Acquisition of L2 Mandarin Rhythm By Russian and Japanese Learners

Yiran Ding[*], *Yanlu Xie*[*], *Jinsong Zhang*[*]

[*] Beijing Language and Culture University, Beijing, China
E-mail: nirvirain@gmail.com, Tel: +86-18810782561

***Abstract*—For Chinese as the second language (CSL) learners with different mother tongues (L1), the developments of their speech rhythm received little attention. Based on the interval-based acoustic rhythm metrics, we compared the speech productions of L2 Mandarin by 15 Japanese and 15 Russian learners with different proficiency level. The data included 103 sentences in read speech by each speaker (3605 sentences in total). Preliminary results showed: a.)During the progress from beginners toward intermediate level, the durational variability decreased in both groups of learners, which indicated acquisition of L2 Mandarin rhythm followed similar developmental paths from more stress-timed toward more syllable-timed; b.)During the progress from intermediate toward advanced level, Russian learners kept kind of stress-timed rhythm, Japanese learners appeared mora-timed rhythm, it indicated the transfer effects were influential at this learning stages.**

I. INTRODUCTION

In the traditional view, speech rhythm should be understood the reoccurrence of perceived certain events in time [1-3]. Nowadays, most languages can be classified into three rhythm categories, stress-, syllable- and mora-timed [4-9].

In the research of L2 prosody, many studies have suggested that the prosody of the target language can be influenced by the learners' native language, usually in the way of transferring L1's to L2 that the learners commonly transfer their prosodic patterns from L1 to L2. Ding [10] pointed out that the rhythm of English spoken by Chinese was usually judged to be intermediate between stress and syllable-timed, Author analysed the potential reason for this situation was the occurrence of epenthesis which was a common phenomenon that Chinese learner added vowels after consonantal finals. Another feature of Chinese English was that unstressed vowels were seldom shortened or reduced and stress vowels were not lengthened. Li and Post [15] showed that the distribution of the vocalic proportions(%V) across the different L1 (English) and L2 speakers groups (German and Mandarin) provided clear evidence of straightforward transfer effect, L2 learners showed intermediate values on opposite trajectories toward the target language. Other studies also provided evidence for transfer effects theory ([15-19]).

Universal developmental paths also play a role in L2 rhythm acquisition. Shport and Ordin [20] showed that Hong Kong and Russian learner of L2 English exhibited similarity in syllabic timing patterns in their L2, although Russian is a stressed-timed and Cantonese is a syllable-timed language. Ordin [16] introduced the acquisition of a stress-timed language English by L2 learners from German and French also followed the similar path from more-syllable towards stressed-timed. Li and Post [15] examined rhythmic patterns in L2 English produced by German and Mandarin learners at either beginner or intermediate proficiency level. The result showed that learners of English with a rhythmically different language (German and Mandarin) followed a comparable development path in durational variability of vowel intervals, toward a higher degree of stress-timing with proficiency growth.

For studies of L2 Mandarin rhythm acquisition, the developmental changes were little studied. Whether the transfer effects dominated the whole progress, or rhythm development followed the Universal developmental paths.

We employed several methods below to study this question: Firstly, we selected the speech data of L2 Mandarin learners from Russia and Japan, the rhythm of their L1 language was stress-timed and mora-timed. We used interval-based metrics to examine the differences in durational variability between Russian and Japanese learners and analysed the potential influence of L1 transfer effects. Secondly, we compared differences in speech productions by learners with varying proficiency levels to evaluate commonalities and difference in rhythm development.

II. METHOD

A. Rhythm metric

Speech was highly encoded in time, and several rhythm metrics have been proposed to capture the timing differences and temporal rhythmic properties. We adopted the widely-used rhythm metrics [ΔC, ΔV, VarcoC, VarcoV, rPVI-C, nPVI-V] ([10], [11], [12], [13], and [14]) to quantify the speech productions of L2 learners. The acoustic meanings and equation were described below.

