A Durational Study of German Speech Rhythm by Chinese Learners

Hongwei Ding[1,2] *and Rüdiger Hoffmann*[3]

[1]School of Foreign Languages, Shanghai Jiao Tong University, China
[2]School of Foreign Languages, Tongji University, China
[3]Institute of Acoustics and Speech Communication, TU Dresden, Germany
hongwei.ding@tongji.edu.cn, ruediger.hoffmann@tu-dresden.de

Abstract

This study focuses on the temporal and metrical features of the German speech produced by Chinese speakers. German is described to be a stress-timed language, while standard Chinese is regarded as a syllable-timed language. It has been suggested that the rhythm of the target language can be influenced by the learners' native language. In this study we conducted an investigation of ten sentences with 18 Chinese students in the low intermediate proficiency level in comparison with six native German speakers. We compared the duration values in terms of pairwise variability indices, and found that most of these Chinese speakers have a lower *nPVI-V* and a higher *rPVI-C* than the German speakers. We illustrate that the conventional duration measures of *nPVI-V* can be influenced by the syllable structures of the utterance and the classification approach of vocalic intervals, and a comparable *nPVI-V* can hardly be expected from different investigations. Furthermore, we argue that duration values alone cannot fully capture the rhythmic patterns of speech because other prosodic parameters such as pitch and energy also join to contribute to rhythmic characteristics of the speech.

Index Terms: durational study, learning German, Chinese learners

1. Introduction

It is well known that Pike [16] and Abercrombie [1] argue that the languages of the world can be classified into two types of rhythm patterns: *a) stress-timed rhythm* and *b) syllable-timed rhythm*. According to this hypothesis, both types of rhythm show rhythmical units of equal duration: *stress-timed* languages tend to have isochronous inter-stress intervals, while *syllable-timed* languages tend to have rather equal syllable durations. Classic examples for a stress-timed language are English and German; while Chinese belongs to syllable-timed languages [13]. The classification of rhythm-classes turned out to be based solely on intuition since a large amount of experiments carried out to provide direct correlates for the isochrony in languages remained without success [17].

In the recent decades many researchers tried to classify languages in other ways. Ramus et al. [17] proposed to divide speech into vocalic and consonantal parts, and to calculate the proportion of the vocalic intervals (*%V*) and the standard deviation of consonantal intervals (ΔC) in a sentence. They showed that stress-timed languages have a higher ΔC and a lower *%V*, whereas syllable-timed languages have a lower ΔC and a higher *%V*. Grabe and Low [9] based their pairwise comparison of successive vocalic and intervocalic intervals and calculated speech rhythm with the *Pairwise Variability Index (PVI)*, which computes the sum of the durational differences between adjacent vocalic or consonantal intervals in an utterance. They found that stress-timed languages have a higher variation in vowel durations, whereas syllable-timed languages (including Mandarin) show a lower variation in vowel length. Lin et al. [13] followed the studies of Ramus et al. [17] and Grabe and Low [9], and measured *%V*, ΔC, normalised variation of the pairs of two adjacent vowel intervals (*nPVI-V*), and raw variation of the pairs of two adjacent consonant intervals (*rPVI-C*) of Mandarin Chinese. Except for *nPVI-V*, all other measures confirmed the auditory impression of Mandarin Chinese being syllable-timed [13].

On the other hand, it has been suggested that the rhythm of the target language can be influenced by the learners' native language [10, 18]. Gut [10] described that L2 German is influenced by L1 of Chinese, English, French, Italian and Romanian in terms of ΔC, *%V* etc. In our previous study [6], we demonstrated that Chinese learners of German in the low intermediate proficiency level have a clearly higher *%V* and roughly higher ΔC than German native speakers, because they insert many vowel epentheses and speak at a slower rate. In this investigation we aim to investigate non-native rhythm of Chinese speakers in terms of *nPVI-V* and *rPVI-C* values, and to discuss the efficiencies and deficiencies of these measures with regard to our investigation material.

