The effect of L2 experience on Mandarin speakers' production of Australian English short and long vowels

Shuting Liu, Olga Maxwell, Janet Fletcher

University of Melbourne
shutingl5@student.unimelb.edu.au, o.maxwell@unimelb.edu.au, j.fletcher@unimelb.edu.au

ABSTRACT

This study explores the effect of L2 experience on Mandarin speakers' production of Australian English short/long vowel pairs: /ɪ - iː/, /ʊ - ʉː/, and /ɐ - ɐː/. Vowel quality (F1/F2 estimates, spectral overlap) and vowel duration are examined. /ɪ - iː/ is found to be produced with the greatest spectral overlap, followed by /ɐ - ɐː/ and then /ʊ - ʉː/, where /ʉː/ is fronted compared to /ʊ/, indicative of the potential influence from Australian English. Durational differences are significant only for /ɪ - iː/. Speakers who have more experience with Australian English (AusE) make greater spectral and durational distinction between the short/long vowels. The results reveal a more nuanced picture with different L2 sources influencing the speakers' vowel production in addition to AusE. This calls for more studies that examine Mandarin L2 English within the 'New Englishes' paradigm.

Keywords: L1 Mandarin, Australian English, L2 experience, vowels, spectral overlap.

1. INTRODUCTION

Adult second-language (L2) speakers are often reported to have difficulties producing L2 vowel contrasts and may rely on partial or different acoustic cues compared to first-language (L1) speakers of the target L2 [1, 2, 3]. L1 Mandarin speakers in the U.S.A. are reported to produce the tense-lax English vowel contrasts, /i - ɪ/, /ɛ - æ/ and /u - ʊ/, which have both spectral (F1/F2) and durational differences, with extensive spectral overlap, and durational contrast maintained for only the high front vowels /i - ɪ/ [4, 5, 6, 7, 8, 9]. Similarly, proficient L1 Mandarin speakers of English, who were born, raised, and educated in Mainland China and mostly exposed to American/British English, are found to spectrally merge tense-lax English vowel contrasts [10, 11, 12], with some studies finding a lack of durational distinction [10, 11] whereas others show the opposite [12]. The term 'China English' (CE) has been used to refer to this emerging English variety, shaped by multiple sources of influence, e.g., language teaching, the media, in China [12, 13, 14, 15].

Extensive research has shown that the phonological system of L1 (acquired very early in life), in combination with L2 experience, critically shapes L2 vowel production [1, 16]. The term 'L2 experience' may be used in diverse ways in the field, but generally indicates the cumulative L2 speech input learners have received during the acquisition process. There is wide agreement that L2 production changes as a function of L2 experience [1, 4]. L2 experience, however, has proved to be difficult to measure [1].

Most previous studies have quantified L2 experience using the index, 'length of residence' (LOR), reported by learners in the format of years of residence in an L2-prevalent environment [1, 4, 17]. LOR, however, may not be able to provide an accurate estimate of the quantity of L2 input learners have received, as it is unlikely that the learners have been exposed to the L2 in a uniform way within a given time interval [1]. Full-time equivalent (FTE) years of L2 input, estimated as LOR multiplied by the proportion of L2 use, is therefore proposed to be a better quantitative measure of L2 experience [1, 18]. L2 experience also varies qualitatively, which may account for the difference or similarity between the performance of L2 learners and has been largely ignored by previous studies [1, 19]. For instance, L1 Spanish speakers, who learned English since childhood and often heard Spanish-accented English, resembled other adult Spanish speakers in producing English /p t k/ with shorter voice onset time compared to L1 English speakers [19].

Research on the effect of L2 experience on L1 Mandarin speakers' English production shows that more experienced learners, measured by LOR in an English-prevalent environment, produced the vowel pairs, /i - ɪ/ and /ɛ - æ/, with greater spectral distinction than less experienced speakers, whereas the two groups did not differ significantly in the temporal distinction they made for these contrasts [4, 17, 20]. The research, however, has been mainly carried out in the U.S.A., where the L2 speakers were exposed to American English, an English variety already familiar to them back in China via education and media, etc. [12, 15]. It is therefore unsurprising that L1 Mandarin speakers of English recorded in China and the U.S.A. had consistent vowel production patterns (i.e., spectrally merged tense-lax vowels), given the same

L1 background and the exposure to shared and similar English varieties. This leads to the question of how exposure to another variety of English, which is less familiar to Mandarin learners from China, would influence their use of acoustic cues associated with the production of varied English vowel categories.