- *Mean_S:* the mean syllable duration
- *%V:* the proportion of the vocalic intervals in total duration
- *ΔV:* the standard deviation (STDEV) of vocalic intervals (VI) duration
- *ΔC:* the STDEV of consonantal intervals (CI) duration.
- *VacroC:* the STDEV of CI duration divided by the mean CI duration and multiplied by 100

$$\text{Vacro}\,C = \frac{100 * \Delta C}{\text{mean}\,C}$$

- *VacroV:* the STDEV of VI duration divided by the mean VI duration and multiplied by 100
- *rPVI-C: t*he raw Pairwise Variability Index(PVI) for Cis

$$rPVI = \frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{m-1}$$

- *nPVI-V:* the normalized PVI for Vis

$$nPVI = 100 * \frac{\sum_{k=1}^{m-1} |d_k - d_{k+1}|}{(d_k + d_{k+1})/2}$$

B. Participants and materials

Russian is considered as stress-timed rhythm[20,24], which has a higher standard deviation of consonantal intervals and relatively lower proportion of vocalic intervals; Japanese is classified as mora-timed rhythm[26], which has a lower standard deviation of consonantal intervals and very higher percentage of vocalic intervals, and the target language (here is Mandarin) is traditionally considered as syllable-timed language.

The data was selected from BLCU-SAIT speech corpus of non-native Chinese [21]. To make sure that materials were comparable and facilitated the metrics measurements with fewer pauses and hesitations, read speech was selected. Reading materials is 103 sentences for every speaker (1304 Chinese characters) which covers 96% syllable types and 97% tri-tone types. The data included 15 Russia learners and 15 Japanese learners of Mandarin (between 19 and 27 yrs. M = 23 yrs.) and 5 Chinese native speakers. All participants came from monolingual families, Russia and Japan learners' participants learned Mandarin as a foreign language through formal instruction in school. We used three levels to grade learners' proficiency. The evaluation of proficiency depended on four factors which were HSK (Chinese Proficiency Test), learning length in total, the period of residence in China and the performance in this recording. Detailed information of each proficiency level described below:

- Beginners: No HSK grade; Less than one-year learning length; recently came to China; Not good performance, hard to correct some misreads through many attempts. (unaccepted misreads: e.g. unexpected long pause, deletion, repeated correction which may jeopardize the alignment between speech and text. Deviations from native speakers' pronunciation were commonly accepted.)
- Intermediate: 3/4 HSK grade (corresponding to B1/2 Level of the Common European Framework of Reference (CEF)); 2-3 years learning length; have lived in China for one year; Good performance, easily correct most of the misreads in the second trial.
- Advanced: 6 HSK grade (corresponding to C2 Level of CEF); at least five years learning length; have lived in China for more than three years; Excellent performance, produced most sentences correctly in the first trial.

C. Analysis

Before computing rhythm metrics, We conducted the annotation in following two steps on Praat [22]: 1) phonetic segmentation of sentence into phonemes and 2) classification of phonemes into vowels and consonants.

In the first step, we adopted the HTK-based SAIT force-aligner to segment speech data into Initials and Finals (demi-syllabic units in Mandarin phonology). The authors corrected automatic annotation manually as accurate as possible by referring to both visual and audio cues.

In the second step, according to traditional Mandarin phonology [23], an Initial is a consonant including (Aspirated) Stops; (Aspirated) Affricates; Nasals; Fricatives; Laterals(Besides, if there is no initial in a syllable, we call it a zero-initial). A Final may be a vowel, a diphthong, a triphthong, or nasalized vowel (which always changes format patterns of a principal vowel). The syllables type we selected in examinations was typical CV syllable where was predominant in Mandarin, and we treated Initials as C (consonant) and Finals as V (vowel).

The duration values were extracted with Praat scripts and the metrics mentioned before were then calculated [9].

III. RESULT

The overview of the metrics averages was described in Table1;

Table1: Scores of rhythm metrics for L2 Japanese Mandarin (JC) with varying proficiency levels, L2 Russian Mandarin (RC) with different proficiency levels, And Native Mandarin (CC);