2. Method

This study followed the same method which was described by Grabe and Low [9] to investigate metrical features of German speech produced by Chinese speakers. The syllable-timed rhythm in the German speech of the Chinese subjects was so striking for their German teachers, it was thus interesting to find out whether these rhythm measures can reflect the perceptual impression of the rhythmic deviation for native German listeners. In order to ensure comparability, the annotation technique used by Grabe and Low [9] was adopted in the investigation. The durations of vowels and intervals between vowels (excluding pauses) in each sentence were measured. Indices of *rPVI* and *nPVI* were calculated according to equation (1) and (2) respectively. (Note: *rPVI-C* and *nPVI-V* are adopted in this paper to represent consonantal *rPVI* and vocalic *nPVI*.)

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

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

where *m* is the number of intervocalic or vocalic intervals in a sentence, and *d* is the duration of the *k*th interval.

10.21437/SpeechProsody.2014-46

With the results of this investigation we endeavour to answer the following questions:

- Do Chinese speakers exhibit different *nPVI-V* and *rPVI-C* from German native speakers? If yes, what factors might account for the differences?
- Can the pairwise variability indices produce comparable values from different investigations? Can these indices fully reflect the perceptual impression of rhythm?

2.1. Subjects

We recruited 18 native Chinese speakers, including 10 males and 8 females, who come from different parts of China, but all of them speak standard Chinese. At the time of speech data collection they had been living in Germany for one month, and they were just enrolled in the German language course for the DSH exam (the German language university entrance exam for foreign students). Their ages ranged from 22 to 28. All of them had learned German for one to one and a half years, and the length of their formal German instructions had been around 1,200 hours. These participants could be classified as having a low intermediate level, they formed a homogeneous group in terms of age, L1 background, motivation, proficiency of the German language, and the length of residence in Germany. Their German teachers found that their German speech was highly syllable-based, which deviated from the stress-timed native German speech, we thus conducted a durational investigation into their L2 speech to test whether *PVI* measures can reflect the deviated rhythm perceived. In order to provide a reference for comparison, we included six German speakers, one was male and five were female speakers. They were between 22-30 years old and were ordinary German native speakers.

2.2. Speech data collection

In order to have certain control of the speech data, reading tasks were used for analysis. The subjects were instructed to read 50 Phondat sentences in German. For the current study only ten read sentences were selected and analyzed, because they include different sentence types and the vowel and consonant percentages vary from sentence to sentence. It is better to concentrate on a small amount of data since the accuracy of annotation is essential for the measurement, which requires much carefulness and patience. Before the recording began, the subjects were given as much time as they needed to read the text to become familiarized with it. Each subject was individually recorded at 16-bit resolution with a sampling rate of 44.1 kHz by a German phonetics expert, who controlled the quality of their production.

2.3. Data analysis

The sentences were first automatically labeled by a trained aligner, and then manually corrected by the first author assisted by a German phonetics expert on Praat [3]. Sentences were first annotated at the phoneme level on the basis of both audio and visual information, and the phonemes were then grouped into vocalic and consonantal intervals. We followed the same criteria described by Grabe and Low [9] in determining the location of vowel-consonant and consonant-vowel boundaries:

- *Vocalic intervals* were characterized by vowel formants, which could contain a monophthong, a diphthong, or more vowels if the formants continue.
- *Intervocalic intervals* were defined as stretch of signals between vocalic intervals, which could include one or more consonants.

Annotation of vowel boundaries were conducted according to the generally accepted criteria [15]. Consonants such as stops, fricatives, affricates and nasals were clearly identifiable from the changes in spectrogram and formant structures. The approach to glides was also based on acoustic criteria, like that in [9]. If a clear change could be observed in the formant structure or in the amplitude as a prevocalic glide, it was excluded from vocalic portions. Otherwise, glides (especially postvocalic glides) were included in the vocalic portion. If there was a gap that was not part of the sound, it was marked as breath, and this breath gap was subtracted from the calculation of the intervals. Any two intervocalic (or vocalic) intervals split by this gap were combined into the same intervocalic (or vocalic) interval. Since we employed sentences as reading material, we computed the pairwise variability in sentences other than in a passage of speech in [9]. This procedure allows us to compare whether there are any differences among different kinds of sentences.

