Convergence of L1 and L2 speech rhythm in Cantonese-English bilingual speakers

Wai Ling Law[1], Olga Dmitrieva[2], Alexander L. Francis[3]

[1]The Chinese University of Hong Kong, Shenzhen, China
[2][3]Purdue University, USA

`lawwailing@cuhk.edu.cn, odmitrie@purdue.edu, francisa@purdue.edu`

Abstract

Previous production studies suggest that first language (L1) speech rhythm can influence second language (L2) speech rhythm, but it remains unclear if the effect is bi-directional, including the influence from L2 to L1. It is also not known how L2 proficiency and amount of L2 use may modulate the interaction between L1 and L2 speech rhythm. Therefore, this study investigated speech rhythm in Cantonese and English productions by twenty native Cantonese-English bilinguals living in Hong Kong. Participants produced segmental near homophones in each language on different days. The rhythm of their Cantonese and English speech was quantified using acoustic measures and the effect of L2 proficiency and use was examined using a detailed language use questionnaire. Results showed that participants with higher English proficiency and use demonstrated rhythmical properties of speech suggesting convergence between L1 and L2 rhythm characteristics. However, when comparing the high and low proficiency groups within each language, the rhythmical properties of Cantonese or English speech were not significantly different. These results support the hypothesis that the convergence pattern reported for L1 and L2 segments extends to the rhythmical properties of L1 and L2, but the effect is not strong enough to determine the direction of influence.

Index Terms: speech rhythm, L1-L2 phonetic interaction, bilingual speech, L2 experience

1. Introduction

Empirical studies have provided ample evidence that the phonetic systems of bilingual speakers' first language (L1) and second language (L2) exhibit bidirectional influences as they reside in a common phonetic space at the segmental level [1], [2]. Extensive L2 experience may lead to restructuring of L1 phonetic categories such that increasing similarity between similar L1 and L2 phones (crosslinguistic convergence) may result [2], [3].

In contrast, at the suprasegmental level, the ways in which L1 and L2 interact with each other are much less clear. Among various suprasegmental properties, speech rhythm is one of the better studied. Previous studies primarily focused on the effects of L1 speech rhythm on L2 (e.g. [4], [5], [6]): Such influence has been documented for both L1 speakers of a syllable-timed language (e.g. Cantonese) learning a stress-timed language (e.g. English) [7], [8], [9], [10] and native speakers of a stress-timed language (e.g. English) learning a syllable-timed language (e.g. French) [10], [11]. In comparison, less is known about the effects of L2 speech rhythm on L1. [12] found that the L1 (Afrikaans) speech rhythm of Afrikaans-Spanish bilinguals

influenced their L2 (Spanish) speech. These bilinguals, however, lived in a Spanish-dominant environment for at least two-thirds of their lives and were strongly dominant in Spanish. Therefore, it is unclear if a similar effect would hold under the conditions of non-dominant L2.

Similar to the documented L1-L2 interactions at the segmental level, L1 speech rhythm's influence on L2 is believed to be mediated by L2 experience, often operationalized as proficiency in and amount of use of L2 [7], [10], [13]. For instance, L1 Korean speakers of L2 English with the greatest amount of L2 experience produced English speech with native-like stress-timing (measured as ratios of the duration of stressed to unstressed syllables) while speakers with moderate L2 experience or less produced English speech as syllable-timed [13]. Investigating L2 English speech produced by L1 French and German speakers, [7] found that L2 speech progresses from syllable-timed toward stress-timed patterns with increased L2 proficiency whether the L1 is rhythmically similar to English (German) or different (French). These findings suggest that if a reverse effect of L2 speech rhythm on L1 is observed, it will also be guided by the amount of L2 use and experience. The speakers of the present study, Cantonese-English bilinguals, present an ideal test case because Cantonese is a typical syllable-timed language [14] while English a typical stress-timed language [6], [15], [16], [17].