Country & Fluency level	*SCORES*							
	Mean_S	*ΔC*100*	*ΔV*100*	*VacroC*	*VacroV*	*rPVI-C*	*nPVI-V*	*%V*
JP								
JC Beginner	0.31	4.92	7.86	53.99	32.99	54.49	37.01	0.74
JC Intermediate	0.28	3.97	6.03	44.00	31.20	43.56	31.97	0.71
JC Advanced	0.24	3.38	4.60	42.18	28.31	38.08	27.40	0.70
RU								
RC Beginner	0.26	8.75	10.80	71.19	42.58	88.24	47.72	0.71
RC Intermediate	0.20	5.28	7.31	50.48	35.12	57.13	40.09	0.70
RC Advanced	0.18	4.01	5.82	48.55	32.73	45.43	36.37	0.68
CN								
CC	0.20	3.08	4.00	40.89	31.76	33.29	34.64	0.65

The observation of Table1 showed that all metrics decreased with increasing proficiency, and also showed that the mean syllable duration of both groups decreased with proficiency growth. Because of the correlation of rate and those metrics, the rate-normalized metrics like Vacro and nPVI have been claimed to be more reliable. In the analysis procedure between proficiency levels, we excluded the raw ΔC and ΔV, but still used rPVI-c because of its particularity [9] and kept in mind the possible influence of speech rate.

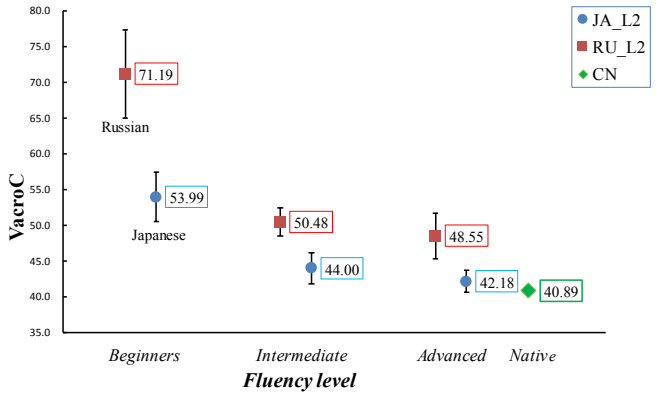

Fig. 1: Values of VacroC in L1 Mandarin and L2 Mandarin produced by Japanese and Russian learners on different proficiency levels. Error bars stand for ±SE.

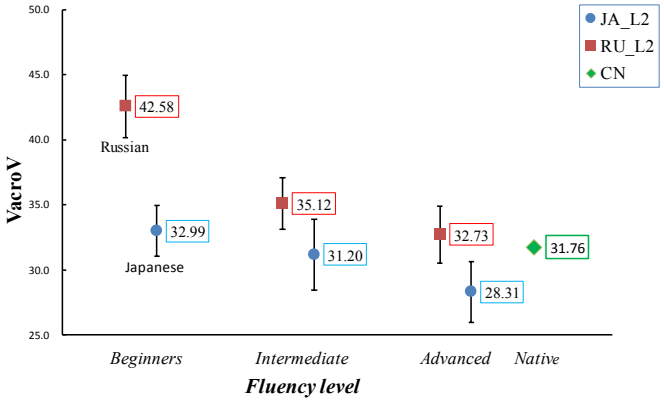

Fig. 2: Values of VacroV in L1 Mandarin and L2 Mandarin produced by Japanese and Russian learners on different proficiency levels. Error bars stand for ±SE.

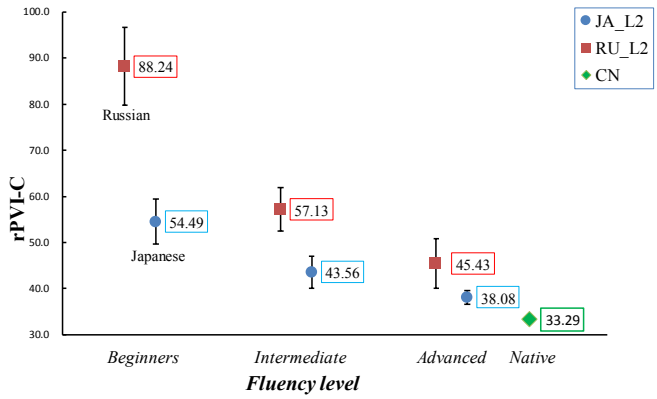

Fig. 3: Values of rPVI-C in L1 Mandarin and L2 Mandarin produced by Japanese and Russian learners on different proficiency levels. Error bars stand for ±SE.