3. Results

The comparison results on *nPVI-V* and *rPVI-C* measures, and speaking rate are presented in the following.

3.1. Pairwise variability indices

Figure 1 demonstrates the data of the six German native speakers (de) shown as small triangles and 18 Chinese speakers (cn) indicated with small squares. *nPVI-V* values are plotted on the horizontal axis against *rPVI-C* values on the vertical axis. The index values of each speaker are the average of ten sentences.

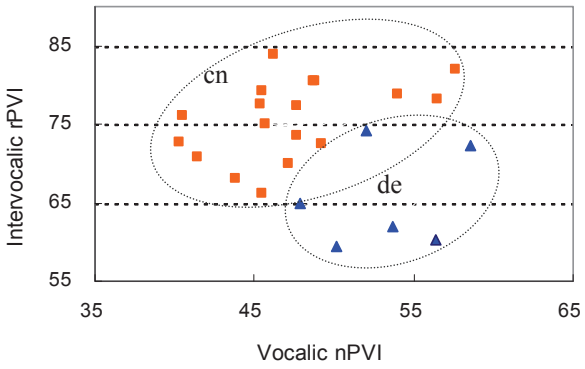

Figure 1: *Vocalic nPVI-V against intervocalic rPVI-C.*

A two-sample independent t-test shows the differences between the Chinese and the German speakers in both *rPVI-C* (t=-3.93, p=0.001) and *nPVI-V* (t=2.515, p=0.020) are significant ($p<0.05$). The Chinese speakers have a higher *rPVI-C* and a lower *nPVI-V* than the German native speakers.

3.2. Speaking rate

It is clear that the Chinese speakers spoke at a slower rate and made more pauses than the native German speakers. Without consideration of pauses or silences, the average speaking rate calculated in phonological syllables per second (syl/s) of the native German speakers and Chinese speakers are 5.95 and 3.90

respectively. A significant negative correlation (r=-0.85) between speaking rate and *rPVI-C* is found for the German speakers, but no such correlation (r=-0.17) can be found for the Chinese speakers.

Table 1: Speaking rate and correlation with *rPVI-C*

	German speakers	**Chinese speakers**
speaking rate (syl/s)	5.95	3.90
standard deviation	0.42	0.26
correlation with *rPVI-C*	-0.85	-0.17

Since the Chinese speakers inserted many epenthesis vowels, they produced much more additional syllables [7]. Speaking rate calculated on the basis of phonetically uttered syllables resulted in a higher rate of 5.27 syl/s (5.27>3.90), which is also negatively correlated with *rPVI-C* (r=-0.567, p=0.014).

4. Discussion

The *PVI* results obtained in this investigation can partly reflect the rhythmic deviance of L2 speech, but may not fully reflect the perceptual impression of speech rhythm.

4.1. Evaluation of *PVI*

The outcome where the Chinese speakers have a higher *rPVI-C* mirrors the result of the previous investigation [6, 10], where the Chinese speakers produced a higher standard deviation of consonantal intervals (ΔC). Since speaking rate is negatively correlated with ΔC in the German language [5], a slower speaking rate results in a higher ΔC, and also produces a higher *rPVI-C*. Moreover, the Chinese speakers can hardly reduce nonstressed vowels, and their *nPVI-V* should be much lower than the German native speakers, which is also indicated in the results. These findings support the results in previous studies [9, 8, 2]. Generally speaking, a higher *rPVI-C* and a lower *nPVI-V* can partly reflect the syllable-timed characteristic of L2 German speech by Chinese speakers.

4.2. Influencing factors

However, there are many factors which can influence the measures of *nPVI-V*, examples in this investigation will be illustrated in the following.

