INTERSPEECH 2023
20-24 August 2023, Dublin, Ireland

# Rhythmic Characteristics of L2 German Speech by Advanced Chinese Learners

*Lindun Ge*[1], *Min Xu*[1], *Hongwei Ding*[2]

[1]Institute of Corpus Studies and Applications, Shanghai International Studies University, China
[2]Speech-Language-Hearing Center, School of Foreign Languages, Shanghai Jiao Tong University, China

lindun.ge@shisu.edu.cn, xumin@shisu.edu.cn, hwding@sjtu.edu.cn

## Abstract

The rhythm of speech produced by advanced Chinese learners can still be problematic. In order to improve their pronunciation, it is necessary to investigate L2 rhythmic deviance. In this study, read-aloud materials of 10 German L1 speakers selected from MULTEXT and parallel recordings of 14 Chinese L2 learners were analyzed by comparing six widely used rhythmic metrics, speech rate, and vowel deletion ratio. A correlation analysis was also conducted between these rhythmic features and the degree of perceived foreign accent. The results indicated that speech rate, varcoC, and vowel deletion ratio were significantly different between L1 and L2 speakers, and that speech rate and vowel deletion ratio had a strong correlation with perceived foreign accent within L2 speech. These findings suggest that complex consonantal intervals and vowel deletion are still challenging for advanced learners and should be given more attention in pronunciation training.

**Index Terms**: non-native speech, rhythm, foreign accent, vowel deletion, Chinese-German L2 learner

## 1. Introduction

As a result of negative transfer from the learners' native language, non-native rhythm in L2 (second language) speech often results in foreign accents and can have a negative impact on intelligibility [1–3]. This rhythmic deviance can be remedied by targeted training. However, rhythm training is usually poorly implemented or even neglected in German language teaching due to insufficient knowledge of non-native rhythmic features [4]. Therefore, research on rhythmic deviance in German as a foreign language is needed.

In the case of Chinese learners of German, previous studies have shown that L2 speech produced by low and intermediate learners was more syllable-oriented than stress-oriented. In comparison with German L1 (first language) speakers, Chinese L2 learners had a slower rate of speech, a higher frequency of vowel epenthesis, and an insufficient reduction of vowels in their German speech [5–7]. As a result, a higher proportion of vocalic intervals (*%V*) and a higher standard deviation of consonantal intervals (*ΔC*) [8] were measured in L2 speech [5]. Such rhythmic deviance may be reduced as L2 acquisition progresses [9, 10]. For example, the occurrence of epenthesis in L2 speech was found to decrease significantly after the learners participated in a three-month German course [7].

However, progress in speech rhythm does not necessarily correlate with progress in overall L2 proficiency or length of L2 exposure. And different features of L2 rhythm (e.g., syllable structure, vowel reduction, final lengthening) have been shown to be acquired at different levels [2]. In fact, even with excellent mastery of vocabulary and grammar, advanced learners of German usually show uneven fossilization of L2 speech rhythm [4]. This suggests that the deviation of advanced L2 rhythm might be heterogeneous in various aspects and might also be different from that of low and intermediate L2 rhythm. Therefore, the rhythmic deviance of advanced Chinese learners' German L2 speech needs to be investigated in detail.

The investigation could be carried out by comparing the rhythmic metrics between L1 and L2 speech. As quantitative measures of speech rhythm, rhythmic metrics (i.e., *%V* and *ΔC* [8], *PVI* [11], *varco* [12]) were first proposed to classify stress-timed and syllable-timed language [13, 14]. Recently, these metrics have been widely used in L2 prosodic research. For example, previous studies demonstrated that rhythmic metrics of L2 English could indicate learners' proficiency and correlate with the level of their foreign accent [15–17]. Some studies even attempted to build accent classifiers based on rhythmic metrics [18–20]. These studies have shown that rhythmic metrics are able to capture the rhythmic features and deviance of L2 speech, and can be used as a reference for rhythm training. Therefore, by comparing rhythmic metrics between L1 and L2 speech and by analyzing the correlation between these metrics and the degree of foreign accent, a detailed understanding of the fossilization of advanced L2 rhythm can be obtained, which can contribute to rhythmic improvement.

