French Rhythm in Read Speech by French Natives and Chinese L2 Learners

Lei XI, Rachid Ridouane
Laboratoire de Phonétique et Phonologie (UMR 7018, CNRS/Sorbonne Nouvelle, Paris)
lei.xi@sorbonne-nouvelle.fr, rachid.ridouane@sorbonne-nouvelle.fr

ABSTRACT

This study examines rhythmic patterns in French sentences read by 10 native French speakers and 21 Chinese learners of French (with different levels of proficiency). We used 9 metrics, both normalized and non-normalized, to compare the rhythmic patterns between natives and learners, and between the 3 groups of learners.

Overall, results from normalized metrics suggest that the rhythmic patterns of natives and learners are rather similar. Indeed, while raw metrics, %V, ΔV, ΔC, and rPVI-C, show significant differences both between natives and learners, and among learners, normalized metrics, VarcoC, VarcoV, and nPVI-V, show no such differences. These results suggest that traditional metrics may not capture changes over stages of rhythm acquisition, especially when L1 and L2 belong to the same rhythm typology. In addition, the CCI (Control/Compensation Index), well-distinguished among learners, shows that as acquisition progresses, learners gain a better timing control ability, and rhythmic pattern becomes more stable.

Keywords: rhythm metrics, L2 rhythm production, French, Mandarin Chinese.

1. INTRODUCTION

Speech rhythm of world languages has traditionally been classified into stress-timed (e.g., English, Dutch), syllable-timed (e.g., French [1], Chinese [2]), and mora-timed (e.g., Japanese) [3, 4, 5]. However, empirical evidence to assess this idealized surface isochrony in different languages is rather weak [6]. Departing from this traditional view, a series of metrics has been proposed in different studies for quantifying rhythmic phenomena based on vocalic and intervocalic interval durations [1, 5, 7]. Even though this quantitative approach has also been criticized [among others 8, 9], it is still a main approach used in the study of cross-linguistic speech rhythm and L2 speech rhythm [10]. In this study, 9 metrics, both normalized and non-normalized, are used to investigate the rhythmic patterns in French sentences produced by French native speakers and 3 groups of Chinese learners of French with different levels of proficiency.

Having a strong command of rhythmic patterns that are characteristic of a native speaker's speech is crucial for improving the intelligibility and reducing the foreign accent of a second language speaker [11, 12, 13, 14]. The ability to produce and perceive the rhythms of speech is therefore of great importance for learning a second language. In recent years, there has been a growing interest in understanding the temporal aspects of non-native speech. Factors such as the type of materials used, the demands of the task, the rate and style of speech, can all impact the rhythmic patterns of second language speech [10, 15]. While previous research has often compared the rhythmic patterns of native speakers to those of second language learners, there has been less focus on how the rhythmic production of a second language changes as the learner progresses.

There have been relatively few studies on Chinese learners of French and their rhythmic patterns. In one study [16], native Chinese speakers learning French as a second language were compared to native French speakers using two metrics, %V and VarcoV, and it was found that the two groups had virtually identical values. Another study [17] found that only the value of %V significantly differed between Chinese learners and native French speakers when examining six different metrics. These studies suggest that Chinese learners may be able to achieve native-like rhythms in French, potentially due to the shared rhythmic patterns between French and Chinese as syllable-timed languages.

This new study aims to further explore these findings by conducting a production experiment with additional rhythm metrics and involving Chinese learners at various proficiency levels. The research aims to answer two questions: 1) How do normalized and non-normalized rhythm metrics capture similarities and differences between French speech produced by learners and natives, and 2) Do Chinese learners' performances vary based on their proficiency levels?

2. METHOD

2.1. Speakers

Ten adult native speakers of French (5 males and 5 females, mean age: 31.3, SD: 8,37) and 21 Chinese learners participated in the production experiment. None of the participants reported any hearing impairments or language impairments. The 21 learners completed a LEAP questionnaire [18] and provided information about their personal backgrounds and language proficiency. Based on their self-reported proficiency in French, the Chinese learners were divided into three groups (as shown in Table 1).

