b)

F2 (Bark)

14   12   10   8

MN American English [i] [u]
AM American English [i] [u]
MN Mandarin Chinese [i] [u]

(c)

F2 (Bark)

14   12   10   8

SH American English [ɪ] [ʊ]
AM American English [ɪ] [ʊ]
SH Mandarin Chinese [i] [u]
SH Shanghai Chinese [ɪ] [ʊ]

(d)

F2 (Bark)

14   12   10   8

MN American English [ɪ] [ʊ]
AM American English [ɪ] [ʊ]
MN Mandarin Chinese [i] [u]

Figure 1. Vowel plots of participants' production of the American English, Mandarin Chinese, and Shanghai Chinese vowels. The upper panel compares Shanghai Chinese (a) and Mandarin Chinese (b) speakers' production of English vowels ([i], [u]) with their production of Mandarin Chinese vowels ([i], [u]), and American English speakers' production of English vowels ([i], [u]). The Lower panel compares Shanghai Chinese (c) and Mandarin Chinese (d) speakers' production of English vowels ([ɪ], [ʊ]) with their production of Mandarin Chinese vowels ([i], [u]), and American English speakers' production of English vowels ([ɪ], [ʊ]). Figure (c) also includes Shanghai Chinese speakers' production of Shanghai Chinese vowels ([ɪ], [ʊ]). SH: Shanghai Chinese speakers; MN: Mandarin Chinese speakers; AM: American English speakers.

speech was recorded individually in a sound-attenuated booth using a Sudotack ST-800 High-Quality Cardioid Microphone connected to a MacBook (64 bit) computer. For the final acoustic analyses, there were (a) [4 (English stimuli) + 2 (Mandarin Chinese stimuli)] * 3 (repetitions) * 20 (Mandarin Chinese participants)] = 360 tokens for the monodialectal Mandarin Chinese group, (b) [4 (English stimuli) + 2 (Mandarin Chinese stimuli) + 2 (Shanghai Chinese stimuli)] * 3 (repetitions) * 20 (Shanghai Chinese participants)] = 480 tokens for the bidialectal Shanghai-Mandarin Chinese group, and (c) 4 (English stimuli) * 3 (repetitions) * 6 (American participants) = 72 tokens for the American English participants. Participants took

a self-paced approach to produce the target stimuli, and they pressed the space key on the keyboard to proceed to the next trial. A 500-ms fixation cross was displayed on the screen between each trial. The stimuli were presented randomly, and so the participants were not likely to know that their vowel production was the target of the study, which was confirmed by a post-experiment debrief where participants expressed they were not aware of the purpose of the experiment. The random presentation of vowel stimuli has been used in numerous studies on L2 speech learning (e.g., Baker & Trofimovich, 2006; Bundgaard-Nielsen et al., 2011; Munro & Derwing, 2008 among many others). But this well-established approach might give rise to production errors due to the unpredictability of the presentation of the stimuli. Therefore, participants were allowed to self-correct speech errors they made during production. Tokens that contained speech errors (approximately 1% of all the tokens) due to the possible priming effect of the random presentation of the vowel stimuli were subsequently excluded from the acoustic data analyses. Participants were allowed to take breaks at their discretion. The experiment lasted approximately 15 min.

Acoustic data analyses

We extracted the vowels of the stimuli for acoustic analyses. The vowel boundaries were determined manually by three phoneticians using Praat (Boersma & Weenink, 2020), based on the start and end points of the periodic waveform of the vowels. The formant values of the vowels were taken as an average from the beginning to the end of the vowel boundaries. Another expert phonetician was invited to check all the vowel boundaries to ensure the labeling was correct. The corresponding duration of the vowels was measured in milliseconds (ms). In order to assess the extent to which the Chinese speakers' production of the English vowels was different from that of native speakers of English, we examined the Euclidean distance between Chinese and American English speakers' production of the English vowels. The use of Euclidean distance is a well-established method to quantify vowel distances across different language conditions in many previous studies (e.g., Chang, 2023; Mora & Nadeu, 2012; Recasens & Espinosa, 2006 among many others). Formant values (F1 and F2) were converted from Hertz to the Bark scale to normalize the intrinsic variation of different speakers' vocal tract lengths (Clopper, 2009). Statistical analysis of the acoustic data was performed in R (Version 3.4.4; R Core Team, 2018) using the *lme4* package (Bates et al., 2015).

Results

Table 1A shows the means of the differences in vowel formants, as measured by the Euclidean distance between Chinese speakers' (Shanghai and Mandarin) and native American speakers' production of the easy and difficult English vowels. To address the first research question, the Euclidean distance data were submitted to a linear mixed-effects model (with "group" (Shanghai vs. Mandarin), "vowel type" (Easy vs. Difficult), and their interaction as the fixed effects, and "participants" and "items" as random effects). The results (Table 2A) showed significant effects of group and vowel type, as well as a significant interaction between group and vowel type. Post

Table 1. Means, SE, 95%CIs for Euclidean distance (A) and duration difference (B) data of Shanghai and Mandarin speakers in easy and difficult English vowel conditions

		Mean	SE	95% CIs
A. Euclidean distance				
Easy	Shanghai	0.87	0.05	[0.76; 0.97]
	Mandarin	1.14	0.07	[0.99; 1.3]
Difficult	Shanghai	1.32	0.11	[1.09;1.54]
	Mandarin	1.33	0.11	[1.1;1.56]
B. Duration difference		**Mean**	**SE**	**95% CIs**
Easy	Shanghai	17.84	1.28	[15.15; 20.52]
	Mandarin	23.44	0.92	[21.47; 25.32]
Difficult	Shanghai	26.16	1.54	[22.93; 29.38]
	Mandarin	43.15	2.5	[35.82; 50.47]