To fill this gap, the present study examines English vowel production of L1 Mandarin speakers who migrated to Australia as adults and had limited exposure to Australian English (AusE) before their arrival to Australia. The L2 production is assumed to be influenced by L1 Mandarin, the mainstream varieties of English the speakers were exposed to before and after the migration (American/British English), and AusE. Given that research on CE commonly factors in the influence of Mandarin and the combined effect from American/British English, characteristics of CE vowels, along with features of Mandarin and AusE vowels, could be used as comparison baselines to reveal the effect of experience with AusE on the production. The effect of L2 experience is also examined quantitatively, using both LOR and FTE as measures. Further, most findings reported to-date are limited to vowels in stressed monosyllabic words, but varied stress patterns may be associated with different acoustic cue use [21, 22]. The present study, therefore, examines three AusE vowel contrasts, /ɪ - iː/, /ʊ - ʉː/, and /ɐ - ɐː/, in trochaic- and iambic-stress pattern words.

1.1. Mandarin and China English vowels

The present study follows [23]'s analysis of Mandarin vowels as a six-vowel system (see Fig. 1 - left), including two high front vowels /i y/, one high back vowel /u/ and an open front vowel /a/. Duration is not phonemically contrastive for Mandarin vowels [23, 24], but it is claimed to be a universally salient cue that could be easily employed to make a vowel distinction [16]. A lack of durational difference in L2 vowel production, therefore, may indicate the influence from L1 Mandarin.

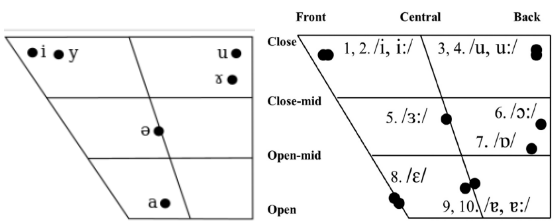

Figure 1: Vowel charts for monophthongs in Mandarin (left, [24]) and China English (right, [12]).

Based on previous research, [12] proposes a ten-vowel inventory for CE (see Fig. 1 - right), with the high front vowel pair /i - iː/ and the back vowel pair /u - uː/ showing a high degree of quality merging. The open central vowel pair /ɐ - ɐː/, in contrast, shows less degree of quality merging, with /ɐː/ being further back than Mandarin /a/ but more front than American English /ɑ/ [9]. All these vowel pairs are contrastive in length.

1.2. Australian English vowels

AusE has 12 monophthongs (see Fig. 2), most of which are distinguished in both quality and quantity [25]. There are six short vowels and six long vowels, with short vowels being approximately 60% the length of the long vowels in the /hVd/ context [26]. The three short/long vowel pairs we selected for the current study, /ɪ - iː/, /ʊ - ʉː/, and /ɐ - ɐː/, have a varied degree of spectral overlap/separation in AusE. Specifically, /ʊ - ʉː/ has the greatest spectral difference, followed by /ɪ - iː/ and then /ɐ - ɐː/, with /ɐ - ɐː/ lacking spectral differentiation [25].

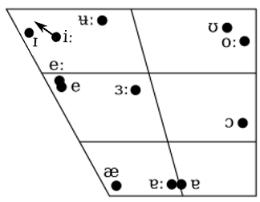

Figure 2: Vowel chart for Australian English ([25]).

1.3. Research questions and hypotheses

The research questions are: 1) What are the durational and spectral characteristics of the short/long vowels produced by L1 Mandarin speakers residing in Australia? 2) What is the effect of experience with AusE on the durational and spectral cues in the L2 vowel production?