4.2.1. Syllable combinations

The *nPVI-V* scores are found to be influenced not only by the structure of the syllables but also by their combinations in our data. This observation supports previous findings [2, 14] that metric scores can be affected by the choice of material. If one pair of syllables exhibits quite different structures or displays quite different stress patterns, their *nPVI-V* values can be higher than other pairs. More various pairs in one sentence can result in a higher *nPVI-V*. Therefore, the mean *nPVI-V* scores in our investigation are quite different from sentence to sentence, which range from 41.98 to 74.18 for the German speakers, and from 37.06 to 55.50 for the Chinese speakers. However, *nPVI-V* values based on sentences are highly correlated between the German speakers and the Chinese speakers with r=0.67 (p<0.05), which explains that *nPVI-V* is partly subject to the syllable combinations in the sentence. For example, in *Wie hast du das gemacht? (How did you do that?)*, schwa in prefix /g@/ between /das/ and /maxt/ makes *nPVI-V* larger for both German and Chinese speakers.

4.2.2. Successive vowels

Successive vowels do not influence *%V*, but can influence *nPVI-V*. The conventional approach in both Ramus [17] and Grabe & Low [9] is to include adjacent heterosyllabic vowels in the same vocalic interval, which makes *nPVI-V* quite different for different sentences. Successive vowels may and may not be separated by a glottalized period.

If glottalization is found between successive vowels, these intervals were treated as silent pauses. We followed the approach in [19], the glottalized part marked between the solid lines in Figure 2 was omitted in the calculation, and the vocalic intervals before and after the glottalization were summed.

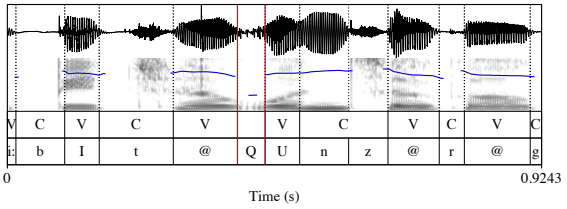

Figure 2: *Waveform, spectrogram and SAMPA annotation of "bitte unsere (please our)" with glottalization between successive vowels by a Chinese speaker.*

Another speaker produced the same phrase without glottalization, as it is shown in Figure 3.

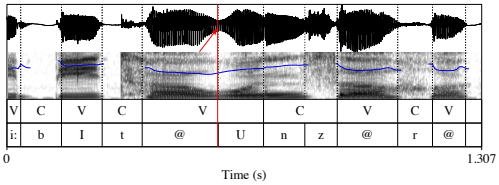

Figure 3: *Waveform, spectrogram and SAMPA annotation of "bitte unsere (please our)" without glottalization between successive vowels by a Chinese speaker.*

To put /@/ and /U/ in one vocalic interval enlarges the *nPVI-V* value greatly for the Chinese speakers. However, both a decrease in pitch and amplitude in the glottalization in Figure 2, and a decrease in the amplitude as pointed out by the arrow in Figure 3 without glottalization can be clearly perceived as a reset in prosody. It is obvious that in the above two figures the duration values of the five vowels are comparable. However, to group these two successive vowels into one vowel interval enlarges *nPVI-V*, which cannot reflect the perceptual impression of the syllable-based rhythm of Chinese speakers faithfully.

4.2.3. Syllabic consonants

Another factor which influences *nPVI-V* values is the annotation of fully reduced vowels in German. One obvious instance is that many word-final syllables in infinite verbs with *-en* are usually pronounced with syllabic consonants. For example, in

Figure 4, schwa @ in word-final with *-en* is reduced, and consonant *s* and syllabic consonant *=n* are combined without any vocalic interval between them.