Table 2. Results of linear mixed-effects models for Euclidean distance (A) and duration difference (B)

Fixed effects	Estimates	SE	95% CIs	t	p
A. Euclidean distance					
(Intercept)	0.87	0.06	[0.75; 0.98]	14.71	<0.001
Group	0.28	0.08	[0.13; 0.42]	3.66	<0.001
Vowel type	0.45	0.08	[0.3;0.6]	6.01	<0.001
Group * vowel type	0.27	0.11	[0.48; 0.58]	2.51	0.013
B. Duration difference					
Fixed effects	**Estimates**	**SE**	**95% CIs**	**t**	**p**
(Intercept)	17.84	1.72	[14.42;21.25]	10.36	<0.001
Group	5.6	2.18	[1.31;9.88]	2.6	0.011
Vowel type	8.32	2.18	[4.03;12.61]	3.82	<0.001
Group * vowel type	11.39	3.08	[5.33;17.45]	3.69	<0.001

hoc one-way ANOVA showed that in the condition of easy vowels, Shanghai speakers had significantly smaller Euclidean distance than Mandarin speakers [$F (1, 38) = 9.43$, $p = 0.004$]. In contrast, the difference in Euclidean distance between the two groups was not significant for difficult vowels [$F (1, 38) = 0.003$, $p = 0.95$].

Table 1B shows the means of the duration difference between Chinese (Shanghai and Mandarin) and American speakers' production for easy versus difficult English
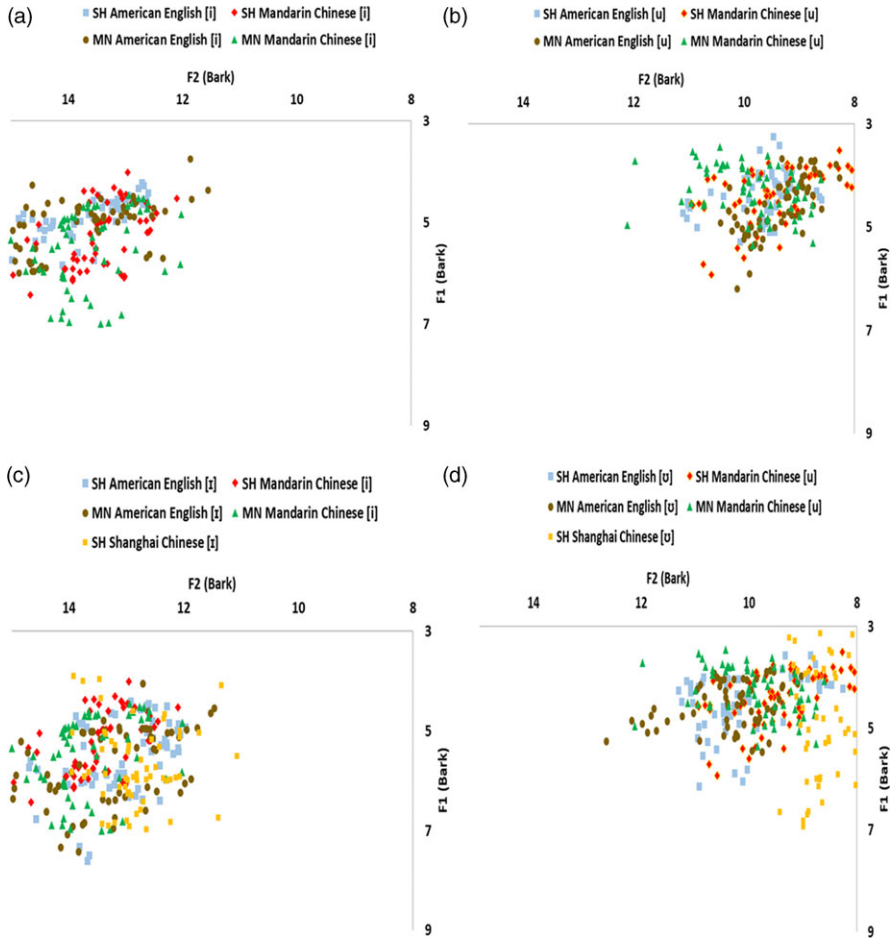
Figure 2. Scatterplots of Shanghai Chinese and Mandarin Chinese speakers' production of the American English vowels ([i], [u], [ɪ], [ʊ]), Mandarin Chinese vowels ([i], [u]) and Shanghai Chinese vowels ([ɪ], [ʊ]). SH: Shanghai Chinese speakers; MN: Mandarin Chinese speakers.

vowels. The duration difference data were submitted to a linear mixed-effects model (with "group," "vowel type" and their interaction as the fixed effects, and "participants" and "items" as random effects). The results of the mixed-effects model (Table 2B) showed significant effects of group and vowel type, as well as a significant interaction between group and vowel type. Post hoc one-way ANOVA showed that Shanghai speakers had significantly smaller duration difference from American speakers than Mandarin speakers for both the easy [F (1, 38) = 12.42, p = 0.001] and difficult [F (1, 38) = 19.73, $p < 0.001$] vowels.