We assume greater influence of AusE if the speakers show the tendency: 1) to make greater spectral separation for /ʊ - ʉː/ than /ɪ - iː/, and the least spectral distinction for /ɐ - ɐː/, with all three vowel pairs contrasting in duration; 2) to produce a more fronted high back vowel, given the presence of the back vowel /u/ and the absence of a high central vowel in Mandarin, CE, or American/British English. In AusE, the /ʉː/ vowel is central/fronted compared to other mainstream English varieties [25].

2. METHOD

2.1. Participants

Eight (4F; 4M) L1 Mandarin speakers were audio-recorded. At the time of recording, they aged 24 to 38 years, with an LOR in Australia between 2-10 years (M = 5.50, SD = 3.16) and a self-reported L2 use proportion between 40%-90%. Four (2F; 2M) speakers had significantly longer LOR (M = 8.75) than the other four (M = 2.75; $p = 0.005$), but their proportional L2 use was not significantly different, so the group with longer LOR also have significantly longer FTE (M = 5.20) than the other group (M = 1.78; $p = 0.013$).

2.2. Materials and procedure

Twelve disyllabic words containing the target vowels were selected for the study. As shown in Table 1, 50% of the words have the main stress on the initial syllable (trochaic, Ss) and 50% have the stress on the final syllable (iambic, sS). These words were embedded in short declarative sentences. The speakers were asked to produce the sentences, displayed on a computer screen one by one, as answers to pre-recorded questions, so that the target words were in narrow focus. The recordings took place in the Recording Studio at the University of Melbourne, Australia, using a Zoom H4N recorder at 44.1kHz. Each carrier sentence was elicited 5 times per speaker, with the sentences presented in randomised order. The four most natural productions were selected for subsequent analyses.

Vowel Contrasts	Stress Pattern	
	Trochaic (Ss)	Iambic (sS)
/ɪ - iː/	Lily - Leena	convince - convened
/ʊ - ʉː/	woman - rumor	afoot - recouped
/ɐ - ɐː/	runner - llama	redone - embalmed

Table 1: Words containing the target vowels.

2.3. Analysis

The recordings (384 sentences) were automatically segmented through WebMAUS [27], followed by manual correction of the boundaries of target words and vowels in Praat [28]. Vowel duration (ms) and formant (F1/F2) estimates (Hz) taken at vowel midpoint were extracted. Vowel duration, F1 and F2 values were fit to separate linear mixed-effect models using the lme4 package in R [29] to examine the effects of vowel pairs, phonemic length, stress pattern, L2 experience and their interactions. Post-hoc analyses were carried out with the *emmeans* function in the emmeans R package [30]. Pillai scores were calculated to explore the degree of spectral overlap (ranging from 0-1 with 0 representing total overlap and 1 for total separation) of the vowel pairs [31]. Spearman correlation tests were run between the speakers' individual Pillai scores and their LOR/FTE data to examine whether longer LOR/FTE predicts less spectral overlap of the vowels.

3. RESULTS

3.1. Vowel duration

Normalised vowel duration values are plotted in Fig. 3. Significant differences were found only for /ɪ - iː/ and /ʊ - ʉː/: 1) /iː/ was significantly longer than /ɪ/ ($t = -6.21$, $p < 0.001$); 2) /ʉː/ was longer than /ʊ/ in trochaic-stress pattern words ($t = -2.55$, $p = 0.011$), but shorter than /ʊ/ in iambic words ($t = 6.53$, $p < 0.001$). Vowels in sS words were significantly longer than in Ss words ($F(1,354) = 128.96$, $p < 0.001$), indicating a final lengthening effect. Overall, the phonemically long vowels (/iː, ʉː, ɐː/) had longer duration than the short vowels (/ɪ, ʊ, ɐ/) ($F(1,354) = 6.46$, $p = 0.011$), and the distinction was greater for the more experienced group than the less experienced group ($F(1,354) = 4.78$, $p = 0.030$).