In addition, vowel deletion could be included in our investigation as a factor contributing to non-native rhythm. Corpus-based studies suggested that German L1 speakers tended to produce /ən/ or /əm/ with a preceding consonant in the form of reduced syllabic consonants /n̩/ or /m̩/ [21], while deletion of schwa /ə/ in this context is rare in Chinese L2 speech [22]. Different frequencies of vowel deletion may result in different syllable structures. Due to fewer vowel deletions, L2 speech could have a higher *%V* and lower *ΔC*, which may result in foreign accents [5, 23]. However, this claim has not yet been confirmed by empirical studies. Therefore, a correlation analysis can be performed between the frequency of vowel reduction and the degree of foreign accent.

Overall, this research aimed to 1) explore the differences in rhythmic features between German speech produced by advanced Chinese learners and German native speakers; 2) investigate the correlation between the degree of foreign accent with appropriate rhythmic metrics.

## 2. Method

### 2.1. Participants

Fourteen native Chinese speakers (7 males, 7 females) were recruited for the current study. They were students at Shanghai International Studies University with a mean age of 23.7 years,


10.21437/Interspeech.2023-1480

ranging from 22 to 29. All speakers are advanced German L2 learners. They majored in German or had gained at least 4×4 scores in the TestDaF exam (the exam consists of four sections: reading, writing, listening and speaking, each of which has a maximum score of five), which served as evidence of the language skills required for university study in Germany. Before German learning, all speakers had 12 years of English learning experience from primary and middle school education.

### 2.2. Materials

In order to obtain speech data for rhythmic measurements and metrics comparison between L1 and L2 speech with fewer mispronunciations, repetitions, and hesitations, a reading task was selected in this research. The reading materials were selected from MULTEXT [24], a multilingual speech database specifically designed for prosodic analysis. The materials consist of 40 different German passages, each containing five thematically related sentences. L1 speech data were taken from the German part of MULTEXT, produced by 10 German native speakers (5 males, 5 females), each reading half of the 40 passages. Fourteen Chinese participants had the task of reading all 40 passages for the production of L2 speech data. The L2 speech data were recorded in a soundproof room at a sampling rate of 44.1 kHz (16-bit). The microphone was placed approximately 15 cm in front of the participant's mouth. Each participant got familiarized with the reading materials one day before the formal recording. Speakers were instructed to read all 40 sentences aloud. Each sentence was read once. Repetitions were required when obvious misreadings, hesitations, or disfluencies occurred. In total, we obtained a parallel speech corpus that was comparable in terms of reading material, recording instructions, quality, and so on.

### 2.3. Annotations

All 40 passages (produced by 10 L1 speakers and 14 L2 speakers) were chosen for vowel deletion analysis, and only passage nine (produced by 5 L1 speakers and 14 L2 speakers) was selected for rhythmic analysis. The wave files were first automatically labeled at both word- and phoneme-level with pre-trained Montreal Forced Aligner [25]. The materials chosen for rhythmic analysis were manually corrected and annotated with reference to previous methods [26]. The first step involved the modification of segmental labels and boundaries, which was carried out in Praat [27]. The segmental labels were checked according to the spectrogram and perceptual impression. Any labels that were inconsistent with actual speech were corrected. Meanwhile, all the segmental annotations were adjusted with reference to changes in waveform, spectrogram, and formants [28, 29].

The next step involved the transformation from separate phonemic segments into vocalic intervals (VIs) and consonantal intervals (CIs), which was conducted by a self-written program. A VI consisted of vowels and /ɐ/ (vocalized allophone of the consonant /ʁ/), while a CI consisted of consonants, syllabic consonants /l̩, m̩, n̩/, and closure period of fricative and plosive consonants. Temporal successive segments of the same type were merged into one interval [15], as shown in Figure 1.

For vowel deletion analysis, all rhyme /ən/ in C+/ən/ or C+/əm/ syllables were temporarily labeled as syllabic consonant /n̩/ or /m̩/ in automatic processing because vowel deletion could not be identified by the algorithm. Manual annotations with Praat were conducted. In accordance with the audio and visual cues, the differentiation between a syllable with schwa /ə/ (syllable with no vowel deletion) or simple syllabic consonants (syllable with vowel deletion) was conducted for each /n̩/ or /m̩/ segment. Any segment referring to /ən/ or /əm/ syllable was split into a nuclear segment and a coda segment.