Level of Proficiency	Gender (M: male, F: female)	Mean Age (SD)	Mean Learning Time (year)
High	3M, 4F	27.00 (2.82)	8.6
Intermediate	2M, 5F	27.14 (2.91)	4
Low	2M, 5F	26.29 (4.15)	1.3

Table 1: Information about the Chinese learners.

2.2. Task and procedure

In this study, participants were asked to read 20 French sentences as part of a speech production task. These sentences were selected to contain a range of syllable lengths (from 7 to 27 syllables) and a variety of pauses and breaks. Each participant was asked to read each sentence at a normal speech rate and with neutral emotion, and to repeat any misread or disfluent sentences before moving on to the next. Each sentence was repeated twice, resulting in a total of 1240 sentences (400 produced by native speakers and 840 produced by Chinese learners). The production data was collected in a sound-proof room at the Laboratoire de Phonétique et Phonologie in Paris, and participants were recorded individually using the Pro Tools program and an AKG C520L wireless head-worn microphone (with the microphone positioned 3 cm away from the mouth). The data was recorded at a 44.1kHz sampling rate and 16-bit quantization.

2.3. Segmentation and measurements

The speech samples were automatically segmented using SPPAS [19] and then manually corrected by the authors. The data was then annotated in Praat [20] in two steps: 1) phonetic segmentation of the sentences into phonemes, and 2) classification of individual phonemes as vowels or consonants. The annotation and classification techniques from [15] were used. Pauses were marked as "#" and excluded from metric calculations. A total of nine rhythm metrics were used in this study (as shown in Table 2). In addition to traditional metrics (such as ΔV, ΔC, %V [5]; VarcoV, VarcoC [7], rPVI, nPVI [1]), values for CCI-V and CCI-C [21] were also obtained. The CCI index (Control/Compensation Index), which is a model for formally representing the rhythmic tendencies of natural languages, can be used to calculate the level of compression (lengthening or shortening) in language production. According to this model, "controlling" languages (such as syllable- or mora-timed languages) should have similar tendencies of vocalic and intervocalic duration fluctuations; while "compensating" languages (such as stress-timed languages) should fluctuate more in vocalic than in intervocalic intervals [22].

Metrics	Description
%V	Proportion of vocalic intervals duration to the duration of the utterance.
Δ	Standard deviation of vocalic (ΔV) or intervocalic (ΔC) intervals duration.
Varco	Coefficient of variation of vocalic (VarcoV) or intervocalic (VarcoC) intervals, i.e., ΔV or ΔC divided by the mean.
rPVI	Raw pairwise variability index for intervocalic intervals, i.e., mean of differences between successive segment intervals.
nPVI	Normalized rPVI for vocalic intervals divided by their sum and multiplied by 100.
CCI	rPVI relativized to the number of segments composing successive vocalic (CCI-V) or intervocalic (CCI-C) intervals.

Table 2: The rhythm metrics examined in this study (with traditional metrics in bold).

A total of 1240 textgrid files of segmented sentences were proceeded in Correlatore [23]. The nine metrics were calculated for each sentence and compared among the four groups.

3. RESULTS

The results for the global raw and normalized metrics [5, 7] are presented first, followed by the pairwise metrics [1, 21].

3.1. Global metrics

3.1.1. ΔC, ΔV and %

The results for the unnormalized variability metrics ΔV and ΔC are depicted in Fig. 1.