Speech rhythm refers to the timing of recurrence of relatively strong and weak elements (isochrony). Researchers have traditionally classified languages into two distinct classes of syllable-timed and stress-timed languages. Stress-timed languages have roughly equal time intervals between stress beats, while syllable-timed languages display more nearly equal temporal intervals between successive syllables [15], [17]. However, subsequent empirical studies did not find support for a strict categorical distinction between syllable-timed and stress-timed languages. Instead, languages can be more or less stress-timed/syllable-timed [16]. In addition, [18] observed that stress-timed languages show greater variation in syllable length and complexity of syllable structure, more variation in phonetic realization of stress, and have more reduced unstressed syllables that are shorter or even absent than syllable-timed languages. Accordingly, some syllables are much more salient than others in stress-timed languages while syllables are roughly equally salient in syllable-timed languages. Based on these observations, rhythm metrics such as pairwise variability index (PVI) [16], which captures durational variability between vocalic intervals have been developed. Stress-timed languages demonstrate greater durational variability than syllable-timed languages [5], [16]. Therefore, the higher the PVI value, the more stress-timed the rhythmic pattern is. The present study used PVI to examine the presence of bidirectional influence of

L1 and L2 speech rhythm in Cantonese-English bilingual speakers.

2. Methods

2.1. Participants

The participants were 20 Cantonese-English bilingual speakers living in Hong Kong (14 women, 6 men) with a mean age of 22 years (range = 18 – 38 years). All of them learned Cantonese from birth and English from age 2 or 3 in school. None of them had lived in a primarily English-speaking country. Only one had travelled to a primarily English-speaking country for a stay of about 2 months. None of them reported knowledge of any additional languages other than elementary proficiency in Mandarin.

Hong Kong is a pervasively bilingual society in which Cantonese is the dominant first language of the community and is used as the primary language at home and for everyday life purposes. English is, in turn, the dominant language used in the governmental, legal, educational and business sectors [19]. Thus, this Cantonese-English bilingual population has a highly uniform bilingual experience and is relatively homogenous in terms of speaker-related factors which are typically identified as important in segmental phonetic interactions such as age of acquisition [20], the quantity and quality of L2 input [21], and language dominance [22].

2.2. Materials and Procedures

All participants completed a questionnaire adapted from the Bilingual Language Profile [23] and the Language Experience and Proficiency Questionnaire [24] to assess the participants' language use and experience in Cantonese and English. Examples of questions related to proficiency are "how well do you speak English?" and "how well do you understand Cantonese?" Participants rated their proficiency on a scale from 0 (not well at all) to 6 (very well). Questions related to language use included evaluation of the percentage of time they use each language in different domains such as with friends and in school/at work. Total use of all languages in any given question had to equal 100% so the percentage of time using each language ranged from 0% (not used at all) to 100% (used all the time).

All speakers then participated in a sentence-reading task in each language on separate days. Participants read 22 bisyllabic near homophones in Cantonese and English embedded in carrier phrases (Cantonese: 我會讀 ___ 俾你聽 "I will read ___ for you to hear" /ŋɔ23 wui23 tok22 ___ pɛi35 lɛi23 tʰɛŋ55/ [25]; English: *I will write ____ to you now*). Near homophones were used to control for factors that were not under investigation but could have an effect on the parameters of interest, such as number of syllables and syllable structure of the stimuli. A "near homophone" pair consists of two words of the respective languages that are phonetically nearly-identical in terms of their segmental composition. An example is 碩士 "Master's" /sɛk6si6/ and *sexy* /ˈsɛksi/. All words had a CV(C)-CV(C) structure. The English list consisted of 16 trochaic and 6 iambic words. An online questionnaire confirmed the familiarity of the words in the context of Hong Kong (a mean rating of 4.68/5). The stimuli were randomized and presented in three blocks separated with breaks. All the spoken responses were digitally recorded in a double-walled sound-attenuated booth with a 44.1 kHz sampling rate and a 16-bit quantization.