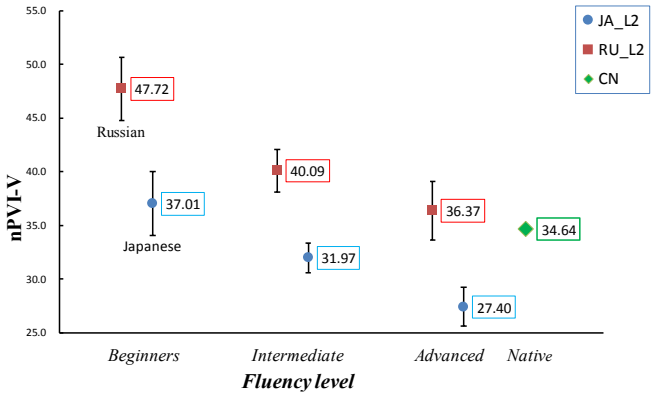

Fig. 4: Values of nPVI-V in L1 Mandarin and L2 Mandarin produced by Japanese and Russian learners on different proficiency levels. Error bars stand for ±SE.

Fig. 1-4 presented the duration variability at three proficiency levels of both groups of the learner. Visual inspection of Figures revealed durational variations of Russian learners were consistently higher than Japanese and native speaker. It might indicate that Russian learner exhibited a higher degree of stress-timed in L2 Mandarin. Observation of Figure 1-4 also revealed that the durational variations in Japanese learners were relatively lower than Russian learner especially in vowel-related metrics, which were surprisingly even smaller than a native speaker at the intermediate and advanced levels. It might indicate the Japanese exhibited a degree of mora-timed in L2 Mandarin yet though they had already reached the advanced level.

IV. DISCUSSION

This study set out to investigate rhythmic development at different proficiency levels in two different rhythmical languages to examine (a) the development of L2 Mandarin measured by various rhythm metrics. (b) Whether the findings can contribute to the L1 transfer hypothesis or whether they point toward a universal development path independent of the L1 backgrounds.

We found that raw and normalized metrics discriminated well between L2 proficiency levels for learners of both groups. We observed that variation in the duration of V and C intervals in L2 was higher on beginners' speech and lower in advanced proficiency level. Therefore we might conclude that rhythm in L2 Mandarin development had a similar path regardless of the L1 background. At the same time, we detected the differences in rhythmic patterns between different L1 groups, the differences of actual rhythmic patterns at a particular development stage revealed L1-specific peculiarities.

For starter, our result showed the changes of tempo, which displayed similarities with existing studies [16]. There were some specific things, occasionally the advanced L2 learners speak faster than intermediate and beginners, but a bit slower than native speakers. It was confirmed in the Japanese group. The mean syllable duration of Japanese L2 learners at advanced level was 0.24s, a little bit slower than native(0.20s) and faster than intermediate and beginners learners. Conversely, Russian

learners showed a rapid speaking rate even faster than native speakers (0.20s) at the advanced level (0.18s). Mandarin is traditionally considered as a tonal language which means the native speakers will commonly pronounce the full length of initials and finals to ensure the full tonal information. As we known Russian is classified as a non-tonal language, learners unintentionally ignore the tonal information during production. With the increasing speak tempo as well as the appearances of slur and vowel reduction, it is not surprising to see the faster tempo than native speaker by advanced Russian learners.

The development of each rhythm metric in a certain development stage showed the influence of L1 was still prominent. At the scene of beginners, Russian learners at beginners level existed an extremely higher variation than Japanese learners and native speakers, particular in consonant-related metrics (ΔC in Table1 VacroC rPVI-C in Figure), and it meant that Russian learner exhibited a higher degree of stress-timed at beginners level. We tried to figure out the possible reason for this phenomenon. The main syllable structure of Mandarin is CV type. The teaching experience showed Russian learners transfer their familiar syllable structure (e.g. CCVCC, CVC) from their L1 to L2 and pronounced the vowel and nasalized coda separately as well as pronouncing cacuminal initials separately which the duration of initials and finals with a nasalized vowel were commonly lengthened. We also observed the durational variability reduced rapidly during the progress from beginners to intermediate.