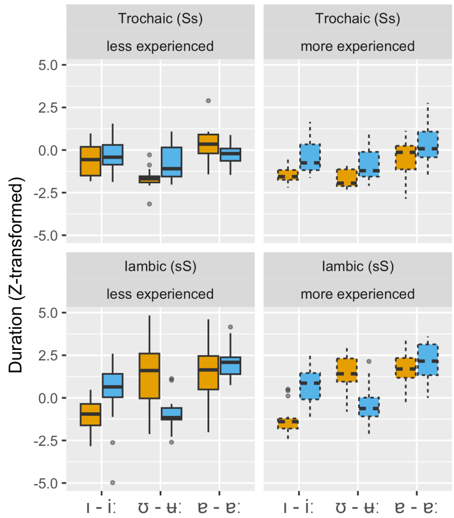

Figure 3: Normalised vowel duration in trochaic- (upper) and iambic-stress (lower) pattern words by less (left) and more (right) experienced speakers.

3.2. Vowel quality

3.2.1. First and second formants

Fig. 4 illustrates the F1/F2 vowel space for the vowels (with Lobanov-normalised F1/F2 values). The three vowel pairs form a triangle in the vowel space, with /ɪ - iː/ being most front along the F2 dimension, followed by /ɐ - ɐː/ and /ʊ - ʉː/ ($F(2,354) = 778.16$, $p < 0.001$). /ʊ - ʉː/ were highest along the F1 dimension, followed by /ɪ - iː/ and /ɐ - ɐː/ ($F(2,354) = 910.94$, $p < 0.001$). The height difference between /ʊ - ʉː/ and /ɪ - iː/ was only significant in sS words ($F(2,354) = 46.46$, $p < 0.001$), which may be an effect of the selected lexical items. Looking at vowels in each pair, no significant F1/F2 differences were found between the front or the open vowels. /ʊ - ʉː/ showed no height difference either, but /ʉː/ was more front than /ʊ/ along the F2 dimension (t = -3.71, $p < 0.001$).

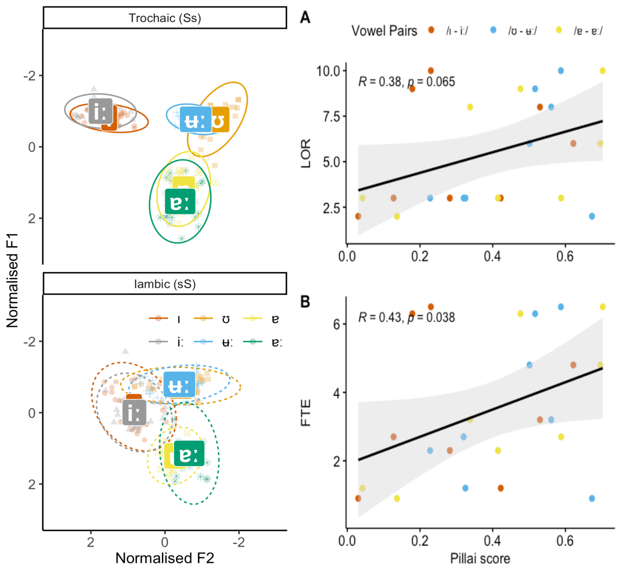

Figure 4 (left): F1/F2 plots with ellipses covering 95% CI from the mean value for each vowel. **Figure 5** (right): Correlation between Pillai scores and LOR (A) / FTE (B).

Fig. 4 also shows that the F1/F2 vowel space is smaller when the stress pattern is iambic compared to trochaic. Specifically, in sS words compared to Ss words: 1) /ɪ - iː/ were more centralised along both F1 (t = -10.92, $p < 0.001$) and F2 dimension (t = 10.39, $p < 0.001$). The F2 centralisation was more prominent for the less experienced group than the more experienced group (t = 3.44, $p < 0.001$); 2) /ɐː/ (but not /ɐ/) was raised along the F1 dimension (t = 3.26, $p = 0.001$); 3) /ʊ - ʉː/ were produced as front as /ɐ - ɐː/ (t = -0.07, $p > 0.05$) along the F2 dimension (t = -8.91, $p < 0.001$).

3.2.2. Spectral overlap

The vowel pair /ɪ - iː/ was produced with the greatest spectral overlap (Pillai = 0.02), followed by /ɐ - ɐː/ (Pillai = 0.08) and /ʊ - ʉː/ (Pillai = 0.10). Vowel category was a significant predictor of the F1/F2 variation for /ɐ - ɐː/ ($p = 0.008$) and /ʊ - ʉː/ ($p = 0.001$), but not for /ɪ - iː/ ($p > 0.05$), confirming that /ɪ - iː/ were produced with the least spectral difference.