2.3. Acoustic analysis

Vowels were manually segmented and their durations were measured using Praat [26]. Vowels were labeled as the portion from the onset of voicing with higher formants visible in the spectrogram to the offset of higher frequency components in the spectrogram, or from the onset to the offset of periodicity in the vocalic portion of the waveform when the boundaries of higher formants were not clear in the spectrogram. We used pairwise variability index (PVI) [5] to quantify rhythmic patterns based on the two vowels in each bisyllabic word. PVI quantifies variability in speech rhythm by calculating the absolute vowel duration difference between successive pairs of vowels, with speaking rate normalized, as shown in formula (1) below:

$$PVI = 100 \times \left[\sum_{k=1}^{m-1} \left| \frac{d_k - d_{k+1}}{\frac{(d_k + d_{k+1})}{2}} \right| / (m-1) \right] \quad (1)$$

where m = number of vowels in utterance and d = duration of the kth vowel.

We chose PVI because it is widely used in current L2 speech studies [4], [5], [27], [28] and has been shown to significantly correlate with expert native speaker ratings of rhythm of L2 speech [29] and to distinguish between beginning and advanced L2 learners [30], [31]. [6] also found that PVI is highly correlated with VarcoV [32], another commonly used rate-normalized metric of vocalic interval duration, and is useful in quantifying the influence of L1 speech rhythm on L2 speech rhythm.

2.4. Statistical analysis

Each participant was assigned to a high or low proficiency group by median split for English proficiency and use. A repeated measures ANOVA was conducted for the dependent variable, PVI with two independent factors: English proficiency and use (high vs. low) and the language factor (Cantonese vs. English). Four planned pairwise comparisons were then conducted to identify the effects of participant grouping factor and language factor which contributed to the significant interactions.

3. Results

3.1. Results of questionnaire

Correlation analysis revealed that English proficiency and English use were significantly correlated with each other, $r = .512$; $p = .021$; df = 18. Therefore, a composite score was calculated by multiplying English proficiency by English use for each participant. Since the possible range of their English proficiency was from 0 (not well at all) to 6 (very well) and that of their English use was from 0% (not used at all) to 100% (used all the time), the possible range of the English proficiency and use composite score was from 0 to 600.

Participants' English proficiency scores ranged from 2 to 5 on a scale of 0 (not well at all) to 6 (very well). Their English use ranged from 0.77% to 35.38%. Therefore, their English proficiency and use composite scores ranged from 2.62 to 176.92. Among the 20 participants, 10 speakers were classified as low English proficiency and use (composite scores 2.62 – 36.92) and 10 as high English proficiency and use (composite scores 46.15 – 176.92). A t-test confirmed that, as intended, the two groups of speakers had significantly different English proficiency and use composite scores, with the lower group averaging 22.38 and the higher group averaging 93.29, $t(18) =$

5.58; $p < .001$.

3.2. Interaction between L1 and L2 speech rhythm

Analysis of the effects of English proficiency and use on PVI showed no significant main effects of proficiency group, $F(1, 18) = .462$, $p = .505$, but there were significant main effects of language, $F(1, 18) = 4.974$, $p = .039$ and a significant interaction between language and proficiency group, $F(1, 18) = 4.903$, $p = .040$. Planned comparisons revealed that for the high English proficiency and use group the mean PVI scores for Cantonese (mean = 49.40, standard error = 6.18) and English (mean = 49.37, standard error = 7.07) were not significantly different, $t(9) = .012$, $p = .991$, but they were significantly different for the low English proficiency and use group (Cantonese: mean = 51.93, standard error = 5.52 vs. English: mean = 43.53, standard error = 8.43), $t(9) = 2.987$; $p = .015$. As shown in Figure 1, the low proficiency and use group maintained a difference between Cantonese and English PVIs, unlike the high proficiency and use group. This suggests a pattern of convergence in PVI as speakers' English proficiency and use increase resulting from a decrease in Cantonese PVI and an increase in English PVI. When the effects of language were tested in planned comparisons, the mean Cantonese PVI scores of the high (49.40) and low English proficiency and use (51.93) groups were not significantly different, $t(18) = .965$, $p = .347$. The mean English PVI scores of the high (49.37) and low English proficiency and use (43.53) groups were also not significantly different, $t(18) = 1.677$, $p = .111$.
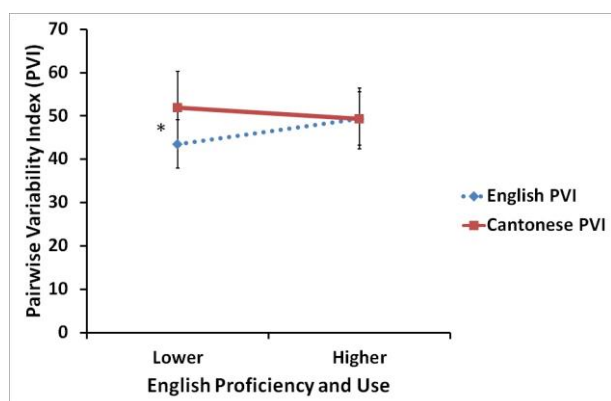