Plots of the easy ([i], [u]) and difficult English vowels ([ɪ], [ʊ]) produced by Chinese and American speakers are presented in Fig. 1. The figures also include Shanghai and Mandarin speakers' production of the Mandarin Chinese vowels [i] and [u] (which are also found in Shanghai Chinese). In addition, Figure (1c) shows

Table 3. Means, SE, 95% CIs for the formants (F1, F2) and duration of the Chinese vowels produced by Shanghai and Mandarin Chinese speakers. SH: Shanghai Chinese speakers; MN: Mandarin Chinese speakers

		Mean	SE	95% CIs
SH Mandarin Chinese [i]	F1 (Bark)	5.25	0.05	[5.15; 5.35]
	F2 (Bark)	13.49	0.1	[13.29; 13.69]
	Duration (ms)	205.93	3.9	[199.34; 212.52]
SH Mandarin Chinese [u]	F1 (Bark)	4.41	0.05	[4.32; 4.5]
	F2 (Bark)	9.49	0.08	[9.34; 9.64]
	Duration (ms)	189.39	3.52	[183.44; 195.34]
MN Mandarin Chinese [i]	F1 (Bark)	5.5	0.07	[5.36; 5.64]
	F2 (Bark)	13.69	0.12	[13.55; 13.83]
	Duration (ms)	207.96	4.02	[201.17; 214.75]
MN Mandarin Chinese [u]	F1 (Bark)	4.28	0.06	[4.16; 4.4]
	F2 (Bark)	9.9	0.1	[9.7; 10.1]
	Duration (ms)	191.2	3.43	[184.35; 197.8]
SH Shanghai Chinese [ɪ]	F1 (Bark)	5.7	0.08	[5.55; 5.85]
	F2 (Bark)	12.83	0.09	[12.65; 13.01]
	Duration (ms)	197.39	2.97	[192.37; 202.41]
SH Shanghai Chinese [ʊ]	F1 (Bark)	4.81	0.07	[4.67; 4.95]
	F2 (Bark)	8.46	0.1	[8.26; 8.66]
	Duration (ms)	187.72	2.56	[183.39; 192.05]

Shanghai speakers' production of the Shanghai Chinese vowels [ɪ] and [ʊ]. Fig. 2 further compares participants' productions of their Chinese vowels with their English vowels. It shows the scatterplots of Shanghai Chinese and Mandarin Chinese speakers' production of the American English vowels ([i], [u], [ɪ], [ʊ]), Mandarin Chinese vowels ([i], [u]), and Shanghai Chinese vowels ([ɪ], [ʊ]).

To address research question 2, we first examined the acoustics of the Chinese vowels produced by Shanghai and Mandarin Chinese speakers in each dialect, detailed in Table 3. The results suggest that for Mandarin Chinese [i], Shanghai Chinese speakers had a lower F1 and F2 than Mandarin Chinese speakers. For Mandarin Chinese [u], Shanghai Chinese speakers had a higher F1 and lower F2 than Mandarin Chinese speakers. In terms of the duration of the two Mandarin Chinese vowels, Shanghai Chinese speakers produced the two vowels shorter than did Mandarin Chinese speakers. In addition, Shanghai Chinese [ɪ] and [ʊ] had higher F1, lower F2, and shorter duration than Mandarin Chinese [i] and [u], respectively.

Hence, the above results demonstrate the acoustic differences between Shanghai and Mandarin Chinese vowels, which could influence Shanghai Chinese speakers'

Table 4. Means, SE, 95%CIs for the comparison between Shanghai and Mandarin Chinese speakers in easy and difficult English vowel conditions with regard to Euclidean distance (ED1, ED2, ED3) and duration differences (Dur1, Dur2, Dur3). ED1/Dur1: the Euclidean distance/duration difference between the Mandarin Chinese vowels and English vowels produced by Chinese speakers. ED2/Dur2: the Euclidean distance/duration difference between Chinese and American speakers' production of the English vowels. ED3/Dur3: the Euclidean distance/duration difference between Shanghai speakers' production of the difficult English vowels and Shanghai Chinese vowels

			Mean	SE	95% CIs
Easy	Shanghai	ED1	0.85	0.04	[0.76;0.93]
		ED2	0.87	0.04	[0.8;0.93]
		Dur1	17.69	1.05	[15.6;19.78]
		Dur2	17.84	1.16	[15.6;20.01]
	Mandarin	ED1	0.9	0.05	[0.8;1.01]
		ED2	1.14	0.04	[1.05;1.24]
		Dur1	12.85	0.91	[11.06;14.65]
		Dur2	23.44	1.16	[21.05;25.82]
Difficult	Shanghai	ED1	0.77	0.06	[0.66;0.89]
		ED2	1.32	0.07	[1.18;1.45]
		ED3	1.87	0.08	[1.71;2.02]
		Dur1	18.17	1.14	[15.9;20.43]
		Dur2	26.16	1.9	[23.29;29.03]
		Dur3	35.29	1.82	[31.69;38.89]
	Mandarin	ED1	0.77	0.04	[0.7;0.84]
		ED2	1.33	0.07	[1.21;1.45]
		Dur1	23.62	1.87	[19.92;27.31]
		Dur2	43.15	1.9	[38.68;47.61]

production of Mandarin and English vowels due to the influence of Shanghai dialect. To gain a deeper understanding of how native Chinese dialect vowel systems affect the production of non-native English vowels among Chinese speakers, we further investigated whether Chinese speakers' production of English vowels is more similar to the production of Chinese vowels in their respective Chinese dialect or to that of American English speakers. We calculated the Euclidean distance and duration difference (ED1 and Dur1) between the Mandarin Chinese and English vowels produced by Chinese speakers, as well as the Euclidean distance and duration difference (ED2 and Dur2) between Chinese and American speakers' production of the English vowels. The data for the easy and difficult vowels conditions were analyzed for Mandarin and Shanghai Chinese groups respectively, and the p value was Bonferroni-corrected. The means are presented in Table 4.