As shown in Fig. 5, the Pillai score from individual speakers for each vowel pair was found to correlate with their FTE positively and significantly ($r = 0.43$, $p = 0.038$), suggesting that more L2 experience with AusE is associated with greater vowel spectral separation. In comparison, the Pillai score was positively correlated with the LOR at a marginally significant level ($p = 0.065$) and with a smaller effect size ($r = 0.38$), suggesting that FTE is a stronger predictor of L2 vowel production variation than LOR.

4. DISCUSSION

L1 Mandarin speakers in this study are found to produce the AusE vowel pair /ɪ - iː/ with extreme spectral overlap and salient durational distinction, which is in line with previous findings [4, 5, 6, 7] and with the features reported for the high front vowels in CE [12, 15]. In contrast, /ʊ - ʉː/ show the greatest spectral separation among the three vowel pairs with /ʉː/ being more fronted than /ʊ/, suggesting an effect of AusE on L2 production. The pair /ʊ - ʉː/ is also produced with significant durational difference but in the opposite direction (/ʊ/ longer than /ʉː/) in iambic-stress pattern words, suggesting that the L2 speakers may still be in the process of acquiring the durational cue for this vowel contrast. /ɐ - ɐː/ is produced with somewhat greater degree of spectral separation than /ɪ - iː/, especially in the iambic-stress condition, exhibiting a pattern more characteristic for CE. This vowel pair, however, is produced without significant durational difference, reflecting the influence from Mandarin. The findings suggest that the examination of Mandarin L2 English within the 'New Englishes' paradigm could be a more useful approach as opposed to simply looking at L1-L2 interaction.

Our study also shows that speakers with more L2 experience, which is better indicated by FTE compared to LOR, make greater distinction between the contrasted vowels by producing less spectral overlap and by manipulating vowel duration compared to less experienced speakers. Stress pattern also influences vowel quality and quantity, where vowels have greater spectral overlap and are produced in a more centralised F1/F2 region with greater duration values in iambic words compared to trochaic words. Future vowel studies need to examine vowel quality in varied stress patterns.

5. CONCLUSION

Mandarin L2 English speakers produce the short/long vowel pairs with a varied degree of spectral overlap and durational contrast, not only reflecting influence from L1 Mandarin and exposure to AusE, but also shows the vowel characteristics of 'China English'. More experienced L2 learners, better indicated by FTE, make greater spectral and durational distinction between the vowels. With some findings potentially limited by the selected speech material and the relatively small data set, this study hopefully could shed light on future research direction.

6. REFERENCES

[1] Flege, J. E., Bohn, O.-S. 2021. The Revised Speech Learning Model (SLM-r). In: Wayland, R. (ed), *Second Language Speech Learning: Theoretical and Empirical Progress*. Cambridge University Press, 3–83.

[2] Munro, M. J. 2021. On the difficulty of defining "difficult" in second-language vowel acquisition. *Frontiers in Communication*, 6, 639398.

[3] Tsukada, K. 2009. Durational characteristics of English vowels produced by Japanese and Thai second language (L2) learners. *Australian Journal of Linguistics*, 29(2), 287–299.

[4] Flege, J. E., Bohn, O.-S., & Jang, S. 1997. Effects of experience on non-native speakers' production and perception of English vowels. *Journal of Phonetics*, 25(4), 437–470.

[5] Wang, X., Munro, M. J. 1998. English vowel production by native speakers of Beijing Mandarin. *The Journal of the Acoustical Society of America*, 103(5), 3089.

[6] Chen, Y., Robb, M., Gilbert, H., Lerman, J. 2001. Vowel production by Mandarin speakers of English. *Clinical Linguistics & Phonetics*, 15(6), 427–440.

[7] Chen, Y. 2006. Production of tense-lax contrast by Mandarin speakers of English. *Folia Phoniatrica et Logopaedica*, 58(4), 240–249.