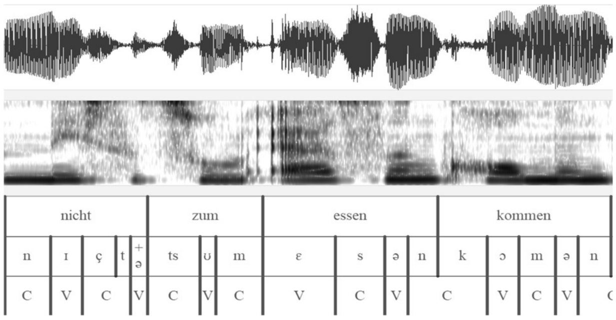


Figure 1. *Segmental and intervocalic annotation of example from L2 speech clip "…nicht zum Essen kommen …".*

### 2.4. Accentedness assessment

In order to investigate the correlation between acoustic metrics and degrees of foreign accent, subjective assessments of L2 speech by each Chinese speaker were carried out.

The assessments were made by two high-proficiency German L2 speakers with abundant experience in teaching German phonetic introductory courses. The 14 wave files by L2 speakers of passage nine were split into 70 audio clips in accordance with the five sentences of every passage. The judges were asked to score each audio clip on a scale from 1 to 5 (from most foreign-accented to most native-like) in reference to the given scoring criteria. After separate scoring, the scores of native-likeness were collected and averaged for each sentence.

### 2.5. Data analysis

#### 2.5.1. Rhythmic metrics analysis

For rhythmic metrics analysis, the inter-pausal unit (IPU), a series of speech segments surrounded by two pauses no shorter than 200 ms [15] was adopted as the basic unit for better data normalization. The duration values of VIs and CIs within all materials were measured. According to Table 1, speech rate and rhythmic metrics were separately calculated for each IPU. Means and standard deviations (SD) of speech rate and each rhythmic metric were calculated for each sentence. A series of t-tests were performed for the metrics between L1 and L2 groups.

Table 1: *Description or formula of rhythmic metrics.*

| Rhythmic metrics | Description or formula |
|---|---|
| **speech rate** | syllables per second |
| ***%V*** | proportion of VIs in the IPU |
| ***ΔC*** | SD of CIs within the IPU |
| ***nPVI-V*** | $= 100 \times \sum_{k=1}^{n-1} \left\| \frac{v_{k+1} - v_k}{\frac{v_{k+1} + v_k}{2}} \right\| /(n-1)$ where $n$ is the number of VIs with the IPU and $v_k$ is the duration of $k^{th}$ VI. |

| | | |
|---|---|---|
| ***rPVI-C*** | $= \sum_{k=1}^{n-1} \lvert c_{k+1} - c_k \rvert / (n-1)$<br>where $n$ is the number of VIs within the IPU and $c_k$ is the duration of $k^{th}$ CI. |
| ***varcoV*** | $= 100 \times \frac{\Delta V}{\mu}$<br>where $\Delta V$ is SD of VIs within the IPU and $\mu$ is the mean duration of VIs within the IPU. |
| ***varcoC*** | $= 100 \times \frac{\Delta C}{\mu}$<br>where $\mu$ is the mean duration of CIs within the IPU. |

### *2.5.2. Vowel deletion analysis*

In order to quantify the tendency of vowel deletion, the vowel deletion ratio (i.e., percentage of vowel deleted syllables within all C+/ən/ and C+/əm/), was calculated for each sentence.

### *2.5.3. Correlation analysis*

Pearson correlation was performed between the scores of native-likeness and various rhythmic metrics, as well as vowel deletion ratios for passage nine of the L2 group.

# 3. Results

## 3.1. Rhythmic metrics

The rhythmic metrics measured are presented in Table 2.