Table 5. Contrasts in Euclidean distance (A) and duration difference (B) of Mandarin and Shanghai speakers in the easy (I) and difficult (II) vowel conditions. ED1/Dur1: the Euclidean distance/duration difference between the Mandarin Chinese vowels and English vowels produced by Chinese speakers. ED2/Dur2: the Euclidean distance/duration difference between Chinese and American speakers' production of the English vowels. ED3/Dur3: the Euclidean distance/duration difference between Shanghai speakers' production of the difficult English vowels and Shanghai Chinese vowels

(I) Easy vowels						
		A. Euclidean distance				
	Contrast	Estimate	SE	95% CIs	t	p
Mandarin	ED1 vs. ED2	0.24	0.06	[0.12; 0.36]	3.84	<0.001
Shanghai	ED1 vs. ED2	0.02	0.05	[-0.12;0.08]	0.37	0.71
		B. Duration difference				
	Contrast	Estimate	SE	95% CIs	t	p
Mandarin	Dur1 vs. Dur2	10.58	1.46	[7.7;13.47]	7.23	<0.001
Shanghai	Dur1 vs. Dur2	0.15	1.47	[-3.04;2.74]	0.1	0.92
(II) Difficult vowels						
		A. Euclidean distance				
	Contrast	Estimate	SE	95% CIs	t	p
Mandarin	ED1 vs. ED2	0.56	0.07	[0.42;0.69]	8.29	<0.001
Shanghai	ED1 vs. ED2	0.54	0.08	[0.37;0.71]	6.4	<0.001
Shanghai	ED1 vs. ED3	1.09	0.08	[0.92;1.25]	12.94	<0.001
		B. Duration difference				
	Contrast	Estimate	SE	95% CIs	t	p
Mandarin	Dur1 vs. Dur2	19.53	2.5	[14.58;24.48]	7.78	<0.001
Shanghai	Dur1 vs. Dur2	7.9	1.56	[4.91;11.07]	5.11	<0.001
Shanghai	Dur1 vs. Dur3	17.12	10.23	[13.82;20.42]	10.24	<0.001

For the easy English vowels, the results [Table 5(I)] showed that ED1 and Dur1 were overall smaller than ED2 and Dur2, respectively, for both Chinese groups (Shanghai and Mandarin). Specifically, for Mandarin speakers, ED1 was significantly smaller than ED2, and similarly, Dur1 was significantly smaller than Dur 2. For Shanghai speakers, the differences were not significant in either Euclidean distance or durational differences. For the difficult English vowels, the results [Table 5(II)] showed that ED1 was significantly smaller than ED2 for both Shanghai and Mandarin Chinese speakers. Similarly, Dur1 was significantly smaller than Dur2 for both Shanghai and Mandarin Chinese speakers. Additionally, for Shanghai Chinese speakers, we also computed the Euclidean distance (ED3) and durational differences (Dur3) between Shanghai Chinese speakers' production of the two Shanghai Chinese vowels (which sound similar to the difficult English vowels) and their production of the two difficult English vowels. ED3 and Dur3

were compared with ED1 and Dur1, respectively, to examine whether Shanghai participants' production of the difficult English vowels was influenced more by Mandarin or Shanghai Chinese. The results [Table 5(II)] showed that ED1 was significantly smaller than ED3, and similarly, Dur1 was significantly smaller than Dur3, indicating that Shanghai participants' production of the difficult English vowels was closer to the corresponding vowels in Mandarin rather than Shanghai Chinese.

Discussion

The present study examined how bidialectalism influences non-native speech production. Particularly, we compared monodialectal Mandarin Chinese with bidialectal Shanghai-Mandarin Chinese speakers in terms of their production of non-native American English vowels classified into two categories of difficulty for Chinese learners of English: easy ([i], [u]) and difficult ([ɪ], [ʊ]). We found that for easy English vowels, Shanghai Chinese was better than Mandarin Chinese speakers in approaching native English speakers with regard to vowel formants and duration. For difficult English vowels, Shanghai Chinese speakers were better in vowel duration but not in vowel formants compared with Mandarin Chinese speakers.

The results suggest that overall, there is a bidialectal advantage for Shanghai Chinese speakers in producing the easy English vowels, but that advantage becomes less apparent for the difficult English vowels, particularly in terms of formant frequencies. The results are in line with the proposal that the bilingual advantage is not broad-based; rather, it is modulated by the degree of difficulty and learnability of the target sounds (Antoniou et al., 2015; Elvin et al., 2018; Escudero et al., 2016; Kopečková, 2016). When the target non-native sounds are "easy," bilingualism could play a positive role in enhancing learning, whereas for learning "difficult" non-native target sounds, bilingualism may not be sufficient to yield high accuracy. The present study extends this proposal to the effect of bidialectalism on non-native speech production.