Figure 1. *The effect of English proficiency and use on Cantonese and English speech rhythm. The figure shows the pairwise variability index (related to speech rhythm) for Cantonese (solid line and squares) and English (dotted line and rhombi) as produced by speakers with lower English proficiency and use (left) and those with higher English proficiency and use (right). Error bars indicate standard error of the mean. The asterisk indicates that the difference was statistically significant.*

4. Discussion

The purpose of the study was to determine whether L1 and L2 speech rhythm exhibit bidirectional influence and whether such influence is modulated by L2 experience. It was predicted that the mechanism of convergence between the segments of L1 and L2 with increased L2 experience reported previously [2], [3] should also be observed between the rhythmic patterns of L1 and L2. This was confirmed in the present study. We demonstrate the convergence between Cantonese and English

PVIs, as a measure of speech rhythm, with an increased English proficiency and use. As English proficiency and use increased, participants produced higher (toward more stress-timed) English PVI values and lower Cantonese PVI (toward more syllable-timed). This constitutes convergence because the higher English proficiency and use group produced PVI values for Cantonese and English that were more like each other, in comparison to the lower proficiency and use group.

It is possible that the L1 and L2 speech rhythm of these bilingual speakers shows convergence because they produce both their L1 (Cantonese) and L2 (English) with the same rhythmic pattern - syllable-timing. This speculation gains support from the results of the current investigation and previous studies. The mean Cantonese PVI scores of the high (49.40) and low English proficiency and use (51.93) groups as well as the mean English PVI scores of the high (49.37) and low English proficiency and use (43.53) groups all fall within the range that previous research has identified for syllable-timed languages [5]. Previous studies have also suggested that Hong Kong English, the English variety spoken in Hong Kong, is syllable-timed [33], [34], similar to other new varieties such as Singapore English [5]. With increased L2 experience, the participants' English speech is still quite syllable-timed, confirming the influence of L1 speech rhythm on L2 speech rhythm evidenced in previous studies (e.g. [5], [6]). Meanwhile, it is important to note that the mean PVI values for Cantonese and English were significantly different for the low English proficiency and use group but not for the high English proficiency and use group. In other words, both the high and low English proficiency and use groups produce the same rhythmic patterns of syllable-timing in Cantonese and English, but the higher group produces crosslinguistic rhythmic patterns that are even more like each other due to a decrease in Cantonese PVI and an increase in English PVI. With higher English proficiency and use, speakers' English production becomes less syllable-timed (toward stress-timing) (cf. see [35] for Canadian English data that indicates the variety as less stress-timed than British English and supports a rhythmic continuum for English).

These findings suggest that with increased L2 experience, L1 and L2 speech rhythmical properties may become more similar, although the direction of phonetic influence remains unclear. This convergence pattern between L1 and L2 speech rhythm is also observed between similar L1 and L2 segments [36]. This indicates influence of L1 and L2 on each other at all levels of linguistic structure: it exists not only at the segmental level, but also at the suprasegmental level. This demands further investigations of other suprasegmental properties in bilingual speech.

5. Conclusions

The present study has investigated the interaction between L1 and L2 speech rhythm in Cantonese-English bilingual speakers and the role that L2 experience plays in such interaction. Our data show that L1 and L2 speech rhythm interact with each other in a fashion that is similar to segmental properties. Participants with higher English proficiency and use demonstrated convergence in their Cantonese and English PVIs such that the speech rhythm patterns of their two languages are more similar to each other in comparison to those with lower English proficiency and use. With higher proficiency and increased amount of use of the L2, restructuring of speech

rhythm in both languages may take place so that they become more similar to each other.