Table 2: *Means and SDs of rhythmic metrics for L1 and L2 groups and group discrimination (t-test).*

| Rhythmic metrics | L1 (n=5) | L2 (n=14) | t-test |
|---|---|---|---|
| | mean ± SD | mean ± SD | |
| ***%V*** | 0.390 ± 0.014 | 0.419 ± 0.040 | - |
| ***ΔC*** | 0.059 ± 0.008 | 0.060 ± 0.005 | - |
| ***nPVI-V*** | 47.15 ± 5.21 | 59.09 ± 4.48 | *** |
| ***rPVI-C*** | 70.85 ± 11.59 | 79.40 ± 10.53 | - |
| ***varcoV*** | 40.45 ± 2.84 | 44.34 ± 3.15 | * |
| ***varcoC*** | 48.56 ± 2.37 | 42.65 ± 2.99 | *** |
| **speech rate** | 14.52 ± 1.73 | 11.44 ± 1.09 | *** |

Note: - = not significant; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

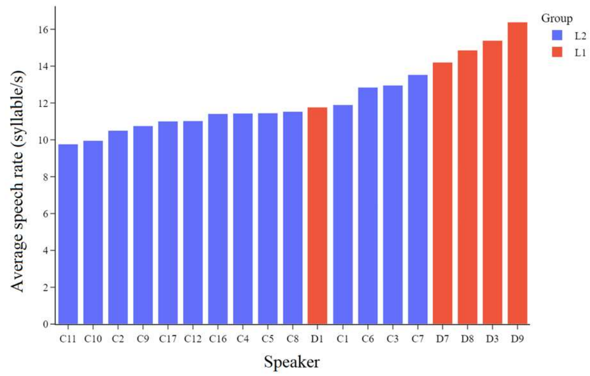


Figure 2. *Average speech rate of L1 and L2 speakers.*

The mean *%V* of the L2 group was slightly higher than that of L1 group, while the *ΔC* value showed no significant difference between the two groups. The *nPVI-V* showed the most considerable discrimination ($p < .001$), with a higher mean value for L2 group. The mean *rPVI-C* was also higher for L2 group, but not significant. The *varcoC* metrics showed significant differences between both groups ($p < .001$), with a higher value for L1 speakers than for L2 speakers. The speech rate of L2 speakers was substantially lower than that of native speakers. As shown in Figure 2, all L1 speakers except for speaker 15 showed a faster speech rate than L2 learners.

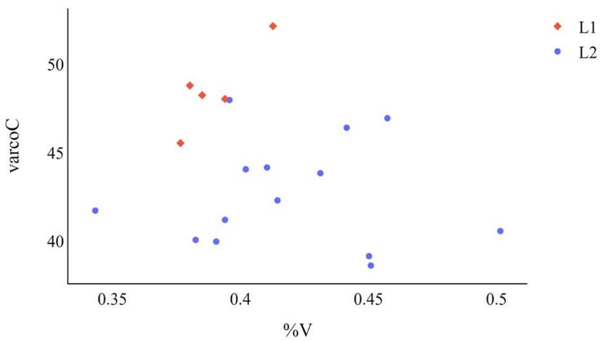


Figure 3. *Measurements of %V and varcoC for each L1 and L2 speaker.*

Figure 3 illustrates the combination of *%V* and *varcoC* for each speaker in both groups, which could roughly differentiate the L1 and L2 speech. Although the holistic distribution of L2 speech was slightly greater on the *%V* axis, there was a large overlap with L1 speech. The traditional *%V-ΔC* coordinate plane could not differentiate the two groups.

## 3.2. Vowel deletion ratio

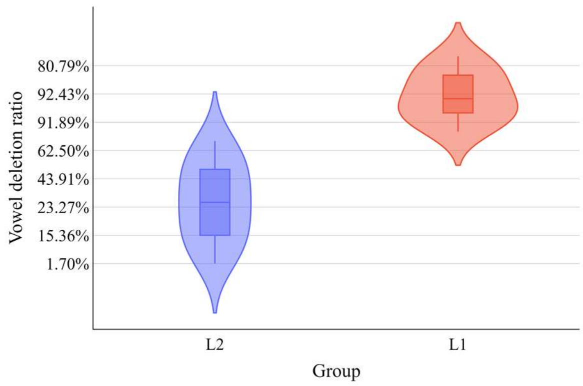


Figure 4. *Violin plot of vowel deletion ratio of L1 and L2 groups.*

As is shown in Figure 4, L1 speakers of German (N=10) produced significantly more vowel deletion (81.55%±8.46%) than L2 speakers (N=14, 33.99%±21.15%) within C+/ən/ and C+/əm/ syllables ($p < .001$). And the distribution of vowel deletion ratios in L1 group was more concentrated than in L2 group.