The result of Japanese learners of L2 Mandarin showed a lower durational variation compared to the Russian learners. Japanese does not have stress or vowel reduction, being a pitch accent language in which the accented vowel is not generally lengthened. It caused lower variabilities of consonantal and vocalic duration, all rhythmic metrics of V and C showed a relatively lower level of durational variations and even reached a lower durational variation in vowel-related metric (nPVI-V, VacroV) than native speakers at the progress of intermediate toward advanced. It meant L2 Mandarin produced by Japanese exhibited a degree of mora-timed. This phenomenon might indicate the L1 transfer also played a role in the intermediate and advanced level. The general difference between Japanese and Mandarin is the rhythm typology, and as a typical mora-timed language, each mora has almost the same duration length. Firstly, in a cross-linguistic study, there is abundant evidence that listeners' preferred word-segmentation strategy which may reflect the rhythm of their native language [26, 27], even if the speaker is completely fluent (but not native) in another language. A similar result in daily teaching experience, we found the distinction between single and compound finals seems harder for normal Japanese learners. The Japanese learner always misapplied the reduction or lengthening strategy, by reducing the duration of initials or lengthening finals to keep the syllable sounds equal, which was more common in advanced learners who have already mastered the manner of articulation and phonology knowledge of neutral tone. The production of L2 speech will sound hasty like mora-timed rhythm. In future teaching, the teacher should emphasize the durational variation of neutral and compound finals of Mandarin. It will help learners to overcome mora-timed features in their speech. The metrics showed the effectiveness to monitor the developmental changes of rhythmic patterns during the progress. If there are in-time feedbacks, it will facilitate the learners to product near-native rhythm.

The results of our study complement the view that L2 rhythm development is fundamentally multi-system [15]. The characteristic rhythm of a language is the product of several systemic properties like syllable structure, vowel reduction or final lengthening because these properties are acquired at different stages in the L2 developmental process, so those properties could be variably affected by L1 transfer or followed universal development paths. Our result showed the acquisition of L2 Mandarin by different L1 learners followed similar developmental paths, from more stress-timed toward more syllable-timed rhythm as the proficiency increased. But the transfer effects were still influential when learners have already reached high proficiency, and their productions still exhibited kind of their L1 rhythmic features.

V. CONCLUSION

In this investigation, we employed the wildly-used rhythmic metrics to investigate native Mandarin, L2 Japanese Mandarin, and L2 Russian Mandarin with comparable reading materials. Our results showed a similar development path of L2 Mandarin rhythm by Russian and Japanese learners that the learners of both groups started with comparably high values and progressed toward low L1 target language values as their proficiency level increased. We also revealed the differences in the certain development stage of each group of learners. These findings reflected the interaction of universal constraints as well as L1 transfer effects on L2 rhythm acquisition. Our results support and complement the multisystem nature of L2 prosodic acquisition in rhythm aspect and can be employed for the teaching of acquisition of near-native rhythm for teachers, for proficiency evaluation as an assistant descriptor in related test and for CAPT developers.

ACKNOWLEDGEMENT

This study was supported by Advanced Innovation Center for Language Resource and Intelligence (KYR17005), the Fundamental Research Funds for the Central Universities (16ZDJ03, 18YJ030006, 19YXC126), and the project of "Intelligent Speech technology International Exchange". Jinsong Zhang is the corresponding author.

REFERENCES

[1] U. Gut and J. Trouvain, *Non-Native Prosody*. Berlin, New York: Mouton de Gruyter, 2008.

[2] J. Yu, "Rhythm Patterns in the Speech of Chinese EFL Learners---a Phonetic Study based on Hangzhou-accented Learners", Ph.D. Zhejiang University, 2013.

[3] D. Abercrombie, *Elements of general phonetics*. Edinburgh: Edinburgh Univ. Press, 2006.

[4] F. Ramus, M. Nespor and J. Mehler, "Correlates of linguistic rhythm in the speech signal", *Cognition*, vol. 73, no. 3, pp. 265-292, 1999.

[5] A. Arvaniti, "Rhythm, Timing and the Timing of Rhythm", *Phonetica*, vol. 66, no. 1-2, pp. 46-63, 2009.