6. Acknowledgements

7. References

[1] J. E. Flege, C. Schirru, and I. R. A. MacKay, "Interaction between the native and second language phonetic subsystems," *Speech Communication*, vol. 40, pp. 467–491, 2003.

[2] I. R. A. MacKay, J. E. Flege, T. Piske, and C. Schirru, "Category restructuring during second-language speech acquisition," *Journal of the Acoustical Society of America*, vol. 110, pp. 516–527, 2001.

[3] J. E. Flege, "The production of "new" and "similar" phones in a foreign language: evidence for the effect of equivalence classification," *Journal of Phonetics*, vol. 15, pp. 47–65, 1987.

[4] U. Gut, "Rhythm in L2 speech," in *Rhythm, Melody and Harmony in Speech. Speech and Language Technology: Studies in Honour of Wiktor Jassem 14/15*, D. Gibbon, D. Hirst, and N. Campbell, Eds.. Poznań: Polskie Towarzystwo Fonetyczne, 2012, pp. 83–94.

[5] E. L. Low, E. Grabe, and F. Nolan, "Quantitative characterizations of speech rhythm: Syllable-timing in Singapore English," *Language and Speech*, vol. 43, pp. 377–401, 2000.

[6] L. White, and S. L. Mattys, "Calibrating rhythm: First language and second language studies," *Journal of Phonetics*, vol. 35, pp. 501–522, 2007.

[7] M. Ordin, and L. Polyanskaya, "Acquisition of speech rhythm in a second language by learners with rhythmically different native languages," *Journal of the Acoustical Society of America*, vol. 138, pp. 533-544, 2015.

[8] D. Gibbon, and U. Gut, "Measuring speech rhythm," in EUROSPEECH 2001 – 7th European Conference on Speech Communication and Technology, September 3–7, Aalborg, Denmark, Proceedings, 2001, pp. 95–98.

[9] H. Lin, and Q. Wang, "Vowel quantity and consonant variance: A comparison between Chinese and English," *Proceedings of Between Stress and Tone*. Leiden.

[10] A. Li, and B. Post, "L2 acquisition of prosodic properties of speech rhythm: Evidence from L1 Mandarin and German learners of English," *Studies in Second Language Acquisition*, vol. 36, pp. 223–225, 2014.

[11] M. Freland-Ricard, "Organisation temporelle et rythmique chez les apprenants étrangers. Etude multilingue" ("Second language learners' temporal and rhythmic organizations: A multilingual study"), *Revue de Phonétique Appliquée*, vol. 118–119 , pp. 61–91, 1996.

[12] A. W., Coetzee, L. García-Amaya, N. Henriksen, and D. Wissing, "Bilingual speech rhythm: Spanish-Afrikaans in Patagonia," in ICPhS 2015 – the 18th International Congress of Phonetic Sciences, August 10 – 14, Glasgow, UK, Proceedings, 2015.

[13] P. Trofimovich, and W. Baker, "Learning second language suprasegmentals: Effect of L2 experience on prosody and fluency characteristics of L2 speech," *Studies in Second Language Acquisition*, vol. 28, pp. 1–30, 2006.

[14] P. P. K. Mok, "On the syllable-timing of Cantonese and Beijing Mandarin," *Chinese Journal of Phonetics*, vol. 2, pp. 148–154, 2009.

[15] D. Abercrombie, *Elements of general phonetics*. Edinburgh: Edinburgh University Press, 1967.

[16] E. Grabe, and E. L. Low, "Durational variability in speech and the rhythm class hypothesis," *Laboratory Phonology*, vol. 7, pp. 515–546, 2002.

[17] K. Pike, *The intonation of American English*. Ann Arbor: University of Michigan Press, 1945.

[18] R. M. Dauer, "Stress-timing and syllable-timing reanalyzed," *Journal of Phonetics*, vol. 11, pp. 51–62, 1983.

[19] M. L. Lai, "Cultural identity and language attitudes – into the second decade of postcolonial Hong Kong," *Journal of Multilingual and Multicultural Development*, vol. 32, pp. 249–264, 2011.