One may argue that the classification of sounds in a specific dialect/language is arbitrary. However, it is important to recognize that this arbitrariness could lead to differences in the acoustic mappings of sounds between one's native language (L1) and the target second language (L2). These differences contribute to varying levels of difficulty and learnability when acquiring non-native speech sounds. The existence of well-known theories, such as the Second Language Linguistic Perception (L2LP, Escudero, 2005, 2009; Escudero & Yazawa, in press), further supports the notion that the classification of sounds based on dialect/language plays a crucial role in understanding the learnability and difficulty of speech sounds. This model recognizes and explains the challenges faced by learners in perceiving and producing non-native sounds due to the acoustic and phonetic differences between their native language and the target language. Therefore, despite the arbitrary nature of sound classification, it is crucial to consider the impact of acoustic mappings and differences between L1 and the target L2 on the learnability and difficulty of non-native speech sounds. These considerations are essential for establishing theoretical

frameworks such as L2LP that can explain and interpret the findings in the context of language acquisition and perception.

Since Shanghai Chinese has the short vowels [ɪ] and [ʊ] that sound similar to the difficult English vowels [ɪ] and [ʊ], plus the fact that these two short vowels in Shanghai Chinese are rather different from Mandarin Chinese [i] and [u], respectively, as detailed in Table 3, one may wonder why Shanghai Chinese speakers did not perform better than Mandarin Chinese speakers in terms of formant frequency accuracy of the two difficult English vowels. A possible explanation is that the bidialectals are fully proficient in two varieties of the same language. According to the L2LP model, they could use either language variety when producing vowels in an additional language. Thus, they may have resorted to their knowledge of Mandarin Chinese when trying to produce the difficult English vowels, as evidenced from the smaller Euclidean distance from Mandarin Chinese vowels (ED1) than from Shanghai Chinese vowels (ED3). This finding echoes Williams and Escudero's (2014a) results, where Northern British listeners' categorization of Dutch vowels was influenced by their knowledge about acoustic patterns of the Standard Southern British English (SSBE) vowels. One of the reasons could be that SSBE is prevalent in British media and education, which means Northern British listeners are regularly exposed to SSBE, even though they may not produce English vowels in a way similar to Southern British speakers (Stuart-Smith, 2007). Such regular exposure may render the Northern listeners' expectation to hear SSBE frequently in daily life, especially in a formal setting such as a university laboratory, resulting in the activation of their SSBE mode of speech perception (Williams & Escudero, 2014a). This further suggests that speech perception is highly dynamic, which is often subject to the modulation of one's expectations and linguistic experiences depending on different contexts (Drager, 2010).

The L2LP model (Escudero, 2005, 2009; Escudero & Yazawa, in press; van Leussen & Escudero, 2015; Yazawa et al., 2020), which applies to both monolingual and bilingual/bidialectal learners, not only posits that monolinguals tend to perceive non-native sounds according to their native phonological categories but also that bilinguals may switch between different language modes when learning, listening to, and speaking in an additional language. More particularly, listeners' knowledge of how to process different dialects or languages is stored in separate perception grammars, each of which could be activated according to the specific language mode the bilinguals are in (Escudero, 2005, 2009; Escudero & Yazawa, in press; Yazawa et al., 2020). Such activation, as a result, could serve to map the incoming non-native speech sounds to either their native or non-native language/dialect (Williams & Escudero, 2014a). As mentioned above, the Shanghai-Mandarin Chinese speakers were fully functional bidialectals, that is, they were proficient in both Shanghai and standard Mandarin Chinese and used these two language varieties on a daily basis. Therefore, both Shanghai and Mandarin Chinese are readily accessible for them as a reference to map onto the incoming non-native English vowels. Moreover, given the predominant status of Mandarin Chinese in media and education all over China, plus the fact that the participants for the present study are students in a Chinese university where the medium of language instruction is Mandarin Chinese, it is likely that such frequent exposure to the standard official language may result in the

Shanghai-Mandarin Chinese participants' activation of their Mandarin Chinese mode when trying to produce the difficult English vowels. This is similar to Williams and Escudero (2014a) where Northern British listeners relied on their knowledge of the SSBE in perceiving non-native Dutch vowels due to the ubiquity of the standard language in media and education. The present findings can thus be seen as an extension of the L2LP theory to the domain of bidialectal non-native speech production, that is, bidialectal speakers can also switch between different modes to map the incoming non-native speech sounds to either their native or non-native language/dialect in speech production.

In terms of vowel duration, Shanghai Chinese speakers' production of the two difficult English vowels was closer to Mandarin vowels rather than Shanghai Chinese vowels, as evidenced from the result that Dur1 was significantly smaller than Dur3, which again suggest that Shanghai speakers could be in their Mandarin Chinese mode when producing those difficult English vowels. This is an interesting result because Shanghai speakers also performed better than Mandarin speakers in producing the difficult English vowels in terms of duration, as Shanghai speakers' Dur2 was shorter than Mandarin speakers' Dur2. Together, these results suggest that even though Shanghai speakers seemed to have been in their Mandarin mode, their production of the difficult English vowels was better than Mandarin speakers in terms of duration. This could be due to Shanghai Chinese speakers' Mandarin vowels being shorter than those of Mandarin Chinese speakers (SH Mandarin Chinese [i]: 205.93 ms; MN Mandarin Chinese [i]: 207.96 ms; SH Mandarin Chinese [u]: 189.39 ms; MN Mandarin Chinese [u]: 191.2 ms; see Table 3), which could be due to the existence of short vowels in Shanghai Chinese with shorter durations than Mandarin Chinese vowels (SH Shanghai Chinese [ɪ]: 197.39 ms; MN Mandarin Chinese [i]: 207.96 ms; SH Shanghai Chinese [ʊ]: 187.72 ms; MN Mandarin Chinese [u]: 191.2 ms; see Table 3). This could provide Shanghai speakers with an advantage in producing the short English vowels [ɪ] and [ʊ] even when they are in their Mandarin Chinese mode, which would explain their higher durational accuracy of the difficult English vowels.