## 3.3. Correlation with native-likeness

As is shown in Table 3, the six rhythmic metrics hardly had a strong correlation with the accentedness scores. Only *nPVI-V* showed a moderate negative correlation ($p < .05$) with accentedness scores. Speech rate and vowel deletion ratio, however, both correlated fairly strongly ($p < .001$) with accentedness assessment. A clear tendency can be seen in Figure 5, that the perceived native-likeness increased with the growth of vowel deletion ratio.

Table 3: *Correlation of score of native-likeness with speech rate, vowel deletion ratio and different rhythmic metrics.*

| **Rhythmic-acoustic metrics** | **Score of native-likeness** | |
|---|---|---|
| | *r*-value | *p*-value |
| **speech rate** | 0.80 | <0.001 *** |
| ***%V*** | -0.19 | >0.05 - |
| ***ΔC*** | -0.28 | >0.05 - |
| ***nPVI-V*** | -0.56 | <0.05 * |
| ***rPVI-C*** | -0.33 | >0.05 - |
| ***varcoV*** | -0.23 | >0.05 - |
| ***varcoC*** | 0.45 | >0.05 - |
| **vowel deletion ratio** | 0.82 | <0.001 *** |

Note: - = not significant; * = $p < .05$; ** = $p < .01$; *** = $p < .001$.

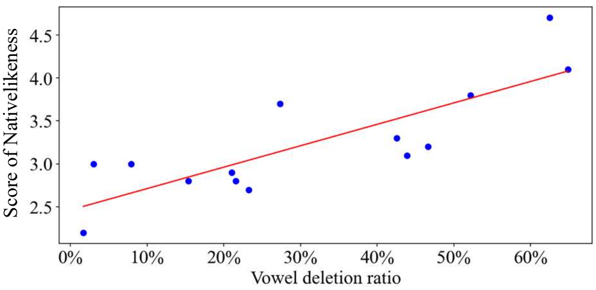


Figure 5. *Values of vowel deletion ratio and scores of native-likeness of L2 speakers with one-way linear regression (r=0.82, p < .001).*

## 4. Discussion

The purpose of this study was to investigate the rhythmic deviance of L2 speech produced by advanced Chinese learners.

Our results indicated that advanced Chinese learners of German made overall progress in L2 rhythm. Compared to the low and intermediate speech in previous research [5], advanced L2 speech showed a declination of *%V* and *ΔC*, which is much closer to L1 speech. This indicates that the progress towards native rhythm may occur with the progress of proficiency level.

The mean *%V* of the L2 group was only slightly higher than that of the L1 group, and the individual values of *%V* showed a large overlap between L1 and L2 speakers. This meant that advanced learners could produce native-like vowel reduction.

However, a native-like speech rate could hardly be achieved even by advanced learners. In our measurements, the speech rates of L2 speakers were generally lower than those of L1 speakers. And a strong negative correlation between speech rate and perceived foreign accent was found in the correlation analysis. This supported the argument that a faster L2 speech rate with advanced non-native rhythm could reduce the perception of a foreign accent [16].

The values of *varcoC* were significantly lower in the L2 group, which is an indication of smaller durational variability within consonantal intervals. Since lower *varcoC* can be measured in Mandarin Chinese as a result of lower consonant complexity in syllables compared to German, the lower *varcoC* in L2 speech can be explained by negative transfer from their native language.

In line with Gut's [22] finding, our research also found a lower mean frequency of vowel deletion in the L2 group. However, this value was found to be highly variable among L2 learners (1.7%–64.9%). Some learners were able to produce syllabic consonants almost as often as L1 speakers, while a few learners made hardly any vowel reductions. This variation may support the view that the acquisition of speech rhythm features does not necessarily correlate with L2 proficiency and is variable across learners.

Furthermore, our analysis also found that the frequency of vowel deletion was significantly correlated with perceived foreign accents, which could confirm Ding and Gut's hypothesis [5, 23]. However, no evidence was found in the correlation analysis that rarer vowel deletion could increase *%V* and decrease *ΔC*. Therefore, the contribution of vowel deletion to accentedness judgments may involve a complex mechanism integrated with other prosodic features, which needs to be further investigated.

Correlation analysis showed that only speech rate and vowel deletion ratio were strongly correlated with a perceived foreign accent, but most of the six rhythmic metrics were not. It might be explained by the unrobustness of these metrics. Previous studies have found that rhythmic metrics are affected by various external factors such as inter-speaker variation, elicitation, syllable composition, material, etc. [30, 31].