[20] G. Yeni-Komshian, J. Flege, and S. Liu, "Pronunciation proficiency in the first and second languages of Korean–English bilinguals," *Bilingualism: Language and Cognition*, vol. 3, pp. 131–150, 2000.

[21] J. Flege, and S. Liu, "The effect of experience on adults' acquisition of a second language," *Studies in Second Language Acquisition*, vol. 23, pp. 527–552, 2001.

[22] M. Antoniou, C. T. Best, M. D. Tyler, and C. Kroos, "Language context elicits native-like stop voicing in early bilinguals' productions in both L1 and L2," *Journal of Phonetics*, vol. 38, pp. 640–653, 2010.

[23] D. Birdsong, L. M. Gertken, and M. Amengual, "Bilingual Language Profile: An Easy-to-Use Instrument to Assess Bilingualism," *The University of Texas, Austin*, 2012. [Online]. Available: https://sites.la.utexas.edu/bilingual/. [Accessed: March 19, 2019].

[24] V. Marian, H. K. Blumenfeld, and M. Kaushanskaya, "The language experience and proficiency questionnaire (LEAP-Q): Assessing language profiles in bilinguals and multilinguals," *Journal of Speech Language and Hearing Research*, vol. 50, pp. 940–967, 2007.

[25] A. L. Francis, V. Ciocca, V. K. M. Wong, and J. K. L. Chan, "Is fundamental frequency a cue to aspiration in initial stops?," *Journal of the Acoustical Society of America*, vol. 120, pp. 2884–2896, 2006.

[26] P. Boersma, and D. Weenink, *Praat: Doing phonetics by computer*. The Netherlands: University of Amsterdam, 2019. [Online]. Available: http://www.praat.org.

[27] P. M. Carter, "Quantifying rhythmic differences between Spanish, English, and Hispanic English," in *Theoretical and Experimental Approaches to Romance Linguistics: Selected Papers from the 34th Linguistic Symposium on Romance Languages (Current Issues in Linguistic Theory 272)*, R. S. Gess, and E. J. Rubin, Eds. Amsterdam, Philadelphia: John Benjamins, 2005, pp. 63–75.

[28] Whitworth, N, "Speech rhythm production in three German-English bilingual families," in *Leeds Working Papers in Linguistics and Phonetics 9*, D. Nelson, Ed. UK: The University of Leeds, 2002, pp. 175–205.

[29] N. Kinoshita, and C. Sheppard, "Validating acoustic measures of speech rhythm for second language acquisition," in ICPhS 2011 – the 17th International Congress of Phonetic Sciences, August 17 – 21, Hong Kong, Proceedings, 2011.

[30] N. Kinoshita, *The Acquisition and Teaching of Japanese Rhythm*. Japan: Waseda University Press, 2011.

[31] N. Kinoshita, and C. Sheppard, "The acquisition of lexical rhythm and duration by Japanese second language learners," *Laboratory Phonology*, vol. 11, pp. 79–80, 2008.

[32] V. Dellwo, "Rhythm and speech rate: A variation coefficient for deltaC," in *Language and language processing: Proceedings of the 38th linguistic colloquium*, P. Karnowski, and I. Szigeti, Eds. Frankfurt: Peter Lang, 2006, pp. 231–241.

[33] J. Setter, "A comparison of speech rhythm in British and Hong Kong English," in ICPhS 2003 – the 15th International Congress of Phonetic Sciences, August 3 – 9, Barcelona, Proceedings, 2003.

[34] J. Setter, "Speech rhythm in world Englishes: the case of Hong Kong," *TESOL Quarterly*, vol. 40, pp. 763–782, 2006.

[35] I. Grenon, and L. White, "Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese," in BUCLD 32 – the 32nd annual Boston University Conference on Language Development, November 2 – 4, Boston, MA, USA, Proceedings, 2008.

[36] J. Flege, "Second-language speech learning: Theory, findings, and problems," in *Speech perception and linguistic experience: Issues in cross-linguistic research*, W. Strange, Ed. Timonium, MD: York Press, 1995, pp. 229–273.