[8] Barkana, B. D., Patel, A. 2020. Analysis of vowel production in Mandarin/Hindi/American-accented English for accent recognition systems. *Applied Acoustics*, 162, 107203.

[9] Gong J., Yang, Z., Bellamy W., Wang F., Ji X. 2020. Effect of perceptual training on the production of English tense-lax vowels by native Chinese speakers. *2020 International Conference on Asian Language Processing (IALP)*. Kuala Lumpur, Malaysia.

[10] Hung, T. N. 2002. Phonological features of Chinese varieties of English. *The 7th English in Southeast Asia Conference*. Hong Kong Baptist University.

[11] Ao, R., Low, E. L. 2015. A description of the Yunnan English accent. *World Englishes*, 34(3), 336–354.

[12] Wang, Q. 2022. An instrumental investigation of the vowel inventory of China English. *Asian Englishes*, 1-22.

[13] Wang, R. 1991. Zhongguo yingyu shi keguan cunzai (China English is an objective reality). *Jiefangjun Waiyu Xueyuan Xuebao* (*Journal of PLA Foreign Languages Institute*), 1, 1–8.

[14] Xu, Z. 2006. Rectifying "Chinese English". In: Hashim, A., Hassan, N. (eds), *Varieties of English in Southeast Asia and Beyond*. University of Malaya Press, 283–291.

[15] Xu, Z. 2010. *Chinese English, features, and implications*. Open University of Hong Kong Press.

[16] Bohn, O.-S. 1995. Cross language speech production in adults: First language transfer doesn't tell it all. In: Strange, W. (ed), *Speech perception and linguistic experience: Issues in cross-language research.* Baltimore: York Press, 279-304.

[17] Flege, J. E., Liu, S. 2001. The effect of experience on adult's acquisition of a second language. *Studies in second language acquisition,* 23(04), 527-552

[18] Flege, J. E. 1991. Age of learning affects the authenticity of voice-onset time (VOT) in stop consonants produced in a second language. *The Journal of the Acoustical Society of America*, 89(1), 395–411.

[19] Flege, J. E., Wayland, R. 2019. The role of input in native Spanish late learners' production and perception of English phonetic segments. *Journal of Second Language Studies*, 2(1), 1–44.

[20] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. *Speech perception and linguistic experience: Issues in cross-language research*, 233-277.

[21] Byers, E., Yavas, M. 2017. Vowel reduction in word-final position by early and late Spanish-English bilinguals*. PLoS ONE*, vol. 12, no. 4.

[22] Silber-Varod, V., Khorshidi, N., Levi, L., Amir, N. 2019. The influence of lexical stress on formant values in spontaneous Hebrew speech. *The 19th International Congress of Phonetic Sciences (ICPhS)*, 3538-3542.

[23] Howie, J. M. 1976. *Acoustical studies of Mandarin vowels and tones,* vol. 6. Cambridge University Press.

[24] Lee, W.-S., Zee, E. 2003. Standard Chinese (Beijing). *Journal of the International Phonetic Association*, 33(1), 109-112.

[25] Cox, F. 2012. *Australian English Pronunciation and Transcription*. Cambridge University Press.

[26] Cox, F. 2006. The Acoustic characteristics of /hVd/ vowels in the speech of some Australian teenagers. *Australian Journal of Linguistics*, vol. 26, no. 2, 147–179.

[27] Kisler, T., Reichel, U. D., Schiel, F. 2017. Multilingual processing of speech via web services. *Computer Speech & Language*, vol. 45, 326–347.

[28] Boersma, P., Weenink, D. 2021. Praat: doing phonetics by computer [Computer program]. Version 6.1.50, http://www.praat.org/.

[29] Bates D, Mächler M, Bolker B, Walker S. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software*, 67(1), 1–48.

[30] Lenth R. 2022. emmeans: Estimated Marginal Means, aka Least-Squares Means. R package version 1.8.0.

[31] Hay, J., Warren, P., Drager, K. 2006. Factors influencing speech perception in the context of a merger-in-progress. *Journal of Phonetics*, vol. 34, no. 4, 458–484.