In general, even with the overall progress in L2 rhythm, the speech rate and duration variability in consonantal intervals were still challenging for advanced Chinese learners of the German language. We also confirmed the importance of vowel reduction for the improvement of L2 pronunciation. However, in the teaching practice of German, vowel reduction has always been ignored. Therefore, Chinese L2 learners of German need to pay more attention to it and practice it consciously in their daily study and communication.

## 5. Conclusion

In this study, we investigated the rhythmic deviance of German speech produced by advanced Chinese learners. We examined the difference in six widely used rhythmic metrics as well as speech rate and vowel deletion ratio between L1 and L2 speech. We found that speech rate, *varcoC*, and vowel deletion ratio were significantly different between the two groups, and that speech rate and vowel deletion ratio had a strong correlation with perceived foreign accent. In order to achieve pronunciation improvement, advanced learners need to pay more attention to complex consonantal intervals as well as vowel deletion.

## 6. Acknowledgements

This work was supported by grants from the Postgraduate Research & Innovation Program of Institute of Corpus Studies and Applications, Shanghai International Studies University, and the major program of the National Social Science Foundation of China (18ZDA293). The corresponding author is Hongwei Ding (hwding@sjtu.edu.cn)

# 7. References


[1] D. Faber, “Teaching the rhythms of English: A new theoretical base,” *International Reviews of Applied Linguistics*, vol. 24, no. 1–4, 1986, pp. 205–216.

[2] J. Anderson-Hsieh, R. Johnson and K. Koehler, “The relationship between native speakers’ judgements of nonnative pronunciation and deviance in segments, prosody, and syllable structure,” *Language Learning*, vol. 42, no. 4, 1992, pp. 529–555.

[3] A. Cutler, D. Dahan and W. van Donselaar, “Prosody in the comprehension of spoken language: A literature review,” *Language and Speech*, vol.40, no. 2, 1997, pp.141–201.

[4] G. Mehlhorn, “Individual pronunciation coaching and prosody,” in *Non-Native Prosody: phonetic description and teaching practice*, J. Trouvain and U. Gut Eds, Berlin, New York: Mouton de Gruyter, 2007, pp. 211–236.

[5] H. Ding and R. Hoffmann, “An investigation of prosodic features in the German speech of Chinese speakers,” in *Prosody and Language in Contact: L2 acquisition, attrition and languages in multilingual situations*, E. Delais-Roussarie, M. Avanzi and S. Herment Eds. Berlin, Heidelberg: Springer, 2015, pp.221–241.

[6] Y. Gao, H. Ding and P. Birkholz, “An acoustic comparison of German tense and lax vowels produced by German native speakers and Mandarin Chinese learners,” *The Journal of the Acoustic Society of America*, vol. 148, no. 1, 2020, pp. EL112–EL118.

*[7]* H. Ding and R. Hoffmann, “An investigation of vowel epenthesis in Chinese learners’ production of German consonants,” in *Proc. INTERSPEECH 2013 – 14th Annual Conference of the International Speech Communication Association*, Lyon, France, Aug. 2013, pp. 1007–1011.

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

[9] J. Zhang and S. Lee, “Acquisition of English speech rhythm by Chinese learners of English at different English proficiency levels,” *Phonetics and Speech Sciences*, vol. 11, no. 4, 2019, pp. 71–79.

[10] B. H. Huang and S.-A. Jun, “The effect of age on the acquisition of second language prosody,” *Language and Speech*, vol. 54, no. 3, 2011, pp. 387–414.

[11] E. Grabe and E. L. Low, “Durational variability in speech and the rhythm class hypothesis,” in *Laboratory Phonology 7*, C. Gussenhoven and N. Warner, Eds. Berlin: Mouton de Gruyter, 2002, pp. 515–546.

[12] V. Dellwo, “Rhythm and speech rate: A variation coefficient for ΔC,” in *Language and Language-processing*, P. Karnowski and I. Szigeti, Eds. Frankfurt am Main: Peter Lang, 2006, pp. 231–241.

[13] D. Abercrombie, *Elements of General Phonetics*. Edinburgh: Edinburgh University Press, 1967.

[14] K. L. Pike, *The Intonation of American English*. Am Arbor: University of Michigan, 1945.