[6] L. Ling, E. Grabe, and F. Nolan, "Quantitative Characterizations of Speech Rhythm: Syllable-Timing in Singapore English", *Language and Speech*, vol. 43, no. 4, pp. 377-401, 2000.

[7] V. Dellwo and Petra. "Relations between language rhythm and speech rate," in *Proceedings of the 15th International Congress of Phonetic Sciences*, Barcelona, 2003, pp.471-474.

[8] V. Dellwo, "Rhythm and Speech Rate: A Variation Coefficient for deltaC" *in Language and Language-Processing*, P. Karnowski and I. Szigeti, Eds. "Frankfurt: Peter Lang, 2006, pp. 231-241.

[9] L. White and S. Mattys, "Calibrating rhythm: First language and second language studies", *Journal of Phonetics*, vol. 35, no. 4, pp. 501-522, 2007.

[10] H. Ding and X. Xu, "L2 English Rhythm in Read Speech by Chinese Students," *Proc. Interspeech 2016*, pp. 2696-2700.

[11] H. Wang, P. Mok and H. Meng, "Capitalizing on musical rhythm for prosodic training in computer-aided language learning", *Computer Speech & Language*, vol. 37, pp. 67-81, 2016.

[12] T. Rathcke and R. Smith, "Speech timing and linguistic rhythm: On the acoustic bases of rhythm typologies", *The Journal of the Acoustical Society of America*, vol. 137, no. 5, pp. 2834-2845, 2015.

[13] M. Ordin and L. Polyanskaya, "Perception of speech rhythm in second language: the case of rhythmically similar L1 and L2", *Frontiers in Psychology,* vol. 6, 2015.

[14] Low, E.L.; Grabe, E.; Nolan, F.: "Quantitative characterizations of speech rhythm: 'syllable-timing' in Singapore English." *Lang. Speech* vol. 43 pp. 377–401, 2000.

[15] A. Li and B. Post, "L2 ACQUISITION OF PROSODIC PROPERTIES OF SPEECH RHYTHM", *Studies in Second Language Acquisition*, vol. 36, no. 02, pp. 223-255, 2014.

[16] M. Ordin and L. Polyanskaya, "Acquisition of speech rhythm in a second language by learners with rhythmically different native languages", *The Journal of the Acoustical Society of America*, vol. 138, no. 2, pp. 533-544, 2015.

[17] P. Carter, "Quantifying Rhythmic Differences between Spanish, English, and Hispanic English", *Theoretical and Experimental Approaches to Romance Linguistics*, pp. 63-75, 2005.

[18] H. Lin and Q. Wang, "Vowel quantity and consonant variance: A comparison between Chinese and English," in *Proceedings of Between stress and tone*, Leiden, 2005.

[19] P. P. Mok and V. Dellwo, "Comparing native and non-native speech rhythm using acoustic rhythmic measures: Cantonese, Beijing Mandarin and English," in *Speech prosody 2008*, pp. 423–426.

[20] I. Shport, "Acquisition of rhythm: evidence from spontaneous L2 speech", *The Journal of the Acoustical Society of America*, vol. 123, no. 5, pp. 3328-3328, 2008.

[21] B. Wu, Y.-L Xie, J.-S Zhang, "The Construction of a Chinese Interlanguage Corpus," *Proc. International Conference on Ococosda*, pp. 1922-1925, 2010.

[22] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. Version 6.0.50, retrieved 03 August 2018 from http://www.praat.org

[23] Z.-J Wu et al, Xian Dai Han Yu Yu Yin Gai Yao, Chinese Language Teaching Press, Beijing, 1992.

[24] L. White, "Acquiring rhythm: A comparison of L1 and L2 speakers of Canadian English and Japanese", In *Proceedings of the 32nd Boston University Conference on Language Development,* 155-166, 2008.

[25] Cao, J.F, "An analysis of Mandarin neutral tone syllables" [In Chinese]. *Applied Acoustics*, 4, 3-8, 1986.

[26] Cutler A, Otake T, "Mora or Phoneme? Further Evidence for Language-Specific Listening [J]". *Journal of Memory & Language*, 33(6):824-844, 1994.

[27] Cutler, Anne, "Listening to a Second Language through the Ears of a First". Interpreting. 5. 1-23, 2000.