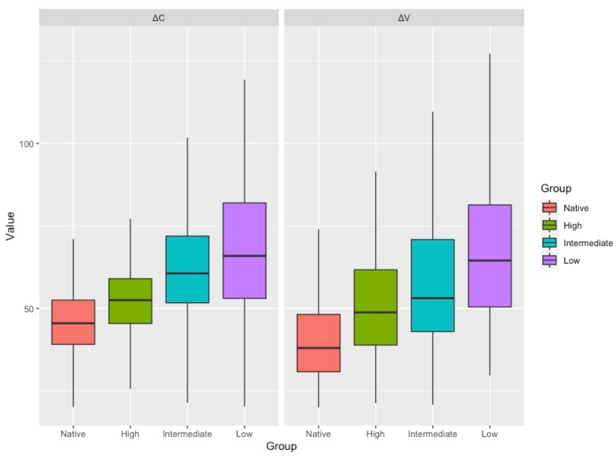

Figure 1: Means and standard errors of the metrics ΔC and ΔV for the four groups.

These results indicate that the learner groups exhibit larger values of ΔV and ΔC than the native group, with proficiency level showing an inverse relationship with Δ . One-way ANOVAs were conducted separately for each metric value, with group of proficiency as the between-subjects factor and the metric value as the dependent variable. The statistical tests revealed significant differences between learners and native speakers for ΔV (F-value = 122.7, $p < 0.001$) and ΔC (F-value = 118, $p<0.001$). To determine whether the metrics differ between the high, intermediate, and low proficiency groups, Tukey post-hoc tests were applied, which showed significant differences between the groups (pairwise comparisons for ΔV and ΔC between high vs intermediate, intermediate vs low, and high vs low: all at $p<0.001$).

The results for %V are presented in Fig. 2. It can be observed that the mean value of %V is larger in low proficiency learners than in intermediate-high proficiency learners (high: 48.40; intermediate: 48.36; low: 50.74). When compared to native speakers (mean %V: 46.39), post-hoc pairwise tests revealed that %V values between learners and natives are significantly different, with the exception of %V ($p = 0.99$) in the intermediate vs high proficiency groups.

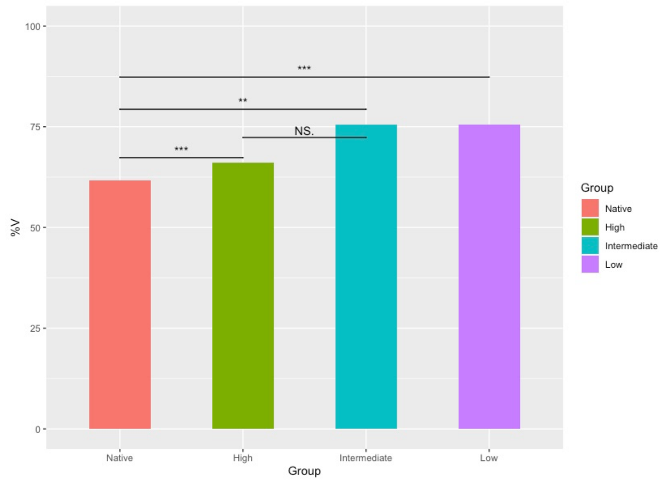

Figure 2: %V values of the four groups and post-hoc tests results.

3.1.2. VarcoV and VarcoC

The normalized variability metrics are presented in Figure 3. The results showed that both native and learner groups had virtually similar values of VarcoV and VarcoC. The mean values of the metrics before and after normalization, shown in Table 3, indicate that the patterns produced by learners were comparable to those of native speakers after accounting for speech rate. The raw and normalized metrics also demonstrated a significant effect of speech rate on learner productions.

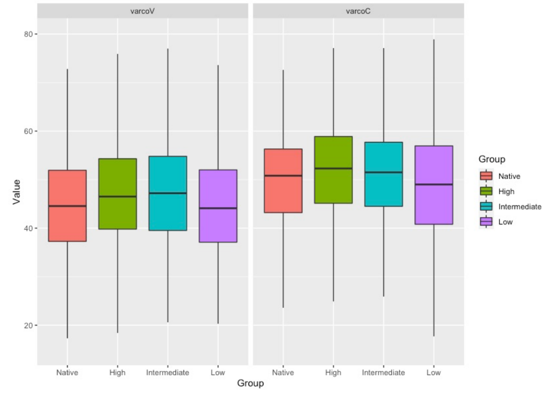

Figure 3: Means and standard errors of the metrics VarcoV and VarcoC for the four groups.