The results are reminiscent of the findings reported in Iverson and Evans (2007) that L2 learners of English tended to have asymmetrical patterns of cue weighting in representing English vowels, that is, those who were accurate in representing one acoustic cue such as duration were not necessarily accurate at other cues such as formant frequencies (Iverson & Evans, 2007). The present study is also consistent with the findings that non-native speech learners may rely on duration as an alternative strategy when they struggle with the spectral characteristics of the target non-native vowels (Bohn, 1995; Bohn & Flege, 1990; Escudero & Boersma, 2004; Escudero et al., 2009), as Shanghai speakers had an advantage (compared with Mandarin speakers) in achieving the durational accuracy of the difficult English vowels, despite their difficulty with achieving accurate production of the vowel formants.

Future research could include other English sounds that are present in Mandarin but not in Shanghai Chinese, for example, the word-final /n/-/ŋ/ distinction. Moreover, more English varieties such as British or Australian English could be included as the target non-native languages to see if the same effects reported in our paper are found in varieties of English with different pronunciations of the target

vowels from those of American English (see for instance Escudero & Chladkova, 2010 for the acoustics properties of American versus Southern British English vowels; and Elvin et al., 2016 for Australian English vowels). Accordingly, an examination of a different cohort of Chinese dialects may lead to more diverse findings, especially regarding the acoustic contrasts with the target English sounds, which could also enhance our understanding of how bidialectalism influences non-native speech production. In addition, the present study used the method of eliciting words in isolation, but for future research, employing methods with greater ecological validity such as words read in the context of a sentence or a story (e.g., Yazawa et al., 2023) to capture natural speech patterns would be beneficial. Additionally, eliciting target vowels from multiple words with different syllabic contexts would enhance the generalizability of the results and promote a more comprehensive understanding of non-native speech production. This has been done in the analysis of native English speech (e.g., Elvin et al., 2016; Williams & Escudero, 2014b) but not so much for non-native English speech (but see Yazawa et al., 2023 where vowels were produced in different consonantal contexts of a story).

Conclusion

The present study presents a unique contribution on how bidialectalism influences non-native speech production. We compared monodialectal Mandarin Chinese with bidialectal Shanghai-Mandarin Chinese speakers in terms of their ability to produce American English vowels, which were classified into easy and difficult categories for Chinese learners of English. The results showed that the bidialectal group had an overall advantage in producing the easy American English vowels [i] and [u] in terms of vowel formants and duration. For the difficult English vowels [ɪ] and [ʊ], both groups experienced the same challenges with vowel formants, but the bidialectals had higher accuracy in vowel duration. The present study thus extends previous bidialectalism research and the L2LP model to the realm of non-native speech production, demonstrating that the bidialectal advantage in non-native speech learning is modulated by cross-linguistic difficulty constraints. Therefore, the present study also contributes to our general understanding and theoretical modeling of how bidialectalism influences second-language acquisition.

Replication package. Data and materials for this article can be found at https://osf.io/a5y49/.

Competing interests. The authors declare none.

References

Abu-Rabia, S., & Sanitsky, E. (2010). Advantages of bilinguals over monolinguals in learning a third language. *Bilingual Research Journal*, **33,** 173–199.

Antoniou, K., Grohmann, K. K., Kambanaros, M., & Katsos, N. (2016). The effect of childhood bilectalism and multilingualism on executive control. *Cognition*, **149,** 18–30.

Antoniou, M., Liang, E., Ettlinger, M., & Wong, P. (2015). The bilingual advantage in phonetic learning. *Bilingualism: Language and Cognition*, **18**(4), 683–695.

Baker, W., & Trofimovich, P. (2006). Perceptual paths to accurate production of L2 vowels: The role of individual differences. *IRAL – International Review of Applied Linguistics in Language Teaching*, **44**, 231–250.

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2015). lme4: Linear mixed-effects models using Eigen and S4. R package version 1.1–7. 2014.

Blom, E., Boerma, T., Bosma, E., Cornips, L., & Everaert, E. (2017). Cognitive advantages of bilingual children in different sociolinguistic contexts. *Frontiers in Psychology*, **8**, 552.

Boersma, P. & Weenink, D. (2020). Praat: Doing phonetics by computer [Computer program]. Retrieved from: http://www.fon.hum.uva.nl/praat/download_win.html.

Bohn, O.-S. (1995). Cross-language speech perception in adults: First language transfer doesn't tell it all. In W. Strange (Ed.), *Speech perception and linguistic experience: Issues in cross-language research* (pp. 279–304). Timonium, MD: York Press.

Bohn, O.-S., & Flege, J. E. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, **11,** 303–328.

Bundgaard-Nielsen, R. L., Best, C. T., & Tyler, M. D. (2011). Vocabulary size matters: The assimilation of second-language Australian English vowels to first-language Japanese vowel categories. *Applied Psycholinguistics*, **32**(1), 51–67.

Chang, Y. H. S. (2023). Effects of production training with ultrasound biofeedback on production and perception of second-language English tense–lax vowel contrasts. *Journal of Speech, Language, and Hearing Research*, **66**(5), 1479–1495.