[15] K. Yazawa and K. Mariko, “A Comparison of Rhythm Metrics for L2 Speech,” in *Proc. Speech Prosody 2022 – 11th International Conference on Speech Prosody*, Lisbon, Portugal, May 2022, pp. 332–336.

[16] N. Polyanskaya, L. Ordin and M. G. Busa, “Relative salience of speech rhythm and speech rate on perceived foreign accent in a second language,” *Language and Speech*, vol. 60, no. 3, 2017, pp. 333–355.

[17] M. Ordin, L. Polyanskaya, C. Ulbrich, “Acquisition of timing patterns in second language,” in *Proc. INTERSPEECH 2011 – 12th Annual Conference of the International Speech Communication Association*, Florence, Italy, Aug. 2011, pp. 1129–1132.

[18] G. Droua-Hamdani, “Design of accent classifier based on speech rhythm features,” *Multimedia Tools and Applications*, 2023, doi: https://doi.org/10.1007/s11042-023-14724-3.

[19] S. Shuju, X. Yanlu, F. Xiaoli and Z. Jinsong, “Automatic detection of rhythmic patterns in native and L2 speech: Chinese, Japanese, and Japanese L2 Chinese,” in *2016 10th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, Tianjin, China, 2016, pp. 1–4.

[20] K. Kyriakopoulos, K. Knil and M. Gales, “A deep learning approach to automatic characterisation of rhythm in non-native English speech,” in *Proc. INTERSPEECH 2019 – 20th Annual Conference of the International Speech Communication Association*, Graz, Austria, Sept. 2019, pp. 1836–1840.

[21] K. J. Kohler, “Articulatory reduction in different speaking styles,” in *Proc. ICPhS 95 – 13th International Congress of Phonetic Sciences*, vol. 2, Stockholm, Sweden, 1995, pp. 12–19.

[22] U. Gut, “Non-native speech rhythm in German,” in *Proc. ICPhS 2003 – 15th International Congress of Phonetic Sciences*, Barcelona, Spain, Aug. 2003, pp. 2437–2440.

[23] H. Ding, R. Jäckel and R. Hoffmann, “A preliminary investigation of German rhythms by Chinese learners,” in *Studientexte zur Sprachkommunikation: Elektronische Sprachsignalverarbeitung 2013*, 2013, pp. 79–85.

[24] E. Campione and J. Véronis, “A multilingual prosodic database,” in *Fifth International Conference on Spoken Language Processing*, Sydney, Australia, Dec. 1998.

[25] M. McAuliffe, M. Socolof, S. Mihuc, M. Wagner, and M. Sonderegger, “Montreal Forced Aligner: Trainable text-speech alignment using Kaldi,” in *Proc. INTERSPEECH 2017 – 18th Annual Conference of the International Speech Communication Association*, Stockholm, Sweden, Aug. 2017, pp. 498–502.

[26] H. Ding and X. Xu, “L2 English Rhythm in Read Speech by Chinese Students,” in *Proc. INTERSPEECH 2016 – 17th Annual Conference of the International Speech Communication Association*, San Francisco, USA, Dec. 2016, pp. 2696–2700.

[27] P. Boersma and D. Weenink. Praat: doing phonetics by computer [computer program]. Version 6.3.09, retrieved 15 March 2023. [Online]. Available: http://www.praat.org

[28] E. Fischer-Jørgensen and B. Hutters, “Aspirated stop consonants before low vowels, a problem of delimitation, - its causes and consequences,” *Annual Reports of the Institute of Phonetics, University of Copenhagen*, vol. 15, 1981, pp. 77–102.

[29] G. E. Peterson and I. Lehister, “Duration of syllable nuclei in English,” *Journal of the Acoustical Socierty of America*, vol. 32, no. 6, 1960, pp. 693–703.

[30] A. Arvanili, “The usefulness of metrics in the quantification of speech rhythm,” *Journal of Phonetics*, vol. 40, no. 3, 2012, pp. 351–373.

[31] L. Wiget, L. White, B. Schuppler, I. Grennon, O. Rauch and S. L. Mattys, “How stable are acoustic metrics of contrastive speech rhythm?” *The Journal of the Acoustic Society of America*, vol. 127, no. 3, 2010, pp. 1559–1569.