Metric	Proficiency			
	Native	High	Intermediate	Low
ΔV	39.90	52.01	57.69	70.48
VarcoV	45.67	47.46	47.36	44.91
ΔC	46.87	56.83	62.03	74.51
VarcoC	50.17	52.00	51.41	50.19

Table 3: Mean values of raw and normalized metrics.

Tukey post-hoc comparisons (Table 4) showed that there was no significant difference in the durational variability of vocalic/intervocalic intervals between learners and native speakers.

Metric	Significance of contrasts		
	Native-High	Native-Intermediate	Native-Low
VarcoV	0.15	0.19	0.92
VarcoC	0.11	0.35	0.98

Table 4: Post-hoc tests for Varco comparison between groups.

3.2. Pairwise metrics

3.2.1. rPVI & nPVI

The pairwise metrics results for nPVI-V and rPVI-C are shown in Fig. 4.

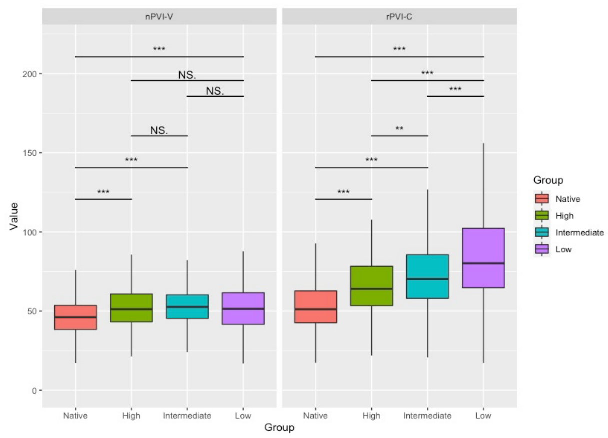

Figure 4: PVI values of the four groups and post-hoc tests results.

Native speakers displayed less variation in nPVI-V values compared to learners. There were no significant differences among the three learner groups (as shown in Fig. 4 on the left), indicating similar performances in successive vocalic intervals after controlling for speech rate. On the other hand, rPVI-C values indicated that lower proficiency was associated with greater temporal variability in successive intervocalic interval productions.

3.2.2. CCI

The average values of CCI-C and CCI-V are depicted in Fig. 5. Native speakers and learners are separated by the bisector. The productions of native speakers and advanced learners are close to the bisecting line, indicating similar controlling tendencies in vocalic and intervocalic durational fluctuations.

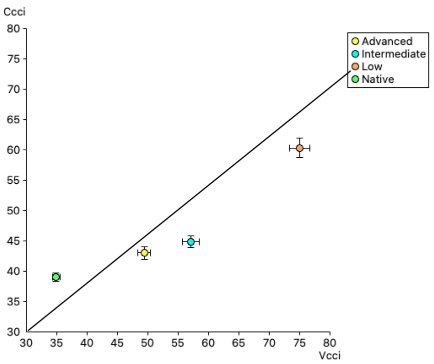

Figure 5: Distribution of CCI values for four groups of subjects.

Native speakers displayed the greatest degree of control, as indicated by the smallest values of CCI-V and CCI-C. In contrast, values for intermediate and low proficiency learners were farther from the bisecting line, indicating a pattern that is more compensating than controlling during speech articulation. Low proficiency learners showed much more fluctuation in both vocalic and intervocalic intervals compared to the other learners.

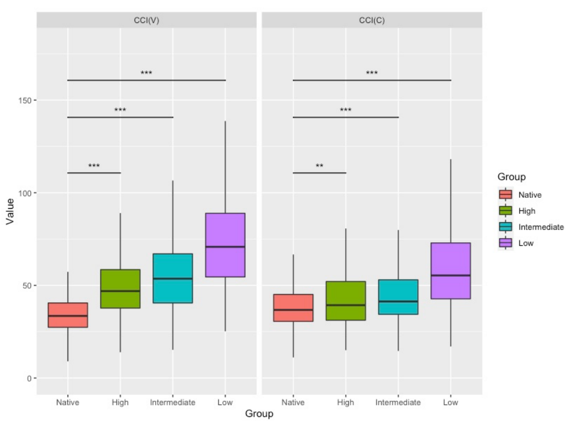

Figure 6: CCI values of the four groups and post-hoc tests results.