Chao, Y.-R. (1967). Contrastive aspects of the Wu dialects. *Language*, **43,** 92–101.

Chen, Y. (2008). The acoustic realization of vowels of Shanghai Chinese. *Journal of Phonetics*, **36,** 629–648.

Chen, Y., & Gussenhoven, C. (2015). Shanghai Chinese. *Journal of the International Phonetic Association*, **45**(03), 321–337.

Chen, Y., Robb, M., Gilbert, H., & Lerman, J. (2001). Vowel production by Mandarin speakers of English. *Clinical Linguistics and Phonetics*, **15,** 247–440.

Clopper, C. G. (2009). Computational methods for normalizing acoustic vowel data for talker differences. *Language and Linguistics Compass*, **3**(6), 1430–1442.

Cohen, S. P., Tucker, G. R., & Lambert, W. E. (1967). The comparative skills of monolinguals and bilinguals in perceiving phoneme sequences. *Language and Speech*, **10,** 159–168.

Colantoni, L., Steele, J., & Escudero, P. (2015). *Second language speech: Theory and practice*. Cambridge: Cambridge University Press.

Drager, K. (2010). Sociophonetic variation in speech perception. *Language and Linguistics Compass*, **4**(7), 473–480.

Elvin, J., & Escudero, P. (2019). Cross-linguistic influence in second language speech: Implications for learning and teaching. In J. Gutierrez-Mangado, M. Martínez-Adrián & F. Gallardo-del-Puerto (Eds.), *Cross-linguistic influence: From empirical evidence to classroom practice* (pp. 1–20). Cham: Springer.

Elvin, J., Escudero, P., Williams, D., & Best, C. T. (2016). The relationship between Australian English speakers' non-native perception and production of Brazilian Portuguese vowels. In *Proceedings of The Sixteenth Australasian International Conference on Speech Science and Technology*, 6-9 December 2016, Parramatta, Australia (pp. 293–296).

Elvin, J., Tuninetti, A., & Escudero, P. (2018). Non-native dialect matters: The perception of European and Brazilian Portuguese vowels by Californian English monolinguals and Spanish–English bilinguals. *Languages*, **3**, 37.

Elvin, J., Williams, D., & Escudero, P. (2016). Dynamic acoustic properties of monophthongs and diphthongs in Western Sydney Australian English. *Journal of the Acoustical Society of America*, **140**(1), 576–581.

Elvin, J., Williams, D., & Escudero, P. (2020). Learning to perceive, produce and recognise words in a non-native language. Linguistic Approaches to Portuguese as an Additional Language, 61–82.

Enomoto, K. (1994). L2 perceptual acquisition: The effect of multilingual linguistic experience on the perception of a "less novel" contrast. *Edinburgh Working Papers in Applied Linguistics*, **5**, 15–29.

Escudero, P. (2005). *Linguistic perception and second-language acquisition: Explaining the attainment of optimal phonological categorization*. LOT Dissertation Series 113, Utrecht University.

Escudero, P. (2009). Linguistic perception of "similar" L2 sounds. In P. Boersma & S. Hamann (eds.), *Phonology in perception* (pp. 151–190). Berlin: Mouton de Gruyter.

Escudero, P., Benders, T., & Lipski, S. C. (2009). Native, non-native and L2 perceptual cue weighting for Dutch vowels: The case of Dutch, German, and Spanish listeners. *Journal of Phonetics*, **37**(4), 452–465.

Escudero, P., & Boersma, P. (2004). Bridging the gap between L2 speech perception research and phonological theory. *Studies in Second Language Acquisition*, **26**(4), 551–585.

Escudero, P., Broersma, M., & Simon, E. (2013). Learning words in a third language: Effects of vowel inventory and language proficiency. *Language and Cognitive Processes*, **28**(6), 746–761.

Escudero, P., & Chladkova, K. (2010). Spanish listeners' perception of American and Southern British English vowels: Different initial stages for L2 development. *Journal of the Acoustical Society of America*, **128**, EL254–EL260.

Escudero, P., Mulak, K. E., Fu, C. S., & Singh, L. (2016). More limitations to monolingualism: Bilinguals outperform monolinguals in implicit word learning. *Frontiers in Psychology*, **7**, 1218.

Escudero, P., Simon, E., & Mitterer, H. (2012). The perception of English front vowels by North Holland and Flemish listeners: Acoustic similarity predicts and explains cross-linguistic and L2 perception. *Journal of Phonetics*, **40**, 280–288.

Escudero, P., Simon, E., & Mulak, K. E. (2014). Learning words in a new language: Orthography doesn't always help. *Bilingualism: Language and Cognition*, **17**(2), 384–395.

Escudero, P., Smit, E. A., & Mulak, K. E. (2022). Explaining L2 lexical learning in multiple scenarios: Cross-situational word learning in L1 Mandarin L2 English Speakers. *Brain Sciences*, **12**(12), 1618.

Escudero, P., & Williams, D. (2012). Native dialect influences second-language vowel perception: Peruvian versus Iberian Spanish learners of Dutch. *Journal of the Acoustical Society of America*, **131**, EL406–EL412.

Escudero, P. & Yazawa, K. (in press). The second language linguistic perception model (L2LP). In Amengual, M. (Ed.), *The cambridge handbook of bilingual phonetics and phonology*. Cambridge, UK: Cambridge University Press.

Grosjean, F. (2021). *Life as a bilingual: Knowing and using two or more languages*. Cambridge: Cambridge University Press.

Hillenbrand, J., Getty, L., Clark, M. & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America*, **97**, 3099–3111.