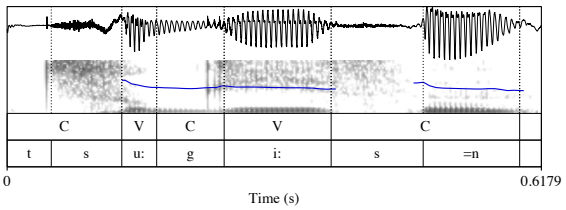

Figure 4: *Waveform, spectrogram and annotation of "zu gießen (to water)" with a syllabic consonant (=n) by a German speaker.*

If a short schwa @ is annotated, the *nPVI-V* can be greater than that with a syllabic consonant. However, a very short weak @ might not be quite different from that with a syllabic consonant perceptually, but the *nPVI-V* values may be quite different.

4.3. Measurements beyond duration

We have observed that various syllable combinations and different annotation approaches of vocalic intervals may lead to various *nPVI-V* scores, which implies that the measures of vocalic and intervocalic intervals may not fully reflect the perception of rhythm. This finding further supports the idea that rhythm metrics such as *PVI* can provide measures of speech timing and variability, but they cannot convey an overall rhythmic impression [2]. It is generally believed that rhythm is the recurring timing patterns of fundamental frequency, syllabic duration, syllabic energy, and spectral dynamics [12]. By comparing the syllable-timed L2 German speech with the stress-timed native German speech, it suggests that not only timing but also fundamental frequency and energy contribute to the rhythmic pattern over time [12]. Cumming [4] also demonstrated that f0 and duration are interdependent in the perception of rhythmic groups in speech and sentence rhythmicity.

German has full as well as spectrally reduced and shortened vowels, and the consequence is a high level of variability in vowel durations. Chinese does not have vowel reduction, and the level of vocalic variability is significantly lower. German employs sentence intonation to express intonational meanings; Chinese uses lexical tones to distinguish words. The difference between a stress-timed German speech by a native speaker and a syllable-timed German speech by a Chinese speaker can be observed in the following two figures, which are vividly displayed with ProZed [11].

In these figures, the pitch contour of the sentence is clearly demonstrated in a continuous dotted line; each circle corresponds to one vocalic or consonantal interval with the vocalic interval marked with corresponding vowels. The level on the vertical y-axis of the circle represents the pitch and the diameter represents the duration of the interval. The unit of pitch has already been normalized to the logarithmic scale $\log_2$(Hz/median), so that speakers of different fundamental frequencies can be comparably displayed on this scale. It is obvious that the vowels have more variability for the German speaker than the Chinese speaker. In Figure 5 the vowels */u:/* in **du** and /@/ in ***ge****macht* are very short in duration, while */a/* in ***gemacht*** is much longer of the German speaker. In Figure 6, except for a very short /@/ in ***ge****macht*, the variations of the other vowels (including the epenthesis /@/ after /t/ in ***haste***) are not so much as those in Figure 5. Not only duration variations but also pitch and energy changes are different. All these prosodic parameters work together to organize speech into rhythmic chunks, which impresses us with different rhythms.

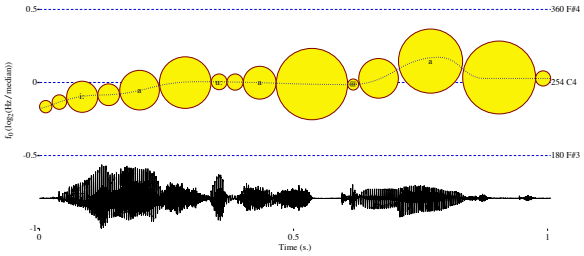

Figure 5: *Prosody display of "Wie hast du das gemacht? (How did you do that?)" produced by a German speaker.*

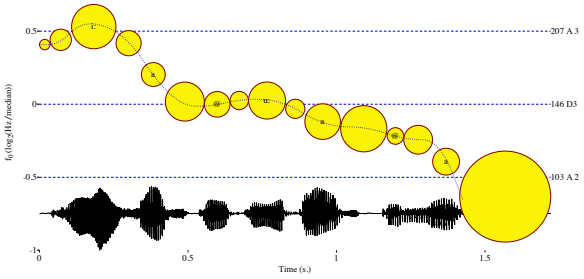

Figure 6: *Prosody display of "Wie hast du das gemacht? (How did you do that?)" produced by a Chinese speaker*