The statistical analysis reveals significant differences between natives and learners (Fig. 6): for CCI-V: F-value=226.3, $p<0.01$; for CCI-C: F-value=77.11, $p<0.01$.

4. DISCUSSION AND CONCLUSION

This study compared the rhythmic patterns of French read speech between native speakers and three groups of Chinese learners with low, intermediate, and high levels of proficiency using nine rhythm metrics.

The results for non-rate-normalized metrics indicated differences among all groups. However, these metrics may not be reliable indicators of differences due to their sensitivity to speech rate [15]. In L2 speech, slower speech rates often result in the realization of more and longer pauses, and thus the formation of more shorter prosodic units. Given that prosodic units are often accompanied by segmental lengthening at boundaries, the insertion of more prosodic boundaries leads to more lengthening, which may contribute to increased variation in duration and larger raw durational metrics [24, 25]. This tendency was more evident in the low proficiency group than in the advanced group. Other L2 rhythm metric studies (e.g. [24, 26]) have also reported similar tendencies in raw metrics. Previous research [15, 27] has demonstrated that %V, VarcoV, and nPVI-V are more robust to variation in speech rate and are effective at discriminating cross-linguistic speech rhythm. In the present study, the VarcoV and VarcoC values for learners in the three proficiency groups approximated the target pattern when the metrics were normalized for speech rate. This may be due to the shared phonological properties of French and Chinese, which are classified as syllable-timed languages: 1) the prevalence of the same CV syllable structure, and 2) the rarity of instances of vocalic reduction [28].

While traditional methods of measuring rhythm have difficulty in accurately assessing the development of rhythm in Chinese speakers learning French at different stages of acquisition, the CCI index does offer some insight into their progress. However, it is not sufficient for fully understanding all issues related to rhythm acquisition. One particular challenge for Chinese learners is their ability to correctly group rhythms, as non-native speakers often have less fluency in their speech and tend to divide an utterance into smaller chunks, resulting in more variability in duration at the segment level.

This research is part of a larger project that also investigates final lengthening and F0 movement. Future publications will provide additional findings on Chinese learners' progress in acquiring French speech rhythm.

5. ACKNOWLEDGEMENTS

This study was supported by the French ANR (ANR-10-LABX-0083, Labex EFL).