Hirosh, Z. & Degani, T. (2018). Direct and indirect effects of multilingualism on novel language learning: An integrative review. *Psychonomic Bulletin & Review*, **25**(3), 892–916.

Iverson, P., & Evans, B. G. (2007). Learning English vowels with different first-language vowel systems: Perception of formant targets, formant movement, and duration. *Journal of the Acoustical Society of America*, **122**(5), 2842–2854.

Jia, G., Strange, W., Wu, Y., & Collado, J. (2006). Perception and production of English vowels by Mandarin speakers: Age-related differences vary with amount of L2 exposure. *Journal of the Acoustical Society of America*, **119**(2), 1118–1130.

Kirk, N. W., Fiala, L., Scott-Brown, K. C., & Kempe, V. (2014). No evidence for reduced Simon cost in elderly bilinguals and bidialectals. *Journal of Cognitive Psychology*, **26**(6), 640–648.

Klein, E. C. (1995). Second versus third language acquisition: Is there a difference? *Language Learning*, **45**, 419–465.

Kopečková, R. (2016). The bilingual advantage in L3 learning: A developmental study of rhotic sounds. *International Journal of Multilingualism*, **13**, 410–25.

Lee, W., & Zee, E. (2003). Standard Chinese (Beijing). *Journal of the International Phonetic Association*, **33**, 109–112.

Leivada, E., Rodríguez-Ordóñez, I., Couto, M. C. P., & Perpiñán, S. (2023). Bilingualism with minority languages: Why searching for unicorn language users does not move us forward. *Applied Psycholinguistics*, **44**(3), 384–399.

Luk, G., & Bialystok, E. (2013). Bilingualism is not a categorical variable: Interaction between language proficiency and usage. *Journal of Cognitive Psychology*, **25**(5), 605–621.

Mora, J. C., & Nadeu, M. (2012). L2 effects on the perception and production of a native vowel contrast in early bilinguals. *International Journal of Bilingualism*, **16**(4), 484–500.

Munro, M. J., & Derwing, T. M. (2008). Segmental acquisition in adult ESL learners: A longitudinal study of vowel production. *Language Learning*, **58**(3), 479–502.

Oschwald, J., Schättin, A., von Bastian, C. C., & Souza, A. S. (2018). Bidialectalism and bilingualism: Exploring the role of language similarity as a link between linguistic ability and executive control. *Frontiers in Psychology*, **9**, 1997.

Patihis, L., Oh, J. S., & Mogilner, T. (2015). Phoneme discrimination of an unrelated language: Evidence for a narrow transfer but not a broad-based bilingual advantage. *International Journal of Bilingualism*, **19**(1), 3–16.

Poarch, G. J., Vanhove, J., & Berthele, R. (2019). The effect of bidialectalism on executive function. *International Journal of Bilingualism*, **23**(2), 612–628.

Recasens, D., & Espinosa, A. (2006). Dispersion and variability of Catalan vowels. *Speech Communication*, **48**(6), 645–666.

Ross, J., & Melinger, A. (2017). Bilingual advantage, bidialectal advantage or neither? Comparing performance across three tests of executive function in middle childhood. *Developmental Science*, **20**(4), e12405.

Scaltritti, M., Peressotti, F., & Miozzo, M. (2017). Bilingual advantage and language switch: What's the linkage? *Bilingualism: Language and Cognition*, **20**(1), 80–97.

Singh, L., Poh, F. L. S., & Fu, C. S. L. (2016). Limits on monolingualism? A comparison of monolingual and bilingual infants' abilities to integrate lexical tone in novel word learning. *Frontiers in Psychology*, **7**, 667.

Stuart-Smith, J. (2007). The influence of the media. In C. Llamas, L. Mullany & P. Stockwell (Eds.), *The Routledge companion to sociolinguistics* (pp. 140–148). New York, NY: Routledge.

Swain, M., Lapkin, S., Rowen, N., & Hart, D. (1990). The role of mother tongue literacy in third language learning. *Language, Culture and Curriculum*, **3**(1), 65–81.

van Leussen, J. W., & Escudero, P. (2015). Learning to perceive and recognize a second language: The L2LP model revised. *Frontiers in Psychology*, **6**, 1000.

Werker, J. F. (1986). The effect of multilingualism on phonetic perceptual flexibility. *Applied Psycholinguistics*, **7**, 141–155.

Williams, D., & Escudero, P. (2014a). Influences of listeners' native and other dialects on cross-language vowel perception. *Frontiers in Psychology*, **5**, 1065.

Williams, D., & Escudero, P. (2014b). A cross-dialectal acoustic comparison of vowels in Northern and Southern British English. *The Journal of the Acoustical Society of America*, **136**(5), 2751–2761.

Yazawa, K., Konishi, T., Whang, J., Escudero, P., & Kondo, M. (2023). Spectral and temporal implementation of Japanese speakers' English vowel categories: A corpus-based study. *Laboratory Phonology*, **14**(1), 1–33.

Yazawa, K., Whang, J., Kondo, M., & Escudero, P. (2020). Language-dependent cue weighting: An investigation of perception modes in L2 learning. *Second Language Research*, **36**(4), 557–581.

Yu, J., Li, A., & Wang, X. (2004). A contrastive investigation of diphthongs between Standard Mandarin and Shanghai accented Mandarin. In International Symposium on Tonal Aspects of Languages with Emphasis on Tone Languages (pp. 229–234). Beijing, China.