5. Conclusion

This paper analyzes L2 German speech with pairwise variability indices and demonstrates that the syllable-timed L2 German speech by Chinese speakers is characterized by a lower *nPVI-V* and a higher *rPVI-C*. However, due to many influencing factors, a comparable *nPVI-V* can hardly be expected from different investigations. We further suggest that other prosodic parameters, such as F0 and energy, can work together with duration to contribute to the acoustic analysis of rhythmicity, which can better reflect the perceptual impressions of speech rhythm. This study also suggests that in the future we should focus more on grouping patterns of prominence with more prosodic parameters than duration to investigate speech rhythm, and it would be more reasonable to link measurements of acoustic parameters in rhythm production with listeners' rhythm perception in the investigation of rhythm, as it is suggested by many researchers in [4, 12].

6. Acknowledgements

The first author is sponsored by Shanghai Social Science project (2011BYY002) and Innovation Program of Shanghai Municipal Education Commission (12ZS030) for this research work. We thank Rainer Jäckel for his help in the data collection.

7. References

[1] Abercrombie, D., "Elements of general phonetics", Aldine: Chicago, 1967.

[2] Arvaniti, A., "Rhythm, Timing and the Timing of Rhythm", Phonetica, 66:46-63, 2009.

[3] Boersma, P. and Weenink, D. "Praat: doing phonetics by computer [Computer program]", http://www.praat.org/, 2013.

[4] Cumming, R. E., "Perceptually Informed Quantification of Speech Rhythm in Pairwise Variability Indices", Phonetica, 68:256-277, 2011.

[5] Dellwo, V. and Wagner, P., "Relations between language rhythm and speech rate", Proc. of the 15th ICPhS, 471–474, 2003.

[6] Ding, H., Jäckel, R., and Hoffmann, R., "A Preliminary Investigation of German Rhythm by Chinese Learners", ESSV2013, 79-85, 2013.

[7] Ding, H. and Hoffmann, R., "A An Investigation of Vowel Epenthesis in Chinese Learners' Production of German Consonants", Interspeech, 1007–10115, 2013.

[8] Gibbon, D. and Gut, U., "Measuring speech rhythm", Eurospeech 2001.

[9] Grabe, E. and Low, E. L., "Durational variability in speech and the rhythm class hypothesis", in C. Gussenhoven, and N. Warner [Ed], Laboratory Phonology, 7:515–546, Berlin: Mouton, 2002.

[10] Gut, U., "Non-native Speech: A Corpus-based Analysis of Phonological and Phonetic Properties", Peter Lang GmbH, 2009.

[11] Hirst, D., "ProZed: A speech prosody analysis-by-synthesis tool for linguists", Proc. of Speech Prosody 2012, 15-18, 2012.

[12] Kohler, Klaus J., "Rhythm in Speech and Language: A New Research Paradigm", Phonetica, 66:29-45, 2009.

[13] Lin, H., and Wang, Q., "Mandarin Rhythm: An Acoustic Study", Journal of Chinese Language and Computing, 17:127-140, 2007.

[14] Loukina, A., and Kochanski, G., and Rosner, B. and Keane, E., "Rhythm measures and dimensions of durational variation in speech", Journal of the Acoustical Society of America, 129:3258-3270, 2011.

[15] Peterson, G. E. and Lehiste, I., "Duration of Syllable Nuclei in English", Journal of the Acoustical Society of America, 32:693-703, 1960.

[16] Pike, K. L., "The intonation of American English", University Press: Michigan, 1945.

[17] Ramus, F., Nespor, M., and Mehler, J., "Correlates of linguistic rhythm in the speech signal", Cognition, 72:1-28, 1999.

[18] Tortel, A., and Hirst, D. "Rhythm metrics and the production of English L1/L2", Speech Prosody, 2010.

[19] White, L., and Mattys, S. L., "Calibrating rhythm: First language and second language studies", Journal of Phonetics, 35:501-522, 2007.