6. REFERENCES

[1] Grabe, E., Low, E. 2002. Durational Variability in Speech and the Rhythm Class Hypothesis. *Papers in laboratory phonology* 7, 515-546.
[2] Lin, H., Wang, Q. 2007. Mandarin Rhythm: An Acoustic Study. *Journal of Chinese Language and Computing* 17(3), 127-140.
[3] Pike, K. 1945. *The Intonation of American English*. University of Michigan Press.
[4] Abercrombie, D. 1967. *Elements of general phonetics*. Edinburgh University Press.
[5] Ramus, F., Nespor, M., Mehler, J. 1999. Correlates of linguistic rhythm in the speech signal. *Cognition* 73, 265-292.
[6] Turk, A., Shattuck-Hufnagel, S. 2013. What is speech rhythm? A commentary on Arvaniti and Rodriquez, Krivokapić, and Goswami and Leong. *Laboratory Phonology* 4(1), 93-118.
[7] Dellwo, V. 2006. Rhythm and Speech Rate: A Variation Coefficient for deltaC. In: Karnowski, P., Szigeti, I., *Language and language-processing*. Peter Lang, 231-241.
[8] Arvaniti, A. 2012. The usefulness of metrics in the quantification of speech rhythm. *Journal of Phonetics* 40, 351-373.
[9] Arvaniti, A. 2009. Rhythm, Timing and the Timing of Rhythm. *Phonetica* 66, 46-63.
[10] Gut, U. 2012. Rhythm in L2 speech. *Speech and Language Technology* 14/15, 83-94.
[11] Kolly, M., de Mareüil, P., Leemann, A., Dellwo, V. 2017. Listeners use temporal information to identify French- and English-accented speech. *Speech Communication* 86, 121-134.
[12] Kolly, M., Dellwo, V. 2014. Cues to linguistic origin: The contribution of speech temporal information to foreign accent recognition. *Journal of Phonetics* 42, 12-33.
[13] Polyanskaya, L., Ordin, M., Busa, M. 2017. Relative Salience of Speech Rhythm and Speech Rate on Perceived Foreign Accent in a Second Language. *Language and Speech* 60(3), 333-355.
[14] Polyanskaya, L., Ordin, M., Ulbrich, C. 2013. Contribution of Timing Patterns into Perceived Foreign Accent. *Elektronische Sprachsignalverarbeitung* 71-79.
[15] Wiget, L., White, L., Schuppler, B., Grenon, I., Rauch, O., Mattys, S. 2010. How stable are acoustic metrics of contrastive speech rhythm? *Journal Acoustical Society of America* 127(3), 1559-1569.
[16] Gabriel, C., Stahnke, J., Thulke, J. 2014. On the acquisition of French speech rhythm in a multilingual classroom: Evidence from linguistic and extra-linguistic data. *Actes de Congrès Mondial de Linguistique Française*, 1267-1283.
[17] Pillot-Loiseau, C., Xie, H. 2018. Transfert rythmique du chinois mandarin au français dans l'apprentissage du Français Langue Étrangère : acoustique et perception. *Actes de Congrès Mondial de Linguistique Française* 46, 09001.
[18] Marian, V., Blumenfeld, H., Kaushanskaya, M. 2007. The Language Experience and Proficiency Questionnaire (LEAP-Q): Assessing Language Profiles in Bilinguals and Multilinguals. *Journal of Speech, Language, and Hearing Research* 50, 940-967.
[19] Bigi, B. 2015. SPPAS - MULTI-LINGUAL APPROACHES TO THE AUTOMATIC ANNOTATION OF SPEECH. *The Phonetician* 111(112), 54-69.
[20] Boersma, P., Weenink, D. 2022. *PRAAT: doing phonetics by computer*.
[21] Bertinetto, P., Bertini, C. 2008. On modeling the rhythm of natural languages. *Proc. of the 4th International Conference on Speech Prosody* Campinas, 427-430.
[22] Droua-Hamdani, G., Selouani, S., Alotaibi, Y. 2014. Rhythm analysis in Arabic L2 speech. *Proc. of the 7th International Conference on Speech Prosody* Dublin, 1007-1011.
[23] Mairano, P., Romano, A. 2010. Un confronto tra diverse metriche ritmiche usando Correlatore. *La dimensione temporale del parlato (Proc. of the V National AISV Congress)* University of Zurich, 79-100.
[24] Gu, W., Hirose, K. 2014. Rhythmic Patterns in Native and Nonnative Mandarin Speech. *Proc. of the 7th International Conference on Speech Prosody* Dublin, 592-596.
[25] Yazawa, K., Kondo, M. 2022. A Comparison of Rhythm Metrics for L2 Speech. *Proc. of the 10th International Conference on Speech Prosody* Lisbon, 332-336.
[26] Ordin, M., Polyanskaya, L., Ulbrich, C. 2011. Acquisition of Timing Patterns in Second Language. *Proc. Interspeech* Florence, 1129-1132
[27] White, L., Mattys, S. 2007. Calibrating rhythm: First language and second language studies. *Journal of Phonetics* 35, 501-522.
[28] Dauer, R. 1983. Stress-timing and syllable-timing reanalyzed. *Journal of Phonetics* 11, 51-62.